ENGINEERING SERVICE CENTER
Port Hueneme, California 93043-4370

TECHNICAL REPORT
TR-2317-ENV

ADAPTIVE LONG-TERM MONITORING AT ENVIRONMENTAL RESTORATION SITES (ER-0629)

Prepared by:

NAVFAC Engineering Service Center – Karla Harre and Tanwir Chaudhry
GeoTrans, Inc. – Rob Greenwald and Weiwei Jian
Environmetrics & Statistics Limited – Charles Davis
Summit Envirosolutions, Inc. – Matt Zavislak
Barbara Minsker (Consultant to Summit Envirosolutions, Inc.)

May 2009

Approved for public release; distribution is unlimited

Printed on recycled paper

FINAL REPORT

Adaptive Long-Term Monitoring at Environmental Restoration Sites (ER-0629)

May 2009

Document Version 1.0

Prepared by:

Naval Facilities Engineering Service Center – Karla Harre and Tanwir Chaudhry
GeoTrans, Inc. – Rob Greenwald and Weiwei Jian
Environmetrics & Statistics Limited – Charles Davis
Summit Envirosolutions, Inc. – Matt Zavislak
Barbara Minsker (Consultant to Summit Envirosolutions, Inc.)

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Services and Communications Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
21-05-2009	Technical Report	04-05-2006 to 21-05-2009

4. TITLE AND SUBTITLE
ADAPTIVE LONG-TERM MONITORING AT ENVIRONMENTAL RESTORATION SITES (ER-0629)

5a. CONTRACT NUMBER
N62473-06-C-3062

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
NAVFAC Engineering Service Center – Karla Harre and Tanwir Chaudhry
GeoTrans, Inc. – Rob Greenwald and Weiwei Jian
Environmetrics & Statistics Limited – Charles Davis
Summit Envirosolutions, Inc. – Matt Zavislak
Barbara Minsker (Consultant to Summit Envirosolutions, Inc.)

5d. PROJECT NUMBER
ESTCP ER-0629

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Facilities Engineering Service Center
1100 23rd Avenue
Port Hueneme, CA 93043

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Environmental Security Technology Certification Program
901 North Stuart Street, Suite 303
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
ESTCP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The primary objective of this ESTCP project was to demonstrate and validate the use of the Sampling Optimizer and Data Tracker software offered by Summit Envirosolutions, at three DoD sites. The three demonstration sites were as follows: Former George Air Force Base Site, Victorville, CA (GAFB site), Former Nebraska Ordnance Plant Site, Mead, NE (NOP site), Camp Allen Landfill Site, Norfolk, VA (Camp Allen site). MAROS was also applied at one of the three demonstration sites.
The Summit Software demonstrated in this ESTCP project provides a set of tools for Long-Term Monitoring Optimization (LTMO), and consists of two major modules: Sampling Optimizer (SO) identifies redundant sampling locations (spatial optimization), or redundant locations and frequencies (spatiotemporal optimization), in historical data. Data Tracker (DT) allows current monitoring data to be reviewed against selected historical data (i.e., the "background data") to identify cases where current data deviate from expectations that are based on the background values and patterns.

15. SUBJECT TERMS
Monitoring, optimization, sampling redundancy, groundwater, optimization software, data redundancy, Summit, HMSL, monitoring well network, spatial analysis, spatiotemporal analysis, sampling cost, sampling error, long term monitoring

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Karla Harre
U	U	U		418	19b. TELEPHONE NUMBER (Include area code) (805) 982-2636

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

TABLE OF CONTENTS iii
LIST OF ACRONYMS vi
ACKNOWLEDGMENTS viii
EXECUTIVE SUMMARY x
1.0 INTRODUCTION 1
1.1 BACKGROUND 1
1.2 OBJECTIVE OF THE PROJECT 2
1.3 REGULATORY DRIVERS 2
2.0 TECHNOLOGY 3
2.1 TECHNOLOGY DESCRIPTION 3
2.2 TECHNOLOGY DEVELOPMENT 10
2.3 ADVANTAGES AND LIMITATIONS OF THE TECHNOLOGY 10
3.0 PERFORMANCE OBJECTIVES 15
4.0 SITE DESCRIPTION 17
4.1 SITE LOCATION AND HISTORY 18
4.2 SITE GEOLOGY/HYDROGEOLOGY 22
4.3 CONTAMINANT DISTRIBUTION 25
5.0 TESTING DESIGN 30
5.1 CONCEPTUAL EXPERIMENTAL DESIGN 30
5.2 BASELINE CHARACTERIZATION 32
5.3 TREATABILITY OR LABORATORY STUDY RESULTS 37
5.4 DESIGN AND LAYOUT OF TECHNOLOGY COMPONENTS 37
5.5 FIELD TESTING 37
5.5.1 Schedule for Software Testing 38
5.5.2 Variations Tested 39
5.5.3 Ease of Use 41
5.5.4 Software Bugs and Evolution of Software Features 41
5.5.5 Testing of Interpolation and Transformation Options in Model Builder 43
5.5.6 Summary of Redundancy Evaluation Results 45
5.5.7 Tracking of Relative Mass and Mass Flux 53
5.5.8 Visualization of Relative Uncertainty 56
5.5.9 Data Tracker Results for Detecting Unexpected Concentrations 57
5.5.10 Import/Export Features 62
5.5.11 Computation Time and Level of Effort 63
5.6 SAMPLING METHODS 66
5.7 SAMPLING RESULTS 66
6.0 PERFORMANCE ASSESSMENT 67
6.1 QUALITATIVE PERFORMANCE OBJECTIVES 67
6.1.1 User Functionality 67
6.1.2 Software Reliability 67
6.1.3 Model Builder Performance 67
6.1.4 Sampling Optimizer Performance 69

6.1.5 Data Tracker Performance 71
6.1.6 Regulatory Acceptance 71
6.1.7 Comparison with MAROS (secondary) 72
6.2 QUANTITATIVE PERFORMANCE OBJECTIVES 72
6.2.1 Model Builder Performance 72
6.2.2 Sampling Optimizer Performance 72
6.2.3 Data Tracker Performance 73
6.2.4 Comparison with MAROS 73
7.0 COST ASSESSMENT 75
7.1 COST MODEL 75
7.2 COST DRIVERS 76
7.3 COST ANALYSIS 78
8.0 IMPLEMENTATION ISSUES 81
9.0 REFERENCES 85

List of Figures

Figure 2-1 General Flowchart of Software Application
Figure 2-2 Overview of the Genetic Algorithm Approach
Figure 2-3 Example of a Tradeoff Curve For Spatial Optimization
Figure 2-4 Cutoff Error Calculator (From Software Manual by Summit Envirosolutions, 2008)
Figure 4-1 Location of Camp Allen Site, Norfolk, VA
Figure 4-2 Location of GAFB Site, San Bernardino, CA
Figure 4-3 Location of NOP Site, Mead, NE
Figure 4-4 Schematic Cross Section, Camp Allen Site
Figure 4-5 Schematic Cross Section, GAFB Site
Figure 4-6 Schematic Cross Section, NOP Site
Figure 4-7 Plume Boundary in Shallow Aquifer, Camp Allen Site
Figure 4-8 Plume Boundary in Deep Aquifer, Camp Allen Site
Figure 4-9 TCE Plume in the OU-1 Upper Aquifer, GAFB Site
Figure 4-10 TCE Plumes (Shallow Aquifer), NOP Site
Figure 4-11 RDX Plumes (Shallow Aquifer), NOP Site
Figure 5-1 Example Model Builder Results for TCE at GAFB site, with Different Combinations of Interpolation Technique and Data Transformation
Figure 5-2 Tradeoff Curve for Spatial Redundancy Evaluation, GAFB Site (TCE Dataset A, Cutoff = 25 ug/l)
Figure 5-3 Comparison of Two Optimal Plans and Baseline Plan, GAFB Site (TCE, Dataset A, Cutoff = 25 ug/l)
Figure 5-4 Spatial Optimization Tradeoff Curves Using Different Cutoff Values (TCE, GAFB Site, Dataset A)
Figure 5-5 Spatiotemporal Optimization Tradeoff Curves Using Different Cutoff Values (TCE, GAFB Site, Dataset A)

Figure 5-6 Spatial Optimization Tradeoff Curves Using Different Population Size and Number of Generations (TCE, GAFB Site, Dataset B)
Figure 5-7 Example of Validation Results Using Reserved Data, GAFB Site (TCE, GAFB Site, Dataset B, Spatial Optimization)
Figure 5-8 Example 1 Regarding Complication for Mass Calculations
Figure 5-9 Example 2 Regarding Complication for Mass Calculations
Figure 5-10 Example of Relative Uncertainty Map, NOP Site (Kriging with Quantile Transformation)
Figure 5-11 Example Data Tracker Plot of an Anomaly Due to Bad Data
Figure 5-12 Example Data Tracker Plot With "Out-of-Bounds" Value That is a Potential Concern
Figure 5-13 Example Data Tracker Plot With "Out-of-Bounds" Value That is Not a Concern
Figure 6-1 Conceptual Example to Illustrate Temporal Interpolation Issue
Figure 6-2 Conceptual Example to Illustrate Spatiotemporal Optimization Issue
Figure 7-1 Cost-Benefit Analysis, Scenario 1
Figure 7-2 Cost-Benefit Analysis, Scenario 2

List of Tables

Table 3-1 Performance Objectives
Table 4-1 Summary of Demonstration Sites
Table 5-1 COCs for Redundancy Analysis and Data Tracking Analysis
Table 5-2 Schedule for Testing the Software
Table 5-3 Summary of Redundancy Evaluation Results, Camp Allen Site
Table 5-4 Summary of Redundancy Evaluation Results, GAFB Site
Table 5-5 Summary of Redundancy Evaluation Results, NOP Site
Table 5-6 General Summary of Time Required to Apply the Software
Table 5-7 Summary of Computation Time versus Number of Wells (Kriging with Quantile Transformation)
Table 5-8 Computation Times for Different Model Types, Camp Allen Site
Table 5-9 Computation Times for Different Model Types, GAFB Site
Table 5-10 Computation Times for Different Model Types, NOP Site
Table 7-1 Estimated Costs to Apply the Software at a Typical DoD Site

List of Appendices

Appendix A Points of Contact
Appendix B Site-Specific Analysis – GAFB Site
Appendix C Site-Specific Analysis – NOP Site
Appendix D Site-Specific Analysis – Camp Allen Site
Appendix E EPA Region V Evaluation
Appendix F MAROS Comparison

LIST OF ACRONYMS

1,1-DCE	1,1-dichloroethylene
2,4-DNT	2,4-dinitrotoluene
AFCEE	Air Force Center for Environmental Excellence
bgs	below ground surface
c12DCE	cis-1,2-dichloroethene
COC	contaminant of concern
CPU	central processing unit
1,2-DCA	1,2-dichloroethane
DoD	Department of Defense
DT	Data Tracker
EM CX	Environmental and Munitions Center of Expertise (formerly the Hazardous, Toxic, and Radioactive Waste Center of Expertise)
EPA	US Environmental Protection Agency
ESTCP	Environmental Security Technology Certification Program
FFA	Federal Facilities Agreement
FS	feasibility study
FY	fiscal year
GA	genetic algorithm
GAC	granular activated carbon
GAFB	former George Air Force Base
Ghz	gigahertz
gpm	gallon per minute
GUI	graphical user interface
LAU	Lower Alluvial Unit
LLU	Lower Lacrustine Unit
LTM	Long Term Monitoring
LTMO	Long Term Monitoring Optimization
MAROS	Monitoring and Remediation Optimization Software
MCLs	maximum concentration limits
MLU	Middle Lacrustine Unit
NAVFAC ESC	Naval Facilities Engineering Service Center
NOP	former Nebraska Ordnance Plant
NPL	National Priorities List
NPV	net present value
O&M	operating and maintenance
OU	operable unit
P&T	pump-and-treat
PC	personal computer
PCE	Tetrachloroethylene
ppb	part per billion
PLZ	Permeable Lacrustine Zone
RDX	Royal Demolition Explosive, or Hexahydro-1,3,5-trinitro-1,3,5-triazine

RI	remedial investigation
RMS	Root Mean Square
SO	Sampling Optimizer
TCE	Trichloroethylene
TNB	1,3,5-trinitrobenzene
TNT	2,4,6-Trinitrotoluene
USACE	United States Army Corps of Engineers
US EPA	United States Environmental Protection Agency
UFU	Upper Fluvial Unit
VC	Vinyl chloride
VOCs	volatile organic compounds

ACKNOWLEDGMENTS

The project team consisted of the following organizations:

- Naval Facilities Engineering Service Center (NAVFAC ESC)
  - Karla Harre, Tanwir Chaudhry

- GeoTrans, Inc. (GeoTrans)
  - Rob Greenwald, Weiwei Jian, Yan Zhang

- Summit Envirosolutions, Inc. (Summit)
  - Matt Zavislak, John Dustman
  - Dr. Barbara Minsker (Consultant to Summit)

- Environmetrics & Statistics Limited (EnviroStat)
  - Charles Davis

NAVFAC ESC led the project, including management of project funds, schedule, and deliverables. GeoTrans was responsible for conducting demonstration analysis for selected sites, coordinating preparation of reports, and coordinating site visits. Summit is the developer of the software and offered training, technical support, and assistance regarding preparation of the reports. EnviroStat prepared data for optimization analysis as well as artificial anomalies for testing of Data Tracker, provided technical support regarding validation of the results, and participated in report preparation and review.

In addition, we would like to acknowledge the participation of the following individuals and/or organizations:

- Dave Becker, US Army Corps of Engineers, Environmental and Munitions Center of Expertise (Omaha) (EM CX) – Mr. Becker participated in numerous conference calls, provided technical support and review of reports, and coordinated the participation of one of the demonstration sites.

- Dave Wilson, USEPA Region V – Mr. Wilson coordinated an additional application of the software by an EPA contractor at a site in Region V, and provided feedback regarding that effort to our project team.

- Mindy Vanderford, GSI Environmental, Inc. – Ms. Vanderford provided guidance and feedback regarding the application of the MAROS software at one of the demonstration sites.

We would also like to thank all of the site personnel and contractors associated with the three demonstration sites associated with this ESCTP project. The demonstration sites were as follows:

- Former George Air Force Base Site, Victorville, CA (GAFB site)
- Former Nebraska Ordnance Plant Site, Mead, NE (NOP site)
- Camp Allen Landfill Site, Norfolk, VA (Camp Allen site)

Their time and efforts are greatly appreciated.

EXECUTIVE SUMMARY

The primary objective of this ESTCP project was to demonstrate and validate the use of the "Sampling Optimizer and Data Tracker" software (the "Summit Software"), offered by Summit Envirosolutions, at three DoD sites. The three demonstration sites were as follows:

- Former George Air Force Base Site, Victorville, CA (GAFB site)
- Former Nebraska Ordnance Plant Site, Mead, NE (NOP site)
- Camp Allen Landfill Site, Norfolk, VA (Camp Allen site)

MAROS was also applied at one of the three demonstration sites. The Summit Software demonstrated in this ESTCP project provides a set of tools for Long-Term Monitoring Optimization (LTMO), and consists of two major modules:

- Sampling Optimizer (SO) identifies redundant sampling locations (spatial optimization), or redundant locations and frequencies (spatiotemporal optimization), in historical data.
- Data Tracker (DT) allows current monitoring data to be reviewed against selected historical data (i.e., the "background data") to identify cases where current data deviate from expectations that are based on the background values and patterns.

Model Builder is an additional component within the software with two functions: one for model fitting, visualization, and analysis (with kriging or inverse distance weighting); and another for visualizing relative uncertainty.

Some of the advantages of the Summit Software demonstrated in this ESTCP project are listed below:

- The Summit Software is the only user-friendly software available (i.e., not a research code) that performs monitoring optimization with mathematical algorithms that provide optimal or near-optimal solutions with high probability.
- A major advantage of the optimization approach utilized in the Summit Software is that it allows sampling redundancy to be evaluated on a system-wide basis, identifying optimal solutions with one, two, three, etc. locations removed, rather than on a well-by well basis such as the redundancy analysis employed in MAROS.

- It is the only LTMO software available that enables users to select multiple site-specific monitoring objectives for the redundancy analysis, thus allowing the tradeoff between the number of samples and the resulting error to be rigorously evaluated.

- Visualizations of the plume for the baseline plan with all samples versus improved plans with reduced number of samples are created within the software, which is not the case with MAROS.

- It is the only LTMO software currently available to incorporate data tracking capabilities to automatically identify unexpected values in recently collected data.

Key results of the project include the following:

- The software was found to be easy to learn and use for a typical DoD analyst or contractor with some experience in monitoring systems, and no bugs or software errors are apparent.

- Kriging interpolation with quantile data transformation, which is one of six modeling combinations available in the software, qualitatively provided the best representation of the plumes.

- Sampling Optimizer provided useful trade-off curves of sampling cost versus the interpolation error that resulted from removing samples.

- For the two sites where spatiotemporal analysis was performed, the potential savings achieved with spatial optimization was far greater than the potential savings achieved with spatiotemporal optimization (no spatiotemporal analysis was performed for the NOP site). Potential savings from spatial analysis, calculated based on number of wells eliminated, ranged from approximately 10% to approximately 67%. A value of approximately 35% appears to be representative. The potential savings from the spatiotemporal analysis, based on number of samples eliminated per year, only ranged from approximately 4% to approximately 17% for the two sites where spatiotemporal analysis was performed. Reasons why overly conservative results may be obtained from the spatiotemporal analysis are discussed in the report.

- The software allows redundancy analysis to be preformed simultaneously for multiple contaminants of concern, and based on the testing that was performed it appears that simultaneous evaluation is preferable to evaluating multiple constituents independently.

- For each demonstration site, the results of the redundancy analysis were validated using the most recently collected data that were reserved for this purpose. For one of the sites (GAFB) additional validation was performed based on data from two additional sampling rounds collected subsequent to the initially reserved data. The general process for

performing the validation was to use the reserved data to make plume maps using the baseline well locations (based on the underlying model from Model Builder), and then to also make plume maps using the reserved data only at the wells in the optimized sampling plans. These maps could then be compared to evaluate if the plume maps based on the optimized sampling plans are reasonable. This validation exercise provided confidence in the results provided by Sampling Optimizer at all three demonstration sites.

- The software includes functionality for tracking relative mass of a contaminant over time, and/or mass flux of a contaminant across a use-specified boundary over time. However, if distribution of sampling locations is not consistent between events, the results of relative mass or mass flux calculations provided by the software may be questionable, for reasons discussed in the report. This limitation also exists for MAROS.

- The relative uncertainty maps provided by Model Builder were not found to be particularly useful because they do not help identify what levels of relative uncertainty are acceptable or are of concern, and also do not address whether one value of uncertainty is acceptable in an area of high concentration but not acceptable in an area of low concentration. It is also not clear how these maps can be used to identify how many new wells might be needed to reduce the uncertainty to an "acceptable" degree, and where to locate those wells. This last item is also a limitation of MAROS.

- Data Tracker identified as "out-of-bounds" the vast majority of artificial anomalies added by EnviroStat for testing of this module, and also identified some actual anomalies.

Based on the application of the software at multiple demonstration sites during this project, some limitations of the software have been identified that could potentially be mitigated by future software improvements, including the following:

- The software interpolates spatially but does not perform interpolations in time, which impacts tracking of mass and/or mass flux and also impacts spatiotemporal redundancy analysis when the sampling locations are not consistent from event to event. The software would be improved if there was a feature to optionally fill in missing values via temporal interpolation.

- In DT, the plots of concentration versus time do not use different symbols to differentiate between the "background data" and the "current data". The software would be improved if different symbols were used.

- In DT, the software does not allow specific historical values to be imported and plotted on graphs but not used for calculation of the prediction limits. The software would be improved if such values could be imported with a flag so that they can be included on concentration versus time plots (with a different symbol) but not used to calculate the

prediction limits.

- The DT portion of the software does a very good job of identifying unexpected values, but does not indicate whether the concentration trend for a specific COC at a specific well is increasing, stable, or decreasing (i.e., as is provided by MAROS). The software would be improved if that functionality was added.

- The software does not include data consolidation or recognition of flags (e.g., for non-detect values). This requires the user to consolidate the data into sampling events during preparation of the SO input files, and to assign "graphing values" for non-detects during preparation of the input files for SO and DT. The software could be improved if this type of functionality was included within the software.

The level of effort and computation time for applying the software at the demonstration sites, and a basis for estimating the costs of applying the software at other sites, are provided in this report. Individual reports detailing the application of the Summit Software at the three demonstration sites are presented as appendices to this report. A User's Guide for the software was finalized as part of this project and was submitted as a separate deliverable to ESTCP. The software and User's Guide are now available for use at government sites by government personnel and their contractors, which was an additional component of this ESCTP project.

1.0 INTRODUCTION

1.1 BACKGROUND

Long-term monitoring (LTM) provides a mechanism for evaluating performance of groundwater remedies and is essential for ensuring protection of human health and the environment. The costs of future monitoring are expected to be substantial, since LTM generally spans many years and is required at a large number of sites. Efficiency of LTM can be improved by the following:

- obtaining only the essential data needed for monitoring current conditions by eliminating redundant sampling locations and/or frequencies;
- using an automated approach to identify values from recently collected data that are not within expectations (based on statistical evaluation of previous values); and
- tracking performance relative to specific metrics (e.g., assessing reductions in overall contaminant mass).

This project demonstrates the application of the "Sampling Optimizer and Data Tracker" software offered by Summit Envirosolutions, Inc. (Summit), which is intended to address the items listed above. We sometimes refer to this software as the "Summit Software".

The evaluation of data redundancy in the Sampling Optimizer (SO) portion of the software uses mathematical optimization, which is unique relative to other LTM optimization (LTMO) software products. This allows sampling redundancy to be evaluated on a system-wide basis (e.g., best solution if one location is removed, if two locations are removed, if three locations are removed, etc.). A key benefit of this approach is that it allows the tradeoff between the number of samples and the accuracy of the resulting plume interpolation to be assessed. This is a significant improvement over the approach for evaluating data redundancy utilized in the Monitoring and Remediation Optimization Software (MAROS) software, which is not based on mathematical optimization. In MAROS, individual wells locations are evaluated for redundancy based on impacts of removing that well alone; consequently, the impact of removing *groups* of wells cannot be assessed and the aforementioned tradeoff cannot be evaluated. Another key benefit of the Summit approach for evaluating data redundancy is that plume visualizations for the baseline plan (i.e., all samples) versus improved plans (i.e., reduced numbers of samples) are created within the software. These comparative visualizations are quite effective for communication with stakeholders and regulators.

The "Data Tracker" (DT) portion of the software identifies values from recently collected data that are not within expectations (based on statistical evaluation of previous values). This is somewhat different in approach and implementation versus MAROS, which evaluates concentration trends over time at individual wells as increasing, decreasing, or stable. Rather

than indicate increasing versus decreasing trends, the intent of the DT portion of the software is to automatically highlight which recently collected data values are unexpected and require further attention. These unexpected values may be due to significant increasing or decreasing trends, or may be due to "bad data" (lab error, sampling error, database error, etc.). DT allows for "expected" time trends to be either stable or smoothly decreasing; the latter is appropriate for monitoring an effective passive remediation system, for example.

1.2 OBJECTIVE OF THE PROJECT

The objective of this project is to demonstrate and validate the use of the "Sampling Optimizer and Data Tracker" software (i.e., the "Summit Software") by applying the software at three Department of Defense (DoD) sites. A secondary objective is to compare the results with MAROS at one site. Another component of the project is to transfer the software and documentation to the government for free use at government sites by government personnel and their contractors.

1.3 REGULATORY DRIVERS

There are no regulatory issues directly associated with this effort, although there has been a general focus in recent years regarding optimization of all facets of remediation including LTM. Application of the software demonstrated in this project is intended to improve the efficiency and assessment of the monitoring well networks and data that are collected in current monitoring events, which will ultimately address regulatory objectives and allow for improved communication between all site stakeholders. Implementation of revised sampling plans suggested for the demonstration sites is not within the scope of this project.

2.0 TECHNOLOGY

2.1 TECHNOLOGY DESCRIPTION

The Summit Software is a set of desktop software tools consisting of two major modules, Sampling Optimizer and Data Tracker.

- *Sampling Optimizer* (SO) identifies redundant sampling locations (spatial optimization), or redundant locations and frequencies (spatiotemporal optimization), in historical data. This module identifies redundancies using a multi-objective genetic algorithm (GA) to obtain monitoring designs that represent optimal tradeoffs among two or more monitoring objectives, such as minimizing the number of samples and minimizing the interpolation error at locations that are removed. The error will generally increase as number of wells decreases, resulting in a tradeoff.

- *Data Tracker* (DT) allows current monitoring data to be reviewed against selected historical data (i.e., the "background data") to identify cases where current data deviate from expectations that are based on the background values and patterns.

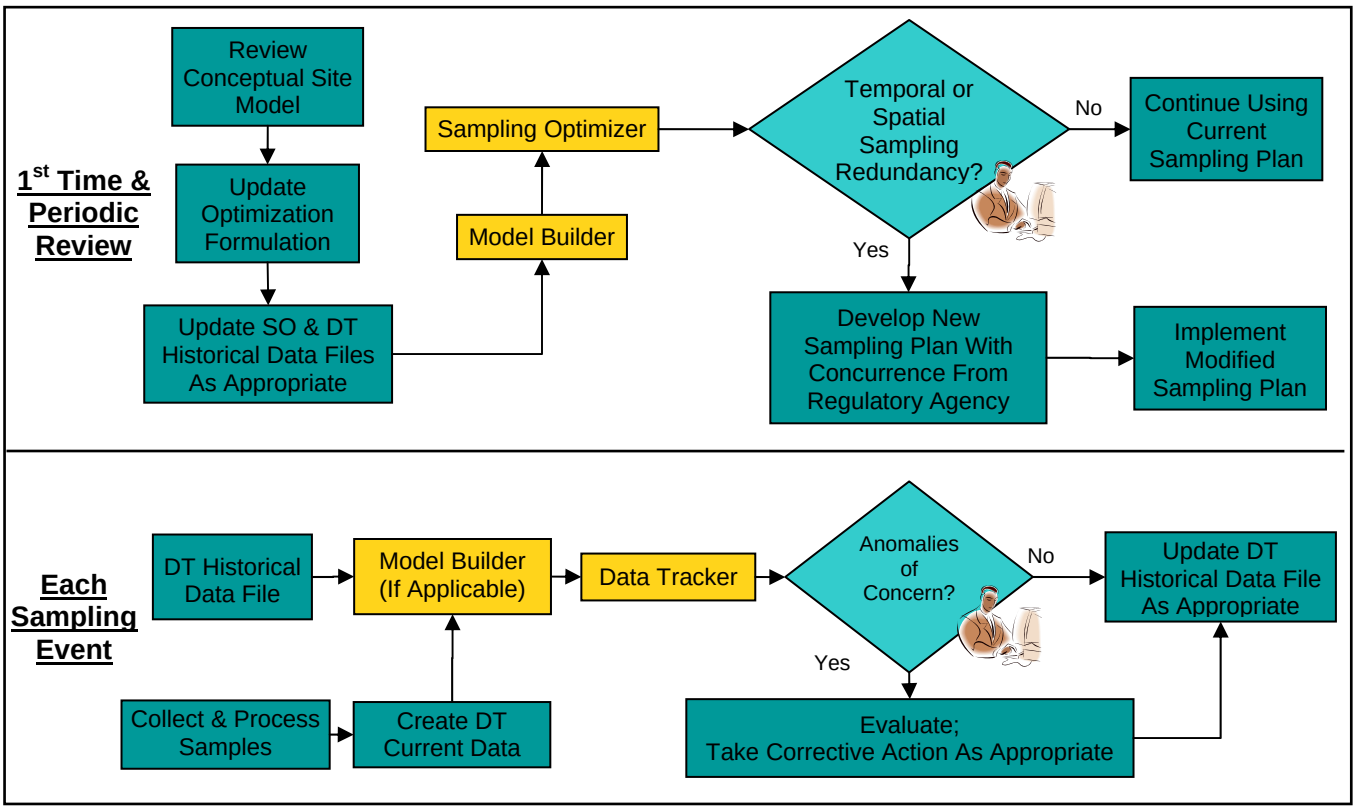
Model Builder is an additional component within the software that is utilized by Sampling Optimizer and, in some cases, by Data Tracker. Model Builder has two sections: one for model fitting, visualization, and analysis (with kriging or inverse distance weighting); and another for visualizing relative uncertainty.

The software module/components listed above (Sampling Optimizer, Data Tracker, and Model Builder) are highlighted on Figure 2-1, which is a general flowchart illustrating the application of the software. The software is applied in several ways:

- During initial optimization and periodic review (upper portion of the figure), Model Builder constructs spatial and/or spatiotemporal models for the measurements of the primary contaminants of concern (COCs). The model identified by Model Builder is then used by Sampling Optimizer to identify optimal sampling plans for subsequent routine monitoring (i.e., with redundancies eliminated).

- After new sampling events (lower portion of the figure) the software can be used during routine monitoring to identify anomalies or departures from expectations. In addition, changes over time in plume mass and/or mass flux across a boundary can be tracked. If the quantity of interest involves modeling for each monitoring event (e.g., contaminant mass based on interpolation) that modeling is provided by Model Builder.

Figure 2-1. General Flowchart of Software Application



Although the software provides specific results as output, there is still an aspect of interpretation required by an analyst, as illustrated on Figure 2-1. With respect to the data redundancy evaluation, the software produces an optimal tradeoff curve (discussed in more detail later), and the analyst must choose specific sampling plans along the tradeoff curve for further evaluation. With respect to the detection of anomalies, the software identifies values in recently collected data that are "out-of-bounds" and provides a graph of concentration versus time for visual review. The analyst must then determine if any response or action is appropriate, such as correction of an erroneous laboratory report or further investigation of a potential new source of contamination. Thus, the corrective action referred in the figure can take many forms depending on the nature of the anomaly.

Additional detail regarding the major modules/components of the software is provided below.

Model Builder

Model Builder is used to perform the geo-statistical functions needed by Sampling Optimizer. It has two major functions:

- model fitting, visualization, and analysis (primary); and
- visualizing relative uncertainty (secondary).

Sampling Optimizer provides users with six possible types of "models", based on the combination of interpolation technique and data transformation:

- Two interpolation technique options
    - Ordinary Kriging
    - Inverse Distance Weighting

- Three data transformation options
    - Quantile
    - Logarithmic
    - None (i.e., no transformation)

Although there are six possible combinations of these options, a user will generally select only one combination; Summit suggests using kriging with quantile transformation. The user can select automated or manual model parameter fitting for either Ordinary Kriging or Inverse Distance Weighting. The automated approach searches through a variety of parameter settings and model configurations to ensure that a good fit to the data is obtained. Regardless of the model type and parameter fitting approach, the user is advised by the software manual to visualize and review the results of the interpolation model used by Model Builder (using plots developed within the software) to assess the reasonableness of the plume representation. This should be done before using that model within Sampling Optimizer.

An additional modeling functionality in Model Builder pertains to calculating "relative mass" and "mass flux" for specific sampling events, based on spatial interpolations:

- For "relative mass", the software calculates mass per unit volume of aquifer. This is useful for comparisons of relative dissolved mass between sampling events, but is not intended to estimate absolute dissolved mass within the plume (because the true volume of water is not rigorously accounted for). The interpolated concentration value at every "cell" in the plume map for that COC at that time period is summed and each cell is approximated to represent the same volume (i.e., incorporating vertical extent and porosity) as every other cell. It would be straight-forward for the software developer to incorporate variations of this approach into the Summit software to meet site-specific objectives.

- For “mass flux”, the software calculates the mass flowing through a cross-section of a site (e.g., across a site boundary). The user defines a cross-section graphically within the software, consisting of several line segments, each of which can have a unique groundwater flow rate. The software multiplies each interpolated concentration value in the cross-section by the specified flow rate, and these individual flux values are summed to create the estimate of total flux travelling through the cross-section.

The “Relative Uncertainty” functionality within Model Builder shows the user an image which illustrates an uncertainty map based on the root-mean-squared (RMS) error value of each pixel in the generated visualizations, based on cross-validation. Note that large differences in concentrations in two nearby wells can lead to high cross-validation errors that may not be a major concern if the plume is thought to be stable and well characterized. If there are locations with high errors where accurate concentration estimates are particularly important (e.g., close to down-gradient exposure locations), these locations may be candidates for additional sampling activities.

Sampling Optimizer

Sampling Optimizer uses a genetic algorithm (GA) to suggest favorable monitoring plan alternatives relative to the base sampling plan (i.e., where one or more of the samples are removed). The software allows spatial optimization and spatiotemporal optimization:

- In *spatial optimization*, the original model is based on one set of sampling data that do not vary in time, and the optimization results are with respect to sampling locations only.

- In *spatiotemporal optimization*, the original model consists of sampling data that vary in space and time, and the optimization results are with respect to sampling location and sampling frequency.

Temporal analysis is a subset of spatiotemporal analysis where well locations cannot be removed.

The GA begins with a randomly-generated set of solutions (in Sampling Optimizer these are called plans), called a population. Solutions from one population are taken and used to form a new population (also called a new generation). This is motivated by a hope that the new population will be better than the old one. Solutions which are selected to form new solutions (offspring) are selected according to their fitness - the more suitable they are the more chances they have to reproduce (fitness is defined in terms of the user-supplied interpolation error metric). After selection, new solutions are created from the selected solutions through operations called mutation and crossover. After the fitness of every solution in a generation has been evaluated, Sampling Optimizer implicitly creates a scatter plot showing the costs and errors for each solution. Sampling Optimizer shows the lower left boundary of the points in the scatter plot, referred to as the Pareto front or tradeoff curve. This plot shows how well the solutions

perform relative to the two objective functions. This process is automatically repeated until an appropriate number of generations have been completed. The user can utilize software defaults for the optimization algorithm (e.g., population size) or can specify values for these parameters in the "GA Settings" screen.

The flow of the GA approach is presented in Figure 2-2. Redundant sampling locations and frequencies are identified using NSGA-II (Nondominated Sorted Genetic Algorithm-II) (Deb, et al. 2002), a well-tested multi-objective optimization algorithm. Genetic algorithms mimic the mechanisms of natural selection in searching for optimal solutions (Goldberg, 1989).

Figure 2-2. Overview of the Genetic Algorithm Approach

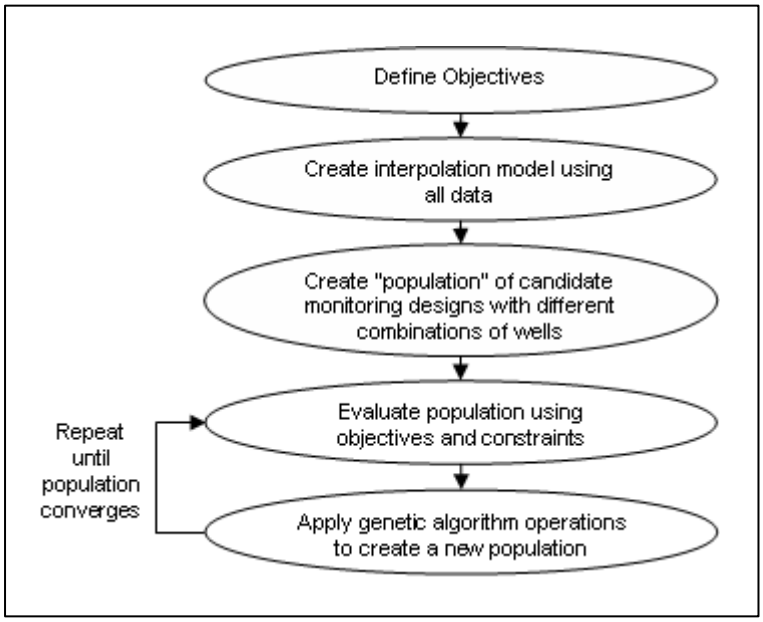


Figure 2-3 presents a conceptual example of a tradeoff curve. In the figure, "Maximum Error" refers to the highest interpolation error that occurs at any location removed from the baseline sampling plan. Each diamond represents the optimal monitoring design for a given level of expenditure (i.e., number of wells). Such solutions are optimal (non-dominated) because no other possible solution is superior in both objectives simultaneously. The advantage of using GAs to solve multi-objective problems is that the entire set of such solutions can be generated in a single optimization run. Figure 2-3 also shows the point of diminishing returns, where additional sampling expenditures result in error reductions that are relatively insignificant. In this example, designs with more than 70 wells yield minimal reduction in error. These figures allow managers to view and quantify the costs and benefits of additional sampling and identify the design that best balances these objectives for their purposes.

Figure 2-3. Example of a Tradeoff Curve for Spatial Optimization

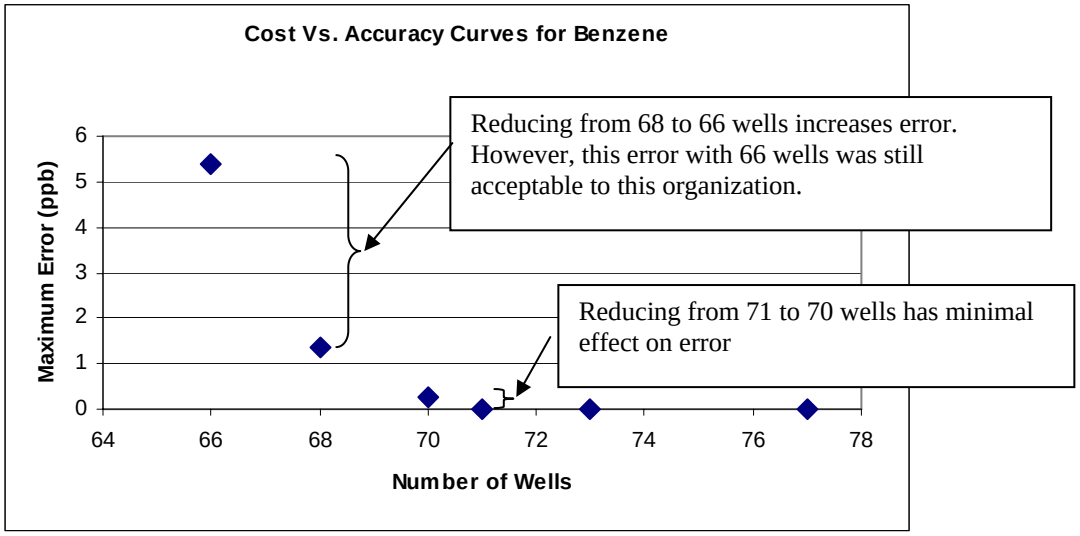
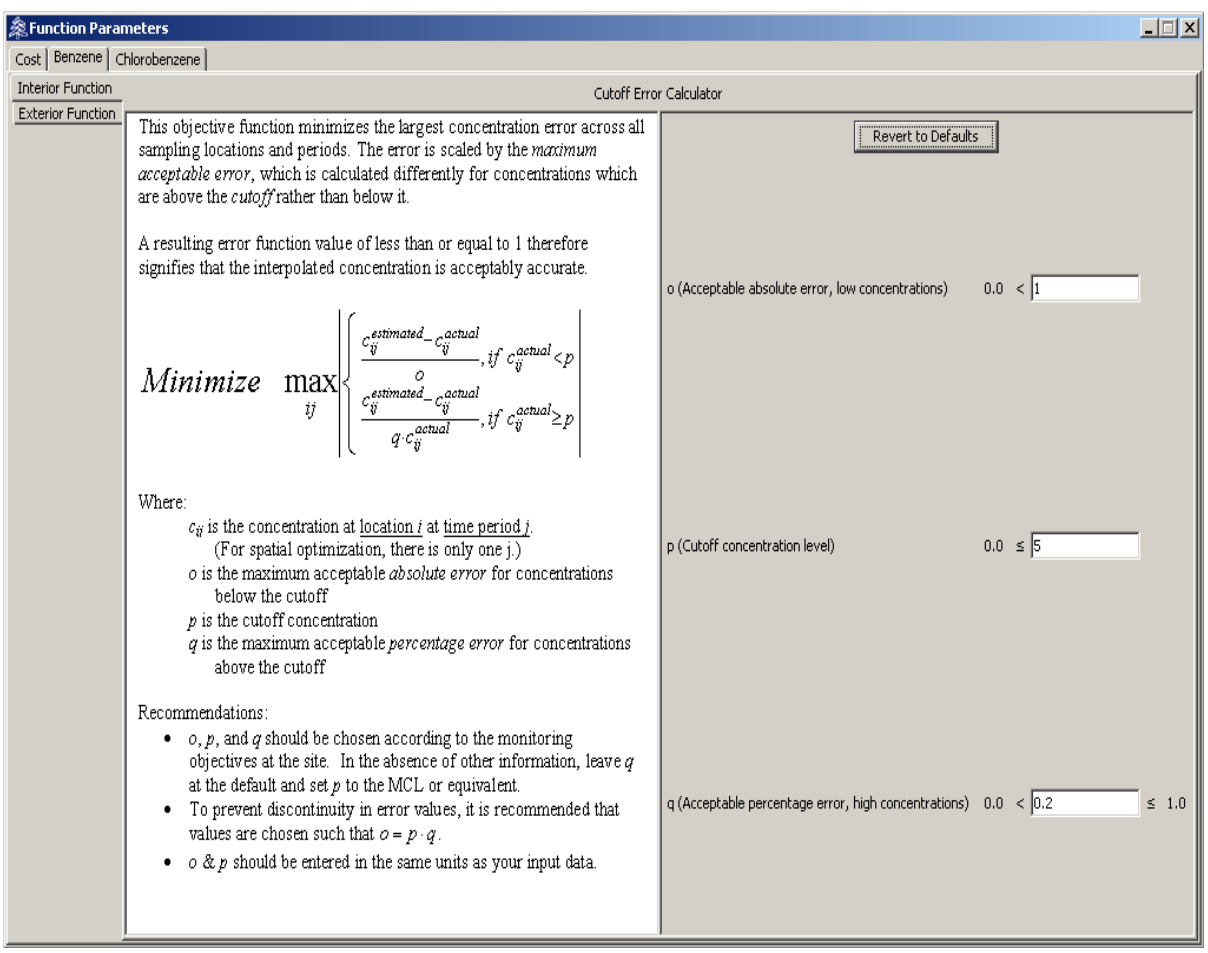


Figure 2-3 is a generic example of a tradeoff curve, where the y-axis represents "error" in units of concentration. The Summit Software actually calculates the "error" as a dimensionless parameter, using a tool called the "Cutoff Error Calculator", which is presented in Figure 2-4.

Figure 2-4. Cutoff Error Calculator (From Software Manual by Summit Envirosolutions, 2008)



The overall purpose of the Cutoff Error Calculator is to come up with an error value that represents the overall similarity of a new sampling plan to the baseline sampling plan. The Cutoff Error Calculator is designed so that error is calculated in a manner that makes deviations between interpolated and actual values more significant in areas of low concentration versus areas of high concentration. This is accomplished as follows:

- The user defines a cutoff concentration (p) for the actual data values that differentiates between "low" and "high" concentrations, and also defines a value for acceptable absolute error (o) for low concentrations.

- When a low concentration data point is removed, error is calculated as the absolute value of the actual value minus the interpolated value, divided by the acceptable absolute error. For example, if the actual value is 4 µg/l (below the cutoff concentration of 5 µg/l), the acceptable absolute error is 1.0, and the interpolated value is 9 µg/l, then the difference between the actual and interpolated value is 5µg/l and the error would be 5 / 1 = 5.

- When a high concentration data point is removed, error is calculated as the absolute value of the actual value minus the interpolated value, divided by a user-supplied percentage (q) of the actual value. For example, if the actual value is 100 µg/l (above the cutoff concentration of 5 µg/l), the percentage input by the user is 20%, and the difference between the actual and interpolated value is 5 µg/l, then the error would be 5 / (0.20 * 100) = 0.25.

Low Concentration	*High Concentration*
actual value = 4 {below cut-off}	*actual value = 100 {above cutoff}*
cutoff value = 5	*cutoff value = 5*
estimated value = 9	*estimated value = 95*
acceptable error = 1	*acceptable percentage error = 20%*
error = |(9 – 4)| / 1 = 5.0	*error = |(100 – 95)| / (100 * 20%) = 0.25*

In these examples, the difference between the actual value and the interpolated value was 5 µg/l in both cases, but in the first case the calculated error is 5.0 whereas in the second case it is only 0.25. This illustrates how the calculation increases the significance of deviation between actual and interpolated values in the lower concentration areas of the plume. Note that with the values of o, p, and q shown in Figure 2-4, if the actual concentration equals the cut-off value p, then the calculated error is the same whether the low concentration or high concentration formula is used. This is desirable for continuity, although the user can supply a value of q that would make the error calculation discontinuous with respect to true concentration.

Data Tracker

Data Tracker reviews current monitoring data against selected historical data (i.e., the "background data") and identifies cases where current data deviate from expectations that are

based on the background dataset. These expectations are formulated as prediction limit bounds, a statistical estimate of the range of values that might be expected in a specific sampling event, based on the background data at that location. Any current data value which is outside the bounds range for that location is reported to the user as being out-of-bounds. DT includes two types of bounds calculators: static and time-dependent decreasing. The software automatically makes a recommendation for the bounds type to be used based on the background data. The default probability content for the prediction limits is 95%; the test for existence of a decreasing trend uses a default significance level of 5%. Advanced users can change these parameter values.

2.2 TECHNOLOGY DEVELOPMENT

The Summit tools were built upon technology development and research at the University of Illinois from 1997 – 2004. Reed et al. (2003), with funding from an Environmental Protection Agency STAR Fellowship, created an automated methodology for setting parameter values for the Nondominated Sorted Genetic Algorithm-II (NSGA-II). A Technology, Research, Education, and Commercialization Center (TRECC, http://www.trecc.org/) project funded by the Office of Naval Research implemented NSGA-II with the automated parameter-setting methodology in the Data to Knowledge (D2K) software development and data mining framework created by the National Center for Supercomputing Applications (NCSA). The resulting software, called Evolutionary Multi-objective Optimizer (EMO), was further developed with funding from BP/Atlantic Richfield to Barbara Minsker, Riverglass Inc., and Hazard Management Systems Inc. (HMSI). HMSI created Sampling Optimizer as a specific application of EMO to long-term monitoring optimization, along with Data Tracker.

Two case studies using Sampling Optimizer were completed in July 2004. Site A had 36 sampling locations and the optimization focused on BTEX characterization, while Site B had 80 sampling locations and focused on benzene characterization. Each study identified roughly 23% redundancy in the sampling locations eligible for removal.

HMSI was purchased by Summit Envirosolutions in December 2006, who have continued to improve and develop the software and documentation with support from BP and this ESTCP project. Summit is currently performing additional several case studies to further evaluate and demonstrate the effectiveness of the software, and summaries of these results will be posted on the Sampling Optimizer website.

2.3 ADVANTAGES AND LIMITATIONS OF THE TECHNOLOGY

Related software and methods for LTM design exist, as follows:

- Geostatistical packages are widely available, including Surfer and Geo-EAS, but these packages create only interpolation models and do not perform data tracking or

optimization.

- Based on current information, GTS (Cameron and Hunter, 2002), the 3-tiered monitoring optimization approach by Parsons (Nobel, 2003), MAROS (Aziz, et al., 2003), and Cost Effective Sampling (Johnson, et al., 1996) perform various spatial and temporal redundancy analyses for LTM. However, they do not perform data tracking for site-wide targets and do not use mathematical optimization. Instead, they use heuristic ("rule-of-thumb") approaches for identifying which samples to remove, which may not identify the optimal sampling plan to best meet the site-specific objectives. Moreover, most of these methods are not yet available as supported software packages (with the exception of MAROS and a limited version of GTS). These methods do not optimize based on removing groupings of samples, and do not yield tradeoff curves based on the results of multi-objective mathematical optimization.

- Herrera and Pinder (1998) and Rizzo et al. (2000) have used Kalman filters for LTMO and model updating. These approaches require the use of transport models for the analysis, which most DOD sites do not have. When transport models are available, Kalman filters can use available data to update the models as new data become available. Herrera and Pinder's approach uses this capability to identify the next location that should be sampled, selecting the location with the most uncertainty in the model predictions, after each sample is collected with an event. This approach assumes that data at a site are collected sequentially, with enough time between each sample to analyze the previous result and use it to determine the next sample location, which is not always the case. Additionally, the sequential sampling approach is not a global algorithm that identifies the best set of locations to maximize the overall reduction in uncertainty. Rizzo et al. couple the Kalman filter with simulated annealing, a global optimization approach, but to our knowledge is not available as a software package.

Some of the advantages of the Summit Software demonstrated in this ESTCP project are listed below:

- The Summit Software is the only user-friendly software available (i.e., not a research code) that performs monitoring optimization with mathematical algorithms that provide optimal or near-optimal solutions with high probability.

- A major advantage of the optimization approach utilized in the Summit Software is that it allows sampling redundancy to be evaluated on a system-wide basis, identifying optimal solutions with one, two, three, etc. locations removed, rather than on a well-by well basis such as the redundancy analysis employed in MAROS.

- It is the only LTMO software available that enables users to select multiple site-specific monitoring objectives for the redundancy analysis, thus allowing the tradeoff between the number of samples and the resulting error to be rigorously evaluated.

- Visualizations of the plume for the baseline plan with all samples versus improved plans with reduced number of samples are created within the software, which is not the case with MAROS.

- It is the only LTMO software currently available to incorporate data tracking capabilities to automatically identify unexpected values in recently collected data.

On the other hand, the Summit Software does not incorporate some features in other software, including: (1) the Summit Software has fewer types of interpolation models than most available geostatistical packages (although the interpolation models available in the Summit Software are well accepted); (2) the data tracking routine in the Summit Software does not include trend analysis (such as performed by MAROS) to indicate if concentration trends at individual wells are increasing, decreasing, or stable; (3) the Summit Software does not incorporate transport models as the Kalman filter approach.

Although the Summit Software performs spatiotemporal optimization, it does so using sequential spatial interpolation and does not include a strictly intra-well temporal interpolation, which likely makes the results of the spatiotemporal optimization overly conservative (this is discussed in more detail later in the report). Another limitation is that the mass and mass flux tracking features in the software are impacted by differences in the number of sampling locations in each event. This is related to the fact that the software does not perform temporal interpolation or extrapolation to fill in missing values in events where specific wells are not sampled. This also will be discussed in more detail later in this report. It is noted that MAROS has the same limitation in this regard.

The Summit Software is expected to provide the following benefits:

- Significant cost savings are expected by eliminating redundant data collection. Costs could be reduced for sampling labor, laboratory analysis, data management, and other elements. Experience suggests that eliminating redundant sampling points can save $500-$1,000 or more per sample in labor and analysis. Additionally, the Sampling Optimizer will help managers identify tradeoffs among multiple monitoring objectives, including identifying when further monitoring expenditures will likely result in minimal benefits.

- New data can be given an initial assessment for significant deviations and other features of interest without substantial labor. Currently, at most sites even a quick visual scan for a few constituents at a few key wells can require several hours for an analyst. Thorough statistical tests could take weeks of labor, but can be readily performed by Data Tracker. This may enable earlier detection and correction of potential problems and/or faster identification of significant changes in the physical system, which results in higher certainty of attaining protectiveness. This benefit will become even greater as emerging sensor technologies produce larger volumes of data to be analyzed and/or more facilities move into LTM in post-closure and/or passive remediation scenarios.

The Summit Software is designed for sites meeting the following criteria which represent limitations to application of the software; other approaches or software products designed for similar purposes will undoubtedly have similar limitations:

- The user defines a cutoff concentration (p) for the actual data values that differentiates between “low” and “high” concentrations, and also defines a value for acceptable absolute error (o) for low concentrations.

- First, for Model Builder and Sampling Optimizer to provide reliable results the site should be in a LTM situation, which may include ongoing active or passive remediation. This implies that the groundwater chemistry should be expected to change smoothly if at all in the foreseeable future, whether or not remediation activities are underway, and that no new sources of potential releases are anticipated.

- The site should have an adequate data history available that is representative of its current groundwater chemistry and status. If spatiotemporal redundancy evaluation using Model Builder and Sampling Optimizer is to be performed, a site should generally have at least four observations obtained over a period of at least two years at most wells under consideration (spatial redundancy evaluation has no such requirement), and eight observations is preferred. Strictly speaking, when using Sampling Optimizer for spatial optimization, only one data value per well for each primary COC is needed for using the software, but a more extended data history is preferable to verify the presence of the requisite LTM situation. For DT the software requires a minimum of four background observations per COC per well.

- The specific number of wells needed depends on the site complexity. In general there should be at least twenty monitoring wells in order to anticipate significant cost reductions from a spatial redundancy analysis (otherwise a significant percentage reduction in sampling locations would only yield a modest cost savings). If one is interested only in using Data Tracker, an adequate data history for each constituent of interest at each well is needed (at least four previous samples to use as background data), but there is no requirement for a minimum number of wells.

- There will generally be one or at most a few primary COCs with respect to which the modeling and optimization for redundancy are performed. Data Tracker is not limited to the COCs used by Model Builder and Sampling Optimizer.

- For efficient use of Data Tracker, future routine monitoring data should be made available in compatible electronic form. This may require the construction of a “data bridge” mechanism is (i.e., a program that reads data in one format and rewrites the data into a desired format).

- One of the assumptions for applying the Summit Software is there are no major discontinuities in the specific aquifer being evaluated with respect to hydraulic connection. Highly fractured media would not meet that assumption. This limitation would of course be true of LTMO software in general, not just the Summit Software. Similarly, sites with extremely large contrasts in hydraulic conductivity (e.g., preferential pathways) might impact the application of the software.

The limitations listed above pertain to the overall technology. Several additional limitations within the software functionality observed during testing are discussed later in the report.

3.0 PERFORMANCE OBJECTIVES

This section provides a summary of the performance objectives stated in the Technical Demonstration Plan for this project, including a conclusion as to whether or not the performance objective was met.

Table 3-1 summarizes the performance objectives for evaluating the Summit Software provided in the Technology Demonstration Plan. To avoid repetition, a detailed discussion is provided for each performance objective in Section 6.0 that explains the criterion, how it was assessed, and the basis for the assessment. Those items are best explained in Section 6.0, after the discussion of testing design and results which are presented in Section 5.0.

Table 3-1. Performance Objectives

Type of Performance Objective	Performance Criteria	Expected Performance (Metric)	Performance Objective Met?
Qualitative	User functionality (primary)	The Summit Software has an acceptable learning curve (e.g., 1-2 days) that will not discourage prospective users and allow users to achieve the intended objectives.	**YES**
	Software reliability (primary)	The Summit Software has no significant errors or bugs remaining by the end of this project.	**YES**
	Model Builder performance (primary)	Model Builder provides a model of spatial and/or temporal variation for each primary constituent of concern at each site that is adequate given the available data.	**YES (spatial)** **PARTIALLY (temporal)**
	Sampling Optimizer performance (primary)	Sampling Optimizer provides reasonable trade-off curves allowing site personnel and other professionals to easily identify optimal monitoring program choices.	**YES (spatial)** **PARTIALLY (spatiotemporal)**
	Data Tracker performance (primary)	Data Tracker enables the easy incorporation of site-specific monitoring and remediation expectations and data objectives along with historical data.	**PARTIALLY**
	Regulatory acceptance (primary)	Results of this ESTCP dem/val will be persuasive to regulatory personnel.	**YES**

Type of Performance Objective	Performance Criteria	Expected Performance (Metric)	Performance Objective Met?
	Comparison with MAROS (secondary)	The Summit Software will be found to be at most modestly more difficult to learn to use, consistent with being much more flexible in incorporating site-specific monitoring objectives. There are no pre-conceptions regarding the comparisons of optimization recommendations to be expected from the two software products.	**YES**
Quantitative	Model Builder performance (primary)	Model Builder provides a model of spatial and/or temporal variation for each primary constituent of concern at each site that is adequate given the available data.	**YES (spatial)** **PARTIALLY (temporal)**
	Sampling Optimizer performance (primary)	Optimized programs identified by Sampling Optimizer in fact permit cost reductions with acceptable losses of information, if appropriate, as anticipated for the large majority of DoD sites.	**YES**
	Data Tracker performance (primary)	Data Tracker responds appropriately to artificially induced anomalies of interest for the particular site.	**YES**
	Comparison with MAROS (secondary)	The Summit Software will be found to be at most modestly more difficult to learn to use, consistent with being much more flexible in incorporating site-specific monitoring objectives. If both products are able to accept the same goals and constraints, results will be similar but slightly different due to small differences due to different optimization methodologies. There are no pre-conceptions regarding the comparisons of optimization recommendations to be expected from the two software products.	**YES**

4.0 SITE DESCRIPTION

Three DoD demonstration sites were selected. Potential sites were initially screened to meet criteria for data history and monitoring network size:

- Data history criteria – A minimum eight monitoring events over at least a two-year period prior to 2006 for most monitoring wells.

- Monitoring network size criteria – At least 20 existing monitoring wells that have been sampled regularly.

In selecting the sites, the project team strove for variety in terms of hydrogeology, nature and extent of contamination, size of the monitoring program, and amount of data history available. There was also a preference, if possible, to select each site from a different agency within the DoD. Furthermore, there needed to be an expressed willingness of the site team to participate in the effort and consider implementation of results. Table 4-1 provides a summary of the demonstration sites.

Table 4-1. Summary of Demonstration Sites

	Norfolk Naval Station Camp Allen Landfill	**Former George AFB (GAFB) OU1**	**Former Nebraska Ordnance Plant (NOP) OU2**
Agency	Navy	Air Force	Army
Location	Norfolk, VA	Victorville, CA	Mead, NE
Geographic Location	East (coastal)	West (arid)	Midwest (plains)
Remediation System	P&T with air stripping for hydraulic containment	P&T started in 1991 and shut down since 2003	P&T with 10 extraction wells
Primary COCs	c12DCE, TCE, VC	TCE	TCE and RDX
Aquifers Evaluated	Shallow and deep aquifers	Upper aquifer	Shallow, intermediate, and deep aquifers
Sampling Frequency	Annual	Semi-annual	Varies by well
Monitoring Network	~70	~50	~220

Figures regarding site location, stratigraphy, and contaminant plumes that are presented in the sections below, for each of the three demonstration sites, are taken from site reports that were provided to the ESTCP project team.

One of the assumptions for applying Model Builder and Sampling Optimizer is that there are no major discontinuities in the specific aquifer being evaluated with respect to hydraulic connection. For instance, highly fractured media would not meet that assumption. At the former GAFB site, there is a natural plume discharge boundary at the eastern edge of the OU-1 Upper Aquifer that is conceptually similar to any other discharge location such as a stream or lake, and that type of boundary does not preclude the application of the software.

4.1 SITE LOCATION AND HISTORY

Camp Allen Landfill, Norfolk, VA ("Camp Allen Site")

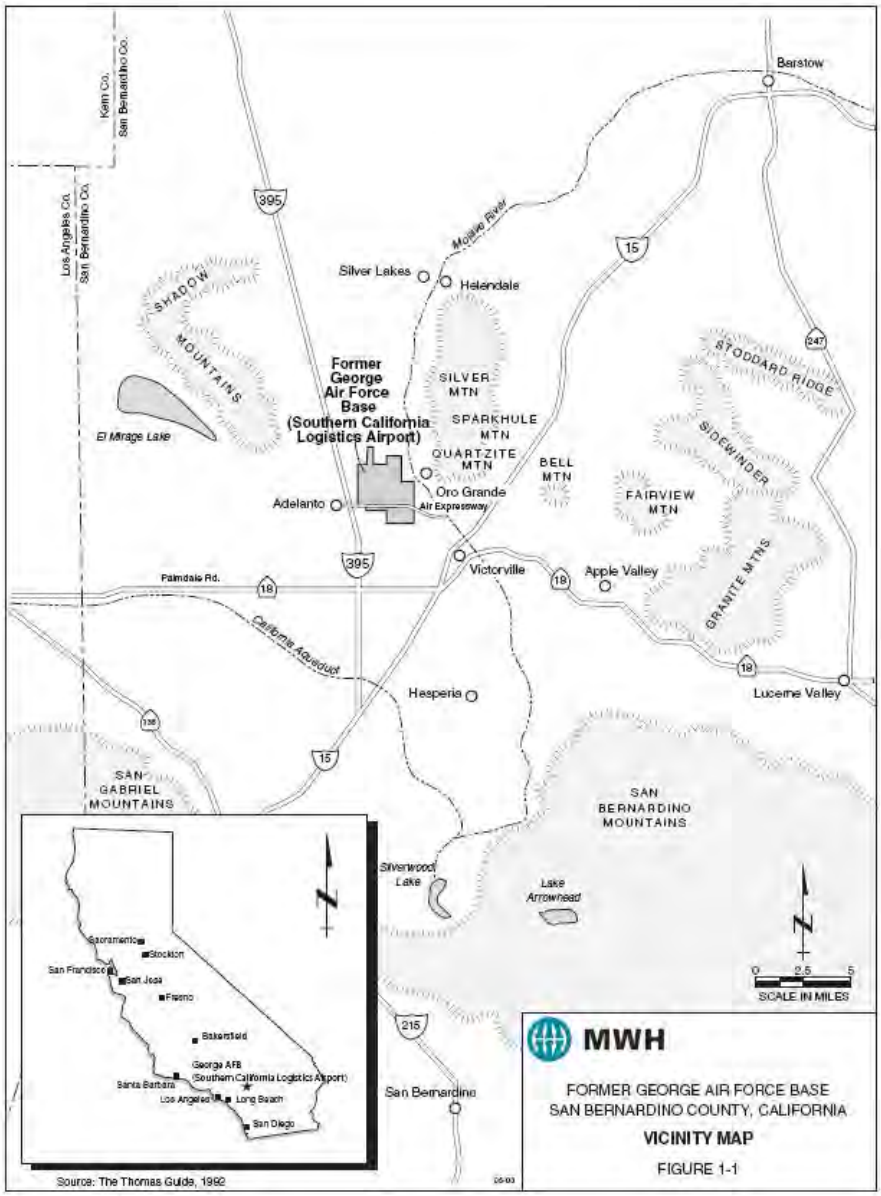
The Camp Allen Landfill is located at the Naval Base Norfolk, VA (Figure 4-1), and is comprised of Landfill Area A (approximately 45 acres), Landfill Area B (approximately 3 acres), and a Salvage Yard located in between. The Site is located in mixed-use urban land. Military facilities are located atop and adjacent to the landfill areas. Area A of the Site was used for the disposal of a variety of wastes. During the early 1940s, landfill operations commenced at the Camp Allen Landfill and continued until about 1974. Portions of the landfill now accommodate the Navy Brig Facility and a heliport. Area B was used to dispose residue and debris resulting from a fire at the Camp Allen Salvage Yard.

Figure 4-1. Location of Camp Allen Site, Norfolk, VA

Former George Air Force Base, Victorville, CA

The former GAFB is located in San Bernardino County, California, approximately 70 miles northeast of Los Angeles, in the Victor Valley portion of the Upper Mojave River Basin (Figure 4-2). This site sits atop the Mojave River Bluffs on the west side of the Mojave River. GAFB was established in the early 1940s and was operational until the early 1990s. Under the Federal Facilities Agreement (FFA), three Operation Units (OUs) were defined. OU-1 consists of the Upper and Lower Aquifer groundwater contaminated with TCE beneath the northeast portion of the base and adjacent off-base areas, and that OU is the focus of this effort.

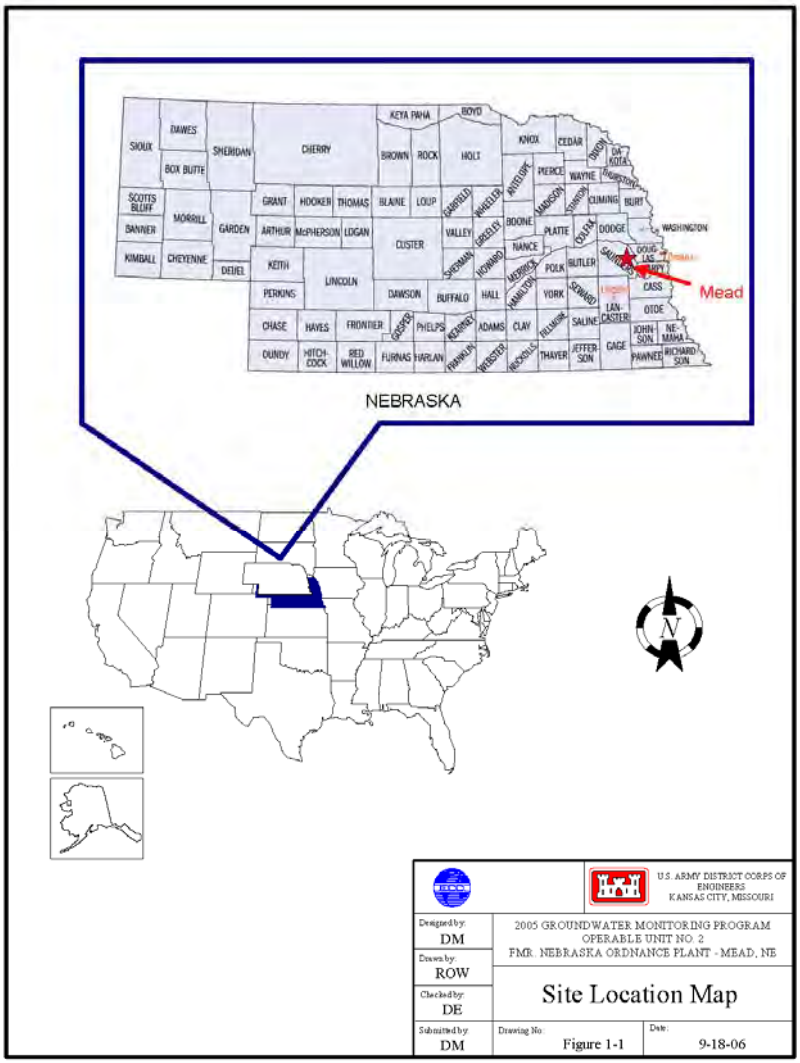
Figure 4-2. Location of GAFB, San Bernardino County, CA



Former Nebraska Ordnance Plant (NOP), Mead, NE

The former NOP occupies approximately 17,250 acres located 0.5 miles south of the town Mead, Saunders County, Nebraska (Figure 4-3). The Site is nearly flat, with a few gentle slopes. Surface water drainage in the eastern portion of the site is generally to the southeast. In the western portion of the site, surface water drains to the southeast, via Silver Creek. During World War II and the Korean Conflict, bombs, shells, and rockets were assembled at the site. The site includes four load lines (LL1 is furthest west and LL4 is furthest east), where bombs, shells, and rockets were assembled; the Burning/Proving Grounds; a Bomb Booster Assemble Area; Administrative Area; an Air Force Ballistic Missile Division Technical Area; and an Atlas Missile Area. According to previous reports, wastewater with explosives from both the load line plant operations and a laundry was discharged into a series of sumps, ditches, and underground pipes. TCE was released from various sources including the Atlas missile site. The site was placed on the U.S. Environmental Protection Agency (EPA) National Priorities List (NPL) of Superfund sites in August 1990 because contamination was identified in the groundwater and the soils at the site, and the releases of contamination from this site is considered to be a potential threat to public health, welfare, and the environment.

Figure 4-3. Location of NOP, Mead, NE



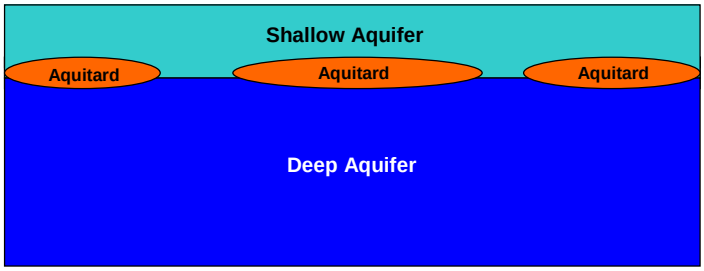
4.2 SITE GEOLOGY/HYDROGEOLOGY

Camp Allen Landfill, Norfolk, VA

The Camp Allen site and surrounding area can be characterized as a former tidal flat associated with the Bousch Creek drainage channel. The area was developed from marine sediments whose major constituents include sands, silts, and clays with considerable amounts of shell material and gravel. The uppermost geologic unit and youngest formation is the Columbia Group; its average thickness ranges from 20 to 50 feet. The unconsolidated sediments are characterized by light-colored clay, sand, and silt. Surficial soils are primarily silts and clays that quickly grade into sands and silts of the Columbia Group. The Yorktown Formation underlies the Columbia Group, and is characterized by coarse sand, gravel, and abundant shell fragments. Regionally, the Yorktown Formation ranges in thickness from 300 to 400 feet.

Two aquifer systems are impacted by the Camp Allen Landfill: the Columbia Group (shallow aquifer), and the Yorktown Formation (deep aquifer). A schematic cross section is presented in Figure 4-4. The shallow aquifer is unconfined. The deep aquifer is separated from the shallow aquifer by a confining clay unit. In the Camp Allen area, a breach and/or ineffective portion of the confining clay unit allows downward migration of constituents from the shallow aquifer to the deep aquifer. Groundwater flow patterns at this site are complex.

Figure 4-4. Schematic Cross Section, Camp Allen Site



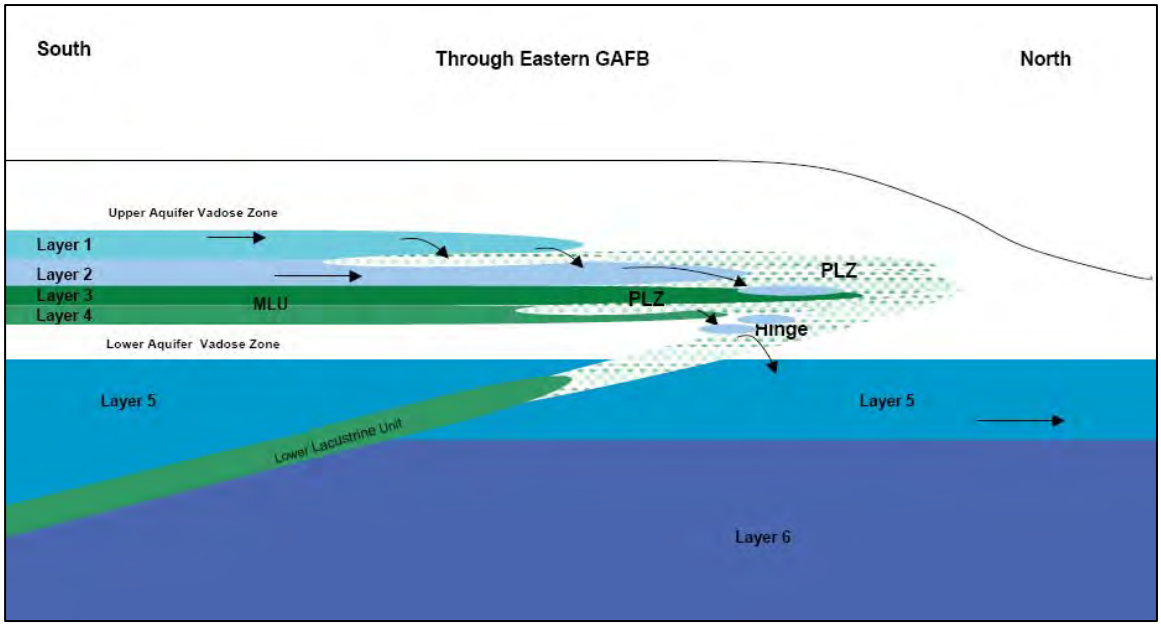
The site team indicated that the plume boundaries have been stable for the past 3-4 years. The remedial measures at the Camp Allen Site consist of groundwater extraction from both aquifers (up to 15 extraction wells total, in both Area A and Area B) and treatment via air stripping. There are approximately 110 monitoring wells in total with some of them used only for water level measurements. Approximately 50 monitoring wells are currently sampled for groundwater quality assessment, of which 45 wells are sampled annually.

GAFB, Victorville, CA

The sediments beneath the site have been divided into four primary units based on their hydrogeologic characteristics. They are the Upper Fluvial Unit (UFU), Middle Lacustrine Unit (MLU), Lower Lacustrine Unit (LLU), and the Lower Alluvial Unit (LAU) for the top to bottom unit respectively. The UFU and LAU contain aquifers termed the upper aquifer and the lower

aquifer, respectively. The upper aquifer is hydraulically separated from the lower aquifer across most of the GAFB site by the MLU, which is a low permeable layer. A schematic cross section is presented in Figure 4-5.

Figure 4-5. Schematic Geologic Cross Section, GAFB Site



The upper aquifer, continuous beneath most of GAFB site, is a semi-confined saturated zone contained within the lower portion of the UFU and is perched on the MLU. There is a north-south trending zone roughly parallel to the Mojave River Bluffs, known as the Permeable Lacustrine Zone (PLZ), which forms the "downgradient edge" of the upper aquifer. The lower aquifer is present beneath the entire base and is related to the area's regional groundwater system. Within the former base area, the Lower Aquifer is semi-confined to unconfined and is encountered at approximately 210 to 250 ft bgs. Groundwater migrates from the upper aquifer to the lower aquifer through the PLZ. There also appears to be some perched water in the upper aquifer.

For this project, only the upper aquifer at OU-1 is of concern. Groundwater in the upper aquifer generally flows to the north and northeast. There has been some accumulation of groundwater due to historical return of water that was extracted and treated via air stripping because the pump-and-treat system which consisted of up to 22 extraction wells has been shut down since 2003.

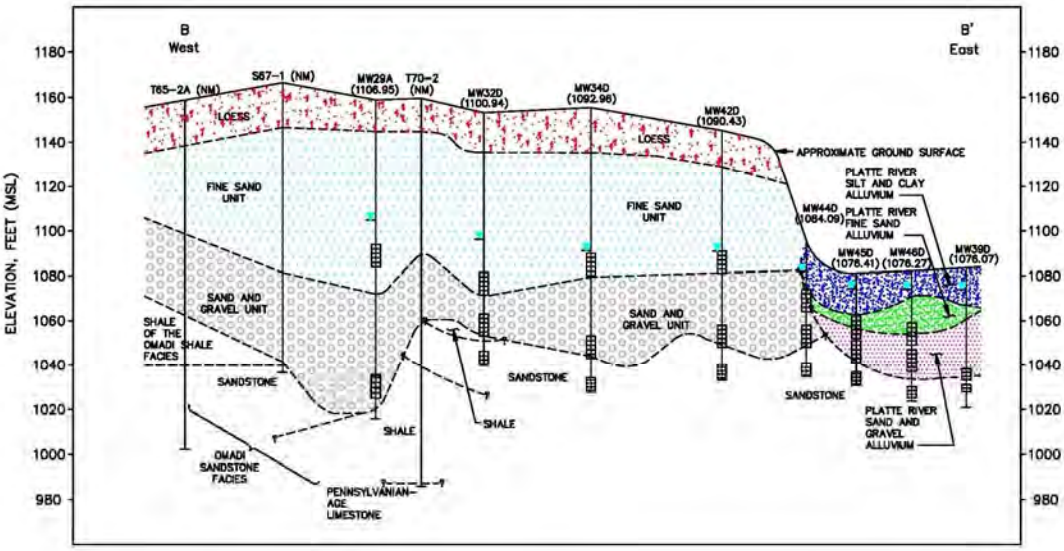
NOP, Mead, NE

The NOP site is located in the Todd Valley, an abandoned alluvial valley of the ancestral Platte River. The thickness of the unconsolidated material above bedrock in the Todd Valley at the site ranges from approximately 81-157 feet. The unconsolidated material consists of topsoil, loess,

and, and gravel of Pleistocene age. The uppermost bedrock unit is the Omadi Shale in the northwest and the Omadi Sandstone in the southeast portions of the site.

Three aquifers are present at the site: the Omadi Sandston aquifer, the Todd Valley aquifer, and the Platte River alluvial aquifer (Figure 4-6).

Figure 4-6. Schematic Cross Section, NOP Site



The Todd Valley aquifer is the first aquifer beneath the site. Towards the Platte River (i.e., towards the east) it grades horizontally into the Platte River alluvial aquifer. The Omadi Sandstone underlies these aquifers, and is part of the bedrock. In places, the Omadi Shale aquitard separates the deeper Omadi Sandstone aquifer from the overlying aquifer(s). Where the Omadi Shale is absent, the Todd Valley aquifer and the Platte River alluvial aquifer are in hydraulic communication with the Omadi Sandstone and behave as a single aquifer without hydraulic barriers. The Pennsylvania Shale aquitard underlies the Omadi Sandstone aquifer.

Monitoring well locations at the Site were established based on regional groundwater flow (generally towards the south and southeast). The water-bearing portions of the unconsolidated material in the Todd Valley are divided into an upper fine sand unit (12-17 feet thick) and a lower sand and gravel unit (17.5-72 feet thick). The upper sand unit is overlain by 4-23 feet of Peoria Loess. The unconsolidated material in the Platte River Valley (i.e., in the immediate vicinity of the Platte River) ranges in the thickness from 39 to 49 feet. Overbank silts and clays ranging from 10-17 feet thick overlie the Platte River alluvial sands and gravels.

The water table surface of the Todd Valley slopes toward the south-southeast with depths to groundwater in the Todd Valley ranging from 6.6 feet to 58.0 feet. A local zone of groundwater discharge is located along the western side of the Platte River floodplain in the southeastern portion of the Site. East of Johnson Creek, the water table surface of the Platte River alluvial

aquifer slopes to the south, paralleling the Platte River Valley with depths to groundwater in the Platte Valley ranging from 0.0-10.2 feet.

4.3 CONTAMINANT DISTRIBUTION

Camp Allen Site, Norfolk, VA

There are nine Constituents of Concern (COCs) which are all volatile organic compounds (VOCs):

Primary COCs	Other COCs
cis-1,2-dichloroethene (c12DCE)	1,2-dichloroethane (12DCA)
Tricloroethene (TCE)	1,1,1-Trichloroethane
Vinyl chloride (VC)	Benzene
	Tetrachloroethane (PCE)
	Toluene
	Xylenes

The plume extent is illustrated for the shallow aquifer in Figure 4-7, and for the deep aquifer in Figure 4-8. More detailed plume maps are included in Appendix D.

Figure 4-7. Plume Boundary in Shallow Aquifer, Camp Allen Site

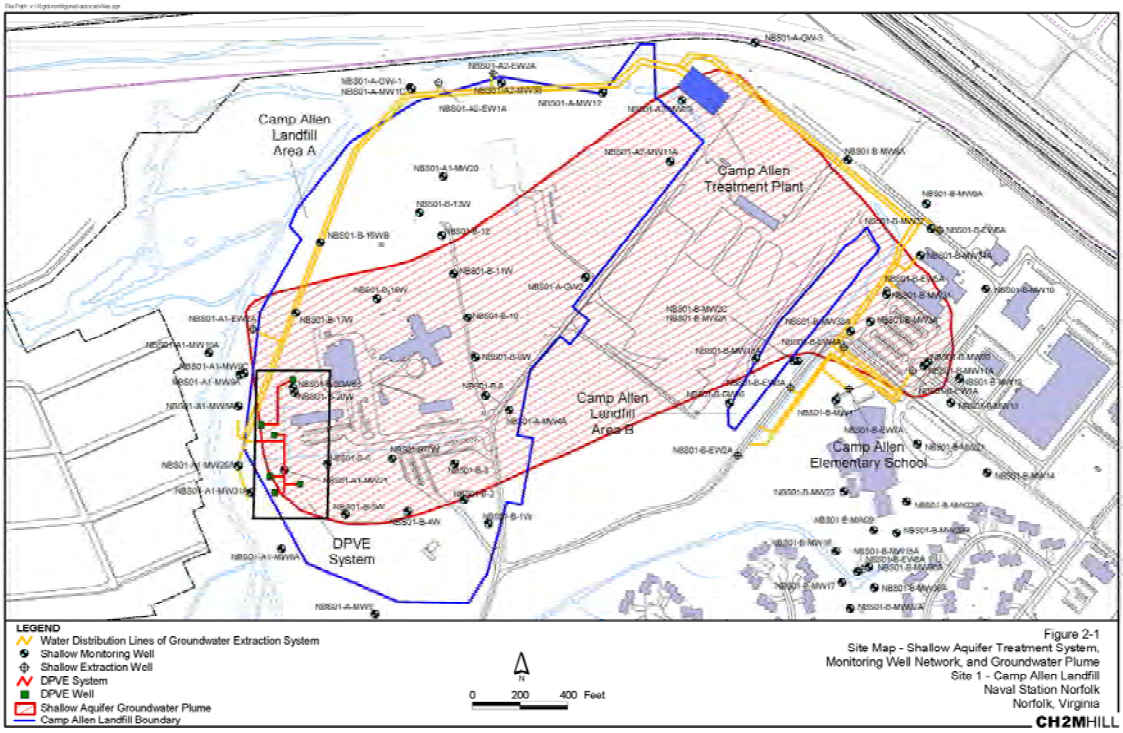
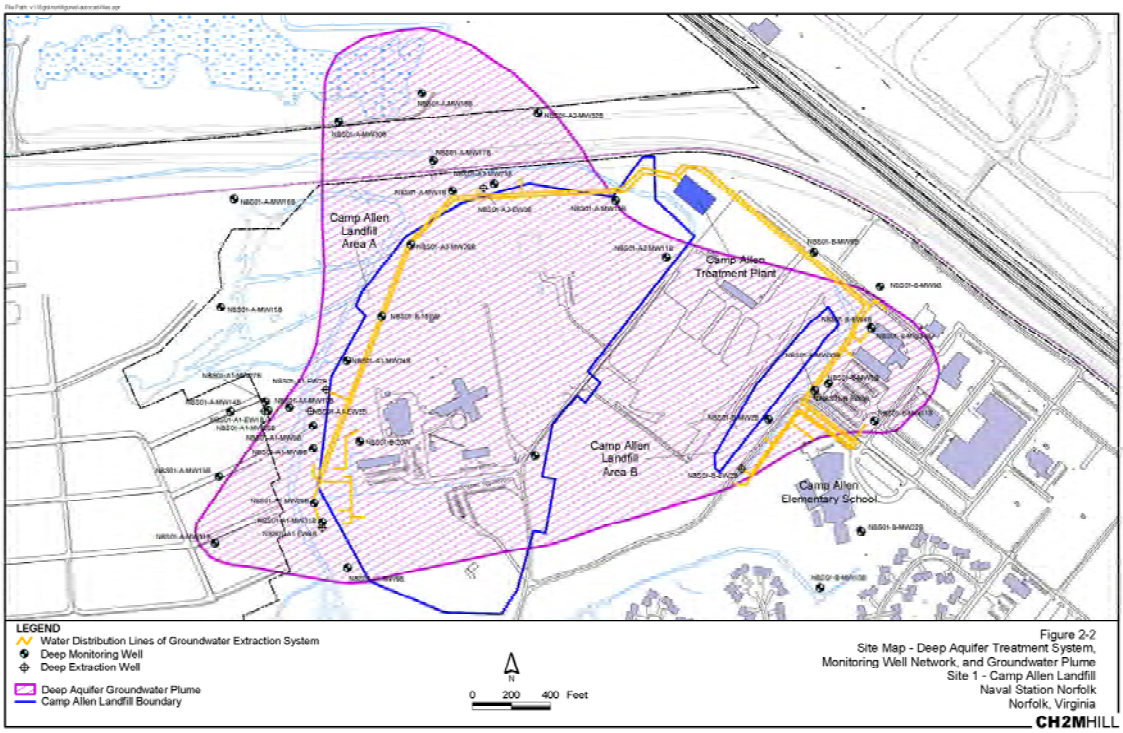


Figure 4-8. Plume Boundary in Deep Aquifer, Camp Allen Site



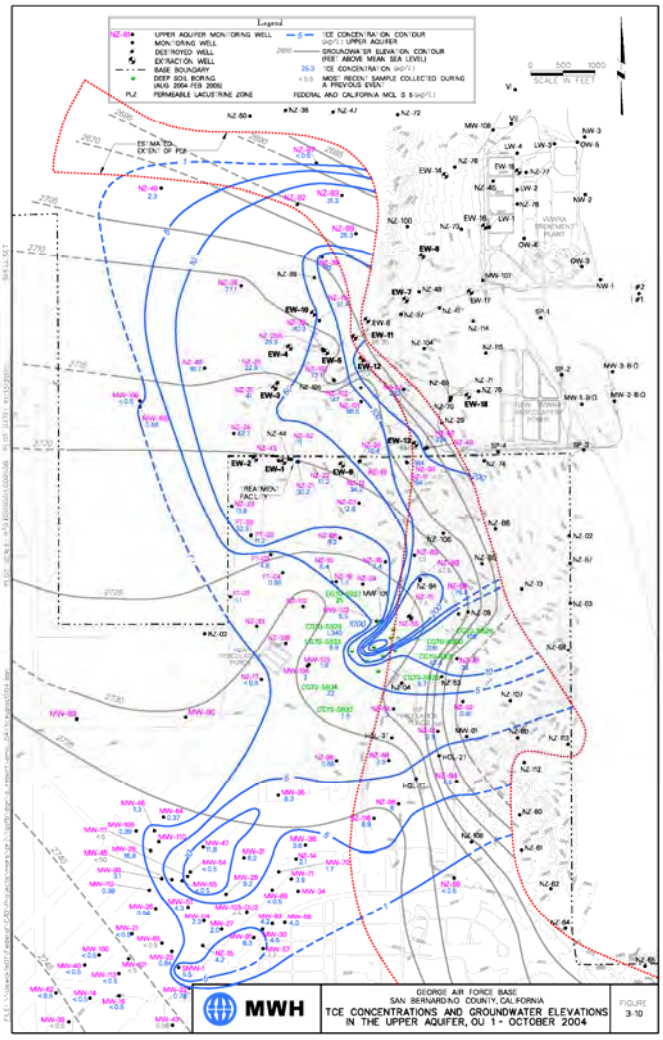
GAFB, Victorville, CA

Contaminants in OU-1 groundwater are:

TCE
c12DCE
PCE
Benzene
Toluene

TCE is the primary groundwater contaminant. Three potential source areas for TCE contamination in the OU-1 area have been identified: FT019c (fire training area); SD025 (industrial storm drain system); and FT082 (burn pit). A map illustrating TCE concentration extent in the shallow aquifer is provided in Figure 4-9. Additional plume maps are provided in Appendix B.

Figure 4-9. TCE Plume in the OU-1 Upper Aquifer, GAFB Site



NOP, Mead, NE

The following VOCs and explosive compounds were identified at the site (primary COCs are indicated with a "*"):

VOCs:

- Trichloroethene (TCE)*
- Methylene chloride;
- 1,2-dichloropropane; and

Explosive compounds:

- Hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX)*
- 1,3,5-trinitrobenzene (TNB)
- 2,4,6- trinitrotoluene (TNT)
- 2,4-dinitrotoluene (2,4-DNT); and

The site generally distinguishes the plumes based on TCE (Figure 4-10) and RDX (Figure 4-11).

Figure 4-10. TCE Plumes (Shallow Aquifer), NOP Site

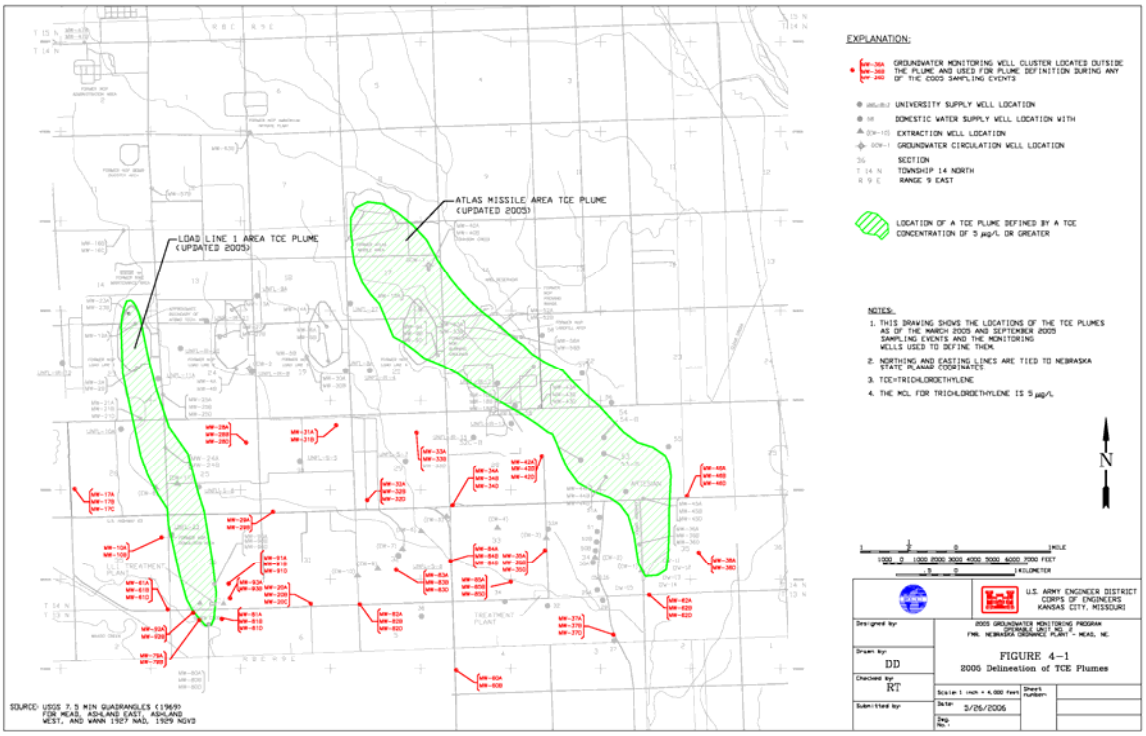
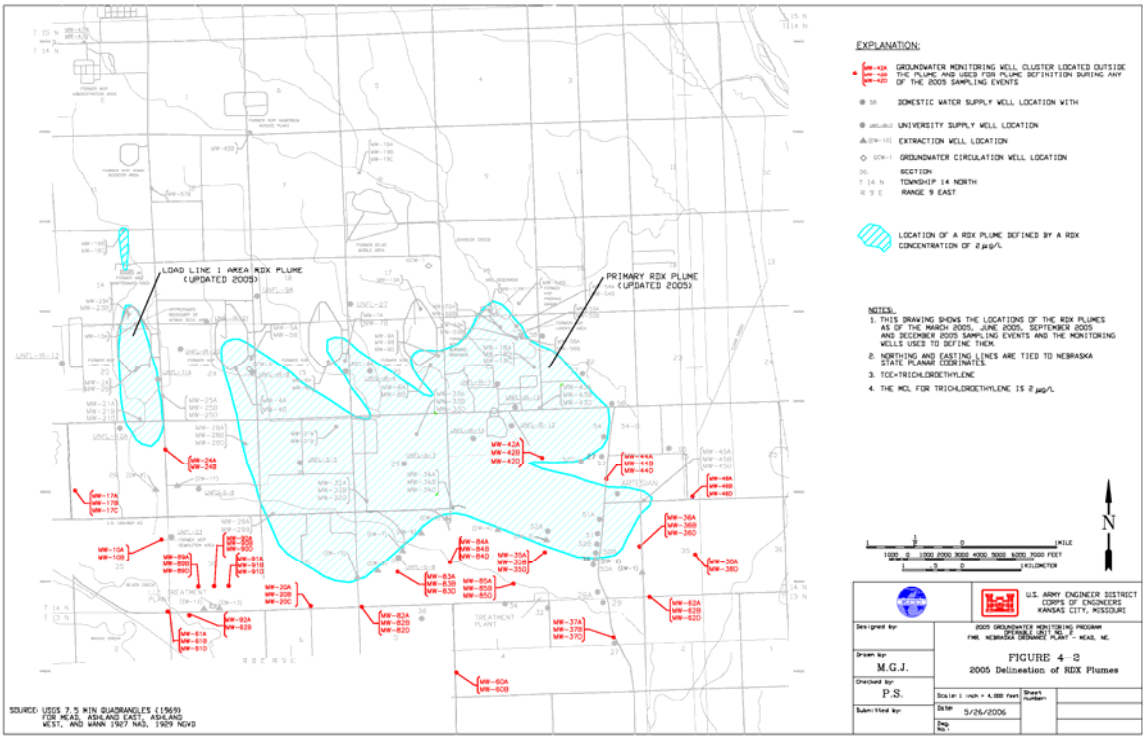


Figure 4-11. RDX Plumes (Shallow Aquifer), NOP Site



The four plumes (or “lobes”) of groundwater contamination identified at the Site are:

- TCE plume with the suspected source from the Atlas Missile Area, which is north of the eastern load lines (LL3 and LL4);
- TCE plume with the suspected source from Load Line 1 (LL1);
- RDX plumes with the suspected sources from LL1, LL2, LL3, and LL4.

According to site reports, the migration of these contaminant plumes is dictated primarily by the southeastward direction of the groundwater flow. The TCE and RDX plumes overlap in two areas: LL1 and LL4. The overlap at LL4 is due to migration of TCE from the Atlas Missile Area. Higher groundwater contamination is found in the upper fine sand units than in the sand and gravel units below. Generally, lower contaminant concentrations are found in the deepest of the three aquifers (the Omadi Sandstone aquifer).

5.0 TESTING DESIGN

This section provides an overview of the testing design for this ESTCP project. Please note that many of the specific details are provided in the following appendices:

- Appendix B Site-Specific Analysis – GAFB Site
- Appendix C Site-Specific Analysis – NOP Site
- Appendix D Site-Specific Analysis – Camp Allen Site

These appendices provide extensive description of the software testing approach and results at each of the demonstration sites.

5.1 CONCEPTUAL EXPERIMENTAL DESIGN

The following general approach was applied for each of the three demonstration sites:

- The ESTCP project team obtained preliminary information from the Site team for review prior to site visit (e.g., reports describing site conceptual model)
- The ESTCP project team conducted a site visit to present an overview of the project and to receive input regarding the optimization formulation from the Site team
- The ESTCP project team developed a preliminary optimization formulation, provided it to the Site team for review, and then finalized the optimization formulation based on feedback from the Site team
- The Site team then provided the most updated version of historical sampling data to EnviroStat in electronic format
- EnviroStat then performed the following activities:
    - Screened historical data to determine if any obvious data quality issues were evident
    - Attempted to resolve any data quality issues with the Site team
    - Reserved the last year of sampling data (which would be used later for validation of Sampling Optimizer results as well as for evaluation of Data Tracker)
    - Provided GeoTrans with Comma Separated Variable (CSV) files, not including the reserved data, that could used as input to the software for evaluation of the Sampling Optimizer functionality

    - Created five alternate versions of the reserved data that incorporated artificial anomalies based on discussions with the Site team (the actual values were also "tweaked" in each alternate version so that identification of the actual version by GeoTrans would be difficult)

- GeoTrans applied the Summit Software to evaluate data redundancy (i.e., the Sampling Optimizer module, in conjunction with Model Builder), and provided the ESTCP project team with a preliminary write-up of results

- EnviroStat then provided GeoTrans with the six versions of the reserved (i.e., current) data for input to Data Tracker (i.e., one was the actual version, and the other five had artificial anomalies as well as "tweaked" values so that identification of the actual version by GeoTrans would be difficult)

- GeoTrans applied the Summit Software to evaluate the Data Tracker functionality, and provided a summary of Data Tracker results to EnviroStat

- EnviroStat then identified which of the six versions of the reserved data was the actual version, and GeoTrans used that dataset to perform validation of the Sampling Optimizer results (i.e., used the more recent data to make plume maps using the baseline well locations, and plume maps based on the optimized sampling plans, to see if the maps based on the optimized sampling plans are reasonable)

- GeoTrans provided the ESTCP team with a write-up of results and conclusions, incorporating sections provided by EnviroStat regarding data preparation and interpretation of Data Tracker results

- After review by the ESTCP team, the write-up was finalized and forwarded to the Site team for their review and feedback, followed by a presentation of results to the Site team via conference call (Camp Allen site, Former NOP site) or in-person meeting (former GAFB site).

In addition to the analysis for each site described above, the following activities were also part of the project design:

- For one of the three sites (GAFB), additional validation of results was performed based on one year of sampling conducted subsequent to the original set of reserved data

- For one of the three sites (Camp Allen) the MAROS software was also applied, so that functionality and results (to the extent possible) could be compared (note that GTS and MAROS will also be applied at the NOP site in a separate effort outside of this project)

- EPA Region V arranged for their contractor to apply the software at one of their sites, and provided feedback to the ESTCP project team by filling out a questionnaire and preparing a brief summary report (see Appendix E)

It was originally envisioned that an additional application of the software would be performed by NAVFAC SE, but that did not occur due to changes in NAVFAC SE personnel.

Throughout the project the Summit Software was applied by a mid-level GeoTrans engineer with no previous LTMO experience, rather than by the software developer. In this manner, the software demonstration would provide a realistic evaluation of its usability by a typical DoD contractor with no prior experience with the software. Before the first site was evaluated, the GeoTrans analyst was provided with basic training for both the Summit Software and the MAROS code. The subsequent two sites were evaluated by a different mid-level GeoTrans analyst who was not provided any training on either software.

5.2 BASELINE CHARACTERIZATION

The steps in the experimental design described in Section 5.1 that might be considered "baseline characterization" are those associated with the following:

- Developing the optimization formulation for each site
- Obtaining and preparing the data for each site

The results of these steps are discussed in more detail below.

Optimization Formulations

The optimization formulation for each of the demonstration sites describes the site-specific optimization objectives and constraints. Each of the site-specific reports (Appendices B to D) contains a section called "Optimization Formulation". Significant results or observations include the following:

- *Meetings*. A meeting was held with site personnel, and during those meetings components of the formulation were discussed. The dates of the meetings were as follows:
    - Camp Allen Site: January 7, 2007
    - GAFB Site: March 16, 2007
    - NOP Site: January 29, 2007

- *Constituents Generally Differed for Redundancy Analysis with SO versus Data Tracking with DT*. During the formulation process the Site teams generally indicated that they focus their attention for mapping the plume on just a few key COCs, and those were the COCs they wanted to evaluate for redundancy using SO. However, for identifying unexpected values in new data via DT, the Site teams indentified a larger set of parameters for evaluation. The differences in COCs between the SO and DT evaluations are summarized in Table 5-1.

Table 5-1. COCs for Redundancy Analysis and Data Tracking Analysis

	COCs for Redundancy Analysis with SO	COCs for Data Tracking with DT
Camp Allen Site	TCE c12DCE VC	TCE c12DCE VC 12DCA 1,1,1-Trichloroethane Benzene PCE Toluene
Former GAFB Site	TCE	TCE PCE c12DCE Benzene Toluene
Former NOP Site	TCE RDX	TCE RDX 1,2-dichloropropane Methylene chloride DNT TNB TNT

- *For Redundancy Analysis, Multiple COCs Can Be Evaluated Simultaneously*. This is an important feature within the software, because it often would not make sense to perform redundancy analysis for different COCs separately. For instance, at Camp Allen there were three COCs considered in the redundancy analysis, and all three are VOCs. If the COCs were evaluated independently for redundancy, the results might suggest eliminating sampling at a location for one COC but not the other COCs. However, that would not ultimately lead to any savings because the well would still have to be sampled, and the same lab analysis would still need to be performed. By simultaneously considering different COCs that type of problem is avoided. The ramification of simultaneously considering multiple COCs is that the error represented on tradeoff curve needs to represent the error for all of the COCs. For the the Camp Allen site, the formulation specified that three COCs be evaluated simultaneously. For the NOP site, the formulation specified that two COCs (TCE, a VOC, and RDX, an explosive) be evaluated simultaneously. At GAFB TCE was the only COC of interest for the redundancy analysis.

- *Redundancy Analysis Can Only Be Performed One Aquifer At a Time.* The software provides spatial representation of plumes in two dimensions. Thus, plume maps for different aquifers, or different horizons of a thick aquifer, must be evaluated separately. For the Camp Allen site, the formulation specified that redundancy analysis be

performed independently for two aquifers (shallow and deep). For the GAFB site the formulation specified that redundancy analysis be performed for one aquifer (shallow). For the NOP site, the formulation specified that redundancy analysis be performed independently for three aquifers (shallow, intermediate, deep).

- *The Formulations for Redundancy Evaluation at Each Site Incorporated Multiple Plumes within an Aquifer.* For each of the demonstration sites, the distribution of contaminants was complex and did not conform to the shape of one simple plume with one distinct source. Rather, there were multiple plumes or "plume lobes" due to multiple contaminant sources. For instance, at the Camp Allen site there is Area A and Area B, and at the NOP site there are distinct plumes associated with multiple load lines. A decision was made during the formulation process at each site to perform the redundancy evaluation for the entire area of interest in each aquifer (i.e., covering multiple plumes or plume lobes) rather than for individual plume lobes. If this was not done, the number of separate evaluations would have increased dramatically, and there would be a risk that samples eliminated for one plume lobe might still be important for a neighboring plume lobe (i.e., there would be a potential inconsistency in the analysis).

- *For Redundancy Analysis, the Site Teams Were More Interested in Interpolation Errors Near the Plume Boundary Than the Plume Interior.* At all three sites, during the formulation process, the Site team indicated it was important to have more accurate interpolation near the plume boundary than in the plume interior. For example, if the MCL for a specific COC is 50 ug/L, the Site team might be concerned with an interpolation error of 20 ug/l near the plume boundary (e.g., interpolated value of 40 ug/l versus actual value of 60 ug/L), but might not be concerned with an interpolation error of 20 ug/L in the plume interior (e.g., interpolated value of 540 ug/l versus actual value of 560 ug/L). The software addresses this concern by calculating a dimensionless error value that places higher weight on interpolation errors in areas of low concentration (discussed previously in Section 2.1). The objective function specification options within the software provide the user with options for defining the cutoff between areas of low concentration (i.e., near plume boundary) and areas of high concentration (i.e., in plume interior), or to explicitly assign wells to each category.

- *Tracking Mass and Mass Flux.* During the formulation process the site teams at the GAFB site and the NOP site identified a desire to utilize the software to track changes in estimates of plume mass over time and/or estimates of mass flux across a boundary. However, it was determined during the application of the software that the results of these calculations are questionable when the distribution of samples in different events is not consistent (see Section 6.1.3 for a conceptual discussion). For the GAFB site the calculations were performed two ways: (1) based on actual samples where some locations were not sampled in some events; and (2) with missing values filled in manually using approximations based on temporal interpolation performed outside the software. For the NOP site, it was determined after the formulation process was complete that inconsistencies in the distribution of sampling locations between events were too severe to use the mass tracking capabilities in the software, even with manual

interpolation of the missing values.

- *Excluded Wells*. During the formulation process, the Site teams for the Camp Allen site and the NOP site provided a list of wells where sampling could not be eliminated. However, the Site team for the GAFB site indicated that all wells could be considered for elimination. Specifying wells that can be not be eliminated is straightforward within the software.

Additional details regarding the formulations are presented in each of the site-specific reports (Appendices B to D).

Obtain/Prepare Data

Obtaining the site-specific sampling data and preparing them for import into the software was another step that was performed prior to the actual testing of the software. Each of the site-specific reports (Appendices B to D) contains a section called "Data Preparation". Significant results or observations include the following:

- *The First Step Should Be to Perform an Initial Review of Data Quality*. EnviroStat performed an initial review of data quality, and discussed any potential issues with the Site team. This included a review of items such as consistency of well names, availability of x-y coordinates in a consistent coordinate system, consistency of reporting limits for non-detect values, and completeness of the electronic data based on hard-copy tables that were included in site reports. The review also included a quick review of aquifer designations for monitoring wells versus the screen intervals and water level elevations, to determine if any issues with the aquifer designations might exist. Furthermore, concentration versus time plots were made to determine if any wells hade "interesting" concentration histories that might be informative (e.g., database errors such as switched values at two wells might be determined with this approach). This step took several days of manual labor per site, and this effort would be a precursor step for application of any type of LTMO software.

- *Input File Format*. The sampling data have to be in a CSV file to import into the software, with the following structure:

    - For Model Builder the format is "Date, SiteID, EastCoordinate, NorthCoordinate, COC1, COC2, …", where SiteID is the well identification.

    - For data tracker the east and north coordinates are optional, and the format of the CSV data is "Date, SiteID, COC1, COC2, …".

  Concentration units have to be consistent over time for each individual COC. Different COCs can have different concentration units. There is no constraint on the exact chemical names.

- *For Spatial Redundancy Analysis using SO, the Input Data Must Be Reduced Into One Discrete Sampling Event.* The baseline data set for spatial redundancy evaluation consists of one data value per sampling location, each of which is assigned the same date value. These data assignments are made outside the software. For this project, the rule applied was to use the latest data value at each location unless the latest value was considered to be too old, in which case the location was not utilized. Other rules are also possible, such as using an average value or a maximum of recent values at each location.

- *For Spatiotemporal Redundancy Analysis using SO, the Input Data Must Be Grouped Into Discrete Sampling Events.* For spatiotemporal redundancy analysis, all samples must be assigned to discrete sampling groups in the data to be imported, such that every sample in that sampling event has the same sampling date within the software. For example, all samples in spring of 2006 might be assigned a value of 03/15/2006. For spatiotemporal data, the time lag between two adjacent sampling events has to be at least quarterly frequency, and the software User's Guide provides guidance for sampling event "frequency alignment". These date assignments must be done outside of the software prior to import.

- *Rules Need to Be Established for Duplicates and ND Values.* The input value for each sample in each event must be a unique value in the input data to the software. Therefore, values from "duplicate" QA/QC samples must be consolidated to a single value outside the software. For this project duplicate sample values were averaged, though other rules could be used (e.g., maximum value). No reporting limits and flags are utilized or allowed within the software, and such flags must be eliminated from the software input. A "graphing value" for non-detects needs to be assigned in the input data to serve as the concentration value for non-detects. Note that different graphing values to replace non-detects might be used for SO versus DT. Considerations regarding replacement of ND values include the following:

    - The selection of a graphing value for NDs can affect the prediction limits calculated by DT. For instance, if all NDs have a reporting limit (RL) of 5 ug/L, assigning a value of 5 ug/L for NDs may results in different prediction limits than assigning a value of 0.05 ug/L for NDs (when combined with other values at the same well). In a few cases, using a very low replacement value introduced spurious downward trends in wells whose data consisted entirely of NDs and "J" values (i.e., quantitative values reported but flagged "J" indicating that they were lower than the typical reporting limit). For DT, using a graphing value for NDs that is approximately half the most common RL for that parameter seems appropriate.

    - *ND Values with Elevated Reporting Limits are Problematic.* This occurs when some samples are diluted, and/or when different analytical methods are used over time. If NDs are replaced with the RL value, or perhaps half the RL value, there can be serious implications for the way data are interpreted in both SO and DT. For instance, if a concentration at a well is "<5 ug/L" in one event and "<50 ug/L" in the next event, assigning a value of 5 ug/L followed by 50 ug/L gives the

appearance of a significantly increasing trend which may be misleading, Similarly, using a value of 50 ug/L rather than 5 ug/L (or less) may cause very different models of plume distribution (and very different plume visualizations) in Model Builder, and may lead to different results in SO. The user must decide during the preparation of input data files (i.e., outside the software) whether to exclude ND samples with high reporting limits, or to include them with consistent rules (e.g., always assign the graphing value to the half the lowest RL for that parameter).

- *There is No Option in the Software to "Read But Not Use" Specific Data Values*. This is a significant limitation of the software, particularly with respect to DT. For instance, if a historical data value is considered "potentially anomalous" the user must decide during preparation of the input data whether or not to include the data value as part of the "background data" for that well. If it is included, it will impact results (such as the prediction limits in DT). However, if it is not included, then the value cannot be included in subsequent software operations, particularly the concentration versus time plots in DT. This issue also has ramifications for application of DT for wells where conditions have changed significantly over time (such as might be caused by slugs of contamination). If early data are excluded because of the impacts those data might have on the prediction limits, then those early data cannot be included on the concentration versus time plots. An improved version of the software would allow an option to "import but not use" such data values. Those values could then be plotted on the concentration versus time plots in DT using a different symbol, but not used in the calculation of the prediction bounds.

5.3 TREATABILITY OR LABORATORY STUDY RESULTS

These items do not apply to this ESTCP project.

5.4 DESIGN AND LAYOUT OF TECHNOLOGY COMPONENTS

The technology demonstrated in this product is a software product. The design and layout of the software was described in Section 2.1, and illustrated on a flowchart on Figure 2-1. Further details are provided in the software user guide, which has been provided as a separate deliverable for this project.

5.5 FIELD TESTING

A summary of key results from the testing of the software is provided in the following sections:

Section 5.5.1:	Schedule for Software Testing
Section 5.5.2:	Variations Tested
Section 5.5.3:	Ease of Use
Section 5.5.4	Software Bugs and Evolution of Software Features
Section 5.5.5	Testing of Interpolation and Transformation Options in Model Builder

Section 5.5.6 Summary of Redundancy Evaluation Results
Section 5.5.7 Tracking of Relative Mass and Mass Flux
Section 5.5.8 Visualization of Relative Uncertainty
Section 5.5.9 Data Tracker Results for Detecting Unexpected Concentrations
Section 5.5.10 Import/Export Features
Section 5.5.11 Computation Time/Level of Effort

More detailed descriptions of the testing and results for the three demonstration sites are provided in the site-specific reports (Appendix B to Appendix D).

5.5.1 Schedule for Software Testing

The schedule for testing of the software is summarized in Table 5-2.

Table 5-2. Schedule for Testing the Software

	2007												2008											
	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D
Camp Allen Site:																								
- Site Visit																								
- Formulation																								
- Software Testing																								
- Apply MAROS																								
Former GAFB Site:																								
- Site Visit																								
- Formulation																								
- Software Testing																								
- Follow-up Evaluation																								
Former NOP Site:																								
- Site Visit																								
- Formulation																								
- Software Testing																								
EPA Testing at One Site																								

5.5.2 Variations Tested

The analysis for each demonstration site incorporated variations that increased the robustness of the testing. These variations included the following:

- Camp Allen Site
    - Evaluated two aquifers (shallow and deep)
    - Evaluated all six combinations of interpolation method and transformation type in Model Builder
    - Evaluated all six combinations of interpolation method and transformation type for spatial redundancy evaluation, and two of the combinations (kriging-quantile and IDW-quantile) for spatiotemporal redundancy evaluation
    - Evaluated three COCs simultaneously for redundancy evaluation
    - Evaluated different values for Population Size in the GA in SO for both spatial and spatiotemporal redundancy evaluation
- GAFB Site
    - Evaluated one aquifer (upper)
    - Evaluated all six combinations of interpolation method and transformation type in Model Builder
    - Evaluated two combinations of interpolation method and transformation type for spatial redundancy evaluation (kriging-quantile and IDW-quantile), and one combination for spatiotemporal redundancy evaluation (kriging-quantile)
    - Evaluated two variations of the baseline data: Dataset A had 55 wells and Dataset B had 47 wells (eight wells that had atypical water levels and screened intervals were removed from Dataset B)
    - Evaluated different combinations of values for Population Size and Number of Generations in the GA in both spatial and spatiotemporal redundancy evaluation
    - Evaluated the Mass Metric and Mass Flux functionality for two cases: one with uneven distribution of samples per event, and one where missing data were filled in manually (outside the software) based on temporal interpolation
    - For spatiotemporal redundancy evaluation, evaluated use of the original dataset versus a modified dataset where missing data were filled in manually (outside the

software) based on temporal interpolation

  - For both spatial and spatiotemporal redundancy evaluation, utilized three different values for cut-off concentration between low concentration (i.e., plume boundary) and high concentration (i.e., plume interior) areas

- NOP site
  - Evaluated three aquifers (shallow, intermediate, and deep)
  - Evaluated all six combinations of interpolation method and transformation type in Model Builder
  - Evaluated all six combination of interpolation method and transformation type for spatial redundancy evaluation
  - Evaluated the difference between considering multiple COCs (i.e., TCE and RDX) simultaneously versus independently for the redundancy evaluation

In addition, for the DT evaluation at each of the three sites, artificial anomalies were added to the actual reserved data by EnviroStat. This included the following types of artificial anomalies:

- Abnormally high concentrations due to a new contaminant source and/or plume migration
- Abnormally high concentrations at individual wells for no apparent reason
- Abnormally low values for a number of samples in the same event
- Switched samples (e.g., due to bottles labeled incorrectly and/or reported incorrectly by the lab)
- Laboratory cross-contamination
- Database errors

These types of anomalies were introduced based on discussions with the Site team at each site regarding the types of anomalies that might be expected. In addition, it turned out that some of these anomalies were present in the actual data from some sites.

5.5.3 Ease of Use

The software was found to be easy to use, based on the application of the software by a mid-level analyst at GeoTrans with no LTMO experience. This was true for a mid-level analyst who received approximately a half-day in-person training on the use of the software (for one of the three demonstration sites), as well as for a mid-level analyst who did not receive training on the software (for two of the three demonstrations sites).

In addition, the EPA group that applied the SO functions of the software (including Model Builder) outside of our project reported that: "The user interface was very easy to use…User's manual was an excellent reference for set-up and execution, and it contained clear directions for navigating dialog boxes, setting parameters, formatting input files, etc…It took only few hours to get comfortable using the software (import/export, model set up, running the program). The user's manual was very helpful in this aspect. It took a few days to fully understand the method, the effects of changes in parameter values, and the results."

5.5.4 Software Bugs and Evolution of Software Features

During the application of the software at the first two demonstration sites, several minor bugs were detected and reported to the software developer. For example, in DT there was an issue with the plotting of the y-axis data labels that was identified. By the end of the demonstration product, no remaining bugs were known to exist.

During this demonstration project several software features were added or improved. These were not due to bugs; rather, they represent evolution of software features. Thus, there were some software features available when the software was applied to the third site (NOP) but not available when the software was applied to the first site (the Camp Allen site). Key features that were added or that evolved during the project include the following:

- With respect to redundancy evaluation, a feature was added to provide a "combined" tradeoff curve within the software when multiple COCs are evaluated simultaneously. The options include "Maximum Error Across COCs" and "Additive Error Across COCs". This option was not available when the Camp Allen site was evaluated, but was available when the NOP site was evaluated.

- With respect to Model Builder, the computation time for generating the kriging models was significantly reduced after the testing at the first site due to programming enhancements (from hours to minutes)

- With respect to redundancy evaluation, a new feature was added regarding the GA so that the software asks the user after each run if he/she wants to seed an additional run with the results from the last run with the population size doubled. First, the user starts with the default values for population size, number of generations, and random seed. After each run is completed the software will ask whether the user wants to "seed this run with the results of the last run, and double the population size" – if yes, the mutation probability will automatically re-adjust for the new population size; the user makes a decision

whether the tradeoff curve changes significantly during the run – if it does, then an additional run is necessary; otherwise the optimization result has converged. This new feature enables the user to perform convergence tests without guessing initial population size, number of generations, and random seed. This feature was not available when the Camp Allen Site and GAFB site were evaluated, and during the evaluation at the GAFB the mid-level analyst had difficulty determining the appropriate population size and number of generations. This new feature was utilized during the evaluation at the NOP site, and the mid-level analysts reported it was very helpful.

- With respect to redundancy evaluation, a feature was added to allow the user to explicitly categorize wells in the software as either plume "interior well" or "exterior well" if desired, as part of the objective function parameter assignment related to error calculation. This new feature was not utilized for the demonstration sites.

- With respect to redundancy evaluation, a feature was added so that the user can specify a cost per sample as part of the objective function parameter assignment related to cost (i.e., rather than using number of wells or number of samples as a surrogate for cost, this allows for a calculation of cost). This feature was used when the NOP site was evaluated.

- With respect to DT, a feature was added so that the software automatically determines whether the prediction limits should be static (i.e., straight lines) or decreasing, based on the background data for the specific COC at the specific well. This new feature was applied for the GAFB site and the NOP site. This feature was not available when the Camp Allen site was evaluated, and for that site only static prediction limits were calculated by the software.

- With respect to the portion of Model Builder that provides maps of uncertainty for plume visualization, the functionality was modified during the project. In the newer version of the software, this feature changed to "Visualize Relative Uncertainty", replacing "Visualize Uncertainty" in the previous version. The idea behind the relative uncertainty is that root mean square error at each pixel comprising the plume image is calculated and then is divided by the arithmetic mean value at that pixel. It is designed to normalize the uncertainty metric so that the values over different sampling events and sites can be compared. This newer version was utilized during the evaluation of the NOP site.

5.5.5 Testing of Interpolation and Transformation Options in Model Builder

Sampling Optimizer provides users with six possible combinations of interpolation technique and data transformation. They are:

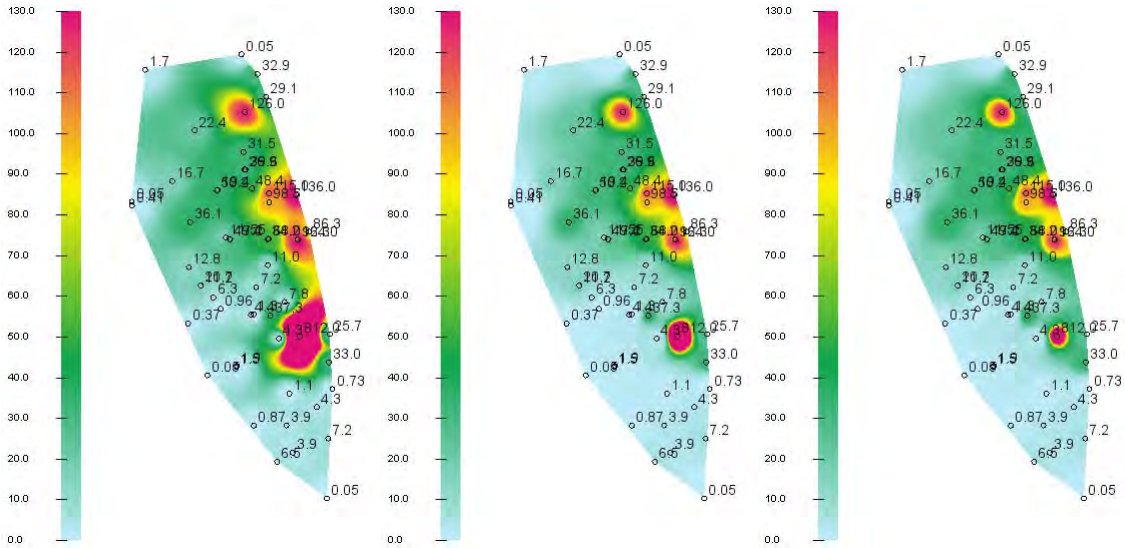
- Two interpolation technique options
  - Inverse Distance Weighting
  - Kriging

- Three data transformation options
  - None (i.e., No transformation)
  - Logarithmic
  - Quantile

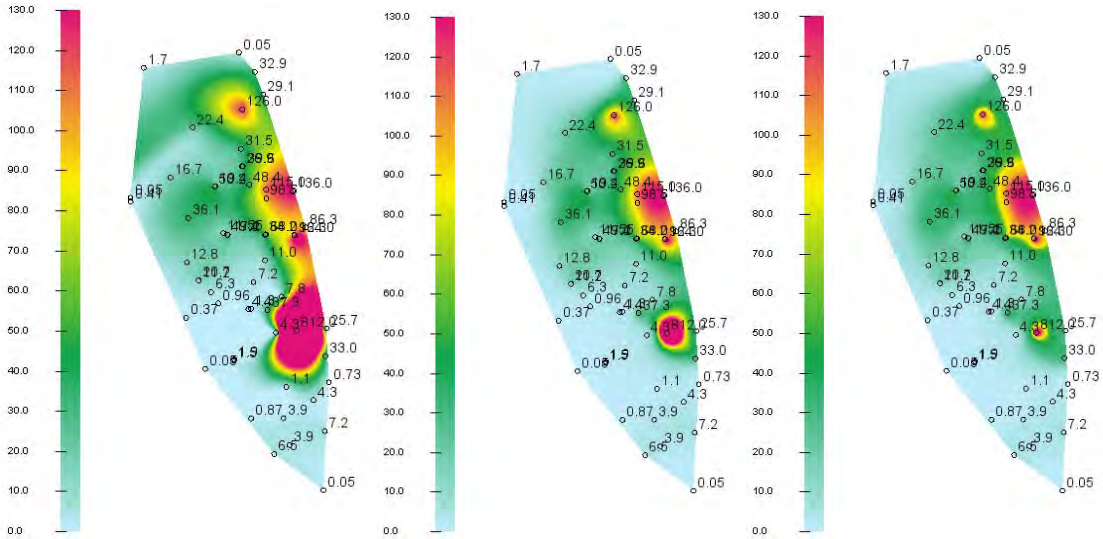
This results in six possible combinations for these basic options. Generally a user will only utilize one combination, and Summit suggests using kriging with quantile transformation. For the ESTCP project, GeoTrans tried all of the combinations at each of the three demonstration sites, and did verify qualitatively that kriging with quantile transformation provided the most reasonable representation of the plume distribution. The details for each site are provided in the site specific reports (Appendix B to D). An illustration from the GAFB site is provided in Figure 5-1, which shows the Model Builder results for TCE in the shallow aquifer for all six combinations of interpolation technique and transformation type.

{this gap is intentional}

Figure 5-1. Example Model Builder Results for TCE at the GAFB site, with Different Combinations of Interpolation Technique and Data Transformation



IDW-None IDW-Log IDW-Quantile



Kriging-None Kriging-Log Kriging-Quantile

Of the six combinations of interpolation and data transformation used, four generally produced better visual representations of the plume. These are:

- Kriging with quantile transformation
- IDW with quantile transformation
- Kriging with logarithmic transformation
- IDW with logarithm transformation

The GeoTrans analysts qualitatively preferred the representation using kriging with quantile transformation at all three demonstration sites. Both IDW with no data transformation and kriging with no data transformation resulted in model outputs that are biased to higher concentrations (i.e., interpolated concentrations on the boundaries of the map are much higher than believed to actually be the case). For example, in the representation labeled "Kriging-None" on Figure 5-1, concentrations in the upper-left corner are modeled to be on the order of 30µg/l, which is clearly higher than the actual measurements in that area which are on the order of 1 µg/l. This representation is not consistent with how most people would choose to interpret the actual data values.

Within Model Builder, the "visualization resolution" has an impact on whether the plume is displayed reasonably. The visualization resolution can be modified in "# of vertical slices for image" and "# of border slices" of "Visualization" settings with Model Builder. The software provides default values for these parameters.

5.5.6 Summary of Redundancy Evaluation Results

Example of Tradeoff Curve and Visualization of Optimal Plan(s)

The software generates a tradeoff curve for each optimization simulation. An example of a spatial optimization tradeoff curve for the GAFB site is presented in Figure 5-2 (many other tradeoff curves are provided in the site-specific reports provided in the appendices). On the tradeoff curve in Figure 5-2 the y-axis represents the sampling cost, which in this example is the number of well locations, and the x-axis is the dimensionless value of error. The sampling plans on the tradeoff curve represent optimal plans for different values of sampling cost (i.e., the plan with the lowest value of error for that sampling cost).

The analyst then selects plans along the tradeoff curve for further consideration. One consideration is obviously the amount of error, which increases to the right on the tradeoff curve. Another consideration is the rate of change in the number of wells, relative to the reduction in error, as one looks to the left on the tradeoff curve. Obviously, one does not want to select a plan on the tradeoff curve when another plan is available with much lower cost (i.e., much lower number of wells) and only slightly more error. For example, on Figure 5-2, plans with 50 or more wells would generally not be favored because there are plans with less than 50 wells and only slightly more error.

On Figure 5-2, two plans selected for further evaluation by the GeoTrans analyst are indicated: Plan 97 (lower error) and Plan 14 (higher error). Plan 97 reduces sampling cost from 55 to 41 (reduction of 25.5%) while Plan 14 reduces sampling cost from 55 to 30 (reduction of 45.5%). The software then allows the user to visualize the resulting plumes from the optimized monitoring network, compared to the plume for the baseline plan, using the same underling model (e.g., kriging with quantile transformation). Plume illustrations for the two selected plans, and for the base sampling plan, are illustrated on Figure 5-3.

Figure 5-2. Tradeoff Curve for Spatial Redundancy Evaluation, GAFB Site
(TCE, Dataset A, Cutoff = 25 ug/l)

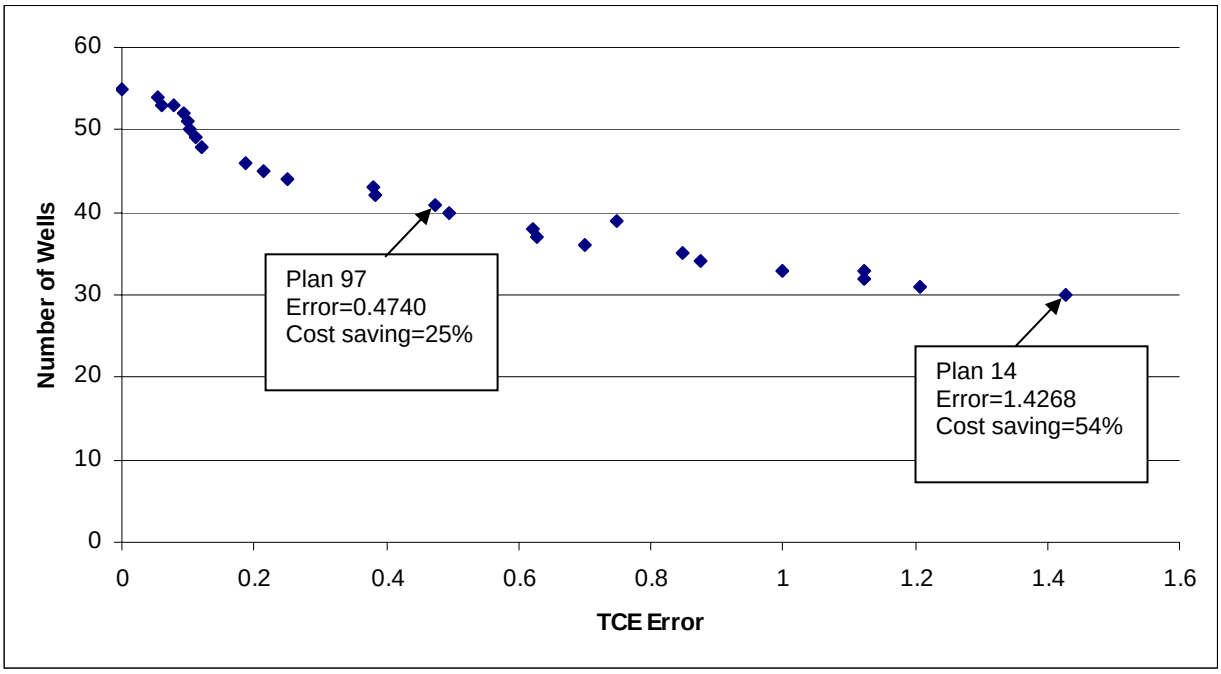
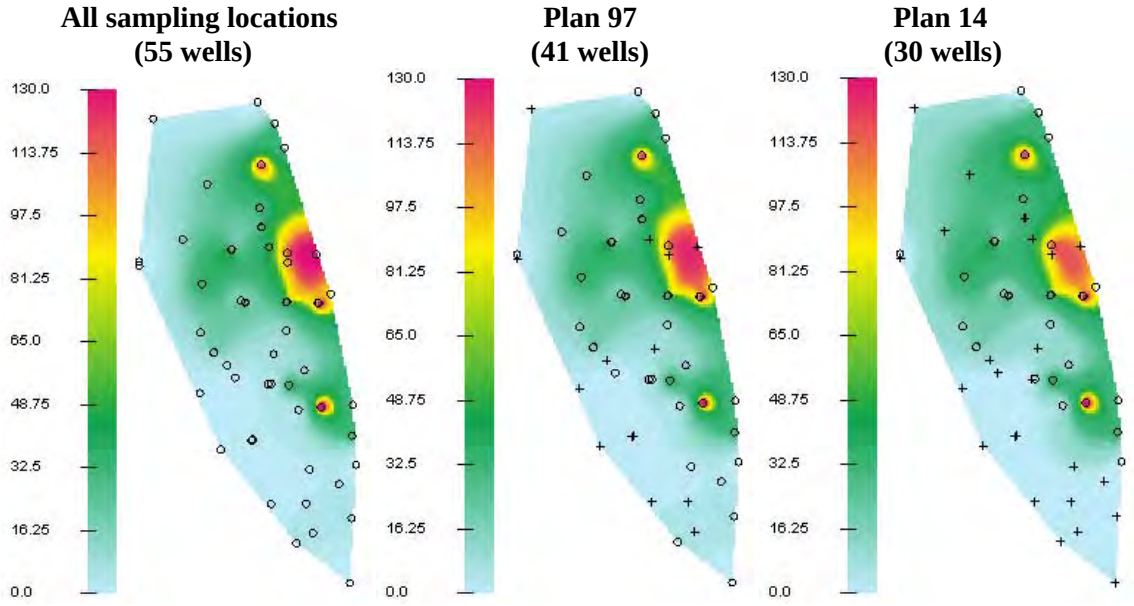


Figure 5-3. Comparison of Two Optimal Plans and Baseline Plan, GAFB Site
(TCE, Dataset A, Cutoff = 25 ug/l)



Note:

1. The symbol "+" indicates wells that are recommended to be removed by the Optimizer, while the symbol "O" denotes wells that are recommended to keep.

After comparing sampling costs (number of wells) and errors on the tradeoff curve, and visually inspecting the plume maps for selected plan(s), the GeoTrans mid-level analyst considered both optimal plans to be acceptable because both are generally similar to the map constructed with all sampling locations. Of course, it would ultimately be up to site stakeholders to decide if either Plan 97 or Plan 14 is acceptable. In this case, when the results were presented to the Site team, they concurred that the plume maps for both Plan 97 and Plan 14 were acceptable to them, and they indicated that they might choose to pursue plans with regulators even further out on the tradeoff curve (i.e., with lower number of wells and higher error).

The spatiotemporal evaluation process is similar, except that the y-axis of the tradeoff curve represents cost per year (or number of samples per year) rather than number of wells. This accounts for the fact that some wells are sampled more frequently than others over time.

Summary of Spatial and Spatiotemporal Results

Redundancy evaluation results are summarized on the following tables:

Table 5-3: Illustration of Redundancy Evaluation Results, Camp Allen Site
Table 5-4: Illustration of Redundancy Evaluation Results, GAFB Site
Table 5-5: Illustration of Redundancy Evaluation Results, NOP Site

It is difficult to fully summarize the results for the three demonstration sites, since the software produces a tradeoff curve for each optimization simulation which provides a family of optimal solutions. In Tables 5-3 to 5-5, the savings versus the baseline data set are presented for different values of normalized error along the tradeoff curve.

Table 5-3. Summary of Redundancy Evaluation Results, Camp Allen Site

Optimization	**# of samples in the baseline model**	**# of samples with the max error per COC of 0.5**	**# of samples with the max error per COC of 1.0**	**# of samples with the max error per COC of 1.5**
Shallow Aquifer				
Spatial	42	17 (59.5% saving)	16 (61.9% saving)	14 (66.6% saving)
Spatiotemporal	N/A	N/A	N/A	N/A
Deep Aquifer				
Spatial	31	28 (9.7% saving)	21 (32.2% saving)	21 (32.2% saving)
Spatiotemporal	21	20.2 (3.8% saving)	17.7 (15.7% saving)	17.5 (16.7% saving)

*note: number of samples is the number of wells (spatial) or the number of samples per year (spatiotemporal)

Aquifers: Shallow and Deep (evaluated separately)
COCs Evaluated: c12DCE, TCE, VC (evaluated simultaneously, errors for each COC added together)
Model: Kriging interpolation with quantile data transformation

Table 5-4. Summary of Redundancy Evaluation Results, GAFB Site

Cutoff Value	Optimization	# of samples in the baseline model	# of samples with the max error 0.5	# of samples with the max error 1.0	# of samples with the max error 1.5
Upper Aquifer with Dataset A					
25	Spatial	55	40 (27.3% saving)	34 (38.2% saving)	30 (45.5% saving)
50		55	42 (23.6% saving)	37 (32.7% saving)	31 (43.6% saving)
100		55	41 (25.5% saving)	36 (34.5% saving)	32 (41.8% saving)
25	Spatio-temporal	108	102 (5.6% saving)	96.33 (10.8% saving)	93.08 (13.8% saving)
50		108	104 (3.7% saving)	96.33 (10.8% saving)	92.83 (14.0% saving)
100		108	102 (5.6% saving)	98.08 (9.2% saving)	94.83 (12.2% saving)
Upper Aquifer with Dataset B**					
25	Spatial	47	36 (23.4% saving)	30 (36.2% saving)	28 (40.4% saving)
50		47	36 (23.4% saving)	31 (34.0% saving)	28 (40.4% saving)
100		47	36 (23.4% saving)	32 (31.9% saving)	29 (38.3% saving)
25	Spatio-temporal	92	88 (4.3% saving)	82.53 (10.3% saving)	79.08 (14.0% saving)
50		92	88 (4.3% saving)	82.53 (10.3% saving)	79.08 (14.0% saving)
100		92	88 (4.3% saving)	82.53 (10.3% saving)	79.33 (13.8% saving)

*note: number of samples is the number of wells (spatial) or the number of samples per year (spatiotemporal)

** Eight wells (MW-102, MW-104, NZ-06, NZ-10, NZ-20, NZ-30, NZ-31, and NZ-32) identified not to be representative of aquifer characteristics are excluded from Dataset B.

Cutoff value is used by the software to differentiate areas of lower concentration from areas of higher Concentration when calculating error (errors in low concentration areas are given more weight)

Aquifers: Shallow
COCs Evaluated: TCE
Model: Kriging interpolation with quantile data transformation

Table 5-5. Summary of Redundancy Evaluation Results, NOP Site

Optimization	**# of samples in the baseline model**	**# of samples with the max combined error of 0.5**	**# of samples with the max combined error of 1.0**	**# of samples with the max combined error of 1.5**
Shallow Aquifer (25 wells are non-removable)				
Spatial	81	54 (33.3% saving)	50 (38.3% saving)	47 (42.0% saving)
Spatiotemporal	N/A	N/A	N/A	N/A
Intermediate Aquifer (25 wells are non-removable)				
Spatial	84	48 (42.9% saving)	43 (48.8% saving)	42 (50.0% saving)
Spatiotemporal	N/A	N/A	N/A	N/A
Deep Aquifer (22 wells are non-removable)				
Spatial	56	35 (37.5% saving)	33 (41.1% saving)	32 (42.9% saving)
Spatio-temporal	N/A	N/A	N/A	N/A

*note: number of samples is the number of wells (spatial) or the number of samples per year (spatiotemporal)

Aquifers: Shallow, Intermediate, and Deep (evaluated separately)
COCs Evaluated: TCE and RDX (evaluated simultaneously, errors for each COC added together)
Model: Kriging interpolation with quantile data transformation

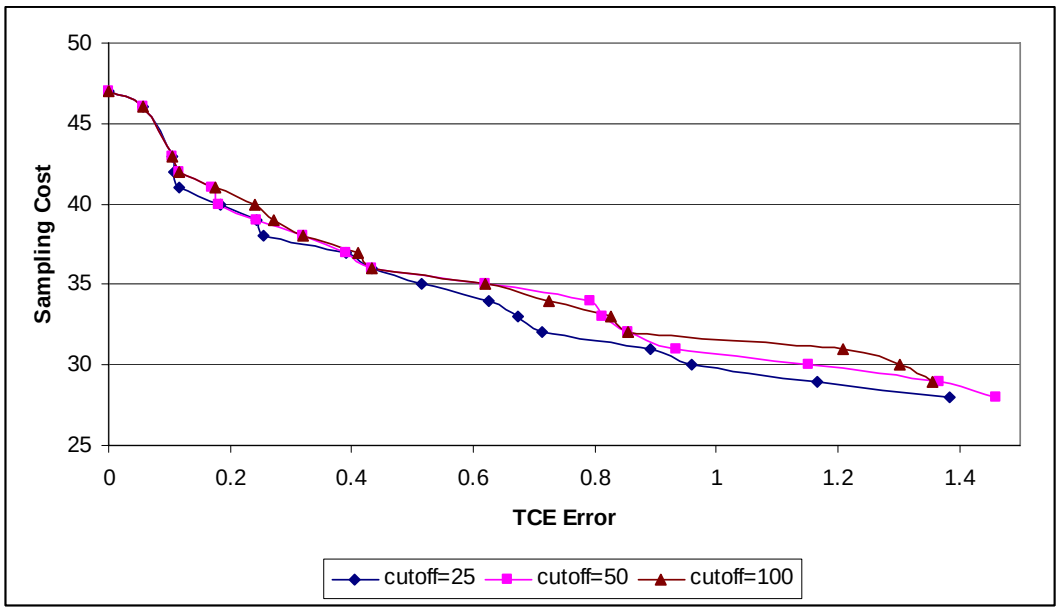
Tables 5-3 to 5-5 clearly indicate that, for the two sites where spatiotemporal analysis was performed, the potential savings achieved with spatial optimization was far greater than the potential savings achieved with spatiotemporal optimization (no spatiotemporal analysis was performed for the NOP site). Potential savings from spatial analysis, calculated based on number of wells eliminated, ranged from approximately 10% to approximately 67%. A value of approximately 35% appears to be representative. The ultimate savings would be even greater if some reduction in sampling frequency was implemented at some of the remaining wells (using some sort of rule to fill in values for wells not sampled in specific events to make plume maps, estimate plume mass, etc.). The potential savings from the spatiotemporal analysis, based on number of samples eliminated per year, only ranges from approximately 4% to approximately 17% at the two sites where spatiotemporal analysis was performed. An underlying conceptual reason for this is discussed in Section 6.1.4.

Impact of Cutoff Value on Tradeoff Curve

As discussed earlier (see Section 2.1), the "cutoff value" is a parameter entered by the user that is used in the calculation of error in the software to distinguish areas of low concentration versus areas of high concentration. The error is calculated differently in each region, such that deviations between interpolated and actual values are given more significance in areas of low concentration than in areas of high concentration. For one of the three demonstration sites (GAFB), the sensitivity of the tradeoff curve to different values of cutoff value was evaluated.

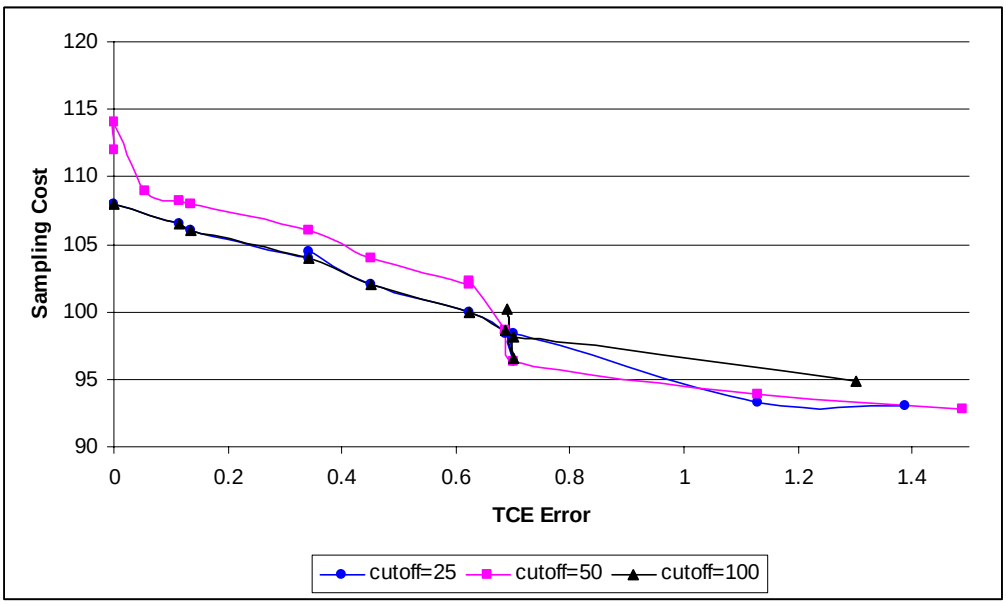
The results for spatial redundancy evaluation are illustrated on Figure 5-4 (based on Dataset B), and the results for spatiotemporal redundancy evaluation are presented on Figure 5-5 based on Dataset A). Keep in mind that the actual values of "error" are not quite comparable when different cutoff values are used.

Figure 5-4. Spatial Optimization Tradeoff Curves Using Different Cutoff Values (TCE, GAFB Site, Dataset B)



**sampling cost on y–axis represents number of wells*

Figure 5-5. Spatiotemporal Optimization Tradeoff Curves Using Different Cutoff Values (TCE, GAFB Site, Dataset A)



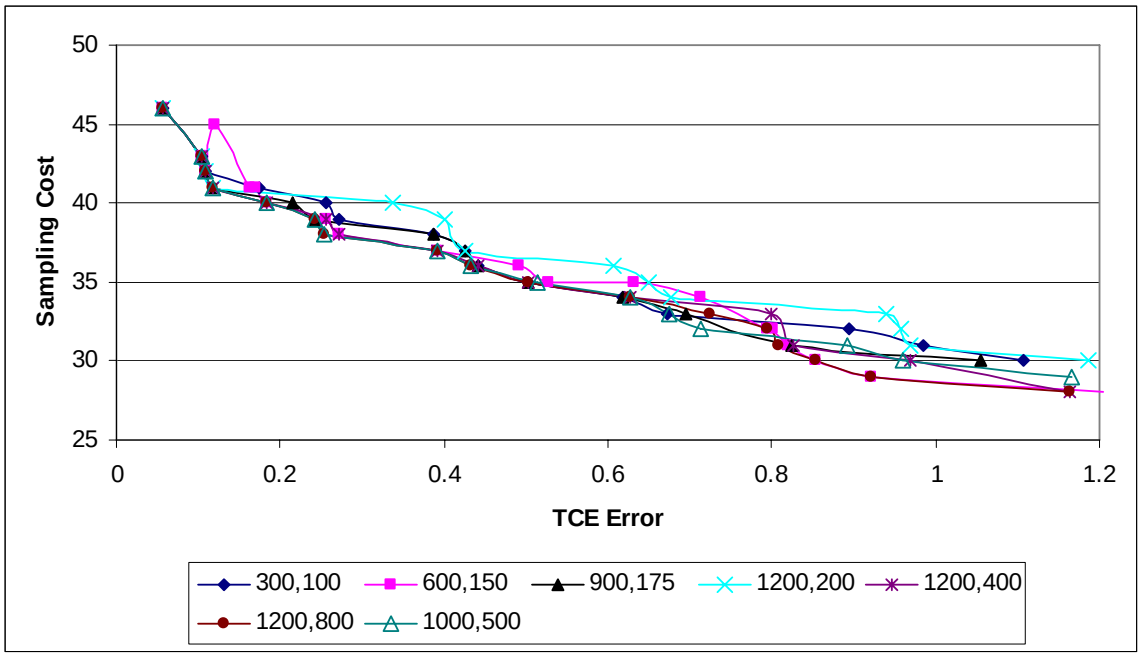
**sampling cost on y–axis represents number of wells*

Based on Figures 5-4 and 5-5 the cutoff value selected does have some impact on the tradeoff curve, though it is relatively minor. Based on these plots, it does not appear that there is a general rule regarding sampling cost versus cutoff value. In some cases the sampling cost is lower for lower cutoff value (for roughly equivalent error) and in some cases it is higher. Thus, the specific selection of cutoff value does not appear to be a significant driver in the results obtained.

Impact of Population Size and Number of Generations on Tradeoff Curve

For the GAFB site, the sensitivity of the tradeoff curve to the input value for population size and number of generations (i.e., GA parameters) was investigated for spatial optimization of Dataset B. The tradeoff curves for different combinations of population size and number of generations are illustrated on Figure 5-6.

Figure 5-6. Spatial Optimization Tradeoff Curves Using Different Population Size and Number of Generations (TCE, GAFB Site, Dataset B)



**sampling cost on y –axis represents number of wells*
first number in legend is population size, second value is number of generations

This illustrates that this combination of parameters does have some impact on the results, but it is a relatively minor impact. For instance, TCE errors for the following optimal solutions with 33 wells are provided below:

POP = 1200	GEN = 400	Error = 0.798383
POP = 300	GEN = 100	Error = 0.672252
POP = 1200	GEN = 200	Error = 0.940216
POP = 1000	GEN = 500	Error = 0.674904

Based on these results, the GeoTrans mid-level analyst noted that it was not at all clear how to assign this combination of parameters. Subsequently, a feature was added to the software to allow for sequential optimization simulations to achieve convergence of the tradeoff curve, based on sequential changes to the population size (discussed in Section 5.4.4).

Evaluating Multiple COCs Simultaneously Versus Independently

A question arises whether one should analyze COCs simultaneously or independently for redundancy in cases where there are two or more COCs to be evaluated. For the NOP site, which had two COCs for the spatial redundancy analysis (TCE and RDX), the GeoTrans mid-level analyst attempted to evaluate this issue. The full details of the analysis are provided in Appendix D. However, the important conclusion was that, if one COC was optimized independently, the resulting sampling plan yielded poor plume representation for the other COC. However, when the two COCs were optimized simultaneously, the resulting sampling plan yielded good plume representation for both COCs. Another scenario that was explored was to optimize both COCs independently, and then finalize the sampling plan by only removing wells that were suggested for removal for both COCs. This yielded acceptable plume representation for both COCs, but resulted in fewer wells removed than the optimal plan with both COCs evaluated simultaneously. The conclusion is that simultaneous evaluation is preferable when there are multiple COCs being considered during the redundancy evaluation.

The EPA Region V group that tested the software reported that they evaluated two parameters independently because these COCs originated from different sources and exhibited distinct footprints that only overlapped at their margins. However, they noted that they may have saved time had they evaluated two COCs simultaneously. Because they did not perform the simultaneously evaluation, it is not possible to compare the results of the two approaches.

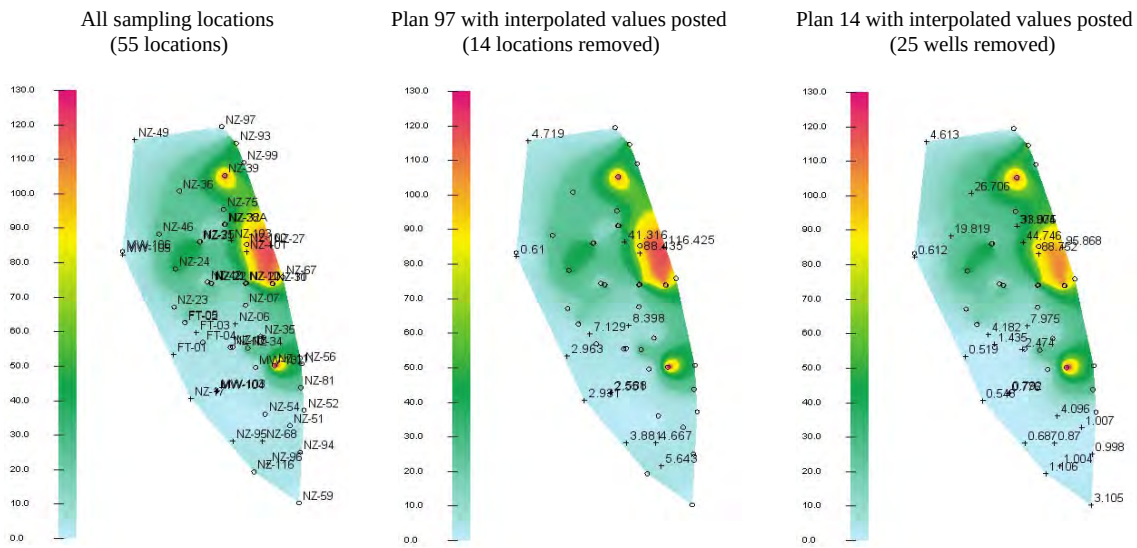
Validation of Results

For each demonstration site, the results of the redundancy analysis were validated using the most recently collected data that were reserved for this purpose. For one of the sites (GAFB) additional validation was performed based on data from two additional sampling rounds collected subsequent to the initially reserved data. The general process for performing the validation was to use the reserved data to make plume maps using the baseline well locations (based on the underlying model from Model Builder), and then to also make plume maps using the reserved data only at the wells in the optimized sampling plans. These maps could then be compared to evaluate if the plume maps based on the optimized sampling plans are reasonable. Details of how this was actually implemented, and the results, are provided in the site-specific reports (Appendix B to D).

An example of the validation results for the GAFB site (spatial optimization, Dataset A) is provided in Figure 5-7 to illustrate the concept. In this example, Plan 97 has fewer locations removed (i.e., lower error) relative to Plan 14, and the well network for Plan 97 provides a

slightly better representation of the plume than Plan 14 using these reserved data (i.e., relative to the map on the left that is based on all the sampling locations and the reserved data). However, the GeoTrans analyst considered the plume representation from both Plan 97 and Plan 14 to be reasonable, and the Site team agreed when they were presented these results.

Figure 5-7. Example of Validation Results Using Reserved Data
(TCE, GAFB Site, Dataset B, Spatial Optimization)



The validation results for all three sites indicated that the optimized plans generally produced acceptable plume representations using the reserved data. In each case, two solutions along the tradeoff curve were evaluated (as in the example above), one with lower error and one with higher error. As expected, the plan with the lower error (i.e., fewer samples removed) performed better with the reserved data than the plan with the higher error (i.e., more samples removed). It is ultimately up to site stakeholders to determine how much error is acceptable. Overall, the validation exercise provided confidence in the results provided by SO.

5.5.7 Tracking of Relative Mass and Mass Flux

The software allows calculation of "relative mass" and "mass flux" for specific sampling events, based on spatial interpolations:

- For "relative mass", the software calculates mass per unit volume of aquifer. Therefore, it is useful for comparisons of relative mass between sampling events, but it is not intended to estimate absolute mass within the plume. The interpolated concentration value at every "cell" in the plume map for that COC at that time period is summed and each cell is approximated to represent the same volume (i.e., incorporating vertical extent and porosity) as every other cell.

- For “mass flux”, the software calculates the mass flowing through a cross-section of a site (e.g., across a site boundary).

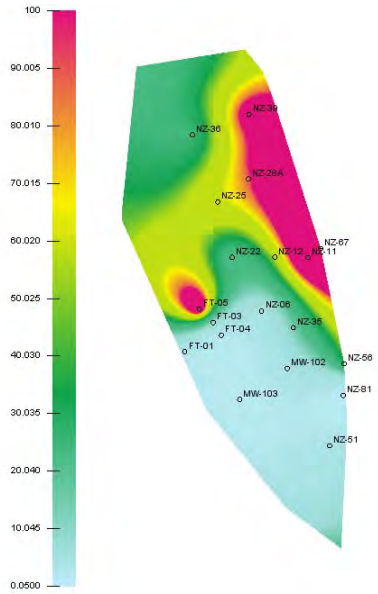
Mass tracking was part of the formulation for two of the three demonstration sites. However, during the evaluation of the GAFB site, complications were discovered as a result of inconsistent distribution of sampling locations over time. This complication is explained below using several examples from the GAFB site

Figure 5-8 illustrates the TCE plume as represented in the software for two different sampling events. Focusing on the extreme northern portion of the maps on Figure 5-8, it is noteworthy that the interpolated concentrations in 6/2001 (right side of the figure) are significantly lower due to the presence of three new monitoring wells in the northeast corner of the site (NZ-97, NZ-93, NZ-99). These wells were not present in 12/2000 or in other previous events. These three additional wells give more accurate information about the actual TCE concentrations in their vicinity. In other words, the interpolated concentrations in 12/2000 are biased due to fewer sampling locations in that area. This causes a problem for calculations of mass or mass flux. Should the map for 12/2000 utilize the data from future events to more accurately represent the northern portion of the site? The software does not interpolate with respect to time, so that type of operation would need to be performed outside the software.

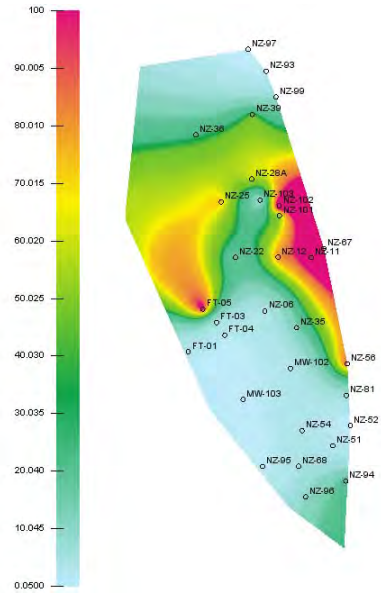
.

Figure 5-8. Example 1 Regarding Complication for Mass Calculations

Interpolated Plume Map, 12/2000

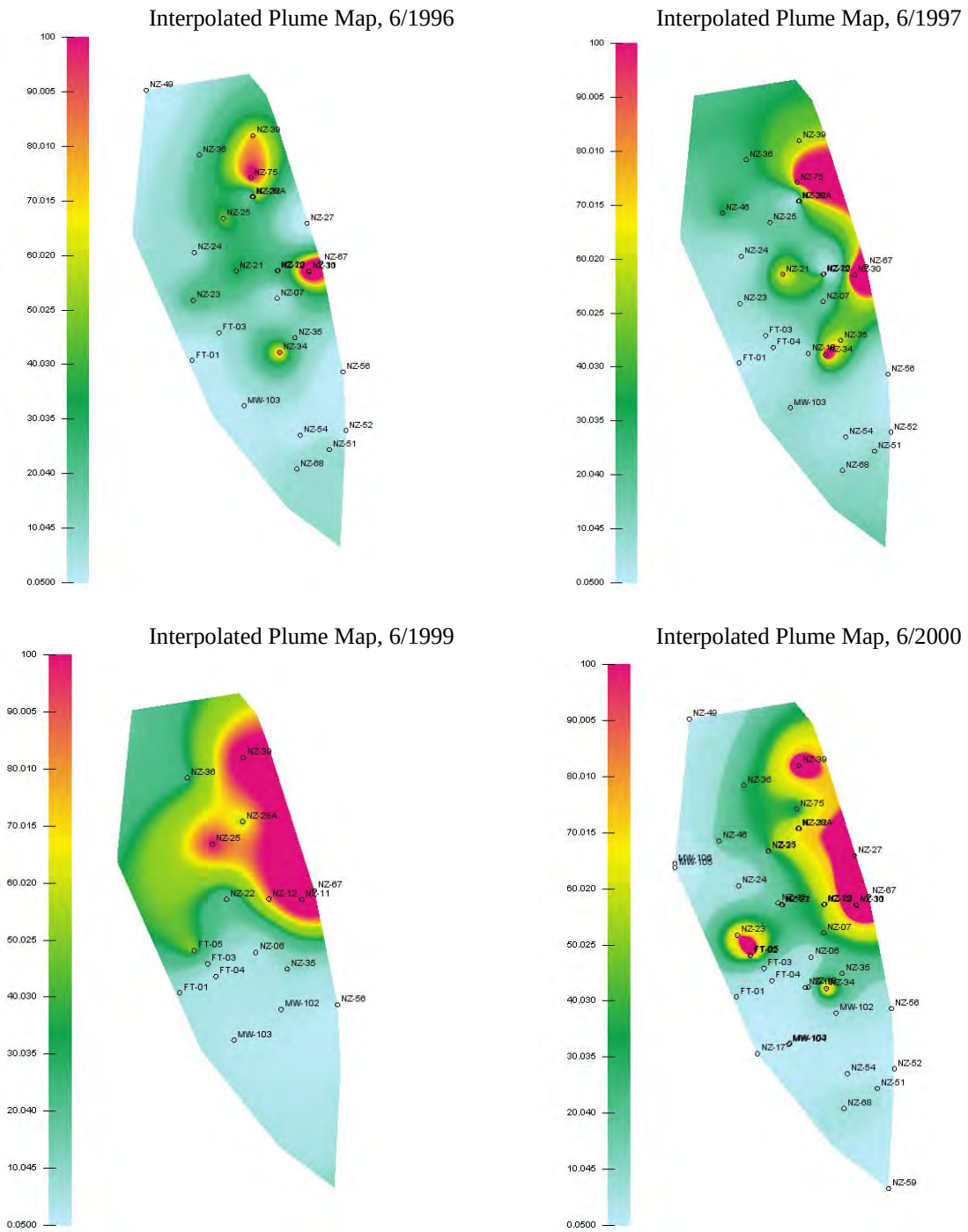


Interpolated Plume Map, 6/2001



Another interesting case is illustrated on Figure 5-9, which involves wells that are sporadically sampled. Note well NZ-49 in the upper left-hand corner in each map on the Figure 5-9.

Figure 5-9. Example 2 Regarding Complication for Mass Calculations



In 6/1996, NZ-49 was sampled, and it appears to be clean. From 1997-1999, NZ-49 was not sampled, but in 6/2000 it was again sampled and was found to be still clean. In this case, it makes perfect sense to assume NZ-49 always had low concentrations in between the 1996 and 2000 samples. However, the plume maps developed by the software for each event are only based on spatial interpolation of data within that event. There is no temporal interpolation, and as a result, the concentrations represented by the software in 6/1997 and 6/1999 illustrate higher concentrations in the northwest part of the site than would likely be expected based on the sampling that occurred before and after those events. A different result would have occurred if values for NZ-49 had been assigned for events where it was not sampled, based on temporal interpolation of the 1996 and 2000 results at that well. At this point the software does not accommodate this.

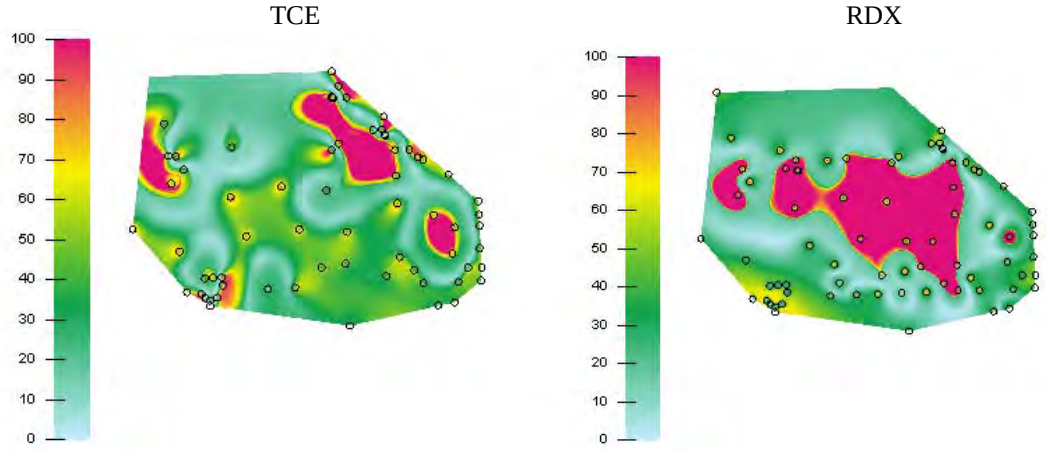
The ramification of these examples is that, if distribution of sampling locations is not consistent between events, the results of relative mass or mass flux calculations provided by the software may be questionable. For GAFB, the relative mass and mass flux estimates over time were tracked two ways: one with uneven distribution of samples per event, and one where missing data were filled in manually (outside the software) based on temporal interpolation. The results with the data filled in were more stable and also considerably different in overall magnitude (results are provided in Appendix B). For NOP, it was determined that the large variation in distribution of sampling locationss over time precluded the reliable application of the mass tracking features in the software. An improved version of the software would provide for a mechanism to use some form of temporal interpolation to fill in missing values before performing the relative mass or mass flux tracking calculations. The software developer indicates that they intend to add this functionality to future versions of the software within approximately one year (outside the scope of this ESTCP Project).

5.5.8 Visualization of Relative Uncertainty

Model Builder can provide maps of uncertainty for plume visualization. In the most recent version of the software, this feature has changed to “Visualize Relative Uncertainty”, replacing “Visualize Uncertainty” in the previous version. The idea behind the relative uncertainty is that root mean square error at each pixel comprising the plume image is calculated and then is divided by the arithmetic mean value at that pixel. It is designed to normalize the uncertainty metric so that the values over different sampling events and sites can be compared.

As an example, Figure 5-10 illustrates the plume uncertainty maps produced by the software for the shallow aquifer at the NOP site, using kriging with quantile transformation.

Figure 5-10. Example of Relative Uncertainty Map, NOP Site
(Kriging with Quantile Transformation)



The GeoTrans mid-level analyst did not find this feature of the software to be useful from a practical standpoint. The units are not straightforward, so although the map illustrates "relative uncertainty", it does not help identify what levels of relative uncertainty are acceptable or are of concern. It also does not address whether one value of uncertainty is acceptable in an area of high concentration but not acceptable in an area of low concentration. It is also not clear how these maps can be used to identify how many new wells might be needed to reduce the uncertainty to an "acceptable" degree, and where to locate those wells.

One possible improvement to this feature in the software might be to provide a map of estimated concentrations and/or a map of "absolute uncertainty" side-by-side with the map of relative uncertainty currently provided. One reason the relative uncertainty maps are hard to interpret is that high relative uncertainty at low concentrations may or may not be important, and conversely, high relative uncertainty in the middle of the plume may also be not very important. This all depends on what the site's goals are. The additional maps would provide the analyst with additional information for making such assessments. However, it would still not fully address concerns raised by the GeoTrans analyst regarding the how these maps can be used to identify how many new wells might be needed to reduce the uncertainty to an "acceptable" degree, and where to locate those wells.

5.5.9 Data Tracker Results for Detecting Unexpected Concentrations

Data Tracker (DT) is a feature of the Summit Software designed to aid in identifying unexpected values in recently collected data. DT compares new ("current") data values with selected background data from that well that are imported into the software. The "background" data used should be representative of values expected in future monitoring. When a current value is not consistent with expectations based on prior data, it is classified as "out-of-bounds". These expectations are formulated as prediction bounds (upper and lower) for the current observation for the COC being evaluated, based on the background data. DT has two types of bounds: static

and time-dependent decreasing (note only static bounds were available in the software when the Camp Allen site was evaluated). If DT finds a statistically significant decreasing trend in the background data, it uses time-dependent bounds; otherwise it uses static bounds. As discussed earlier, DT allows at most one measurement per COC per well per date, and therefore duplicate values must be reduced to one value (e.g., average or maximum value) in the input data. Also, as discussed earlier, Data Tracker does not read or utilized flags such as for non-detects, so the user must assign values to the non-detects when preparing the input data.

The software divides the current data into two tables: "in-bounds" and "out-of-bounds". The user can generate concentration versus time plots for each of the wells in each of those tables, for each COC. The values flagged as "out-of-bounds" presumably merit further evaluation of the nature and/or cause of the unexpected value. The GeoTrans mid-level analyst categorized the out-of-bounds values into various categories, such as the following categories used for the NOP site:

A. The current concentration is much higher than the background data and above cleanup goal, more likely bad data than plume migration.

B. The current concentration is much lower than the background data which were above cleanup goal, could be bad data.

C. The current concentration is higher than background data and above cleanup goal but following an increasing trend.

D. The current concentration is lower than background data but following a decreasing trend.

E. The current concentration is out of bounds, but not a concern.

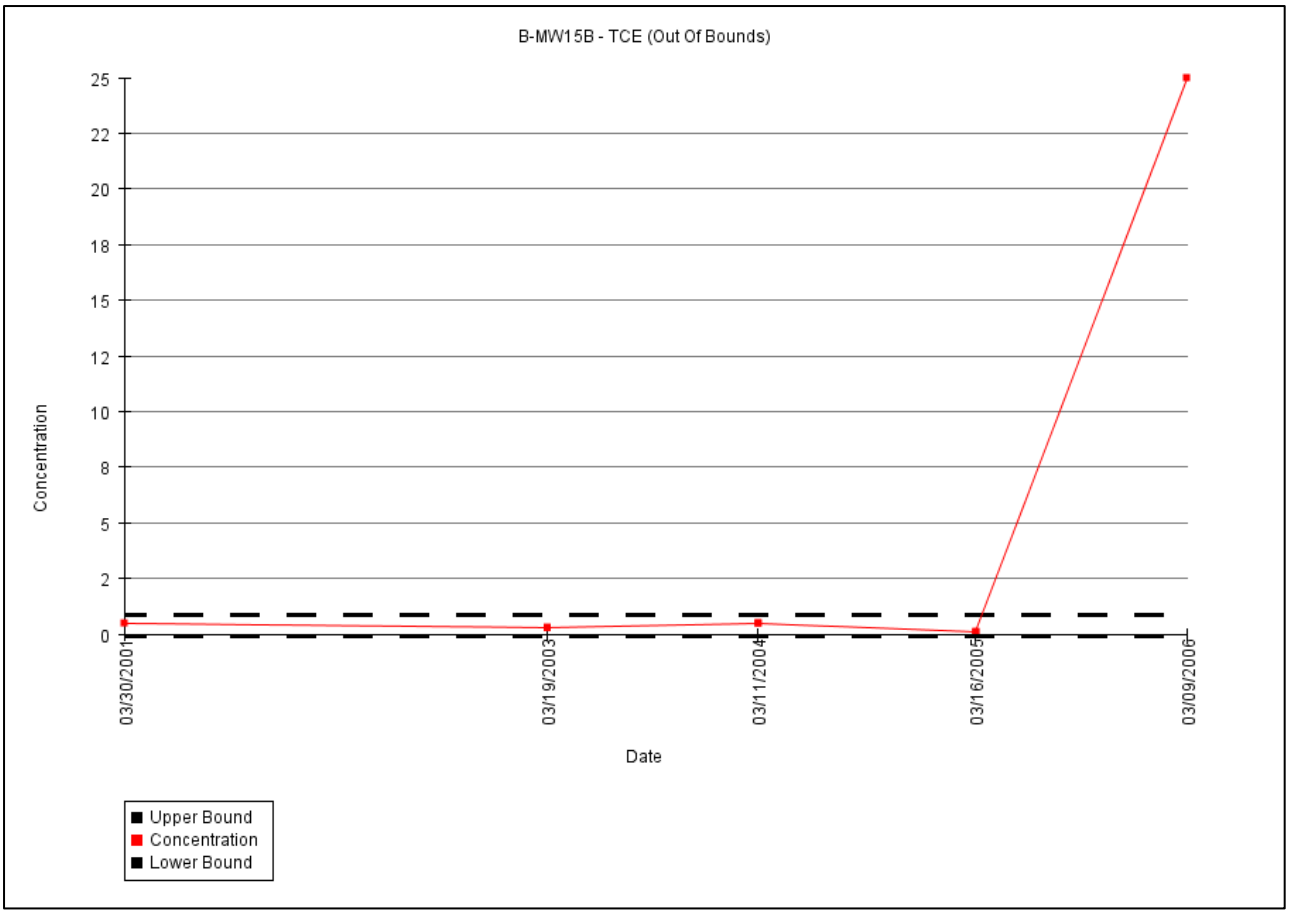
F. Current concentrations are higher than previous data, more likely plume migration than bad data.

The full details of the DT analysis at each site are provided within each site report (Appendix B to D). Overall, DT identified the vast majority of artificial anomalies added by EnviroStat, and some actual anomalies. Some examples are provided below.

An example of DT detecting anomaly that is due to "bad data" is illustrated in Figure 5-11, which is from the Camp Allen site. On Figure 5-11, the latest data value for TCE at well B-MW15B is anomalously high. There were similar "out-of-bounds" values for other COCs at the same well. Upon further inspection, based on these out-of-bounds values, it became apparent that the data from co-located wells B-MW15A and B-MW15B were reversed. Interestingly, DT identified only the errant low values at B-MW15A for one of the tracked constituents, because the data histories for the other tracked constituents had enough low values to make the lower prediction limits quite low. Regardless, DT served its function by identifying a potential issue that merited further attention. However, DT does not have the functionality to realize that multiple COCs are out-of bounds at the same well; that connection must be made by the analyst.

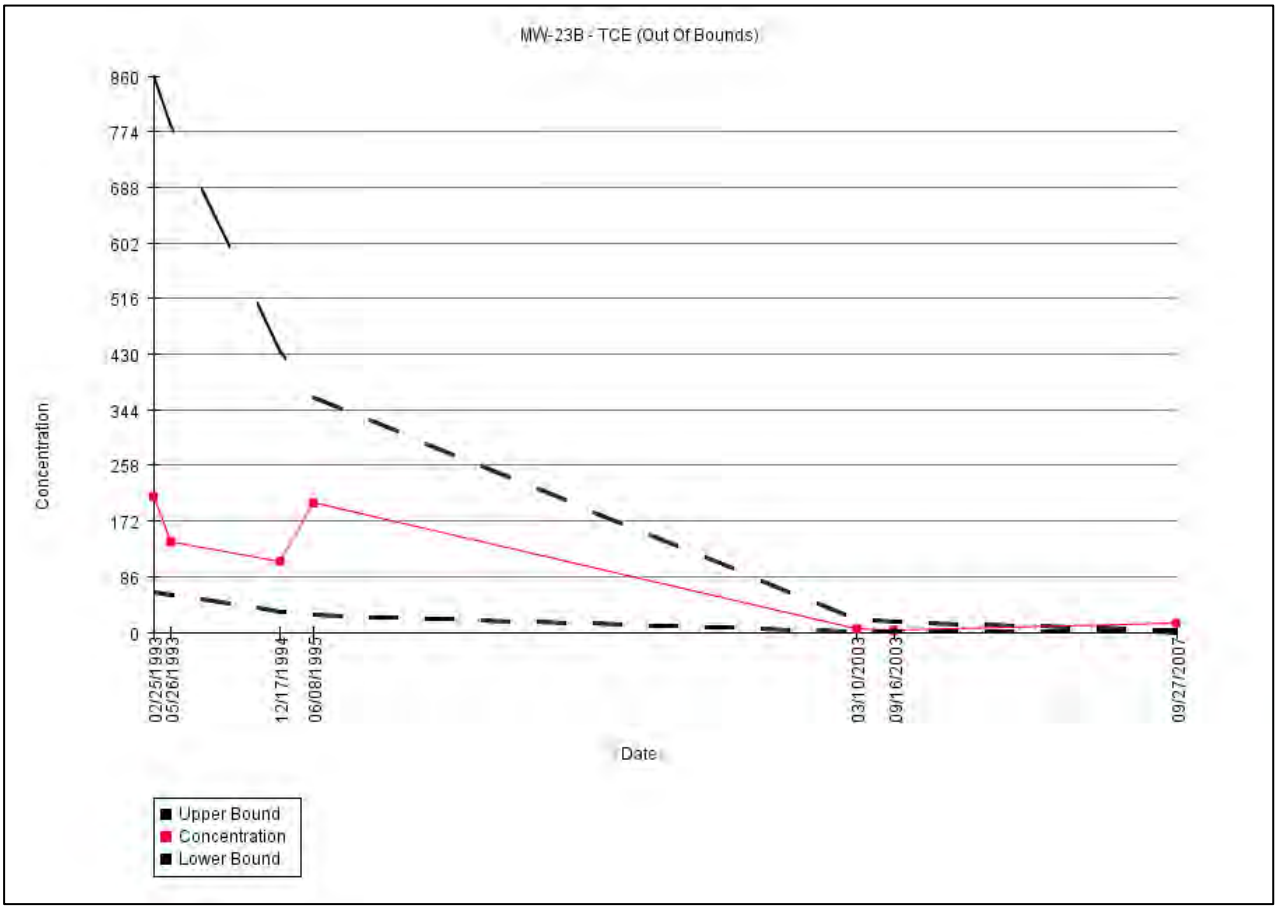
Similarly, the software does not have a mechanism for identifying that groups of wells in the same general location have similar "out-of-bounds" values. Again, it is up to the analyst to make that connection.

Figure 5-11. Example Data Tracker Plot of an Anomaly Due to Bad Data



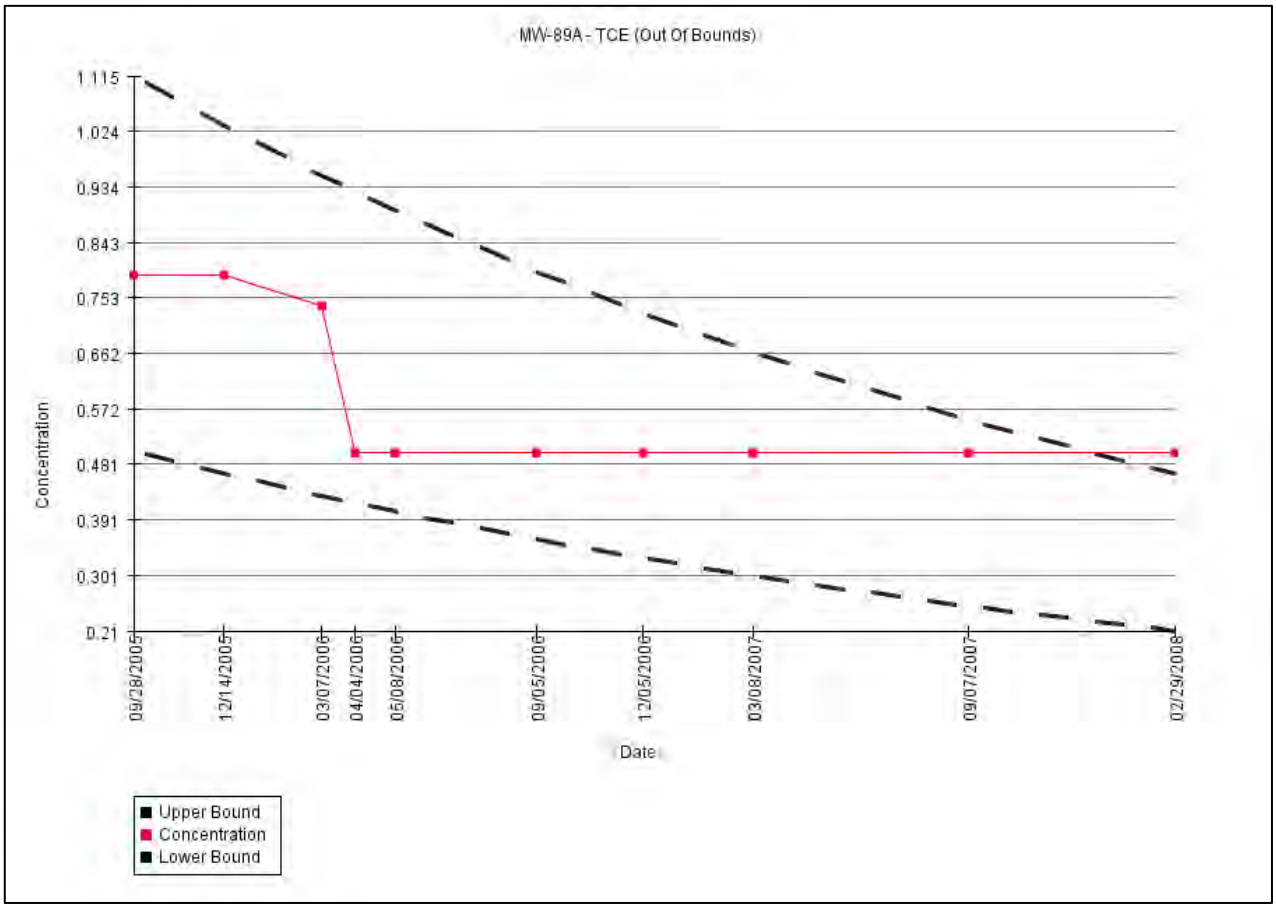
DT successfully identified many artificially introduced anomalies that would be of concern. An illustration is provided in Figure 5-12, which is from the NOP site. This well location is at the toe of a TCE plume, and in this case EnviroStat added an artificially high value at the well in the "current" data to represent potential plume migration. DT successfully flagged the value as "out-of-bounds". The software does not provide any interpretation regarding location of the point, so the analyst must have knowledge of the location to conclude that this is potentially "plume migration at the toe of the plume".

Figure 5-12. Example Data Tracker Plot With "Out-of-Bounds" Value
That is a Potential Concern



Many of the out-of-bounds values were immediately characterized by the GeoTrans mid-level analyst as "not-of-concern" because all of the concentrations are so low. An example of such a case is illustrated in Figure 5-13, which is from the NOP site. On Figure 5-13, the last point, which is current data, is above the upper prediction limit. However, the prediction limits show a decreasing trend because of the initial three values in the background data, which were all reported as "J" values. The remaining values were all non-detect, though with varying detection limits. All the non-detects were assigned a graphing value of 0.5 ug/L. Although the current data value is out-of-bounds, the analyst quickly categorized this as "not-of-concern" because all the values are below any concentration that is of interest.

Figure 5-13. Example Data Tracker Plot with "Out-Of-Bounds" Value That is Not a Concern



* First three values were "J" values, remaining values were non-detect an assigned as 0.5 ug/L

Limitations noted regarding DT include the following:

- In DT, the plots of concentration versus time do not use different symbols to differentiate between the "background data" and the "current data". The software would be improved if different symbols were used.

- In DT, the software does not allow specific historical values to be imported and plotted on graphs but not used for calculation of the prediction limits. If some historical values are considered potentially anomalous, those values have to either be included (such that prediction limits are impacted) or completely ignored. The software would be improved if such values could be imported with a flag so that they can be included on concentration versus time plots (with a different symbol) but not used to calculate the prediction limits.

- The DT portion of the software does a very good job of identifying unexpected values, but does not indicate whether the concentration trend for a specific COC at a specific well is increasing, stable, or decreasing. The software would be improved if that functionality was added. As currently configured, increasing trends may or may not

exceed prediction limits, particularly if the background data are frequently updated (i.e., the prediction limits will continue to increase due to the increasing background values).

Also, determining when and how to update the background data is not straightforward. During the follow-up analysis for the GAFB site, the GeoTrans mid-level analyst concluded there was not a compelling reason to not update the previous background data with the 2006 data prior to using DT on the newer 2007 data. However, EnviroStat pointed out that with continual updating of background there may be cases with a slowly increasing trend where each individual observation would be "in-bounds", because the bounds would widen with each additional observation that is added to the background. DT does not currently indicate which wells have increasing trends, it simply flags "out-of-bounds" values. This leads to a basic question "What should DT bring to our attention?". If the goal is for DT to be sensitive to increasing trends (i.e., indicate the new values are "out-of-bounds"), then not updating the background values would be preferred. However, that approach would ultimately require periodic updates of the background data.

5.5.10 Import/Export Features

The software allows the input data to be imported as CSV files, which can be easily prepared with commonly used software (e.g., MS-Excel). More details regarding Data Preparation were provided earlier (Section 5.2). The software allows the user to export the following files:

- The plume maps, uncertainty maps, and variogram charts (for kriging model type only) can be exported as image files (.png files)

- A tradeoff curve for each COC can be exported as an image file

- A file containing, for each optimization sampling plan on the tradeoff curve, which wells are recommended to be "on" or "off", the maximum concentration error for each COC, and the sampling cost (i.e., number of wells which are on) can be exported as a CSV file

- For each optimization sampling plan on the tradeoff curve, an individual listing of which wells are on and which wells are off can be exported as a CSV file (for spatiotemporal analysis the frequencies are also exported)

- For each potential optimization sampling plan on the tradeoff curve (and for the current sampling plan), a plume map can be exported as image file for each COC with symbols indicating which wells are on and which wells are off, with "+" indicating wells that are recommended to be removed from the monitoring network and "o" indicating remaining active wells.

The software does not allow for download to mapping formats such as ESRI shape files. This was noted as a limitation by the EPA Region V group that applied the software.

5.5.11 Computation Time and Level of Effort

A summary of the amount of time it takes to apply the software is indicated in Table 5-6. This is primarily an indication of the computation time, though the data preparation task is primarily associated with manual labor. The computation time provided for Model Builder, spatial optimization, and spatiotemporal optimization are for each problem simulated (e.g., each aquifer was a separate problem for the demonstration sites). Also, additional time beyond the computation time is required to interpret results.

Table 5-6. General Summary of Time Required to Apply the Software

Task	Time*	Comments
Data Cleanup, Screening, and Formatting	Several days (labor)	Similar effort is needed to apply any LTMO software; effort primarily manual labor
Model Builder	Minutes**	More time for kriging and higher spatial resolution
Spatial Optimization	Minutes to Hours**	Computation time increases with "population size" and "generation number" for GA
Spatio-Temporal Optimization	Hours to Days**	Computation time increases with "population size" and "generation number" for GA
Data Tracker	Minutes to hours (computation)	Preparation of data and plotting results requires most of the time

*for tasks where computation time is indicated, additional time is required for interpretation of results
**computation time per problem (e.g., per aquifer)

A more detailed summary of computation time required for the Model Builder and redundancy evaluation at the different demonstration sites is presented in Table 5-7. On that table, the results are based on kriging and quantile transformation, which as discussed was the preferred model with respect to the representation of the plume. The ranges in computation time are due to some of the other parameters such as number of vertical slices (resolution) and the GA parameters such as population size. It is also noted that the programming of Model Builder was improved after the evaluation of the Camp Allen site, leading to faster computation speeds for the other two sites.

Table 5-7. Summary of Computation Time Versus Number of Wells
(Kriging with Quantile Transformation)

Site Name	**Aquifer**	**# of samples per year**	**Computation Time***		
			Model Builder (kriging interpolation)	Sampling Optimizer	
				SO	STO
Camp Allen Site	Shallow	42 for SO	2-4 hours**	30-60 minutes	N/A
	Deep	31 for SO 21 for STO	2-4 hours**	30-60 minutes	5-6 hours
GAFB	Upper	55 for SO 108 for STO	20 minutes	10-30 minutes	several days
NOP	Shallow	81 for SO	~5 minutes	15-50 minutes	N/A
	Intermediate	84 for SO	~5 minutes	15-50 minutes	N/A
	Deep	56 for SO	~5 minutes	15-50 minutes	N/A

N/A = Not Analyzed; SO = Spatial Optimization; STO = spatiotemporal optimization
**Pentium 4, 3.2 GHz*
***the programming was changed to increase speeds after the Camp Allen site was evaluated*

Tables 5-8 To 5-10 provide further detail for each site, illustrating the difference in computation time as a result of the model type (i.e., combination of interpolation method and data transformation). Kriging results in higher computation times, though that increase was largely mitigated by programming improvements after the evaluation of the Camp Allen site.

Table 5-8. Computation Times for Different Model Types, Camp Allen Site

Interpolation	**Data Transformation**	**Computation Time***		**Comment Regarding Model Quality**
		Model Builder	Sampling Optimizer	
Inverse Distance Weighting	None	10-20 minutes	20 minutes for SO; N/A for STO	Bad
	Logarithm	10-20 minutes	20 minutes for SO; N/A for STO	Reasonable
	Quantile	10-20 minutes	20 minutes for SO; 5-6 hours for STO	Reasonable
Kriging**	None	2-4 hours	30-60 minutes for SO; N/A for STO	Bad
	Logarithm	2-4 hours	30-60 minutes for SO; N/A for STO	Questionable model for VC
	Quantile	2-4 hours	30-60 minutes for SO; 5-6 hours for STO	**Best**

N/A = Not Analyzed; SO = Spatial Optimization; STO = spatiotemporal optimization
**1GB RAM PC (Pentium 4, 3.2 GHz)*
***before the kriging interpolation algorithm was improved*

Table 5-9. Computation Times for Different Model Types, GAFB Site

Interpolation	**Data Transformation**	**Computation Time***		**Comment Regarding Model Quality**
		Model Builder	Sampling Optimizer	
Inverse Distance Weighting	None	5-10 minutes	N/A	Bad
	Logarithm	5-10 minutes	N/A	Acceptable
	Quantile	10-15 minutes	10-30 minutes for SO N/A for STO	Good
Kriging**	None	10-15 minutes	N/A	Bad
	Logarithm	10-15 minutes	N/A	Good
	Quantile	20 minutes	10-30 minutes for SO; Several days for STO	**Best**

*N/A = Not Analyzed; SO = Spatial Optimization; STO = spatiotemporal optimization **2GB RAM PC (Pentium 4, 3.2 GHz)*
***after the kriging interpolation algorithm was improved*

Table 5-10. Computation Times for Different Model Types, NOP Site

Interpolation	**Data Transformation**	**Computation Time***		**Comment Regarding Model Quality**
		Model Builder	Sampling Optimizer	
Inverse Distance Weighting	None	5 minutes	10-20 minutes for SO	Bad
	Logarithm	5 minutes	10-20 minutes for SO	Acceptable
	Quantile	5 minutes	10-20 minutes for SO	Acceptable
Kriging**	None	5 minutes	15-50 minutes for SO	Bad
	Logarithm	5 minutes	15-50 minutes for SO	Good
	Quantile	5 minutes	15-50 minutes for SO	**Best**

N/A = Not Analyzed; SO = Spatial Optimization; STO = spatiotemporal optimization
**2GB RAM PC (Pentium 4, 3.2 GHz)*
***after the kriging interpolation algorithm was improved*

The EPA Region V group that tested the software reported that they only performed spatiotemporal redundancy evaluation, and that each optimization simulation took approximately four to six hours. This is consistent with the results provided above, which is that spatiotemporal redundancy evaluation requires hours to days of computation time for each problem solved.

5.6 SAMPLING METHODS

No samples were collected by the ESTCP project team as part of this project. The data that were utilized were from sampling results previously obtained by the demonstration sites under their site-specific sampling plans.

5.7 SAMPLING RESULTS

Again, no samples were collected by the ESTCP project team as part of this project. The data that were utilized were from sampling results previously obtained by the demonstration sites under their site-specific sampling plans.

6.0 PERFORMANCE ASSESSMENT

6.1 QUALITATIVE PERFORMANCE OBJECTIVES

6.1.1 User Functionality

The expected performance metric is that the Summit Software has an acceptable learning curve (e.g., 1-2 days) that will not discourage prospective users and will allow users to achieve the intended objectives. The purpose of this performance objective is to indicate if a typical data analyst (e.g., site personnel and/or their consultants) will be able to apply the technology being demonstrated. This was evaluated in this project by using a mid-level data analyst with no previous LTMO experience to apply the software at the demonstration sites. For the first site, a mid-level analyst was provided a one-day training session by the software developer prior to using the software. For the second and third demonstration sites, a different mid-level analyst was used, and that person relied only on the software documentation plus phone support by the software developer. Based on the application of the software at all three demonstration sites, this performance objective was met.

6.1.2 Software Reliability

The expected performance metric is that the Summit Software has no significant errors or bugs remaining by the end of this project. The purpose of this performance objective is to identify if there are any reliability issues associated with general future use of the software. This was evaluated in this project by testing the software at three different sites, and analyzing multiple scenarios at each site. Any bugs that were identified were reported to the software developer, who then fixed the problems such that the performance objective was met.

6.1.3 Model Builder Performance

The expected performance metric is that Model Builder provides a model of spatial and/or temporal variation for each primary COC at each site that is adequate given the available data. The purpose of this performance objective is to indicate if the modeling component of the software is adequate, since the subsequent mathematical optimization performed by the Sampling Optimizer portion of the software is based on the underlying model developed within Model Builder. This was evaluated in this project by testing the various Model Builder options at the three demonstration sites. For instance, as discussed earlier, Model Builder allows for two different interpolation techniques and three different data transformation techniques, for a total of six combinations. The reasonableness of all six combinations was evaluated at each of the three sites. In each case, the analyst reviewed the plume map generated by the software for each combination of interpolation and transformation, and qualitatively assessed the reasonableness of the plume map.

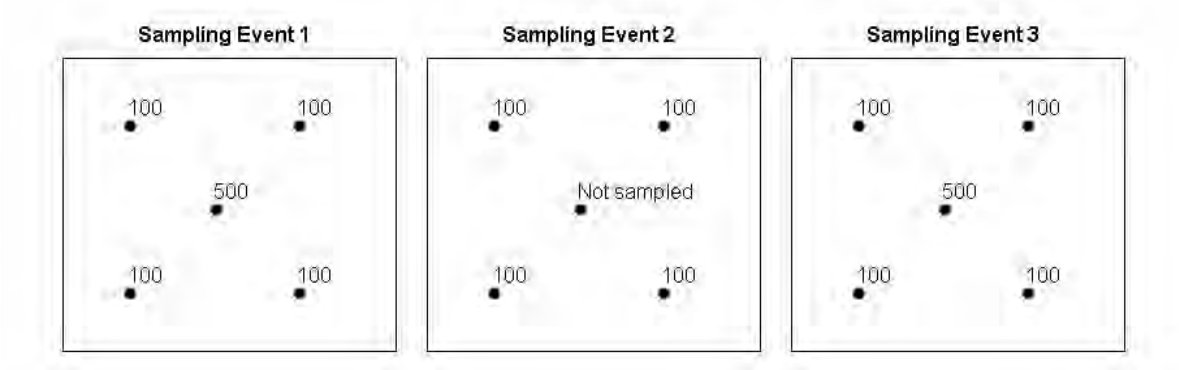
The qualitative performance objective for Model Builder was met with respect to modeling of spatial variation. With respect to spatial variation, the analyst established that the combination of kriging with quantile transformation qualitatively provided the most reasonable representation of the data, which is consistent with the recommendations in the software manual. In general, using no data transformation resulted in unacceptably poor representation of the data.

The qualitative performance objective for Model Builder was only partially met with respect to modeling of temporal variation. Model Builder addresses temporal changes in concentration in support of several aspects of software functionality, such as:

- Spatiotemporal optimization for reducing sampling redundancy (in both space and time)
- Calculation of relative mass or mass flux in different sampling events, based on interpolation of sampling results in each event

It became apparent during implementation of the software at the demonstration sites that, for each of these software functions listed above, Model Builder performs a series of spatial interpolations over time, but does not perform any interpolation with respect to time. This is easily explained with a conceptual example, illustrated in Figure 6-1.

Figure 6-1 Conceptual Example to Illustrate Temporal Interpolation Issue



In Figure 6-1, the only difference between the three sampling events is that the middle point is not sampled in Sampling Event 2. The software only performs spatial interpolation within each event. Therefore, in Sampling Event #2, the software will interpolate values inside the four actual samples based only on those four values. Since those values are each 100, presumably the software will interpolate a value of 100 at each location inside the four actual values. However, it seems unlikely that an environmental scientist provided with the data illustrated in Figure 6-1 would consider that the most reasonable approach. Rather, a temporal interpolation at the location not sampled in Sampling Event 2 based on the other two events would appear to be more appropriate, or perhaps some combination of temporal and spatial interpolation. More specifically, an environmental scientist would likely conclude that that the value at the location not sampled in Sampling Event 2 is probably close to 500, and then perform subsequent spatial interpolation in the remaining area accordingly.

Because of this issue, the GeoTrans analyst concluded that the qualitative performance objective regarding Model Builder is only partially met with respect to temporal variation. Temporal variation is modeled adequately only if the sampling locations are consistent between sampling events, or if the user fills in missing values in specific events manually prior to use of the software, based on considerations including temporal interpolation. Otherwise, the representation of temporal variability may not be adequately modeled within Model Builder.

6.1.4 Sampling Optimizer Performance

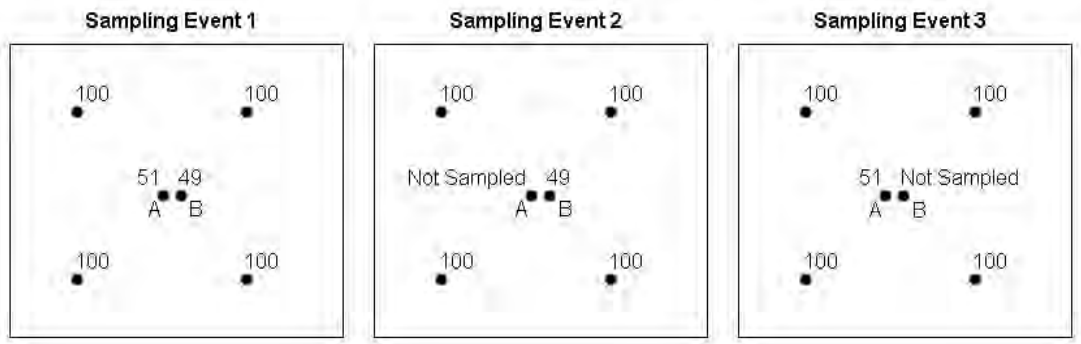
The expected performance metric is that Sampling Optimizer provides reasonable trade-off curves allowing site personnel and other professionals to easily identify optimal monitoring program choices. The purpose of this performance objective is to indicate if the software successfully addresses the key tradeoff between sampling cost and sampling error (i.e., error will increase as the number of samples is decreased). This was evaluated in this project by asking a mid-level analyst to solve spatial optimization formulations at all three demonstration sites, and to solve spatiotemporal optimization problems if possible (this was ultimately performed at two of the three demonstration sites).

Based on the application of the software at all three demonstration sites, the qualitative performance objective for Sampling Optimizer was met for spatial optimization, but was only partially met for spatiotemporal optimization. In all cases the software did produce trade-off curves that identify optimal monitoring program choices. In fact, The EPA Region V group noted that a particularly important feature of the software is “the ability to evaluate/produce an ensemble of potential sampling plans: this is not common in other software developed for LTM purposes”.

However, in the case of spatiotemporal optimization, the GeoTrans analyst noted that the results seemed overly conservative because the software tended not to eliminate wells entirely from the recommended sampling plans. This was potentially not reasonable given the corresponding spatial optimization results which suggested that many wells could be eliminated with acceptable levels of error. Details regarding specific results for demonstration sites that pertain to this issue were presented in Section 5.5.6. However, further explanation of an underlying issue associated with spatiotemporal optimization is provided below.

Spatiotemporal results are inherently more conservative than the spatial results alone, because the optimization problem has additional constraints. However, the fact that the software does not perform temporal interpolation of data provides an additional component of conservatism. The reason why the spatiotemporal optimization can be overly biased towards not eliminating wells is illustrated using a conceptual example presented in Figure 6-2. This figure illustrates data values from three sampling events.

Figure 6-2. Conceptual Example to Illustrate Spatiotemporal Optimization Issue



On Figure 6-2. note the locations labeled "A" and "B" in the middle of each event. In Sampling Event 1, both of those locations are sampled. In Sampling Event 2 location "A" is not sampled, and in Sampling Event 3 location "B" is not sampled. It appears based in inspection that locations "A" and "B" appear to be redundant. It is likely that the missing value at "A" in Sampling Event 2 is close to 51, and likely that the missing value at "B" in Sampling Event 3 is close to 49. Furthermore, it would not cause significant interpolation error if either "A" or "B" was permanently eliminated, as long as a value at the other well was available.

However, the software would likely not reach the same conclusion. Spatiotemporal optimization, as implemented in the software, provides a tradeoff curve of sampling cost versus sampling error. The general manner in which error is calculated in the software was discussed in Section 2.1. In a spatiotemporal optimization, the error calculated by the software for each potential sampling plan is the maximum error at any sampling location in any sampling event, calculated at locations/times where a sample is removed, *based on spatial interpolation using the remaining samples in that event*. As discussed with respect to Model Builder, no temporal interpolation is performed within the software to address the missing values as part of these calculations. Thus, in the example in Figure 6-2, the software would generally not eliminate location "A" because it would lead to a large error in Sampling Event 3 at location "A" (because no value would be assumed at location "B" during that event). Similarly, the software would not eliminate location "B" because it would lead to a large error in Sampling Event 2 at location "B" (because no value would be assumed at location "A" during that event). Thus, the ultimate result from the software would likely include continued sampling at both wells at some frequency, rather than eliminating one of the two wells.

This underlying issue will only be significant when well locations are not consistent for each sampling event. However, it is very often the case that at least some sampling events will have different sampling locations than others (due to different sampling frequencies, new wells, abandoned wells, and wells that cannot be sampled in specific events due to logistics). Because the software does not provide for temporal interpolation, those missing values are not adequately represented within the spatiotemporal optimization process. The assignment of these missing values can be assigned external to the software, but that is potentially a labor-intensive process. Moreover, assigning values to missing data using any form of interpolation involves making subjective judgments about the temporal smoothness of the data; as a result, a temporal optimization will be evaluating some combination of the actual temporal redundancy and the

temporal redundancy of the subjective judgments, rather than evaluating only the actual temporal redundancy. It seems likely that this issue would be problematic to any approach to temporal optimization that one might attempt to use with datasets containing numerous missing values.

Therefore, in summary, the software does produce reasonable tradeoff curves for spatial optimization, but in some cases where sampling locations are not consistently sampled, the tradeoff curves for spatiotemporal optimization may be overly conservative with respect to elimination of wells.

6.1.5 Data Tracker Performance

The expected performance metric is that Data Tracker enables the easy incorporation of site-specific monitoring and remediation expectations and data objectives along with historical data. The purpose of this performance objective is to indicate if Data Tracker includes functionality to track recently collected data versus site-specific objectives and historical data. This was evaluated in this project by testing Data Tracker at the three demonstration sites, using historical data (or in some cases a subset of the historical data) for the "background data" and recently collected data to evaluate the tracking functionality.

The qualitative performance objective for Data Tracker was partially met. Data Tracker was found to be easy to use. Data Tracker does allow tracking of recently collected concentration data against statistically computed bounds that are calculated from the historical (i.e., background) data. The software indicates which of the recently collected data values are "out-of-bounds" relative to expectations that are calculated from the background values. However the software does not appear to track any quantity with respect to "remediation expectations" or "site-specific objectives" as stated in the performance objectives. With respect to concentrations, the software does not indicate which values are above or below remediation goals. With respect to other parameters that can be tracked by the software, such as mass and mass flux, the software will provide a table of historical versus current values, but it does not compare these to any site-specific remediation goals for those parameters. Therefore, the software does effectively track recently collected data versus historical data, and does allow quantities such as relative plume mass and mass flux across a site boundary to be tracked, but does not specifically address "site-specific remediation expectations" within the software. In fairness, one should note that tracking progress toward site-specific remediation goals did not arise in discussions of objectives with site personnel at any of the three demonstration sites.

6.1.6 Regulatory Acceptance

The expected performance metric is that the results of this ESTCP dem/val will be persuasive to regulatory personnel. The purpose of this performance objective is to indicate if the results provided by the software, particularly those associated with eliminating data redundancy, might be acceptable to regulators. This was evaluated in this project by receiving feedback from site personnel associated with the three demonstration sites. In each case to date where presentations of results have been made to the installation (Camp Allen and GAFB sites), site personnel

indicated that the types or tradeoff curves produced by the software for evaluating redundancy (based on mathematical optimization), in conjunction with the comparison of plume visualizations with and without redundant data that are produced by the software, would be convincing.

6.1.7 Comparison with MAROS (secondary)

The expected performance metric is that the Summit Software will be found to be at most modestly more difficult to learn to use compared to MAROS. The purpose of this performance objective is to indicate if the Summit Software is significantly more difficult to learn to use than MAROS. This was evaluated in this project by applying both software products at the first site, conducted by the same analyst, with equivalent training in both software products. Based on this effort, the performance objective was met. The analysts reported that both software products were equally easy to learn and implement. In the case of applying MAROS, there was some initial confusion about how to utilize some of the input parameters (current plume length, distance from source, etc.) for the specific problem being solved, but those were easily addressed by corresponding with the software developer. Similarly, there were some minor questions about what values to assign for some of the Summit Software parameters, and those were easily addressed by corresponding with the software developer (some of those issues were subsequently addressed by improvements made to the software and added to the User's Manual during the remainder of the project). More detail regarding the comparison between the Summit Software and MAROS is presented in Appendix F.

6.2 QUANTITATIVE PERFORMANCE OBJECTIVES

6.2.1 Model Builder Performance

The expected performance metric is that Model Builder provides a model of spatial and/or temporal variation for each primary constituent of concern at each site that is adequate given the available data. The purpose of this performance objective is to indicate if the "modeling" component of the software is adequate, since the subsequent mathematical optimization performed by the Sampling Optimizer portion of the software is based on the underlying "model" developed within Model Builder. This was evaluated in this project by testing the various Model Builder options at the three demonstration sites, and making visual comparisons. It is not clear that there is a more quantitative manner to evaluate this objective. Thus, the conclusions are the same as stated for the qualitative performance objective for Model Builder (see Section 6.1).

6.2.2 Sampling Optimizer Performance

The expected performance metric is that the optimized programs identified by Sampling Optimizer in fact permit cost reductions with acceptable losses of information, if appropriate, as anticipated for the large majority of DoD sites. The purpose of this performance objective is to

indicate if sampling plans were recommended that reduce cost with acceptable levels of error. This was evaluated in this project by testing Sampling Optimizer at the three demonstration sites. Based on those efforts, this performance objective was met. At each of the threes demonstration sites, sampling plans were selected from the tradeoff curves (sampling cost versus error) with substantially lower cost that the analyst felt had acceptable error (see results presented in Section 5.5.6). Furthermore, at each of the three sites validation was performed on a subsequent data set to determine if the errors that might result from the reduced sampling in the recommended plans were acceptable, and in each case it was confirmed that a plan recommended by the software with significantly reduced sampling resulted in acceptable error when applied to the reserved dataset. The validation exercise provided confidence in the results provided by SO.

6.2.3 Data Tracker Performance

The expected performance metric is that Data Tracker responds appropriately to artificially induced anomalies of interest for the particular site. The purpose of this performance objective is to indicate if Data Tracker will detect certain types of unexpected data values that might occur due to human error, database software error outside of the LTMO software, or abrupt changes in field conditions. This was evaluated in this project by having EnviroStat create artificial anomalies in consultation with site and other personnel at each of the three demonstration sites. Those data were provided to the software analyst from GeoTrans who used Data Tracker in a single-blind evaluation (i.e., the GeoTrans analyst was provided six variations of the actual "current data" with no external clues as to which was correct and which had artificial anomalies. EnviroStat subsequently evaluated the extent to which the GeoTrans analyst was able to detect these artificial anomalies with the software, and provided a summary write-up for each site. Based on those efforts, this performance objective was met. The vast majority of artificial anomalies were detected (more details provided in Section 5.5.9 and in Appendix B to D).

6.2.4 Comparison with MAROS

The expected performance metric is that: 1) the Summit Software will be found to be at most modestly more difficult to learn to use, consistent with being much more flexible in incorporating site-specific monitoring objectives; and 2) if both products are able to accept the same goals and constraints, results will be similar but slightly different due to small differences due to different optimization methodologies. There were no pre-conceptions regarding the comparisons of optimization recommendations to be expected from the two software products.

With regards to ease of use, GeoTrans personnel received training for both the Summit and MAROS tools by the software developers. The Summit training was a little more than a half-day, and the MAROS training was approximately a half-day hands-on training that was part of a two-day LTMO conference. Both software products were similarly easy to install and could be learned and used by people with similar training and qualifications. Both User's Guides are comprehensive and clearly presented. Similarly, to get historical site data into the required format for input to the software was no more significant for one software versus the other software; the user simply needs to follow the instructions regarding the input structure and

requirements for that software product. These modifications took on the order of minutes to several hours for each software product for someone experienced with performing such operations in MS Excel or MS Access.

With regards to results, comparisons of specific recommendations provided by each software package are difficult, as explained in detail in Appendix F. Key comparison observations that can be made include the following:

- The primary advantage of the Summit Software is that the redundancy evaluation is based on mathematical optimization which allows sampling redundancy to be evaluated on a system-wide basis (e.g., best solution if one location is removed, if two locations are removed, if three locations are removed, etc.). A key benefit of this approach is that it allows the tradeoff between the number of samples and the accuracy of the resulting plume interpolation to be assessed. This is a significant improvement over the approach for evaluating data redundancy utilized in MAROS, which is not based on mathematical optimization. In MAROS, individual well locations are evaluated for redundancy based on impacts of removing that well alone; consequently, the impact of removing groups of wells cannot be assessed and the aforementioned tradeoff cannot be evaluated.

- The Summit Software approach to data redundancy evaluation provides plume visualizations for the baseline plan (i.e., all samples) versus improved plans (i.e., reduced numbers of samples) within the software. These comparative visualizations are quite effective for communication with stakeholders and regulators. However, these maps can only be exported as image files, thus, it is difficult to directly import these image files into other software packages such as Surfer, ArcGIS, and AutoCAD. MAROS does not include such plume visualizations.

- The Summit Software has a Data Tracker module that indicates if new data are "in-bounds" or "out-of-bounds" relative to expectations, based on previous data at that well. This functionality is useful but is not present in MAROS. MAROS indicates if the concentration trend at a well is increasing, decreasing, or stable. This functionality is also useful, but is not present in the Summit Software.

- Both software products suffer from some similar limitations. With respect to areas of uncertainty, neither software package provides specific recommendations (i.e., number of new wells and/or locations of new wells) to reduce the uncertainty. With respect to mass calculations, neither software performs temporal interpolation or extrapolation to fill in missing values in events where specific wells are not sampled. As a result, mass or mass flux results will have higher variability and uncertainty for events with fewer samples, different spatial distribution of samples, and/or for events where key wells (e.g., wells with high concentrations) are not sampled.

Additional comparison with MAROS is planned using data from the NOP site, but that comparison will be performed outside the scope of this ESTCP project.

7.0 COST ASSESSMENT

This section addresses the costs of implementing the use of this LTMO software at a DoD site and the potential cost savings that might result. The cost benefit is primarily associated with the data redundancy functionality of the software. There is some additional, but minor, cost saving potential associated with the DT functionality because unexpected values in recent sampling results can be quickly detected in a semi-automated manner.

7.1 COST MODEL

The software is free for use at DoD sites. Furthermore, since the software runs on standard desktop computers no capital purchases are required. Therefore, the cost of implementation is the estimated cost of applying the software at a typical site, and perhaps some minor training costs for initial use.

For the demonstration project, approximately $60,000 per site was allocated for testing the software. However, this is far in excess of what would be required for a typical site. This is because many potential variations were addressed during this project to allow for robust testing of the software, as detailed in Section 5.5 of this report. In Table 7-1, estimates are provided for applying the software at a typical DoD site assuming that the redundancy evaluation will be performed spatially rather than spatiotemporally.

{this gap is intentional}

Table 7-1. Estimated Costs to Apply the Software at a Typical DoD Site

Cost Element	Estimated Level of Effort	Estimated Cost
Start-Up		
Software Cost	Free	$ 0
Software download	1 hr @ $100/hr	$ 100
Training/learning	16 hrs @$100/hr	$1,600
	Subtotal	**--------->$1,700**
Redundancy Evaluation (Periodic)		
Per Site:		
Formulation	Lump sum	$5,000
Data Prep	24 hrs $100/hr	$2,400
Import Data Into Software	2 hrs @$100/hr	$ 200
	subtotal	**--------->$7,600**
*Per Plume Evaluated**:		
Model Builder	2 hrs @ 100/hr	$ 200
Optimization **	24 hrs @ 100.hr	$2,400
Interpret Results and Write Up	20 hrs @ 100/hr	$2,000
	subtotal	**--------->$4,600 (per plume)**
Data Tracker		
First Time:		
Develop Initial Background Data File	(Part of Data Prep listed above)	$ 0
Each Year:		
Evaluate Need to Update Background	16 hrs@100/hr	$1,600
	subtotal	**--------->$1,600 (per year)**
Each Event:		
Create CSV File for New Data	2 hrs @100/hr	$ 200
Import Data and Run DT	1 hrs @100/hr	$ 100
Export Charts, Print Charts, Interpret	5 hrs @100/hr	$ 500
	subtotal	**--------->$800 (per event)**

**each plume may consist of multiple primary COCs, but each aquifer or aquifer horizon where the plume is represented with a different map would be treated as separate plume*

***assumes several variations will be attempted such as changing the Model Builder algorithm or the list of excluded wells*

7.2 COST DRIVERS

The cost estimates provided in Table 7-1 are rough estimates based on the testing performed as part of this demonstration project. Some cost drivers that would potentially impact the cost of applying the software are provided below:

- *Formulation Task Will Depend on Number of People and Need for Additional Meetings.* The formulation of the LTMO problem to be addressed (e.g., which COCs to evaluate, which wells cannot be excluded, what rules to use for graphing values, etc.) is an up-front task associated with the use of the software. For most sites, this effort could be addressed during a periodic site meeting, such that all required personnel are present and no additional travel is required. For such sites, the effort may be less than the $5,000 estimated in Table 7-1, but may approach that amount considering additional time for multiple individuals to spend on this issue during the meeting. For other sites, where

periodic meetings do not occur, this effort could typically be done via conference call and/or email exchange for within the $5,000 estimated in Table 7-1. However, if a special meeting is required for this task which involves travel and labor for multiple individuals, the cost could be higher than the $5,000 estimated in Table 7-1. However, that is not expected for most applications of the software.

- *Data Preparation Cost Will Depend on the Quality of the Site Data.* During the data preparation step, site data are converted into CSV files that can be imported into the software. This includes an input file for the redundancy evaluation, and also a "background data" file for the DT evaluation (which is a subset of the historical data for each well). Obviously, the level of effort will depend on the format of the site data, and the extent to which the site data have previously been screened for data quality issues. For most sites, historical data are already available electronically, and reformatting those data into the proper format for input into the software is a simple exercise within commonly used software (e.g., MS-Excel). However, if some site data are not in digital format (as was the case for some of the data at one of the three demonstration sites) then those data may need to be entered into digital format, which could increase the data preparation cost. The estimate provided in Table 7-1 of $2,400 for data preparation assumes the data are available electronically, allows for fairly detailed screening of the data for potential data quality issues, and assumes that only minor data quality issues will be discovered (e.g., inconsistent well names and/or missing well coordinates). However, if more substantial data quality issues are determined, data preparation costs could be higher. An example would be if the screening of data quality raised questions regarding the aquifer designation of multiple monitoring wells, such that a review of the overall conceptual site model is subsequently required.

- *Redundancy Evaluation Costs Depend on the Number of Plumes.* The costs estimated for redundancy evaluation using the SO functions in the software have two components: per site costs (such as formulation and data preparation) and per plume costs (for the execution of the GA and subsequent review of optimal solutions along the tradeoff curve). For instance, if there are three different aquifers for which plume maps are interpreted, the optimization will need to be performed for three distinct problems. Thus, the "per plume" estimate of $4,600 in Table 7-1 needs to be multiplied by the number of plumes being evaluated (i.e., this component of the cost estimate would be a total $13,800 for three distinct aquifers, assuming each aquifer containing one overall monitoring network to be optimized).

- *Spatiotemporal Optimization Requires Much More Computation Time Than Spatial Optimization.* Table 7-1 assumes that spatial optimization will be applied rather than spatiotemporal optimization, for reasons discussed in other portions of this report. However, if spatiotemporal optimization is performed, the computation time will increase. This may or may not increase the actual cost of the effort, as long as computer time is not a charged quantity.

As discussed in Section 5-5.11, the computation time is somewhat impacted by the number of wells and the model type (kriging versus inverse distance weighting), but these variations should

not significantly impact the costs estimated in Table 7-1. It is also noted that the labor cost estimates in Table 7-1 are approximations that may differ from site to site.

7.3 COST ANALYSIS

A cost-benefit analysis for applying this LTMO software must account for the costs of applying the technology and the cost savings likely to be realized. The estimated costs of applying the technology were presented in Table 7-1. The costs savings will result from reduced labor and analysis associated with the elimination of some sampling.

The actual costs and savings are subject to many site-specific factors such as the number of aquifers, the number of wells, the cost of sampling, the cost of laboratory analysis, and many other factors. Since these factors vary from site to site, examples are provided below to illustrate how the costs and savings can be evaluated.

For the first scenario, the following assumptions are made:

- Evaluate a 10-year monitoring horizon
- The LTMO costs are based on the values estimated in Table 7-1
- There is only one aquifer, and the redundancy evaluation is performed on one comprehensive plume
- The redundancy evaluation is performed once at the beginning of the 10-year period, and again in year 6 of the 10-year period. To support the second periodic evaluation, the full set of monitoring wells is sampled for one of the two events in year 5 (i.e., year five only has half the savings associated with reduced amount of wells sampled). For simplicity, we assume the same level of sampling reduction after this second round of optimization versus the original baseline number of wells.
- The wells are sampled twice per year
- There are 60 total samples per sampling event in the current monitoring plan (i.e., 120 samples per year)
- The cost of collecting a sample, plus the laboratory cost for analysis, is $800
- The spatial LTMO analysis eliminates 35% of the sampling locations (representative results as described in Section 5.5.6)
- Future costs are discounted to present day dollars using a 10-year discount rate of 2.6% as per OMB (www.whitehouse.gov/omb/circulars/a094/a94_appx-c.html).

Other minor savings might occur because fewer duplicate samples and QA/QC samples (e.g., trip blanks and field blanks) may be required, but those details have not been included. Also, the additional costs of evaluating the current data for unexpected values without using data tracker are hard to quantify and are not included.

The cost benefit analysis for this scenario is summarized on Figure 7-1. The net present value of the LTMO costs is approximately $50,000 and the net present value of the LTM savings is approximately $285,000. Thus, the net savings is over $230,000 over 10 years. This relatively modest savings is because this scenario includes only one aquifer with a total of 60 wells, and a relatively low cost of $800 per sample for sampling labor, analysis, and validation.

Figure 7-1. Cost-Benefit Analysis, Scenario 1

Scenario 1											
	Assumptions										
1	# of Plumes										
5	Frequncy of Periodic Evaluation (in year)										
2	Sampling Frequency (per year)										
60	Total # of Samples per Event (all plumes combined)										
$800	Cost per Sample (labor plus analysis)										
35%	Savings for Redundancy Analysis										
2.60%	Discount Rate (OMB 10 year value)										
10	Monitoring Time Frame (in years)										
	Year										
	1	2	3	4	5	6	7	8	9	10	**Total**
LTMO Costs:											
Start-Up	$1,700										$1,700
Periodic Evaluation	$12,200					$12,200					$24,400
Data Tracker	$1,600	$3,200	$3,200	$3,200	$3,200	$3,200	$3,200	$3,200	$3,200	$3,200	$30,400
subtotal (non-discounted)->	$15,500	$3,200	$3,200	$3,200	$3,200	$15,400	$3,200	$3,200	$3,200	$3,200	$56,500
subtotal (discounted)->	$15,500	$3,119	$3,040	$2,963	$2,888	$13,545	$2,743	$2,674	$2,606	$2,540	**$51,617**
LTM Savings:											
Optimized Sampling Plan	$33,600	$33,600	$33,600	$33,600	$16,800	$33,600	$33,600	$33,600	$33,600	$33,600	$319,200
subtotal (discounted)->	$33,600	$32,749	$31,919	$31,110	$15,161	$29,553	$28,804	$28,074	$27,363	$26,669	**$285,001**
							net benefit (discounted) ----->				***$233,384***
Notes:											
Costs are based on estimates in Table 7-1											
Costs for start-up assumed to be $1,700 for training and learning											
Costs for Periodic Evaluation assumed to be $7,600 per site plus $4,600 per plume evaluated											
Costs for Data Tracker analysis assumed to be $800 per event plus $1,600 per year to assess need to update background data											
Potential savings from reduction in number of trip blanks, field blanks, duplicate samples, etc. not included											
Less savings assumed in year 5 - assume sampling at all wells in one of the two events to use for periodic redundancy evaluation											

A second scenario is the same as the first scenario except for the following:

- There are three aquifers instead of one, each with 60 wells, which has the following ramifications:
    - The redundancy analysis include three plumes rather than one plume

- The total number of samples per event in the current monitoring plan is 180 instead of 60 (i.e., 3 aquifers rather than 1 aquifer)

- The cost of collecting a sample, plus the laboratory cost for analysis, is $1,500 rather than $800 (perhaps more parameter types to be analyzed and/or more difficult sampling conditions)

With these changes, as summarized on Figure 7-2, the net present value of the LTMO costs is approximately $70,000 and the net present value of the LTM savings is approximately $1,500,000. Thus, the net savings is over $1.4 million over 10 years. This is a very substantial net benefit.

Figure 7-2. Cost-Benefit Analysis, Scenario 2

Scenario 2											
	Assumptions										
3	# of Plumes										
5	Frequncy of Periodic Evaluation (in year)										
2	Sampling Frequency (per year)										
180	Total # of Samples per Event (all plumes combined)										
$1,400	Cost per Sample (labor plus analysis)										
35%	Savings for Redundancy Analysis										
2.60%	Discount Rate (OMB 10 year value)										
10	Monitoring Time Frame (in years)										
	Year										
	1	2	3	4	5	6	7	8	9	10	**Total**
LTMO Costs:											
Start-Up	$1,700										$1,700
Periodic Evaluation	$21,400					$21,400					$42,800
Data Tracker	$1,600	$3,200	$3,200	$3,200	$3,200	$3,200	$3,200	$3,200	$3,200	$3,200	$30,400
subtotal (non-discounted)->	$24,700	$3,200	$3,200	$3,200	$3,200	$24,600	$3,200	$3,200	$3,200	$3,200	$74,900
subtotal (discounted)->	$24,700	$3,119	$3,040	$2,963	$2,888	$21,637	$2,743	$2,674	$2,606	$2,540	**$68,909**
LTM Savings:											
Optimized Sampling Plan	$176,400	$176,400	$176,400	$176,400	$88,200	$176,400	$176,400	$176,400	$176,400	$176,400	$1,675,800
subtotal (discounted)->	$176,400	$171,930	$167,573	$163,326	$79,594	$155,154	$151,222	$147,390	$143,655	$140,014	**$1,496,257**
							net benefit (discounted) ----->				***$1,427,348***
Notes:											
Costs are based on estimates in Table 7-1											
Costs for start-up assumed to be $1,700 for training and learning											
Costs for Periodic Evaluation assumed to be $7,600 per site plus $4,600 per plume evaluated											
Costs for Data Tracker analysis assumed to be $800 per event plus $1,600 per year to assess need to update background data											
Potential savings from reduction in number of trip blanks, field blanks, duplicate samples, etc. not included											
Less savings assumed in year 5 - assume sampling at all wells in one of the two events to use for periodic redundancy evaluation											

As stated earlier, many of the parameters used in these scenarios will vary from site to site. The cost analysis approach summarized in Figures 7-1 and 7-2 can be applied to any such set of parameters. This simple spreadsheet approach can be used to screen sites for potential benefits that might be realized from applying the LTMO software. For instance, for sites with few monitoring locations and infrequent sampling, the potential savings will be limited. However, the cost-benefit examples provided above clearly indicate that net savings of millions of dollars are possible across the universe of DoD sites.

8.0 IMPLEMENTATION ISSUES

The following implementation issue merit discussion:

- Software availability and documentation
- Ease of use
- Key limitations of the current software
- Regulatory issues

Each of these issues is discussed below.

Software Availability and Documentation

The anticipated end-users for Sampling Optimizer and Data Tracker include government personnel and support contractors managing groundwater monitoring programs. A copy of the software executable and user's guide is available on the Summit Sampling Optimizer website (http://www.samplingoptimizer.com/) for free and immediate download by government employees and educational users (those accessing with ".gov", ".mil" and ".edu" extensions). The software and user's guide were previously submitted as a separate deliverable under this ESTCP project. Input data files from this project that can be used as sample data have also been included on the website. This website will also be linked to the ESTCP and Federal Remediation Technology Roundtable websites.

The Summit website provides a form for contractors to government sites to fill out to obtain a license file and download link for the software. Contractors will be required to provide evidence that the software will be used at a government site (e.g., a government work order or letter from government personnel) and the license will limit the software to only be able to work with data from that site. Also, contractors will need to renew the software license annually for continuing use of the software. The free software license does not include technical support or training, which can be purchased separately (further information is available on the Summit website). Other private sector users will be able to purchase a commercial license to the software as needed. Note that this procedure is similar to those employed for other software packages such as RACER and GMS.

Ease of Use

The software was found to be easy to use, based on the application of the software by a mid-level analyst at GeoTrans with no LTMO experience. This was true for a mid-level analyst who received training on the use of the software (for one of the three demonstration sites), as well as for a mid-level analyst who did not receive training on the software (for two of the three demonstration sites).

In addition, the EPA group that applied the SO functions of the software (including Model Builder) outside of our project reported that: "The user interface was very easy to use…User's manual was an excellent reference for set-up and execution, and it contained clear directions for navigating dialog boxes, setting parameters, formatting input files, etc…It took only few hours to get comfortable using the software (import/export, model set up, running the program). The user's manual was very helpful in this aspect. It took a few days to fully understand the method, the effects of changes in parameter values, and the results."

Key Limitations of the Current Software

The software has some limitation that will impact the use of the software by end-users. Key limitations (which have already been discussed in previous portions of this report) are indicated below.

- The software interpolates spatially but does not perform interpolations in time. This impacts the tracking of mass and/or mass flux when the distribution of sampling is not consistent from event to event. It also impacts the performance of spatiotemporal redundancy analysis, resulting in more conservative results than spatial redundancy analysis when the sampling locations are not consistent from event to event. The software would be improved if there was a feature to optionally fill in missing values via temporal interpolation.

- In DT, the plots of concentration versus time do not use different symbols to differentiate between the "background data" and the "current data". The software would be improved if different symbols were used.

- In DT, the software does not allow specific historical values to be imported and plotted on graphs but not used for calculation of the prediction limits. If some historical values are considered potentially anomalous, those values have to either be included as background data (such that prediction limits are impacted) or completely ignored. The software would be improved if such values could be imported with a flag so that they can be included on concentration versus time plots (with a different symbol) but not used to calculate the prediction limits.

- The DT portion of the software does a very good job of identifying unexpected values, but does not indicate whether the concentration trend for a specific COC at a specific well is increasing, stable, or decreasing. The software would be improved if that functionality was added.

- The software does not include data consolidation or recognition of flags (e.g., for non-detect values). This requires the user to consolidate the data into sampling events during preparation of the SO input files, and to assign "graphing values" for non-detects during preparation of the input files for SO and DT. The software could be improved if this type of functionality was included within the software.

- Plume visualization for both Model Builder and Optimizer also allows users to change the zoom scale and color scale. The color scale is a linear scale allowing users to define the minimum and maximum concentrations for each COC. Then the software can plot the plume maps in color based on the minimum and maximum concentrations defined. However, it does not provide an option for a logarithmic scale, thus, for sites with a very big range in concentration, it cannot plot both high-end concentrations and low-end concentrations with sufficient detail (though multiple plots with different ranges could be made independently).

Several other minor limitations were noted by the EPA Region V group that applied the software (see Appendix E). The first bullet listed above has the most profound implications for future use of the software. For instance, since there will generally be an uneven distribution of sampling in different events, the use of spatiotemporal redundancy evaluation may not be advisable in most cases, because the results will be more conservative than those obtained using spatial redundancy evaluation (with respect to elimination of wells). An additional consideration is that spatiotemporal redundancy evaluation requires far greater computation time than spatial redundancy evaluation (for the demonstration sites, it required days for spatiotemporal simulations versus hours for spatial simulations). Thus, a prudent approach to applying the software for reduction of redundancy might be as follows:

- Perform spatial optimization rather than spatiotemporal optimization

- Determine if eliminated well locations in one or more of the recommended plans are reasonable and acceptable

- Qualitatively specify a sampling frequency for remaining locations, based on where changes in concentration are expected and/or are of greatest concern

- Develop rules for estimating the values at locations not sampled in a specific event for developing plume maps and/or for performing mass calculations (e.g., latest value, moving average of latest values, etc.)

This approach allows the user to utilize the most powerful and beneficial aspect of the software, which is the application of mathematical optimization in conjunction with multiple objectives to develop a tradeoff curve for evaluating spatial redundancy.

The EPA Region V group that applied the software utilized a slightly different approach, as described in Appendix E. They performed only spatiotemporal analysis, but first filled in missing values manually via temporal interpolation at wells. The spatiotemporal optimization was executed several times with different scenarios, including (a) different variations of the input data set, and (b) different variogram parameters. They then post-processed the results of each scenario, outside the software, to determine the fraction of times a sampling frequency was selected for each well. They assumed the sampling frequency recommended the highest percentage of time was the most robust result for that well. This approach is likely sub-optimal and requires substantial computation time. The EPA group did not perform any spatial optimization, so comparison of spatial versus spatiotemporal optimization results cannot be

evaluated. However, it is very likely that the approach recommended in the bullets above (i.e., based on spatial optimization) would likely suggest greater sampling reductions and would also require much less computation time and effort.

Regulatory Issues

Regulatory approval regarding the implementation of LTMO results provided by the software primarily pertains to the results of redundancy evaluation (i.e., the SO results). Interaction with regulators regarding implementation of results at the three demonstration sites was not a specific part of this ESTCP project. However, in each case to date where presentations of results have been made to the installation (Camp Allen and GAFB sites), site personnel indicated that the types or tradeoff curves produced by the software for evaluating redundancy (based on mathematical optimization), in conjunction with the comparison of plume visualizations with and without redundant data that are produced by the software, would be convincing.

Obtaining regulatory acceptance of the software will require two major steps: 1) increasing awareness of LTMO in general, and awareness of this software in particular, within the regulatory community; and 2) making site-specific requests to regulators for modifying an LTM program based on results of the software. With respect to the first item, the ITRC was briefed on the project approach and software applicability at the beginning of the project, and was subsequently briefed on the results of the project. In addition, the Final Report will be circulated to Kathy Yager (EPA OSRTI) and Dave Wilson (EPA Ground Water Forum), and they will be asked for their assistance in providing one or more forums where the results of this project can be presented .to the regulatory community at large (e.g., EPA NARPM meeting or Ground Water Forum meetings). With respect to the second item, the project team has offered to assist each of the demonstration sites with regulatory issues associated with LTMO, but no such assistance has been requested to date. For example, the site team at the former GAFB site indicated they would like to perform further analysis on their own, using the software, before presenting results to regulators in the form of a revised LTM plan. Also, given the long schedule of our project and the fact that the most data recent data at each site were reserved for validation in our project, the sites would be advised to repeat the analyses using up-to-date data before incorporating the results into an LTM program revision proposal. Obtaining ‘regulatory acceptance” of the software will ultimately require that LTM modifications based on software recommendations be brought before site-specific regulators, but that has not yet occurred.

9.0 REFERENCES

1. Aziz, J. J., M. Ling, H. S. Rifai, C. J. Newell, and J. R. Gonzales. 2003. *MAROS: A Decision Support System for Optimizing Monitoring Plans*. Ground Water vol.41, no. 3, pg. 355-367.

2. Cameron, K., and P. Hunter. 2002. *Using Spatial Models and Kriging Techniques to Optimize Long-Term Groundwater Monitoring Networks: A Case Study*. Environmetrics, 13, 629-656.

3. Deb, K., et al. 2002. *A Fast and Elitist Multi objective Genetic Algorithm*. NSGA-II, IEEE Trans. Evol. Computation, 6(2), pg182-197.

4. Goldberg, D. E. 1989. *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison-Wesley, Reading, MA.

5. Herrera, G. S., and G. F. Pinder. 1998. *Cost-Effective Groundwater Quality Sampling Network Design*. Proceeding XII International Conference on Computational Methods in Water Resources, Vol. I, Crete, June.

6. Johnson, V. M., M. N. Ridley, R. C. Tuckfield, and R. A. Anderson. 1996. *Reducing the Sampling Frequency of Ground Water Monitoring Wells*. Environmental Science and Technology, 30, 355-358.

7. Nobel, C. 2003. *Applying a Three-Tiered Approach to Monitoring Network*. Seventh International Battelle In Situ and On-site Bioremediation Symposium, June.

8. OMB (2008). *Discount Rates for Cost-Effectiveness, Lease Purchase, and Related Analyses.* OMB Circular No. A 94, Appendix C (www.whitehouse.gov/omb/circulars/a094/a94_appx-c.html)

9. Reed, P, Minsker, BS, and Goldberg, DE (2003). *Simplifying Multiobjective Optimization: An Automated Design Methodology for the Nondominated Sorted Genetic Algorithm-II.* Water Resources Research 39(7):1196.

10. Rizzo, D. M., D. E. Dougherty, and M. Yu. 2000. *An Adaptive Monitoring and Operations System (aLTMOsTM) for Environmental Management*. In Building Partnerships, Proceeding. 2000 ASCE Joint Conference on Water Resources Engineering and Water Resources Planning & Management, R. H. Hotchkiss and M. Glade (eds.), American Society of Civil Engineers, Reston, VA, CD-ROM (ISBN 0-7844-0517-4).

11. Summit Envirosolutions (2008) *Sampling Optimizer and Data Tracker: Reference Manual.*

APPENDICES

Appendix A: Points of Contact
Appendix B: Site-Specific Analysis – GAFB Site
Appendix C: Site-Specific Analysis – NOP Site
Appendix D: Site-Specific Analysis – Camp Allen Site
Appendix E: EPA Region V Evaluation
Appendix F: MAROS Comparison

APPENDIX A:

POINTS OF CONTACT

POINT OF CONTACT Name	ORGANIZATION Name Address	Phone Fax E-mail	Role in Project
Ms. Karla Harre	U.S. Navy NAVFAC ESC Code 414, 1100 23rd Ave, Port Hueneme, CA 93043	Phone: (805) 982-2636 Fax: (805) 982-4304 Email: karla.harre@navy.mil	PI
Mr. Tanwir Chaudhry	U.S. Navy NAVFAC ESC Code 414, 1100 23rd Ave, Port Hueneme, CA 93043	Phone: (805) 982-1609 Fax: (805) 982-4304 Email: tanwir.chaudhry@navy.mil	Co-PI
Mr. Robert Greenwald	GeoTrans, Inc. 2 Paragon Way, Freehold, NJ 07728	Phone: (732) 409-0344 Fax: (732) 409-3020 Email: rgreenwald@geotransinc.com	Technical Consultant, Contractor
Mr. Weiwei Jian	GeoTrans, Inc. 2 Paragon Way, Freehold, NJ 07728	Phone: (732) 409-0344 Fax: (732) 409-3020 Email: weiwei.jian@geotransinc.com	Technical Consultant, Contractor
Mr. Matthew Zavislak	Summit Envirosolutions, Inc. 1217 Bandana Boulevard North St. Paul, MN 55108	Phone: (651) 842-4248 Fax: (651) 647-0888 Email: mzavislak@summite.com	Software Developer
Dr. Barbara Minsker	Independent Consultant 2508 Waterbury Place Champaign, IL 61822	Phone: (217) 417-4198 Fax: (217)-333-4050 Email: barbara.minsker@gmail.com	Technical Consultant, Contractor
Dr. Charles Davis	EnviroStat 3468 Misty Court Las Vegas, NV, 89120	Phone: (702) 456-8994 Fax: (702) 456-8114 Email: charles.davis@envirostat-nv.com	Technical Consultant, Contractor

APPENDIX B:

SITE-SPECIFIC ANALYSIS – GAFB SITE

ESTCP Project: LTMO Optimization Software Tools
Write-Up for Former George Air Force Base Site

Sections

Background Site Information **2**

LTMO Software Demonstrated **4**

Optimization Formulation **5**

Data Preparation **7**
Spatial Optimization (SO) Dataset 9
Spatio-Temporal (ST) Data 9
Data Tracker (DT) Data 10
Other Data Preparation Observations 10

Sampling Optimizer (Including Model Builder) **11**
Brief Overview of Functionality 11
Observations Regarding Use of Sampling Optimizer Including Model Builder 13
Computation Time for Model Builder and Optimizer 15
Spatial Analysis Results 16
Spatio-Temporal Analysis Results 26
Uncertainty Analysis 34

Data Tracker Results (Reported by Charles Davis, EnviroStat) **35**
Overview and Summary 35
Details: Preparing Datasets with Artificial Anomalies 38
DT-GeoTrans Results 39
Discussion of DT Results 44
Concerning the PL Confidence Levels 45

Mass Flux and Mass Metric Results **46**
Mass Flux Calculations 46
Discussion Regarding Mass Flux 51
Mass Metric Results 53
Discussion Regarding Mass Flux With Incomplete Data Filled in Manually 54

Model Validation **63**
Model Validation for Spatial Analysis 64
Model Validation for Spatio-Temporal Analysis 67

LTMO Follow-up: George Air Force Base Site **70**
Additional Validation of Spatial and Spatio-Temporal Optimization Results 70
Data Tracker Results 77
Mass Flux and Mass Metric Results 97

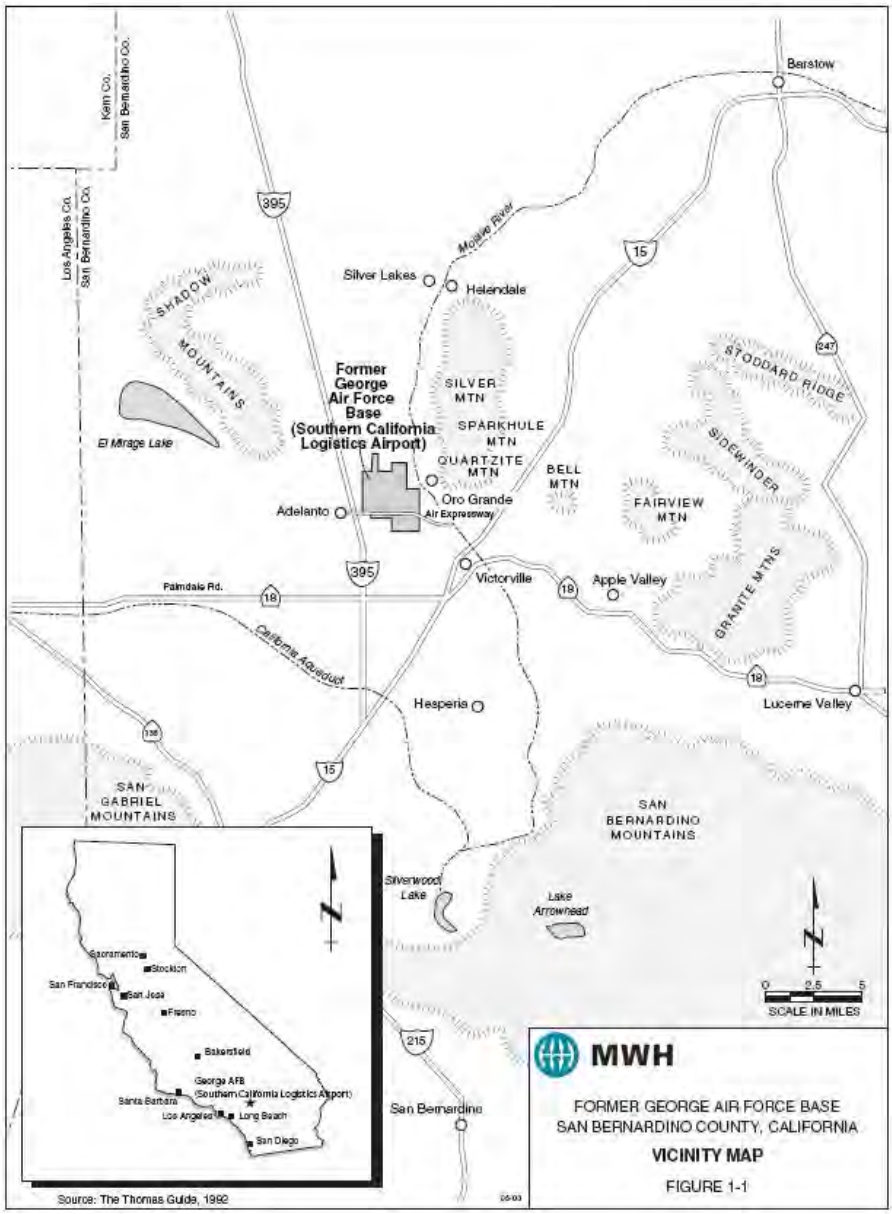
Suggested Approach for Utilizing our Results **115**

ESTCP Project: LTMO Optimization Software by Summit
Write-Up for Former George Air Force Base Site

Background Site Information

The site, the former George Air Force Base (GAFB), is located in San Bernardino County, California, approximately 70 miles northeast of Los Angeles, in the Victor Valley portion of the Upper Mojave River Basin. This site sits atop the Mojave River Bluffs on the west side of the Mojave River.

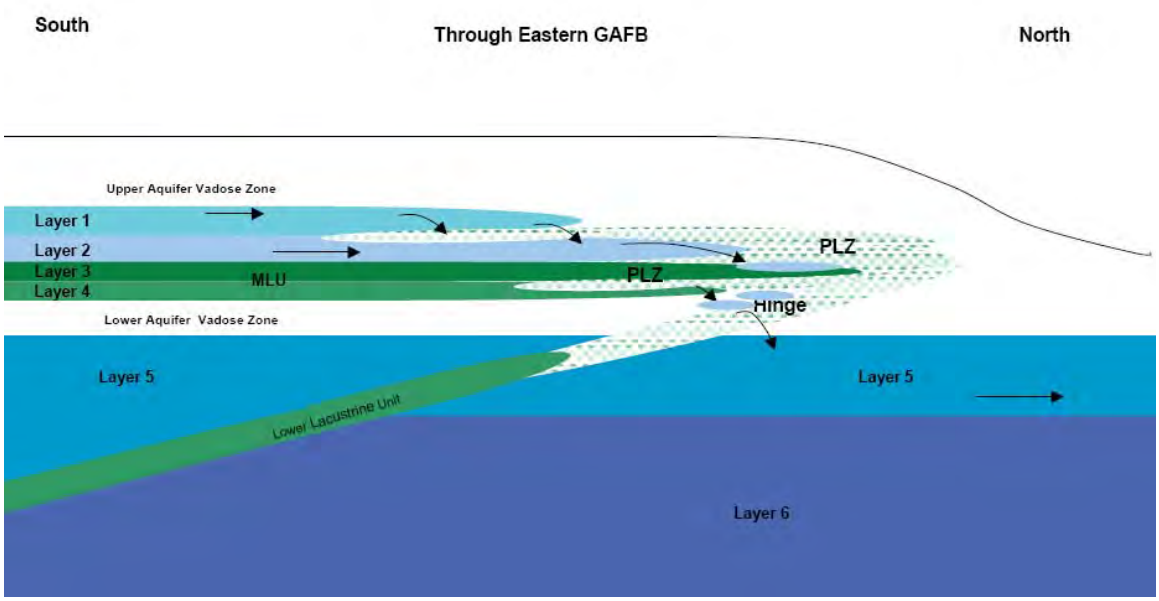
Location of George Air Force Base Landfill, San Bernardino County, CA



GAFB was established in the early 1940s and was operational until the early 1990s. Under the Federal Facilities Agreement (FFA), three Operation Units (OUs) were defined. OU-1 consists of the Upper and Lower Aquifer groundwater contaminated with TCE beneath the northeast portion of the base and adjacent off-base areas, and that OU is the focus of this effort.

The sediments beneath the GAFB site have been divided into four primary units based on their hydrogeologic characteristics. They are the Upper Fluvial Unit (UFU), Middle Lacustrine Unit (MLU), Lower Lacustrine Unit (LLU), and the Lower Alluvial Unit (LAU) from top to bottom. The UFU and LAU contain aquifers termed the Upper Aquifer and the Lower Aquifer, respectively. The Upper Aquifer is hydraulically separated from the Lower Aquifer across most of the GAFB site by the MLU, which is a low permeability layer. The Upper Aquifer is a semi-confined saturated zone contained within the lower portion of the UFU and is perched on the MLU. According to the 2005 Hydrogeologic Conceptual Site Model Report by MWH Americas, Inc. (MWH), the Upper Aquifer is continuous beneath most of GAFB site. There is a north-south trending zone roughly parallel to the Mojave River Bluffs, known as the Permeable Lacustrine Zone (PLZ), which forms the "downgradient edge" of the Upper Aquifer. Groundwater migrates from the Upper Aquifer to the Lower Aquifer through the PLZ. There also appears to be some perched water above the saturated portion of the Upper Aquifer. Groundwater in the Upper Aquifer generally flows to the northeast. There has been some historical groundwater mounding due to historical return of water that was extracted and treated via air stripping. This pump-and-treat system, consisting of up to 22 extraction wells, has been shut down since 2003.

Schematic Geologic Cross Section (N-S), Former George AFB Site



For this project, only the Upper Aquifer within OU-1 was evaluated. Contaminants in OU-1 groundwater are:

TCE	(MCL = 5 µg/l)
c12DCE	(MCL = 70 µg/l)
PCE	(MCL = 5 µg/l)
Benzene	(MCL = 5 µg/l)
Toluene	(MCL = 1000 µg/l)

TCE is the primary groundwater contaminant. Three potential source areas for TCE contamination in the OU-1 area have been identified: FT019c (fire training area); SD025 (industrial storm drain system); and FT082 (burn pit). The ROD objectives are listed below:

- To prevent exposure to contaminated groundwater that poses a risk greater than 1×10^{-6}.
- To reduce the TCE contamination in the groundwater beneath the northeast portion of the base and adjacent areas to below the federal allowable MCL of 5 µg/l.
- To eliminate or reduce the potential for further migration of the existing TCE plume in groundwater.

The groundwater monitoring goal at OU-1 is to ensure compliance with remedial action objectives, and to provide water level and analytical data for use in assessing the overall effectiveness of the remedy. There have historically been four groundwater monitoring events conducted annually for OU-1: two events for the Basewide Groundwater Monitoring Program, and two events for the OU-1 Lower Aquifer Focused Monitoring Program. Approximately, 100 monitoring wells are sampled during Basewide Groundwater Monitoring and over 10 monitoring wells are sampled during Lower Aquifer Focused Monitoring events. The purpose of the focused monitoring in the Lower Aquifer was to determine whether "shut down" extraction wells in the Lower Aquifer should be turned back on. These focused events have been discontinued.

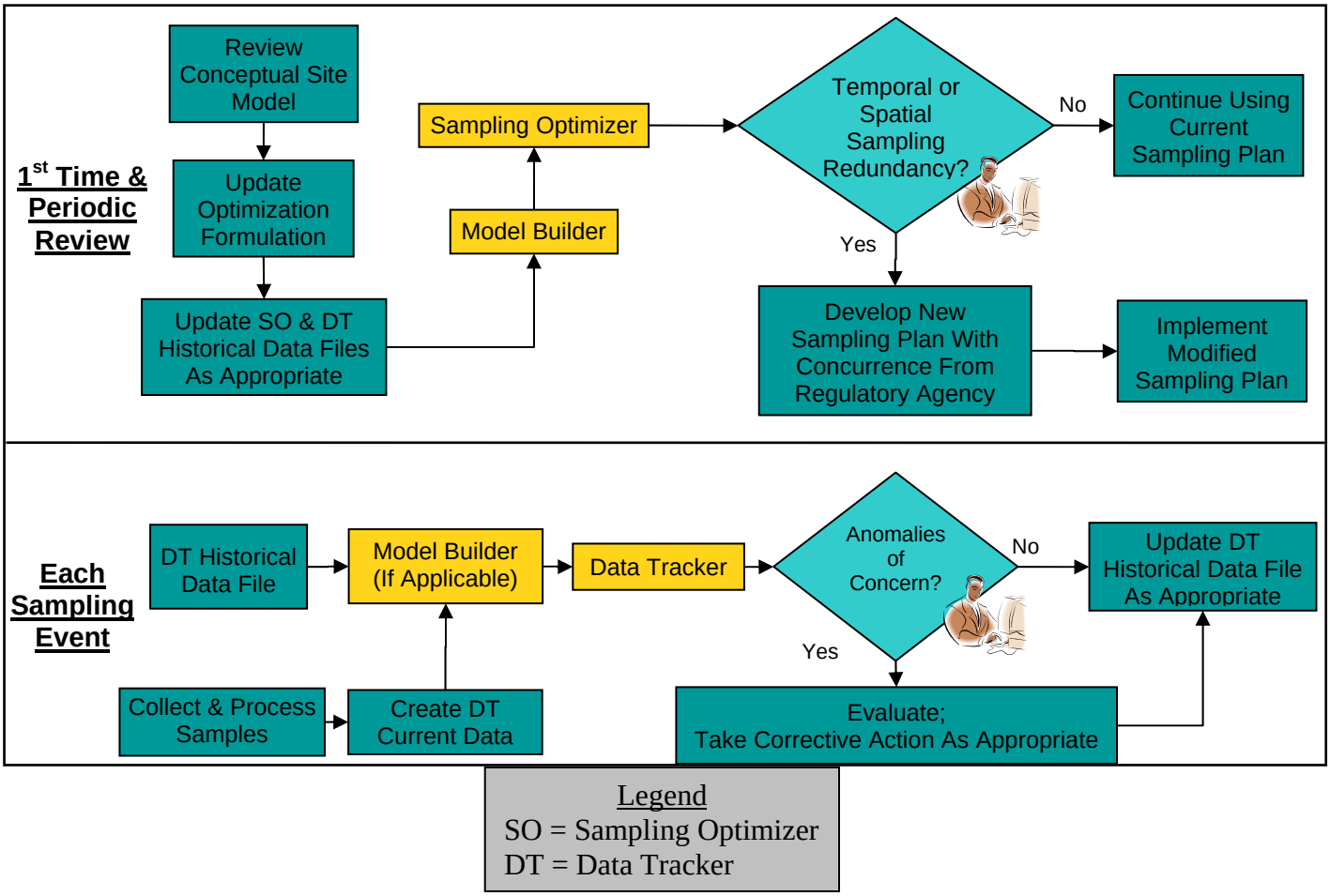
LTMO Software Demonstrated

The goal of this project is to demonstrate and validate the use of the Sampling Optimizer tools (Summit Envirosolutions, Inc.) to reduce cost and improve effectiveness of long term monitoring at DoD sites. Two major modules comprise the Summit software: Sampling Optimizer and Data Tracker.

- Sampling Optimizer identifies redundant sampling locations and/or frequencies in historical data. This module identifies redundant sampling locations and/or frequencies using a multi-objective genetic algorithm to obtain monitoring designs that represent optimal tradeoffs among two or more monitoring objectives, such as minimizing the number of samples and minimizing the concentration error (i.e., error typically increases as number of wells decreases, resulting in a tradeoff)
- Data Tracker allows current monitoring data to be reviewed against historical data to identify cases where current data deviate from expectations that are based on the historical values and patterns.

Model Builder is an additional component within the software that is utilized by the Sampling Optimizer, and in some cases, by Data Tracker. Model Builder has two sections: one for model fitting, visualization, and analysis (with kriging or inverse distance weighting), and another for visualizing relative uncertainty. A general flowchart of the software modules is presented below.

General Flowchart of Software Modules



Optimization Formulation

The project team visited the site on March 15, 2007, and a meeting with site personnel was conducted on March 16, 2007. During the meeting, several issues and considerations regarding optimization objectives and constraints were discussed. These include:

- Optimization should be conducted only for TCE in the Upper Aquifer of OU-1.
- There is a desire among site personnel for the frequency of sampling to be reduced to annually, with even less frequent sampling at some wells.
- It was found that TCE at OU-1 in the Upper Aquifer has been leaking down to the Lower Aquifer through the PLZ. Site personnel are interested in knowing the amount of mass migrating downward and tracking that mass in time.
- For Data Tracker, utilize the following constituents:

    - TCE
    - PCE
    - c12DCE
    - Benzene
    - Toluene

- Concentrations fluctuating over time at some monitoring wells suggest the existence of pulse releases to groundwater, and this needs to be considered with respect to concentration trends (i.e., Data Tracker).

- For Data Tracker, any increasing TCE concentrations that might occur in the Upper Aquifer deserve attention because they would indicate potential new sources.

- Site personnel are particularly interested in tracking concentration trends downgradient of two known source areas, and would like to see a trend toward attenuation of the plume over the long term, although local increases in concentrations may occur in the interim as plume "hot spots" shift.

- Site personnel would like to focus on the accuracy of the knowledge of concentrations at the plume edge, particularly concentrations that are near the MCL. Larger error tolerance can be accepted within the plume.

- Site personnel indicated they would allow all monitoring wells to be considered to be removable, and they would also consider adding new wells where current data are insufficient.

- A ROD amendment and a long-term monitoring plan are expected at the end of 2008.

The Optimizer provides "tradeoff curves" associated with two competing objectives:

- Minimize the number of sampling points and (optionally) the sampling frequency at retained sampling points; and

- Minimize the maximum concentration errors (emphasizing the importance of errors near the plume boundary) that result from removing specific sampling locations.

The error computation built into the objective function in the software accounts for the different significance of errors near the plume boundary versus the plume interior, placing greater emphasis on the significance of errors near the plume boundary. The user defines a cutoff concentration between low values (i.e., plume boundary) and high values (i.e., plume interior). For locations where the actual value is below the cutoff, the error is calculated as the difference between actual and estimated concentration, divided by the "acceptable error for low concentrations" for the specific parameter. The "acceptable error for low concentrations" is frequently assigned as the MCL. For locations where the actual value is above the cutoff, the error is calculated as the difference between actual and estimated concentration, divided by a specified percentage of the actual value. Details are provided in the software documentation. This function scales the difference between observed and estimated values by a different amount for different contaminants in the plume boundary area (e.g., by the MCL for each contaminant), and also diminishes the importance of errors in high concentration regions.

Several combinations of values for the "cutoff" and the "acceptable percentage error for high concentrations" were applied:

	TCE (µg/l)
Acceptable error level for low concentrations	5
Cutoff between low and high concentrations	25, 50, and 100
Acceptable percentage error for high concentrations (%)*	20% with cutoff of 25 µg/l 10% with cutoff of 50 µg/l 5% with cutoff of 100 µg/l

**percentage chosen such that there is continuity of the error function at the cutoff value*

As stated earlier, all wells were allowed to be candidates for removal.

Data Preparation

Data were received from the Former GAFB personnel in an EXCEL file containing the following information:

- Well details including ID, northing and easting, measuring point elevation and top and bottom screen depths relative to that measuring point;
- Historical water level data for March 1993-2006; and
- VOC data for July 1986-2006.

Only upper aquifer data were requested and supplied. Several other reports were also supplied. The data files were complete and consistent; no problems such as differing well IDs in different files or reports, missing coordinates, or missing screen depths, etc. were found. The VOC data were in a standard format (one row per constituent per well per event). EnviroStat used this data file to prepare CSV files to be used as input to the Summit Monitoring Tools as well as other information displays for use by the project team:

- First, a master EXCEL file containing several worksheets was created. The constituents of interest (TCE for SO; TCE, c12DCE, PCE, Benzene, and Toluene for DT) were arranged into columns keyed to a common well and sampling date. Separate worksheets were created for the historical (1986-2005) and current (2006) data. Worksheets were created for the well location and depth information and for the water level histories.

- Summaries of data availability were prepared and included in the master file, as were Time Series Plots (TSPlots) of TCE data (July 1986-2005) at wells with "interesting" histories. TSPlots of water level (March 1993-2005) were included as well.

- Summit personnel had noticed during the site visit that certain wells had been excluded from the groundwater level contouring. The notation on the potentiometric surface maps was that these wells were screened deep in the upper aquifer. These wells were given special attention in the TSPlots of water level and in the plots of screened interval depth prepared by EnviroStat and included in the master data file.

- Data were reviewed for inclusion or exclusion with regard to use for by Sampling Optimizer (SO) and Data Tracker (DT). The master data file includes all data in one pair of columns (value and "less-than" flag) for each VOC, along with an additional set of columns containing the EnviroStat recommendations for data to actually be used.

- A version of the master file excluding the current (2006) VOC and water level data was supplied to GeoTrans and the rest of the project team.

The master file was then used to prepare the CSV files of historical data. Two versions were prepared for each of SO (spatial optimization only) and SO-st (spatiotemporal optimization): one including the wells with particularly deep screens, and one excluding those wells. Some further minor data massaging was needed at this step. Field duplicate values were averaged with regular sample values, with some adjustment where one value was a nondetect and one not or where the two values were nondetects with different reporting limits. Nondetects (NDs) with typical reporting limits were replaced by a low "graphing value" (selected to be 0.05 µg/l). NDs with elevated reporting limits (RLs) were omitted because they provide ambiguous information. Three wells with latest sampling dates prior to 2004 were omitted from all historical datasets. In many cases, data from sampling events prior to 1994 were reported simply as "ND" with no RL given. The data reporting procedures apparently changed that year. Accordingly, EnviroStat recommended that data obtained prior to 1994 be systematically excluded from SO-st and DT use. Only TCE was included in the SO and SO-st CSV files.

Preparation of the background/historical data CSV files for DT went through several iterations. Initially, based on discussions with Summit personnel, the CSV files contained three columns for each COC: value, less-than flag, and "use?". There are several issues here, reflecting the continuing evolution of the DT software. One is that, at the time, it was anticipated that DT would acquire the capability of handling NDs itself rather than requiring that they be handled by the user *a priori*. The handling of NDs in DT would be different from that in SO and SO-st, particularly with the advent of time-dependent bounds in DT.

Also, there are some cases where a "slug" of a COC passed through a well, so there is a question of just which historical values should be included in the background data; this issue had arisen in discussions with the site personnel. One anticipated update to DT would allow for historical values to be included in the file for plotting, but not included in the background data. A draft update to the Reference Manual was obtained during this period, which reduced the three columns to two (value or "<RL" in the first, "use?" in the second), and the CSV files was revised accordingly. The functionality to handle the "use?" column has not yet been implemented, however, so GeoTrans converted the file prepared in anticipation of these updates to a standard (one column per COC) format using one approach. EnviroStat later provided another version using another approach. The issues involved in these different approaches (ND treatment and selection of background data) and the differing results obtained are discussed further in the DT section of this report.

A CSV file containing only TCE, with aligned dates, was then prepared for use with DT-mm (mass metric tracking within DT). This was prepared using the two-column convention; GeoTrans converted it to the one-column format. Finally, CSV files of the current data, both as they are and with anomalies artificially added, were prepared for testing DT. This is described in detail in the DT section of this report.

The data preparation effort was quite straightforward for GAFB. These steps are mostly not specific to the Summit Monitoring Tools. The only data preparation item specific to the Summit Tools is an artificial alignment of sample dates to designate events for SO, SO-st, and DT-mm.

Spatial Optimization (SO) Dataset

For the Spatial Optimization, the latest values from each well were used, as long as the sampling was in 2004 or 2005. There are only three wells whose latest sampling values are earlier than 2004. They are:

- NZ-109, sampled only once, during 2003, with values similar to nearby MW-104, NZ-17, and FT-01;

- NZ-33, sampled four times up to April 2000, with values similar to nearby FT-01, MW-104, and NZ-17;

- NZ-43, sampled twice up to April 2000, with values similar to nearby NZ-42 (sharp decrease between the two measurements), and earlier value similar to NZ-22.

Consequently, an "A" version of the data, excluding these three wells (NZ-109, NZ-33, and NZ-43) was developed, with a total number of 55 wells.

The issue of widely varying screen elevations and water levels was considered, as discussed previously. There was some feeling that there may be multiple perched units rather than one continuous "Upper Aquifer". In particular, GAFB has previously excluded eight wells (MW-102, MW-104, NZ-06, NZ-10, NZ-20, NZ-30, NZ-31, and NZ-32) from their water level contouring. All these wells except NZ-10 have atypical (generally low) water levels. These wells were not included by GAFB in the Upper Aquifer plume map contouring because they are screened in the MLU/PLZ and are believed to represent perched water moving from the Upper Aquifer to the Lower Aquifer. Consequently, a "B" version of the dataset, excluding the aforementioned three wells in addition to these eight wells, was also developed, with a total number of 47 wells.

Finally, as noted previously, the "graphing value" (0.05 µg/l) was assigned for observations reported as nondetects (NDs).

Spatio-Temporal (ST) Data

The following points summarize the ST data information:

- Dataset A and Dataset B were created according to the same procedures discussed above for spatial optimization.

- All data prior to September 1994 were omitted. There are relatively few such data available, and those are limited to 1986 and a few values in 1987, 1991, 1993, and early 1994. More importantly, there are several measurements from this early period reported as "ND" with no reporting limit. The data reporting practice changed after September 1994, and most reporting limits decreased thereafter.

- The dates for data taken from "January through June" and "July through December" were normalized to 4/1/yy and 10/1/yy, respectively. This generally was consistent with sampling patterns. In a few cases there were two samples taken during the same half-year, which were averaged. The number of wells actually present during any particular semi-annual "event" varies greatly; as such, plume maps based on only data from single consecutive events may not be comparable. (Also, field duplicates, where present, were averaged with regular sample values.)

As with SO, the graphing value utilized for NDs was 0.05 µg/l, and NDs with relatively high RLs were excluded.

Data Tracker (DT) Data

For Data Tracker, all data prior to September 1994 were omitted because there were relatively few such data, and more importantly many of the measurements are reported simply as "ND" with no reporting limit. Items of note include the following:

- For values reported as non-detect (e.g., "<1.0") the value imported into DT was set to the detection limit for the first approach; see the discussion in the DT section of this report.

- All of the elevated values of toluene and benzene occur in six adjacent wells along the SE fringe of the site during one sampling event (29-30 April 2002). The other COCs are at low levels at five of these. These values were excluded from the DT backgroundl data.

- TCE at NZ-24 in 2004 and 2005 was considerably higher than that of previous years (starting from 1986). It was recommended that these values be removed from the background data, so that if high values persist they will pop up as anomalies. However upon further discussion with the project group, it was suggested that data tracker should be tried using all historical data as background data first (since that's what most people would do) and then experiment with excluding some data to see how things change. The consequences of different background data selection for NZ-24 are discussed in the DT section of this report to follow.

Other Data Preparation Observations

The GeoTrans team provides the following additional observations regarding data preparation:

- The sampling data have to be in a CSV file to import into the software. For Model Builder the format is "Date, SiteID, EastCoordinate, NorthCoordinate, COC1, COC2, …", where SiteID is the well identification For DT the east and north coordinates are optional, and the format of the CSV data is "Date, SiteID, COC1, COC2, …".

- Concentration units have to be consistent over time for each individual COC; the Summit Tools do not check units.

- Different COCs can have different concentration units.

- There is no constraint on the exact chemical name (whatever user enters is OK, which is not the case with MAROS).

- No detection limits and flags are utilized or allowed. However, a "graphing value" for non-detects needs to be assigned to serve as the concentration value for non-detects in the data that are going to be imported. This is similar to a process in MAROS where user enters "detection limit" for non-detects. The user must decide *a priori* whether to include or delete "high non-detects" (i.e., non-detects with reporting limits higher than actual data values). The sensitivity of data transformation results to the graphing value was not investigated.

- For SO (but not DT) all samples must be assigned to a sampling group in the data to be imported, such that every sample in that sampling event has the same sampling date within the software. For spatiotemporal SO, the time lag between two adjacent sampling events has to be at least quarterly frequency. The software User's Guide provides guidance for sampling event "frequency alignment". This data consolidation is done outside of the software prior to import. This is less flexible than MAROS, in which such data can be treated as separate events or consolidated as part of one "sampling event" defined by the user within the software.

Sampling Optimizer (Including Model Builder)

Brief Overview of Functionality

Sampling Optimizer provides users with six possible combinations of interpolation technique and data transformation. They are

- Two interpolation technique options
    - Inverse Distance Weighting (IDW)
    - Kriging

- Three data transformation options
    - None (i.e. No transformation)
    - Logarithmic
    - Quantile

This results in six possible combinations for these basic options, though the user will generally apply only one combination. Generally a user will only utilize one combination, and Summit suggests using kriging with quantile transformation. For the ESTCP project, GeoTrans tried all of the combinations, and did in fact determine during their use of the software that kriging with quantile transformation provided the most reasonable representation of the plume distribution (discussed in more detail later).

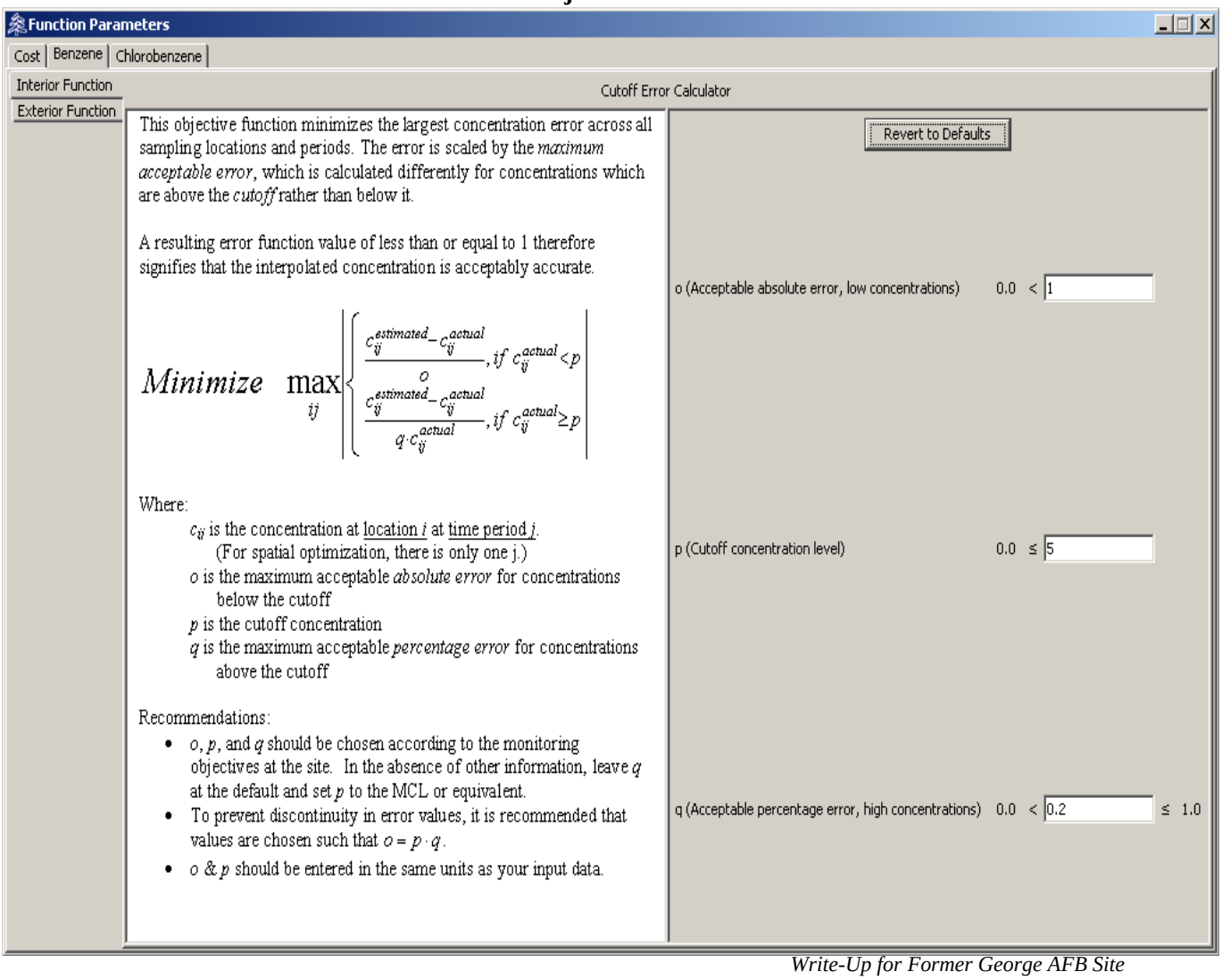
The Model Builder component of the software provides model fitting, visualization, and analysis functions, as well as maps of relative uncertainty. Within Model Builder the user defines options for the parameters of the interpolation technique selected by the user for the Optimizer. Both automated and manual model parameter fitting are supported for Kriging, while the user must manually specify the power to be used for IDW. Within Model Builder the user also specifies desired changes to the defaults on the "Model Builder Settings" screen, such as the number of

vertical slices that defines the resolution of the image. If the data imported into the Sampling Optimizer has multiple events, Model Builder provides visualization for each event. The Optimizer module uses the model parameters specified within Model Builder.

Sampling Optimizer uses a genetic algorithm optimization approach to suggest favorable monitoring plan alternatives relative to the base sampling plan (i.e., where one or more of the samples are removed). In spatial optimization, the original model is based on one set of sampling data that do not vary in time, and the optimization results are with respect to sampling locations only. In spatiotemporal optimization, the original model consists of actual data that vary in space and time, and the optimization results are with respect to sampling location and sampling frequency. Temporal analysis is a subset of spatio-temporal analysis where wells cannot be removed. The user can utilize software defaults for the optimization algorithm (e.g., population size) or can specify values for these parameters in the "GA Settings" screen (further discussion of these parameters to follow).

For this site the "errors" were calculated by using the "Cutoff Error Calculator" option for the objective function provided within Optimizer (the other option is the "Percentage Error Calculator"). The Cutoff Error Calculator incorporates a function (displayed below) for calculating error associated with samples that are removed.

Error Objective Calculator



The overall purpose of Error Calculator is to compute an objective function value that represents the overall similarity of a new sampling plan to the baseline sampling plan. The Cutoff Error Calculator is designed so that "error" is calculated in a manner that makes deviations between interpolated and actual values more significant in areas of low concentration than in areas of high concentration. This is accomplished as follows:

- The user defines a cutoff concentration (p) for the actual data values that differentiates between low concentrations versus high concentrations, and also defines a value for Acceptable absolute error (o).

- When a low concentration data point is removed (i.e., below the cutoff), error is calculated as the absolute value of the actual value minus the interpolated value, divided by the acceptable absolute error. For example, if the actual value is 5 µg/l (i.e., below the cutoff concentration of 10 µg/l) and acceptable absolute error is 1.0, and the difference between the actual and interpolated value is 5 µg/l, then the error would be 5 / 1 = 5.

- When a high concentration data point is removed (i.e, above the cutoff), error is calculated as the absolute value of the actual value minus the interpolated value, divided by a percentage (q) of the actual value, where q is specified by the user. For example, if the actual value is 100 µg/l (i.e., above the cutoff concentration of 10 µg/l) and the percentage input by the user is 10%, and the difference between the actual and interpolated value is 5 µg/l, then the error would be 5 / (0.10 * 100) = 0.5.

In these examples, the difference between the actual value and the interpolated value was 5 µg/l in both cases, but in the first case the calculated error is 5.0 whereas in the second case it is only 0.5. This illustrates how the calculation increases the significance of deviation between actual and interpolated values in the lower concentration areas of the plume.

Observations Regarding Use of Sampling Optimizer Including Model Builder

The following observations were made by GeoTrans based on application of Sampling Optimizer (including Model Builder) in conjunction with both spatial and spatio-temporal analysis:

- The software is very easy to use. However, the post-software analysis of results can take quite some time to analyze the resulting monitoring plans, especially when multiple COCs are present. This was not the case for GAFB, which had only one contaminant considered for optimization.

- The software allows the user to easily save a project and re-open it later. However, some updates to the software made during the project prevented previously saved projects from opening.

- The software allows the user to enter "run titles" which are used as a part of the file names when exporting the results for both Model Builder and Optimizer.

- The software uses a "seed" value to start the random number generator used in many calculations. A genetic algorithm does not guarantee an optimal solution, just one that has high probability of being close to optimal, and a different solution may be obtained if

a different seed value is utilized. The software uses default seed values that are fixed for both Model Builder and Optimizer, which ensures that the same results can be obtained by different users with the same parameter settings. The user can change the seed manually if desired.

- Model Builder provides visualization of plume concentrations and plume uncertainty estimates. It gives users the option to post either well names or measured concentrations on the visualizations. This is a useful feature not available in MAROS.

- The "visualization resolution" (the user controls this by defining the number of vertical and border slices for the image) has a big impact on whether the plume generated by Model Builder can be correctly displayed. The visualization resolution can be modified in "# of vertical slices for image" and "# of border slices" of "Visualization" settings with Model Builder. We used the default value for # of vertical slices, and we increased the default value of 10 for # of border slices to 50 to create a larger margin at the border.

- For the SO module, the "Well Constraint" feature allows the user to specify the maximum sampling frequency and the minimum sampling frequency for each well. For spatial analysis, this feature allows the user to specify which wells cannot be removed from the system (which may be specified in the optimization formulation). This can also be useful for abandoned wells which may be part of the historical data but cannot be sampled in the future, by specifying such wells as "always off".

- The software currently has one general type of objective function available, allowing the user to enter an acceptable error level for lower concentration points, the cut-off concentration between high and low concentration points, and the acceptable error percentage for high concentration points. This was consistent with the formulation for GAFB.

- A "population size" (utilized for the genetic algorithm) of 1,000 for SO was recommended by Summit to ensure that "good" solutions can be found. We attempted a variety of population sizes (described later)

- Plume visualization for both Model Builder and Optimizer also allows users to change the zoom scale and color scale. The color scale is a linear scale allowing users to define the minimum and maximum concentrations for each COC. Then the software can plot the plume maps in color based on the minimum and maximum concentrations defined. However, it does not provide an option for a logarithmic scale, thus, for sites with a very large range in concentrations, it cannot plot both high-end concentrations and low-end concentrations with sufficient detail (though multiple plots with different ranges could be made independently). This was not a big issue for GAFB, because the range in TCE concentrations was relatively small.

- The software allows the user to export the following files:

    - The plume maps, uncertainty maps, and variogram charts (for kriging models) can be exported as image files (.png files).

    - A tradeoff curve for each COC can be exported as an image file.

- A file containing, for each optimization sampling plan on the tradeoff curve, which wells are recommended to be "on" or "off", the maximum concentration error for each COC, and the sampling cost (i.e., number of wells which are on) can be exported as a CSV file.

- For each optimization sampling plan on the tradeoff curve, an individual listing of which wells are on and which wells are off can be exported as a CSV file (for spatio-temporal analysis the frequencies are also exported).

- For each potential optimization sampling plan on the tradeoff curve (and for the current sampling plan), a plume map can be exported as image file for each COC with symbols indicating which wells are on and which wells are off, with "+" indicating wells that are recommended to be removed from the monitoring network and "o" indicating remaining active wells.

Computation Time for Model Builder and Optimizer

The computation time depends on the size of the dataset (e.g., number of wells) and model type selected (e.g., kriging versus inverse distance, plus the type of data transformation). Computation time for Model Builder also increases significantly with the increase in resolution (i.e., number of vertical slices for image), and computation time increases with increased population size for Optimizer. An estimate of the amount of labor and computation time it takes to apply Model Builder and Optimizer to evaluate the TCE plume for GAFB is as follows:

- EnviroStat spent several days evaluating the data, preparing data files for import into the software, and preparing additional preliminary analyses and reports, such as data availability summaries, time series plots (TSPlots) of water level and TCE concentration at various wells, and spatial plots of well screen depths and water level. The data supplied by GAFB were in the standard "one record per value (regular or field duplicate) per well per COC per date" format. No problems were found with inconsistent SiteIDs, missing location coordinates, or the like.

- The data provided by EnviroStat were already in the correct format required by the software. Any revisions to these data prior to import only took a few minutes.

- Importing the concentration data into the Summit software took seconds.

- Next, it took minutes to enter the facility ID and choose the model type (IDW or kriging) and data transformation type (quantile, log, or none).

- For the GAFB site, which has on the order of 50 sampling locations, computation time for visualizing a plume in Model Builder was as follows:

Interpolation Method	Data Transformation	Approximate Time Required for Visualization of Plume in Model Builder (minutes)	Qualitative Description of the Quality of Plume Representation
Inverse Distance	Quantile	10-15 minutes.	Good
	Logarithm	5-10 minutes	Acceptable
	None	5-10 minutes	Bad

Kriging	Quantile	About 20 minutes	**Best**
	Logarithm	10-15 minutes	Good
	None	10-15 minutes	Bad

**2 GB RAM PC with Pentium 4, 3.2 GHz*

- After performing Model Builder, it took several minutes to set up the well constraints and objective function for Optimizer.

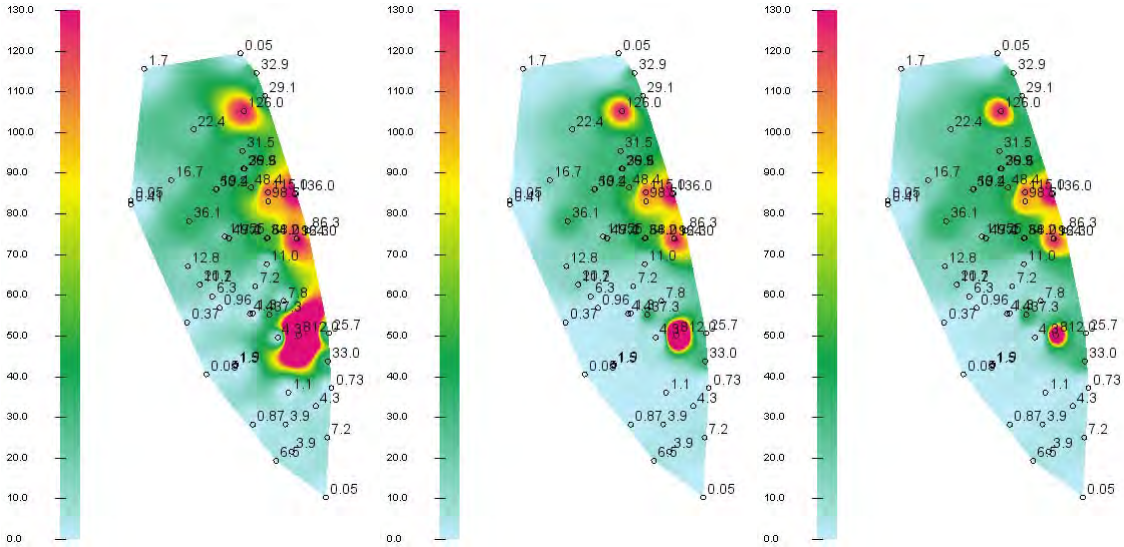
- Spatial optimization with Sampling Optimizer took approximately 10-30 minutes depending primarily on the population size for the genetic algorithm. We used population sizes of 300 (default), 600, 900, and 1200 (discussed in more detail later).

- Computation time for spatio-temporal optimization ranged from 1.5 to 7.5 days, depending on the combination of population size and number of generations (discussed in more detail later)

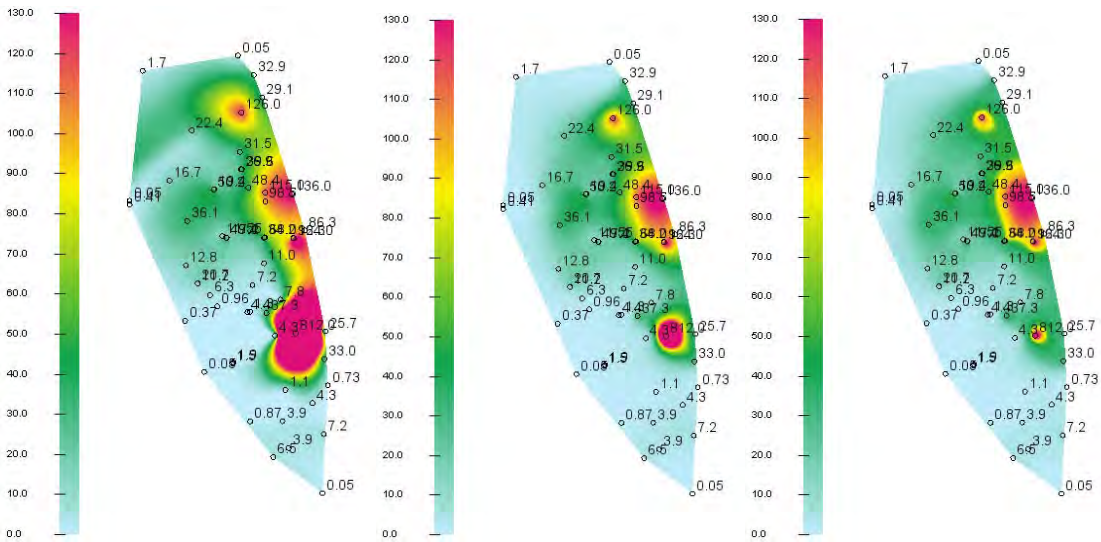
- Finally, reviewing plume maps for the potential plans and exporting them to image files took from several minutes to up to several hours.

Spatial Analysis Results

All six combinations for data interpolation and transformation methods were first applied to dataset A for the Upper Aquifer and were visualized in Model Builder. The figures below illustrate the plume maps generated by each of the above-mentioned methods for dataset A.



IDW-None IDW-Log IDW-Quantile



Kriging-None Kriging-Log Kriging-Quantile

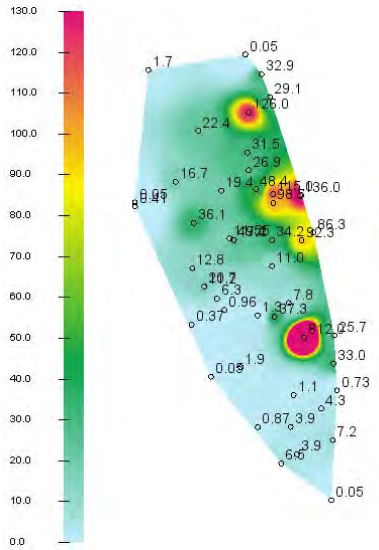
Observations made regarding visualizing the generated plume maps are summarized below:

- Of the six combinations of interpolation and data transformation used, four generally produced better visual representations of the plume. These are:
    - Kriging with quantile transformation,
    - IDW with quantile transformation,
    - Kriging with logarithmic transformation and
    - IDW with logarithm transformation

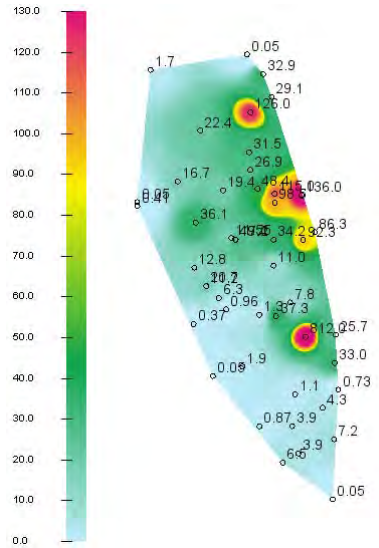
- Both IDW with no data transformation and kriging with no data transformation resulted in model outputs that are biased to higher concentrations, i.e., interpolated concentrations on the boundaries of the map are much higher than believed to actually be the case. For example, in the figure above labeled Kriging-None, concentrations in the upper-left corner are modeled to be on the order of 30µg/l, which is clearly higher than the actual measurements in that area which are on the order of 1 µg/l. This representation is not consistent with how most people would choose to interpret the actual data values. Therefore, these two combinations for interpolation and transformation were eliminated from the optimization modeling.

The four "good" data interpolation methods, selected from visualizing plume maps using dataset A, were then applied to dataset B which excluded the eight possibly problematic monitoring wells. The results are shown in the figures below.

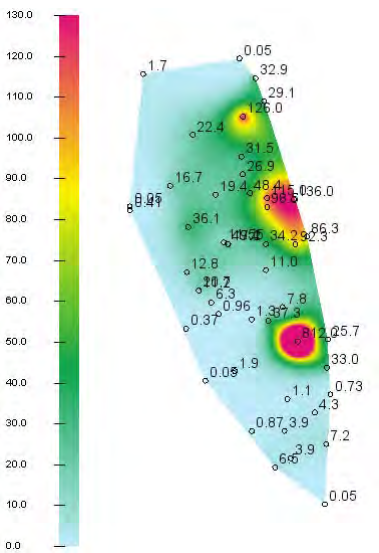
{this gap is intentional}



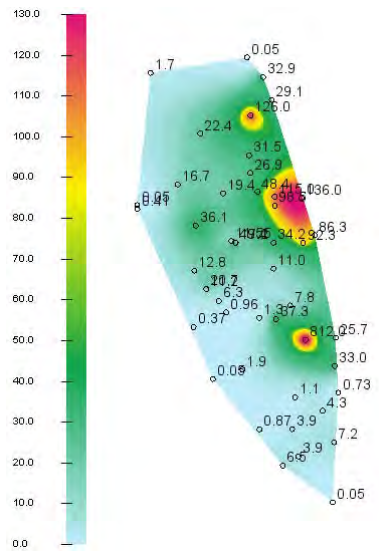
IDW-Log



IDW-Quantile



Kriging-Log



Kriging-Quantile*

Comparing the figures presented above, the following observations are made:

- The Kriging-Quantile method exhibits a smoother plume without abrupt changes in concentration contours
- With the Quantile transformation the zone with high concentration values interpolated around a data point with a high data value is smaller than with the Log transformation.

- The interpretations for datasets A and B are very similar, indicating that the elimination of the eight questionable data points from dataset B does not cause vastly different interpolations.

Our analyst qualitatively ranked the combinations for interpolation technique and data transformation from best to worst, as follows

- Kriging with quantile transformation
- IDW with quantile transformation
- Kriging with logarithm transformation
- IDW with logarithm transformation
- Kriging with no data transformation
- IDW with no data transformation

The Kriging with Quantile method was considered to be the best because:

- It interpolated the size of high concentration as our analyst would interpolate them

- It transitions from high to low concentrations in a manner preferred by our analyst

Therefore, the Kriging-Quantile method was then applied to optimization modeling. Three possible cutoffs between high and low concentrations were specified in the formulation:

TCE Cutoff Concentrations, Acceptable Error for Low Concentrations, and Acceptable Percentage Error for High Concentrations

	TCE (µg/l)
Acceptable error for low concentrations	5
Cutoff between low and high concentrations	25, 50, and 100
Acceptable percentage error for high concentrations (%)	20% with cutoff of 25 µg/l 10% with cutoff of 50 µg/l 5% with cutoff of 100 µg/l

One important factor in setting up the optimizer model is determining the appropriate ‘population size” and “number of generations”. The population size is the number of sampling plans that the Genetic Algorithm is working with. Each generation a new population is evaluated. Summit indicated that, as the population size increases the number of generations should also be increased (although the appropriate number of generations is also related to the number of decision variables). From running a number of cases with dataset B, using the Kriging-Quantile method, it was observed that the population size as well as the number of generation makes a difference in the results. The table below presents the results from a variety of combinations of population size and number of generations (the sampling cost is represented by the number of wells).

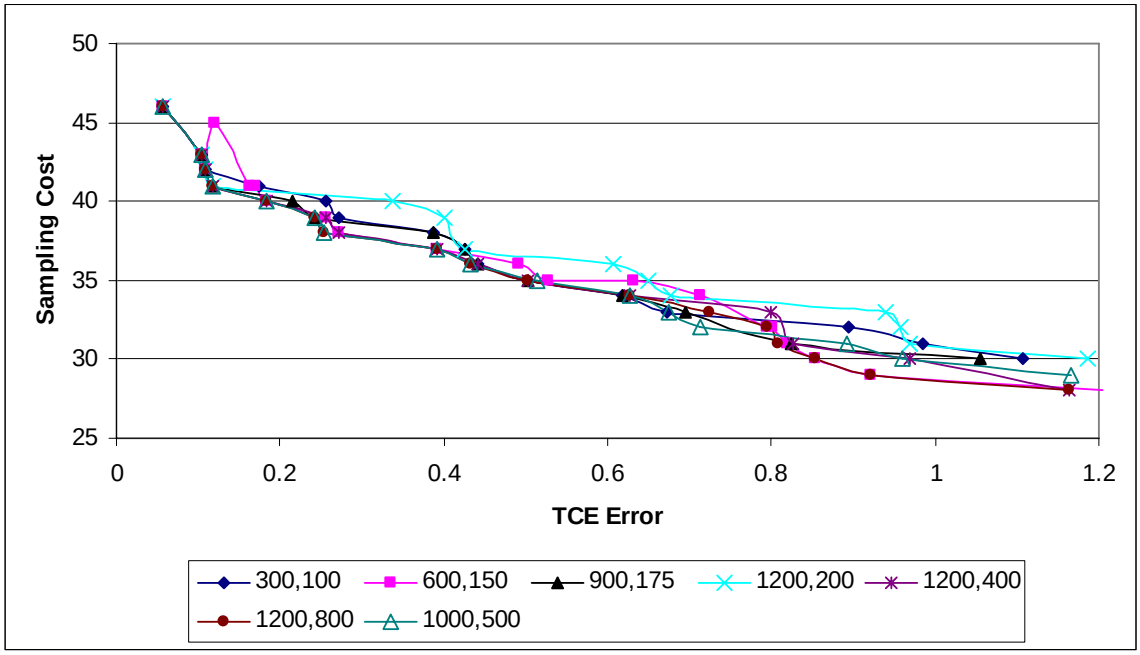
TCE Error Versus Sampling Cost (number of monitoring wells) for Varying Population Sizes and Number of Generations (Pop=population size, Gen= number of generations)

Pop=300, Gen=100		Pop=600, Gen=150		Pop=900, Gen=175		Pop=1200, Gen=200	
TCE Error	Sampling Cost	TCE Error	Sampling Cost	TCE Error	Sampling Cost	TCE Error	Sampling Cost
0.055533	46	0.055533	46	0.055533	46	0.055533	46
0.104349	43	0.104349	43	0.104349	43	0.104349	43
0.108283	42	0.108283	42	0.108283	42	0.108283	42
0.175368	41	0.120832	45	0.116712	41	0.116712	41
0.255859	40	0.163822	41	0.214439	40	0.336252	40
0.271113	39	0.169482	41	0.242646	39	0.400337	39
0.386819	38	0.184522	40	0.387434	38	0.425296	37
0.425444	37	0.255332	39	0.425444	37	0.60601	36
0.441726	36	0.272728	38	0.441726	36	0.649847	35
0.503608	35	0.390915	37	0.503608	35	0.676136	34
0.617262	34	0.490667	36	0.617262	34	0.940216	33
0.672252	33	0.526841	35	0.694743	33	0.957665	32
0.894984	32	0.630626	35	0.82499	31	0.9701	31
0.984538	31	0.713043	34	1.055877	30	1.186365	30
1.108126	30	0.795266	32				
		0.801702	32				
		0.819201	31				
		0.854375	30				
		0.921471	29				
		1.204237	28				

Pop=1200, Gen=400		Pop=1200, Gen=800		Pop=1000, Gen=500	
TCE Error	Sampling Cost	TCE Error	Sampling Cost	TCE Error	Sampling Cost
0.055533	46	0.055533	46	0.055533	46
0.104349	43	0.104349	43	0.104349	43
0.108283	42	0.108283	42	0.108283	42
0.116712	41	0.116712	41	0.116712	41
0.184522	40	0.184522	40	0.184522	40
0.255332	39	0.242646	39	0.242646	39
0.272728	38	0.25423	38	0.25423	38
0.390915	37	0.390915	37	0.390915	37
0.441009	36	0.433236	36	0.432691	36
0.503608	35	0.503608	35	0.513941	35
0.62745	34	0.626606	34	0.626816	34
0.798383	33	0.723597	33	0.674904	33
0.826085	31	0.793732	32	0.712752	32
0.969324	30	0.809028	31	0.891708	31
1.163182	28	0.854375	30	0.959392	30
		0.921471	29	1.165209	29
		1.163182	28		

In the table above, the minimum number of wells that allow an error less than 1.0 is highlighted for each combination of population size and number of generations, and that value varies from 29 to 31. This illustrates that this combination of parameters does have some impact on the results. A graphic representation of the tradeoff curve for the data presented above is as follows:

Tradeoff Curves for Various Combinations of Population Size and Number of Generations



It can be observed that, given acceptable error of less than 1.0, higher population size does NOT guarantee better plans. For instance, compare the TCE errors for the following optimal solutions with 33 wells:

POP = 1200	GEN = 400	Error = 0.798383
POP = 300	GEN = 100	Error = 0.672252
POP = 1200	GEN = 200	Error = 0.940216
POP = 1000	GEN = 500	Error = 0.674904

Based on the plot above, we selected to use a population size of 1,000, and number of generations of 500 for all subsequent runs. However, this is not a general result, and we note that it is not at all clear how to make a decision on this combination of parameters. In a research setting, one might try an even greater variety of combinations for population size, number of generations, and the seed (we used the default seed). However, most practitioners will not have the desire or patience to perform such an exercise.

We applied the following procedures for spatial optimization:

- All the plans with errors less than 1.50 on the tradeoff curve were considered
- Potential plan(s) based on the tradeoff curve were qualitatively chosen, and the TCE plume maps for those plans were then compared to plume maps based on the total number of locations in the base plan.

Spatial optimization was performed for a variety of interpolation techniques and "cutoff values", dataset A and dataset B, as summarized below. The table below lists the number of optimal plans and the minimum number of wells within error scale of 0.5, 1.0, and 1.5 for TCE, for each optimization simulation.

Summary of "Optimal Plans" Identified by Optimizer – Spatial Optimization

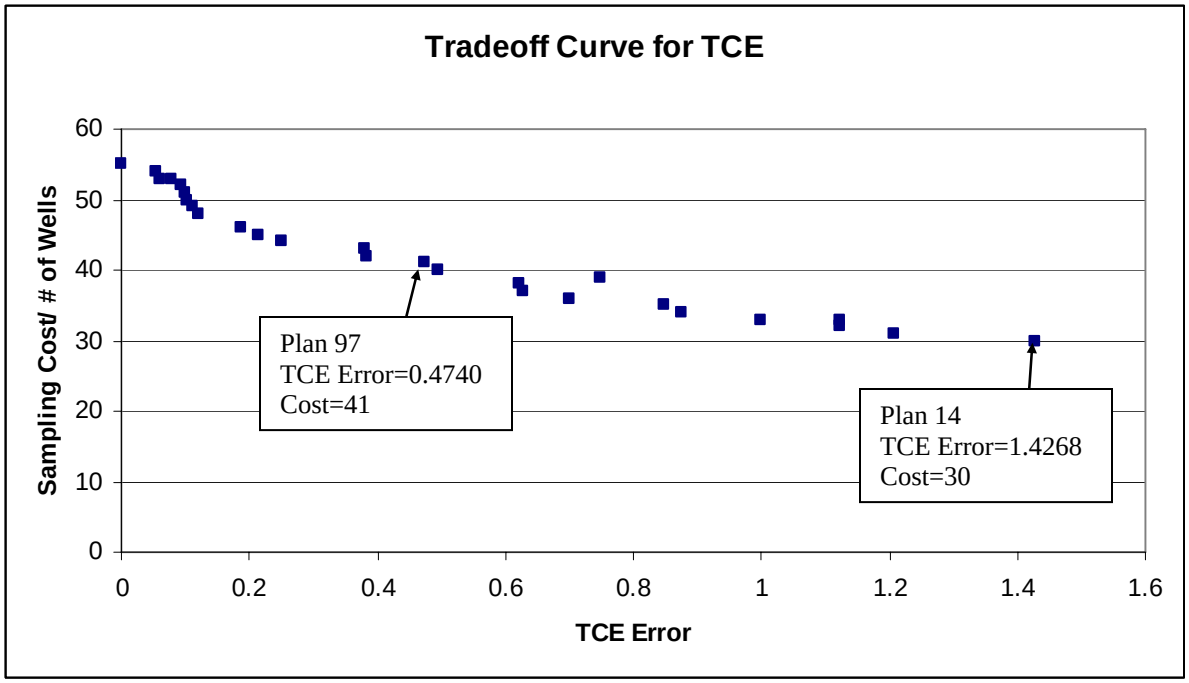
Cutoff Value	**Interpola-tion**	**Data Transform-tion**	**# of Plans w/ Errors < 0.5 for TCE**	**# of Plans w/ Errors < 1.0 for TCE**	**# of Plans w/ Errors < 1.5 for TCE**	**Min Cost w/ errors < 0.5 for TCE**	**Min Cost w/ errors < 1.0 for TCE**	**Min Cost w/ errors < 1.5 for TCE**
Upper Aquifer with Dataset A (Baseline Model has a total of 55 Wells)								
25	Inverse Distance Weighting	Quantile	9	14	17	46	41	38
50			8	13	18	47	42	37
100			8	12	14	47	43	40
25	Kriging	Quantile	15	21	26	40	34	30
50			12	17	23	42	37	31
100			12	17	21	41	36	32
Upper Aquifer with Dataset B (Baseline Model has a total of 47 Wells*)								
25	Inverse Distance Weighting	Quantile	8	13	17	39	33	28
50			7	11	15	39	35	30
100			7	11	13	39	35	32
25	Kriging	Quantile	9	15	17	36	30	28
50			9	14	17	36	31	28
100			9	13	16	36	32	29

** The eight wells (MW-102, MW-104, NZ-06, NZ-10, NZ-20, NZ-30, NZ-31, and NZ-32) identified not to be representative of aquifer characteristics are excluded from dataset B.*

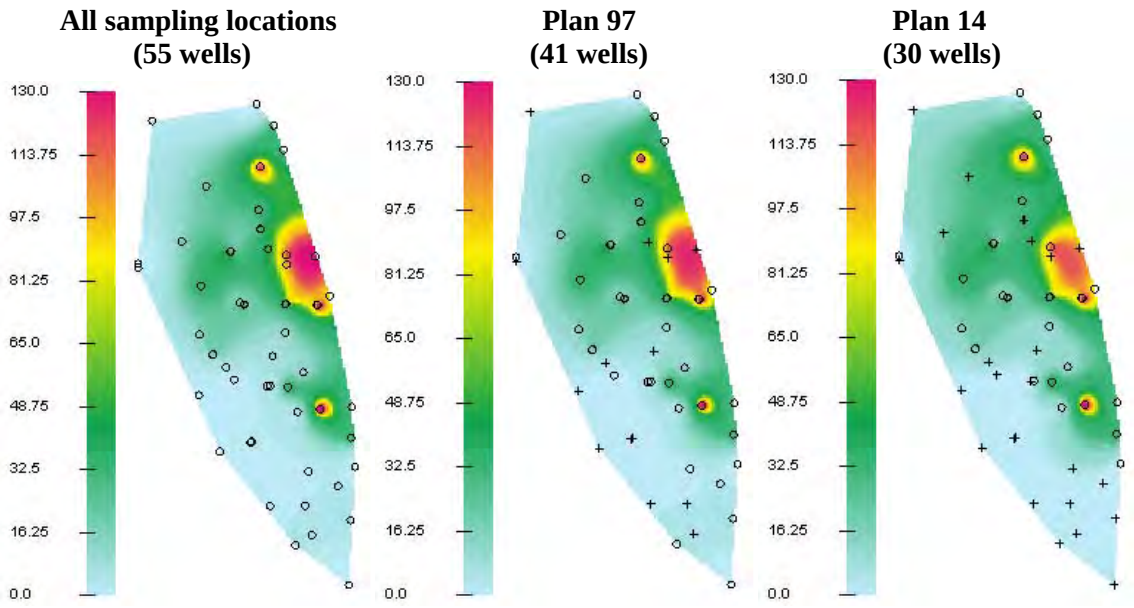
From this table, it can be concluded that Kriging-Quantile method performed better than IDW-Quantile in that it generally offers more plans that satisfy each error criterion, and also offers plans with lower sampling cost for each error criterion. In addition, optimized plans generated from Dataset B are cheaper than plans from Dataset A, but this is biased by the fact that there are originally eight fewer monitoring wells in Dataset B. Furthermore, the sampling costs vary slightly with the cutoff values. For example, for Dataset A with Kriging-Quantile, the minimum sampling costs associated with errors less than 1.0 were 34, 37, and 36 for cutoff values of 25, 50, and 100 respectively. This indicates that the value selected for the cutoff between low concentrations and high concentrations does have some impact on the results, although that impact appears to be minor.

The figure below shows the cost-error tradeoff curve using a cutoff of 25 µg/l and kriging with quantile transformation for Dataset A. Only optimal plans with errors less than 1.50 for TCE are included.

Tradeoff Curve for TCE (Dataset A, K-Q Method, Cutoff=25 µg/l)



Plan 97 reduces sampling cost from 55 to 41 (25.5%) while Plan 14 reduces sampling cost from 55 to 30 (45.5%). Plume illustrations for these plans, versus the base sampling plan with all locations, are as follows:



Note:

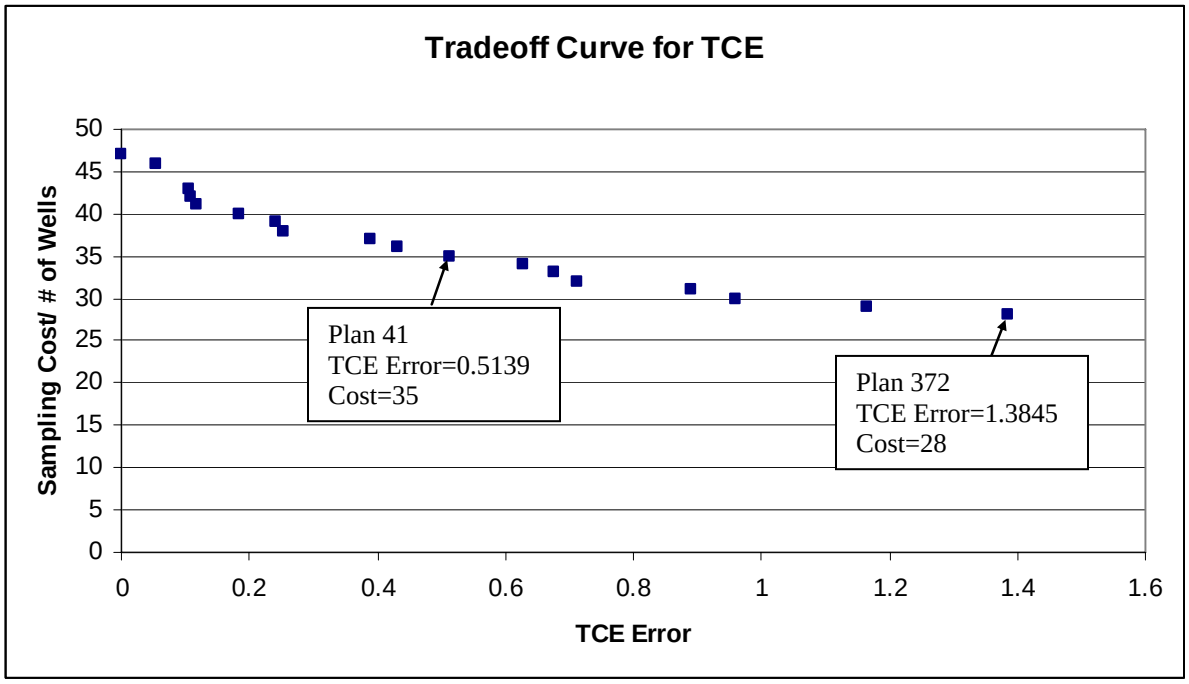
1. The symbol "+" indicates wells that are recommended to be removed by the Optimizer, while the symbol "O" denotes wells that are recommended to keep.

Comparing sampling costs (number of wells) and errors on the tradeoff curve, and visually inspecting the plume maps for selected plan(s), plan 14 was considered by our analyst to be acceptable because it is generally similar to the map with all sampling locations for both the higher concentration areas and the lower concentration areas. Of course, it would ultimately be up to site stakeholders to decide if either Plan 97 or Plan 14 is acceptable. The table below lists the wells recommended to be removed for these two plans (note that those denoted with asterisks are among the eight problematic wells previously mentioned and eliminated in Dataset B).

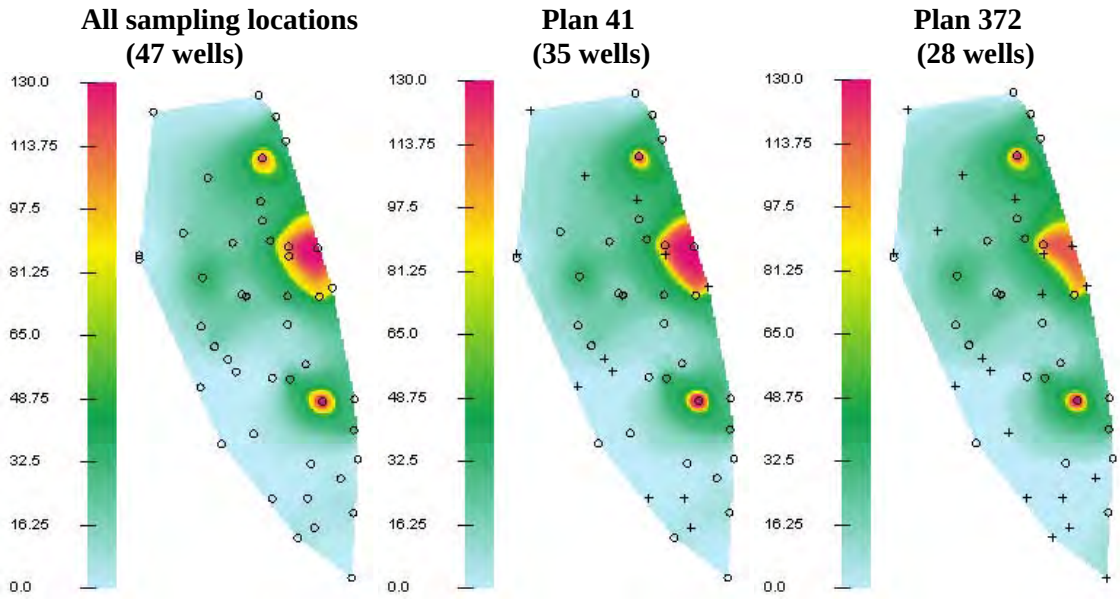
Wells recommended to be removed, Plan 97		**Wells recommended to be removed, Plan 14**	
FT-01	NZ-103	FT-01	NZ-28A
FT-03	NZ-17	FT-03	NZ-32*
MW-103	NZ-27	FT-04	NZ-36
MW-104*	NZ-49	MW-103	NZ-46
MW-105	NZ-68	MW-104*	NZ-49
NZ-06*	NZ-95	MW-105	NZ-51
NZ-101	NZ-96	NZ-06*	NZ-54
		NZ-10*	NZ-59
		NZ-101	NZ-68
		NZ-103	NZ-94
		NZ-116	NZ-95
		NZ-17	NZ-96
		NZ-27	

The tradeoff curve for dataset B (8 fewer wells in baseline sampling plan) is presented below:

Tradeoff Curve for TCE (Dataset B, K-Q Method, Cutoff=25 µg/l)



Plan 41 reduces sampling cost from 47 to 35 (25.5%) while Plan 372 reduces sampling cost from 47 to 28 (40.7%). Plume illustrations for these plans, versus the base sampling plan with all locations, are as follows:

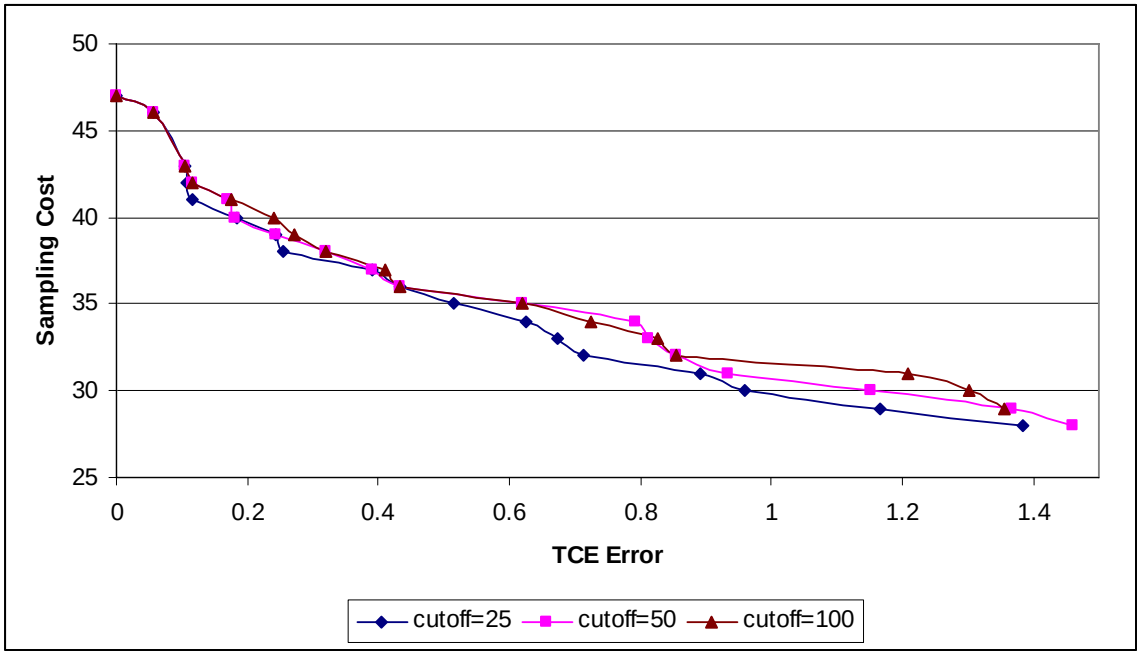


In Plan 372 the high concentration area in the middle of the eastern boundary is somewhat distorted versus the baseline sampling plan with all sampling locations. Nevertheless, Plan 372 is reasonably similar to the baseline sampling plan representation. The greatest difference is in the "plume interior" along the eastern boundary, and as explained earlier, the error calculation intentionally de-emphasizes errors in the plume interior. Our analyst considered Plan 372 to be acceptable. Of course, it would ultimately be up to site stakeholders to decide if either Plan 41 or Plan 372 is acceptable. The table below lists the wells recommended to be removed for these two plans.

Wells recommended to be removed, Plan 41		**Wells recommended to be removed, Plan 372**	
FT-01	NZ-49	FT-01	NZ-46
FT-03	NZ-67	FT-03	NZ-49
FT-04	NZ-68	FT-04	NZ-51
MW-106	NZ-75	MW-103	NZ-59
NZ-101	NZ-95	MW-106	NZ-67
NZ-36	NZ-96	NZ-101	NZ-68
		NZ-116	NZ-75
		NZ-12	NZ-95
		NZ-27	NZ-96
		NZ-36	

The figure below illustrates how the cutoff value between the low concentration area and high concentration area impacts the tradeoff curve for dataset B, using the Kriging-Quantile method. Although a small difference between the three curves is found (i.e., the cutoff=100 tends to have somewhat more expensive plans), the overall difference is not too substantial.

Tradeoff Curves Obtained from Three Cutoff Values (25, 50, and 100 µg/l), Dataset B.



Spatio-Temporal Analysis Results

For a reliable analysis, spatio-temporal analysis with the Sampling Optimizer has the following requirement for the input data:

- For a specific COC at a specific well to be included in the dataset to be analyzed, there must be at least four samples at that well for that COC. The software will warn the user if there are only four to 7 seven samples for a COC at a specific well by indicating that the data may be insufficient for reliable analysis (i.e., eight or more samples per well is preferable).

- For a specific COC to be analyzed as a part of a specific sampling event, there must be 15 samples of a COC for that sampling event. The software will warn if there are only 15 to 19 samples for that event by indicating that the data may be insufficient for reliable analysis (i.e., 20 or more samples per event is preferable).

The following tables indicate warnings provided by the software for this site because the number of samples for a well was less than eight or the number of samples for an event was less than 20 (i.e., permissible but perhaps insufficient for a reliable analysis).

Software Warnings: Dataset A

Location ID	# of samples	Event	# of samples
FT-02	7	12/30/1999	17
MW-104	5	12/30/2000	19
MW-105	7		
MW-106	6		
NZ-101	6		
NZ-103	7		
NZ-116	6		
NZ-17	6		
NZ-42	5		

Software Warnings: Dataset B

Location ID	# of samples	Event	# of samples
FT-02	7	12/30/1999	15
MW-105	7	12/30/2000	17
MW-106	6		
NZ-101	6		
NZ-103	7		
NZ-116	6		
NZ-17	6		
NZ-42	5		

The spatio-temporal analysis was run first by disregarding these warnings. Then, as an experiment, Dataset A was modified by adding data to eliminate the warnings. Two methods were used for creating such data. The first one involved interpolating data points by averaging data from the nearest sampling event that occurred before and after the event with insufficient data. For example, suppose for a specific sampling location there is a value of 10 µg/l in June 1999 and a value of 20 µg/l in June 2001, but there is no sampling value from June 2000. A value of 15 µg/l could be assigned for June 2000, i.e., (10+20)/2=15. The second method involved assigning a concentration to wells with relatively consistent concentrations over other sampling events. For instance, if a well has many “non-detect” values in other sampling events, a value consistent with “non-detect” could be assigned for events where the well was not sampled.

The following table summarizes the three datasets evaluated spatio-temporally.

	# of COC's	# of sampling locations	# of sampling events	# of samples	Sample date range
Dataset A	1	54	22	650	10/1/1994-10/1/2005
Dataset B	1	46**	22	566	10/1/1994-10/1/2005
Modification of Dataset A*	1	54	22	672	10/1/1994-10/1/2005

**Includes data added by our analyst, as described above.*
*** NZ-111 was omitted from analysis due to insufficient number of sampling events*

It was not clear what the best combination of population size and number of generations to assign for the genetic algorithm. The sensitivity of results to these parameter values was evaluated by assigning three combinations of population size and number of generations for the Spatio-temporal analysis, using dataset A and a cutoff value of 25 µg/l:

- Population of 500, with 250 for number of generations
- Population of 500, with 400 for number of generations
- Population of 1000, with 500 for number of generations

The table below presents the results obtained from these three combinations, using the kriging-quantile interpolation method.

TCE Error Versus Sampling Cost for Varying Population Sizes and Generation Numbers for Dataset A. (Pop=Population size, Gen=Generation number)

Pop=500, Gen=250		Pop=500, Gen=400		Pop=1000, Gen=500	
TCE Error	Sampling Cost	TCE Error	Sampling Cost	TCE Error	Sampling Cost
0	112	0	108	0	108
0.134347	110.25	0.113759	106.5	0	112
0.343129	104.45	0.134347	106	0.053682	107
0.623421	102.95	0.343129	104	0.113759	106.2
0.623421	102.95	0.343129	104.5	0.113759	106.5
0.685482	99.11667	0.450566	102	0.134347	106
0.701029	97.11667	0.623421	100	0.134347	106.3333
1.130668	94.28333	0.685482	98.33333	0.343129	104
		0.701029	96.33333	0.343129	104.5
		0.701029	98.33333	0.450566	102
		1.130668	93.33333	0.623421	100
		1.388378	93.08333	0.623421	103.7
				0.685482	98.33333
				0.685482	100.8333
				0.701029	96.33333
				0.701029	97.58333
				1.130668	93.33333
				1.388378	93.08333
				1.487893	96.83333

Note: *The computation time for the above three combinations was 1.5 days, 3 days, and 7.5 days respectively, 2 GB RAM PC with Pentium 4, 3.2 GHz*

The results from the first combination (Pop = 500, Gen = 250) provided fewer choices for optimal plans, and also generally yielded more expensive plans for similar values of TCE error. The other two combinations for population size and number of generations yielded results that were generally similar, but the middle combination (Pop = 500, Gen = 400) had much lower computation time (3 days versus 7.5 days). It is not clear how the user will know in advance what combination for these parameters works best, and since the computations take days, experimentation may not be practical for many cases. Note that all of the spatio-temporal simulations described above were performed using kriging with quantile transformation, since that model was preferred by our analyst. Using IDW rather than kriging would lower the computation time. For instance, our analyst performed a simulation with 500 for population size and 250 for number of generations, using IDW and quantile transformation. The spatio-temporal optimization took approximately 12 hours with IDW, versus approximately 36 hours for kriging with quantile transformation.

It is also noted from the tabular results presented above that the optimizer appears to have included sub-optimal plans in the results. For example, results from (Pop = 1000, Gen = 500) gives plans with the same TCE error but with two different costs. It is clear that the higher cost with the same TCE error should be eliminated as sub-optimal during the optimization process. The Summit technical team suggested further increasing the number of generations to get rid of these bad plans, and our analyst verified this. Also, the user can simply eliminate these in their presentation of results. However, doing so also results in longer computation time.

A series of spatio-temporal simulations were then performed for different cutoff values on each of the three datasets (A, B, and modified A) using 500 for population size and 400 for number of generations. Each simulation utilized kriging with quantile transformation.

Summary of Optimal Plans Identified by Optimizer for Spatio-Temporal Optimization

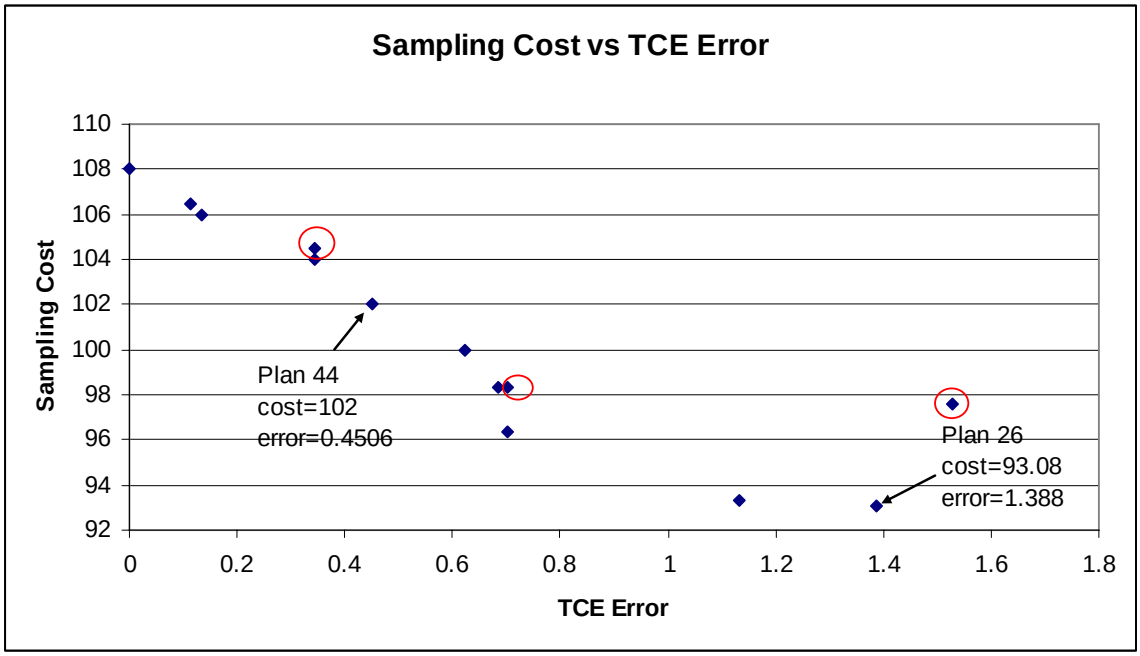
Dataset	**Cutoff Value (µg/l)**	**Interpolation & Transformation**	**# of Plans w/ errors < 0.5 for TCE**	**# of Plans w/ errors < 1.0 for TCE**	**# of Plans w/ errors < 1.5 for TCE**	**Min. Cost w/ errors < 0.5 for TCE**	**Min. Cost w/ errors < 1.0 for TCE**	**Min. Cost w/ errors < 1.5 for TCE**
A	25	Kriging-Quantile	5	9	11	102	96.33	93.08
	50		7	11	13	104	96.33	92.83
	100		5	10	11	102	98.08	94.83
B	25	Kriging-Quantile	5	8	11	88	82.53	79.08
	50		4	6	9	88	82.53	79.08
	100		4	6	8	88	82.53	79.33
Modified A	25	Kriging-Quantile	5	11	13	104.2	98.58	93.58

** The minimum cost for the spatio-temporal optimization is calculated as the average number of samples per year, resulting in decimal values for some cases.*

The tradeoff curve for Dataset A, using cutoff value of 25µg/l, is illustrated below. After comparing the plans, visually inspecting the plume maps for selected plan(s), and comparing the

plume maps with the baseline model, two potential plans- Plan 44 and Plan 26 were identified as promising plans. The sampling cost (i.e., average number of wells per year) is reduced from 108 to 102 (5.56%) for plan 44, and from 108 to 93.08 (13.81%) for plan 26 but with somewhat more error.

Spatio-Temporal Tradeoff Curve for TCE (Dataset A, K-Q Method, Cutoff=25)



**The circled points correspond to what appear to be sub-optimal plans included in the Optimizer results relative to other optimal plans that were identified by Optimizer.*

The table below lists the recommended sampling frequency for these plans, the wells that are recommended to be removed (i.e., always "off"), and the maximum error for each plan.

Spatio-Temporal Analysis Recommendations (Dataset A, Cutoff = 25 µg/l)

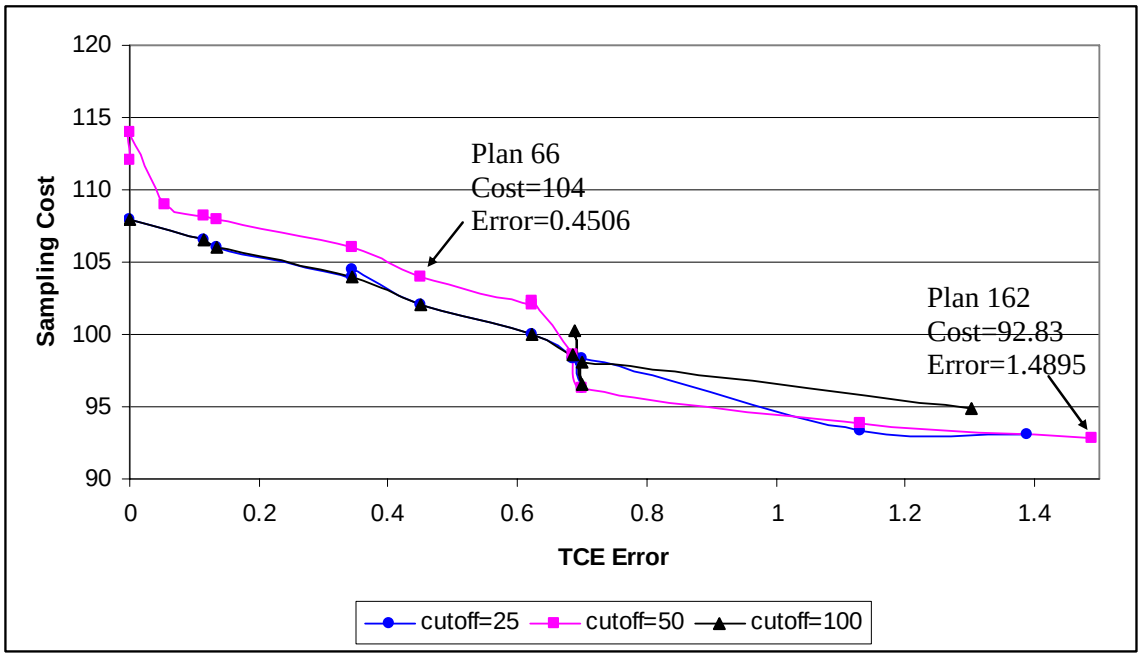
Plan #	Recommended Sampling Frequency						Max TCE Error
	Semi-Annually	Annually	Every 2 Years	Every 3 Years	Every 4 Years	Off	
44	All the remaining wells					MW-104 MW-106 NZ-17	0.4506
26	All the remaining wells		NZ-51	NZ-96	MW-103	FT-04 MW-104 MW-106 NZ-17 NZ-95	1.388

The results were examined to ascertain if the wells recommended to be turned off in the spatio-temporal analysis correspond to wells recommended to be removed for the spatial optimization, given a similar amount of error. After reviewing the results, it was found that all wells recommended to be turned off in the spatio-temporal evaluation were also recommended for removal in the spatial analysis, except that MW-106 and MW-105 are exchanged as "on" versus

"off". This might be attributable to the fact that these two wells are very close to each other, and turning off either of them may have similar effect on the outcome. In addition, the wells recommended for less frequent sampling in Plan 26 were also recommended to be removed in the spatial analysis. This perhaps suggests that the spatial analysis may be robust and the spatio-temporal analysis, which is more complicated than the spatial analysis, may not be worthwhile to perform.

Next, the impact of using different cutoff value (i.e., between "low values" and "high values") was evaluated. The software calculates the error differently for concentration above the cutoff versus below the cutoff. The figure below shows the tradeoff curves for Dataset A using cutoff values of 25, 50, and 100 µg/l. From this figure, it is observed that for TCE error less than 0.7, the costs associated with plans of cutoff value of 50 µg/l are generally higher than for the other two cutoff values, whereas for errors between 0.7 and 1.5, the costs associated with plans of cutoff value 100 µg/l are generally higher than for the other two cutoff values. The lowest cost solutions were generally found with the lowest cutoff value, as was the case for the spatial analysis. It is not clear if this is a general result.

Spatio-Temporal Tradeoff Curves Using Cutoff Values of 25, 50, and 100 µg/l (Dataset A)



The spatio-temporal recommendations for cutoff value of 50 µg/l (Dataset A) are as follows:

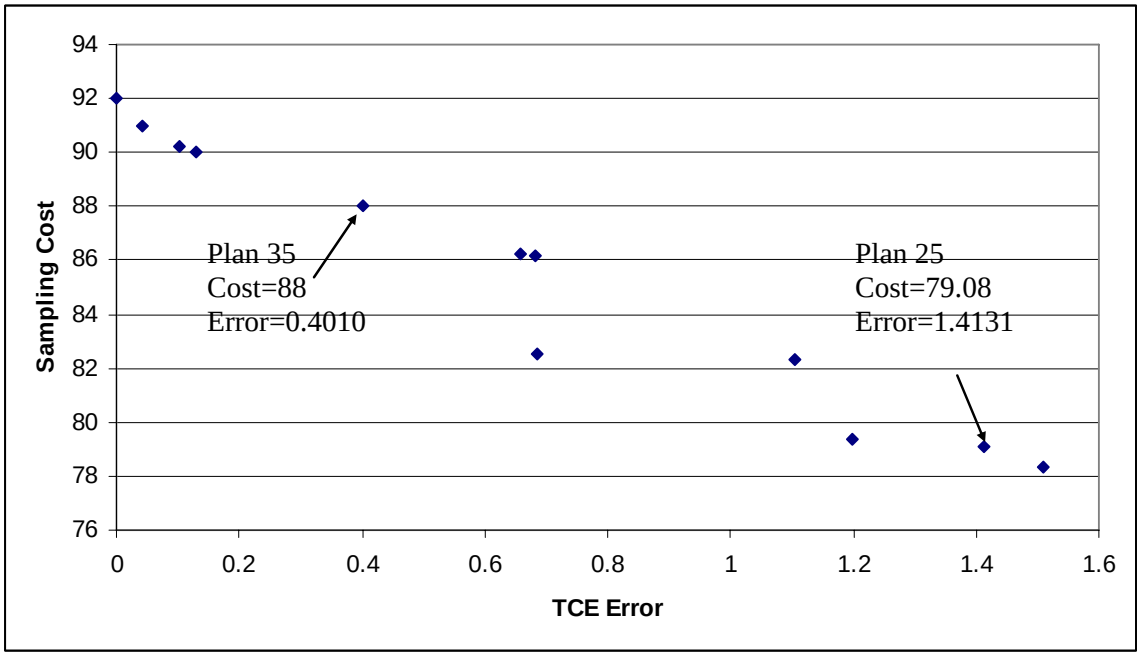
Spatio-Temporal Analysis Recommendations (Dataset A, Cutoff = 50 µg/l)

Plan #	Recommended Sampling Frequency						Max TCE Error
	Quarterly	Semi-Annually	Annually	Every 2 Year	Every 3 Year	Off	
66	NZ-96	All the remaining wells				MW-104* MW-106* NZ-17*	0.4506
162		All the remaining wells		FT-04	NZ-96	MW-104* MW-106* NZ-17* NZ-51 NZ-54 NZ-95*	1.4895

Note: *the wells donated with * are those recommended to be also turned off for cutoff value=25*

The tradeoff curve for Dataset B, using cutoff value of 25 µg/l, is illustrated below. After comparing the plans, visually inspecting the plume maps for selected plan(s), and comparing the plume maps with the baseline model, two potential plans- Plan 35 and Plan 25 were identified as promising plans. The sampling cost (i.e., average number of wells per event) is reduced from 92 to 88 (4.35%) for plan 35, and from 92 to 79.08 (14.04%) for plan 25 but with somewhat more error.

Spatio-Temporal Tradeoff Curve for TCE (Dataset B, K-Q Method, Cutoff=25 µg/l)



The table below lists the recommended sampling frequency for these plans, the wells that are recommended to be removed (i.e., always "off"), and the maximum error for each plan.

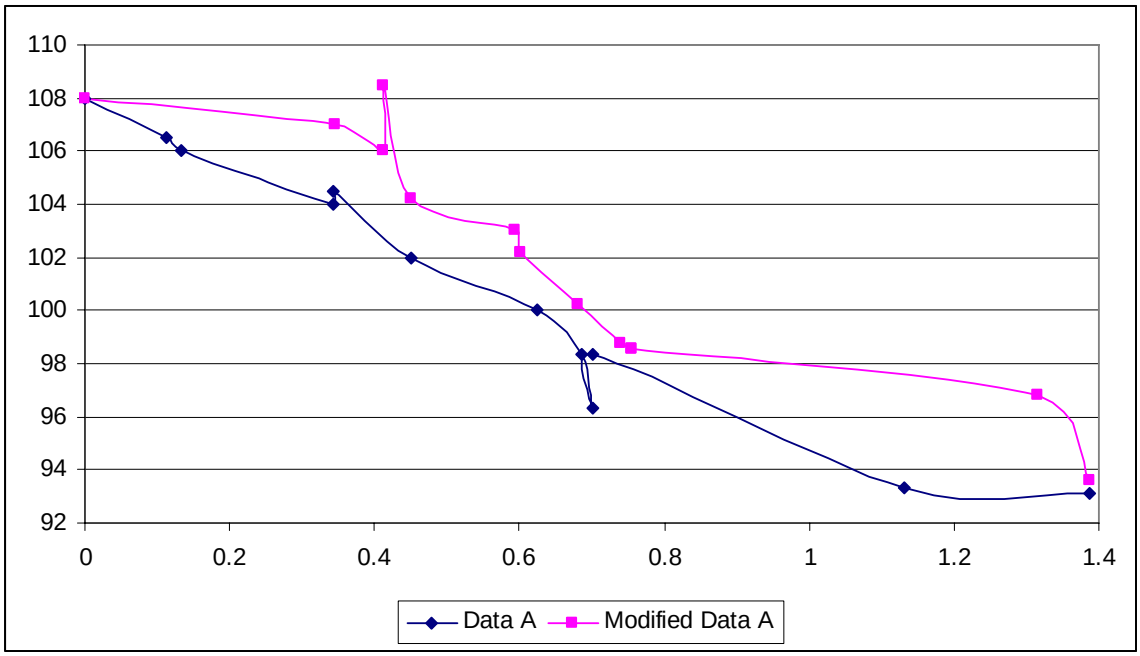
Spatio-Temporal Analysis Recommendations (Dataset B, Cutoff = 25 µg/l)

Plan #	Recommended Sampling Frequency						Max TCE Error
	Semi-Annually	Annually	Every 2 Year	Every 3 Year	Every 4 Year	Off	
35	All the remaining wells					MW-106 NZ-17	0.4010
25	All the remaining wells		NZ-51	NZ-96	MW-103	FT-04 MW-106 NZ-17 NZ-95	1.4131

Comparing the outcomes of optimized sampling frequency from Dataset A (presented earlier) and Dataset B (table presented above), it is observed that plans suggested by the software with maximum errors less than 0.5 and 1.5 are identical except that one of the eight problematic wells (MW-104) is not included in the optimal result for dataset B.

Next, spatio-temporal evaluation was performed for Modified Dataset A (i.e., with some values added manually so that there were at least 8 samples per well and at least 20 wells per event). A cutoff value of 25µg/l was utilized. The figure below compares the tradeoff curves. The sampling costs associated with Modified Dataset A are higher, though it is not clear why.

Spatio-Temporal Tradeoff Curves for Dataset A vs. Modified Dataset A (Cutoff = 25 µg/l)



The table below lists the recommended sampling frequency for plans with maximum error of 1.5, for Dataset A versus Modified Dataset A. These results are very similar. The only difference involves the MW-105 and MW-106. For Dataset A, MW-106 was suggested for removal,

whereas for Modified Dataset A nearby well MW-105 was recommended to be sampled every 2 years. Therefore adding data points to dataset A did not result in significant changes of the optimal plans generated by the software at this part of the tradeoff curve.

Comparison of Recommended Sampling Frequency for Dataset A and Modified Dataset A

Plan w/ Max Error 1.5	Recommended Sampling Frequency					
	Semi-Annually	**Annually**	**Every 2 Year**	**Every 3 Year**	**Every 4 Year**	**Off**
Dataset A	All the remaining wells		NZ-51	NZ-96	MW-103	FT-04 MW-104* MW-106 NZ-17 NZ-95
Modified Dataset A	All the remaining wells		MW-105 NZ-51	NZ-96	MW-103	FT-04 MW-104* NZ-17 NZ-95

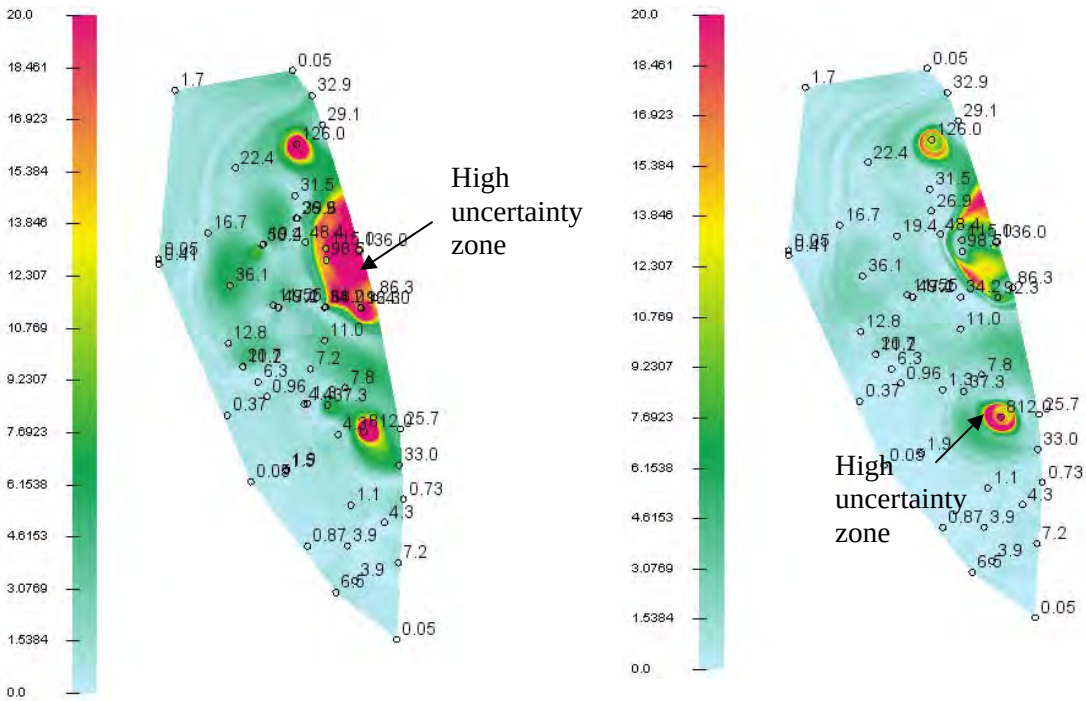
It is important to note that the spatial optimization results suggest removing more wells than the spatio-temporal optimization results, suggesting that the spatio-temporal evaluation may be inherently more conservative. This likely results from the manner in which error is calculated for spatio-temporal analysis, which is the maximum spatial error at any sample in any event. Since some events have fewer wells than the total number of wells in the spatial analysis, spatial errors caused by removing samples in those events would tend to result in larger spatial errors within those events, and therefore larger overall errors. Thus, the spatio-temporal analysis would remove fewer locations than the spatial analysis.

Uncertainty Analysis

Model Builder can provide maps of uncertainty for visualization. The figure below illustrates the uncertainty maps using kriging and quantile transformation (K-Q), for cutoff of 100 µg/l. The areas with high uncertainty (in red) basically overlap those of high concentration. GeoTrans reports that it is not clear to them how to make practical use of these results. By comparing the uncertainty maps between Dataset A and B, it is observed that the plume map for dataset A has higher uncertainty particularly along the eastern boundary. It is not clear why these results are so different. Summit indicates this feature is being modified, such that a different calculation will be used for the third demonstration site.

{this gap is intentional}

Examples of Concentration Uncertainty Visualization



Dataset A, K-Q, cutoff=100 Dataset B, K-Q, cutoff=100

Note: 1.Posted values are measured concentrations at each location.
2. The color scale bar indicates the uncertainty in units of concentration (µg/l).

It is also not clear to the GeoTrans analyst how these uncertainty maps can be used to specifically indicate what an acceptable amount of uncertainty is, whether that varies spatially, and how the software can be used to determine how many new wells might be need to reduce the uncertainty to an "acceptable" degree, and where to locate those wells.

Data Tracker Results (Reported by Charles Davis, EnviroStat)

Overview and Summary

Data Tracker (DT) is a feature of the Summit Monitoring Tools designed to aid site personnel in identifying anomalies of potential interest while they are sorting through voluminous periodic monitoring data reports. DT has two modes: tracking data values for individual COCs at individual wells, and tracking other types of metrics that are computed using more than one COC and/or more than one well. Both functionalities are involved in the demonstration/ validation project at the former GAFB. This section deals with the first (and probably most common) function. The "Mass Metric" discussion (later in this document) deals with the second function.

For tracking individual COCs at individual wells, the idea is that DT should assist the process of screening values by comparing them with background data. A well/COC combination for which the new ("current") value is not consistent with one's expectations based on prior data should be

flagged for evaluation. These prior expectations are formulated as prediction bounds (upper and lower) for the current observation. DT has two modes: static and time-dependent. If DT finds a statistically significant decreasing trend in the background data, it uses time-dependent bounds; otherwise it uses static bounds.

DT functions as follows:

- Historical data are first manually screened to remove atypical values that are not representative of current conditions, such as erratic values that are not repeated, nondetects (NDs) with unusually high reporting limits (RLs), or in some cases early data where concentrations have changed over the years. Several examples are discussed in detail below. The idea is that the remaining "background" data should be representative of values to be expected in future monitoring. Future values inconsistent with the background data will be flagged for inspection. This is done separately for each COC and each well. The user may set the minimum number of background observations; DT requires an absolute minimum of four background observations per well/COC combination. The user must average field duplicates with regular measurement values at this stage, or otherwise handle field duplicates; DT allows at most one measurement per COC per well per event. In the current version of DT, the user must supply a nominal value for each ND. In its initial analyses of the GAFB data, GeoTrans simply used the reporting limit for the NDs. Actual sample collection dates are used; there is no need to align dates across wells in this module of the software.

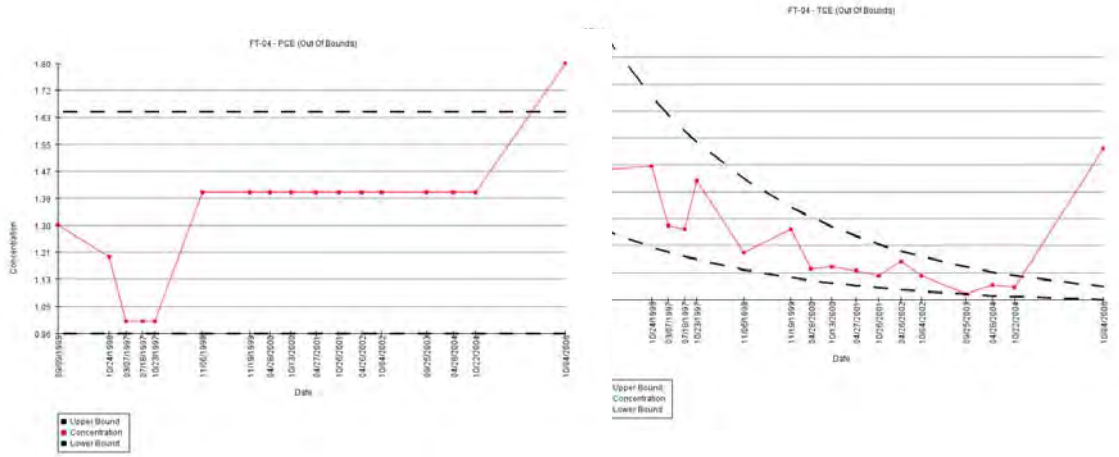
- The background data form one dataset, stored in CSV format. Historical values that are not included in the background data are not included in this dataset, at least for the version of DT used at GAFB, and therefore are omitted from any plots that are produced; discussion of this issue follows.

- Prediction limits (PL bounds) are calculated for each well/COC combination with at least the minimum number of background values. These can be either static or time-dependent. If static, they are computed as nominal two-sided 95% PL bounds using the original data. If time-dependent, they are computed as nominal 95% PL bounds based on an exponentially decaying fit to the background data. The selection of type of bound is made automatically by DT. If a statistically significant decreasing trend is detected, the time-dependent option is used; otherwise static bounds are computed. DT will not provide time-dependent increasing bounds. The nominal prediction confidence is user-configurable; with the default 95% prediction confidence, so long as the assumptions are reasonably met, only around 5% of observations from truly steady-state or truly exponentially decaying processes should be out-of-bounds. In the static case the nominal confidence level is based on assumptions of (a) steady-state variation, (b) normal distributions of data, and (c) uncensored data; in the time-dependent case it is based on assumptions of (a') steady-state variation about a decreasing straight line on the log scale with (b') normally distributed errors and (c') uncensored data. In actuality, of course, real data often have trends, outliers, NDs, and non-normal distributions; hence the nominal confidence level is somewhat approximate.

- When data for one or more new monitoring events arrive, another CSV file of these current data is prepared. DT compares each current value with its PL bounds. It prepares two tables, one of the well/COC combinations that are in-bounds for all dates in the current data file, and the other of well/COC combinations that are out-of-bounds for at

least one measurement in the current data file. The user can then click on a well/COC combination in either table to see a Time Series Plot (TSPlot) of the historical and current data along with the bounds. The numerical result and the TSPlot can be saved for future reference or embedding in documents.

Below are two examples of TSPlots with out-of-bounds data; the one on the left uses static bounds, and that on the right time-dependent bounds.



Simply flagging a data value as in- or out-of-bounds is not in itself a decision or action. The intent of these tables is to allow the user to concentrate on the out-of-bounds values, hence focusing attention on the well/COC combinations likely to require attention or action. Since the bounds are two-sided, in some cases one will find values that are lower than anticipated, and in others that are higher than anticipated. Either might indicate anomalies of interest; it is left to site personnel to make the actual decisions regarding each flagged value.

The operation of DT was tested using GAFB data for TCE, c12DCE, PCE, benzene, and toluene. Six versions of the reserved (2006) data were prepared and used with background data from 1994-2005. One version contains the actual 2006 data, unaltered except for averaging of field duplicate values and replacing NDs with a value. Artificial anomalies were introduced into the other five datasets by EnviroStat (blind to GeoTrans who performed the DT analysis with the software), following plausible scenarios of interest at GAFB. Descriptions of these scenarios (see below) were presented *a priori* to and discussed with site personnel and members of the project team not directly involved in using the software.

DT identified three out-of-bounds situations in the actual, unaltered data: a jump in TCE from recent values around 1.0 up to 8.7 (µg/l) in FT-04; a large increase in TCE in NZ-93; and a value of 1.1 for benzene where most prior data had been NDs at <0.4. It did not identify an increase in toluene to 3.4 when most background data had been NDs at <1.1; one elevated value in the background data was enough to make the upper prediction bound slightly higher than that value. (See the general discussion of background data selection to follow.)

DT also produced four false alarms with the actual, unaltered data. These occurred when all data, both background and current, were various mixtures of NDs and "J" values (from analyses for which an actual value is reported, even though it is less than the RL stated at the time). An example of this situation is described in detail below.

With the modified datasets DT was successful in identifying the major artificial anomalies. It could not detect some smaller ones in situations where the background data were quite variable, regardless of any choices that might be made regarding the selection of background data. In addition, there were several situations in which the success of DT was sensitive to the decisions made in preparing the background data. This is an issue of major interest; it is discussed in some detail below, and recommendations for the use of DT are provided.

Details: Preparing Datasets with Artificial Anomalies

These datasets were generated from the 2006 data received from GAFB. Two wells were sampled during 2006, but only twice previously; these wells were excluded from both the background and new data files. Otherwise, there were two events during 2006 (Spring and Fall); some wells have data from one, some from the other, and some from both.

As a preliminary step, the few field duplicates were averaged with their partner sample data values. Also, "J" flags were ignored. The six versions of the 2006 data were labeled Apple, Banana, Cherry, Grape, Lime, and Orange. Orange is the original GAFB 2006 data.

The first step in creating the other five versions was to "jitter" all values (except NDs) slightly, to avoid having, say, all values except the Grape dataset value being identical for a given well, which would be an obvious clue that the Grape value was the artificial one. This was accomplished by multiplying each original value by a random number between 0.75 and 1.25, then rounding the result the same way as the original value has been rounded. Artificial anomalies were then added to the five versions, following the scripts previously proposed.

Narrative descriptions of the artificial anomalies follow. Various combinations of these were included, to varying degrees, in the five datasets.

a. A new release/source develops in the area of the new percolation ponds, affecting MW-103 and MW-104 first, then MW-102, NZ-18, then NZ-35. (Recall that this project uses only upper aquifer data.)

b. The lab swaps the TCE and PCE values for all samples taken in the second week of April.

c. Benzene and Toluene start showing up in low concentrations in MW-105, MW-106, NZ-46, NZ-31, NZ-25, and in trace concentrations in NZ-24 and NZ-42, increasing from April to October. This would represent a completely new source or release.

d. TCE concentrations in a pair of neighboring wells depart from their historical patterns in opposite ways, with NZ-32 increasing and NZ-28a decreasing.

e. Increasing trends in TCE begin at all of NZ-116, NZ-96, NZ-68, and NZ-94.

f. A hot spot develops in the Dozer Scar Site (NZ-07) involving not only TCE but also atypically high concentrations of PCE and c12DCE.

g. TCE increases in two of the three of NZ-06, NZ-07, and NZ-10, decreasing in the other.

h. TCE is quite high at one or more historically low and/or ND wells in April, but returns to typical values by October; trace levels of PCE follow along.

i. Conversely, TCE at one or more historically high wells drops to ND in April but resumes its typical high values in October.

j. TCE values are 20 times too low for all wells sampled during the third week of October.

These ten scenarios were assigned to the five datasets as shown in the following table.

Anomalies Introduced to GAFB Reserved Data

	Anomaly scenario									
Dataset	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**
Apple	a	b		d		f			i	j
Banana	a		c		e	f		h		
Cherry		b	c							j
Grape			c	d	e		g		i	
Lime	a			d	e		g	h		
Orange										

DT-GeoTrans Results

GeoTrans made subjective judgments about each value flagged by DT as to whether or not it was truly anomalous, and provided a recommended interpretation or action. In its interpretations GeoTrans took into account the cleanup goals for the specific COCs, historical data as well as the data values from the other 2006 event (where there was one), and values at neighboring wells. In characterizing the out-of-bounds values, GeoTrans assigned one of six descriptions, as follows.

A: Current concentration is much higher than the historical data and is above the cleanup goal, more likely bad data than plume migration.

B: Current concentration is much lower than the historical data which were above the cleanup goal, could be bad data.

C: Current concentration is higher than historical data and above the cleanup goal but following an increasing trend.

D: Current concentration is lower than historical data but following a decreasing trend.

E: Current concentration is out of bounds, but not a concern. No action is required.

F: Concentrations higher than previous data, more likely plume migration than bad data

GeoTrans provided a spreadsheet for each dataset listing the out-of-bounds well/COC combinations, the cleanup goal, and its interpretation. It also provided the plots given by DT. In many cases the DT output was sufficient to allow the GeoTrans team to identify spatial scenarios involving adjacent wells and/or two COCs at the same well.

Two software bugs were discovered during close examination of the DT output, one involving slight rounding inaccuracies in the plots, and the other involving distinguishing between the two-tailed confidence levels used for the prediction bounds and the one-tailed significance levels used

for testing whether or not time-dependent bounds should be used. Summit was alerted to those bugs and has subsequently corrected them.

Detecting the artificial anomalies

The following discussion of the GeoTrans results is organized by anomaly (a through j) rather than dataset (Apple through Orange), since the GeoTrans results were essentially unrelated to the dataset(s) in which a particular anomaly appeared.

a. This involves a new source arising in the vicinity of the percolation ponds, raising TCE levels in several wells in that vicinity. The artificial data values for this anomaly were in most cases large and easy for DT to detect. The spatial patterns allowed the GeoTrans team to conjecture a new source arising in the vicinity. The only issue arising is that of the background data selection for NZ-56; this is discussed further below.

b. This involves a clerical error in the lab, in which TCE and PCE values are swapped for all samples taken during the second week of April. DT was reasonably successful at identifying this anomaly. In general the ambient concentrations of TCE tend to be higher than those for PCE, and as a consequence when only one of the pair swapped was detected, it was the PCE value; the TCE prediction bounds, where there were elevated TCE values, tended to be wider and often included the values used as PCE nondetect values.

One curiosity is that the typical recent RL for TCE has been 0.5, whereas that for PCE has been 1.4. As a consequence, and due to choice made by GeoTrans to substitute the RL for NDs, in several cases the swap was detected and flagged even though all values involved were NDs! At present, the DT input format is “single-column”, so that the program does not know which values are actual values and which are NDs; if the values were reported with the “<” notations, the cause of these out-of-bounds values would be even more readily identifiable.

c. This involves benzene and toluene beginning to show up in the area around MW-105. Although the artificial values are low, the background values are quite low and/or NDs, so DT and GeoTrans had no difficulty in detecting them and identifying the spatial pattern involved.

d. This involves TCE concentrations in two neighboring wells with somewhat elevated values departing from their past patterns in opposite directions, NZ-32 going up and NZ-28a going down. DT had mixed success in detecting this anomaly. For NZ-32 the background data are too variable and the prediction bounds are too wide; this anomaly is not large enough to be detected using that background data, regardless of background data selection with currently available background data. For NZ-28a, on the other hand, the background data consist of two distinct regimes: lower values earlier and higher but decreasing values later. If all data are used to produce the prediction bounds, they are quite wide and the anomalies are not detected. If only the later values are used, DT elects time-dependent decreasing bounds, and the anomalies are detected. This is discussed further below.

e. This involves increasing trends in TCE in the vicinity of NZ-116. DT did not find these trends; they were too small relative to the variability in the background data.

f. This involves a hot spot developing in the Dozer Scar Site (NZ-07), with a new source containing TCE, PCE, and c12DCE. DT found the increases in PCE and c12DCE, for which the background data were nearly all NDs (there was one J value), despite having anomaly values only around 2 to 2.4. It did not, however, find the greater increase in TCE. As with NZ-28a in

scenario d, it would have been more successful with a more judicious selection of the background data.

g. This involves TCE changes in NZ-06, NZ-07, and NZ-10, with NZ-07 decreasing to ND and the others increasing. DT did not find these. In the case of NZ-06 the background data were too variable. For NZ-10 there was an outlier (atypical measurement) in the background data; if it had been screened out, the anomaly would have been detected. And for NZ-07, as discussed in scenario f, the tighter time-dependent bounds provided by the deleting the earlier portion of the background data would have allowed the anomaly to be noticed, but with all background data included the bounds were too wide.

h. This involves values at two historically low wells (NZ-51 and NZ-97) becoming suddenly rather high in April, returning to typical values in October, for TCE with PCE following along. One occasionally sees such patterns, for example, when sample collection bottles are mis-labeled. DT found three of the four easily. The fourth, PCE in NZ-51, would have required careful selection of background data to produce time-dependent decreasing bounds. This is discussed further below.

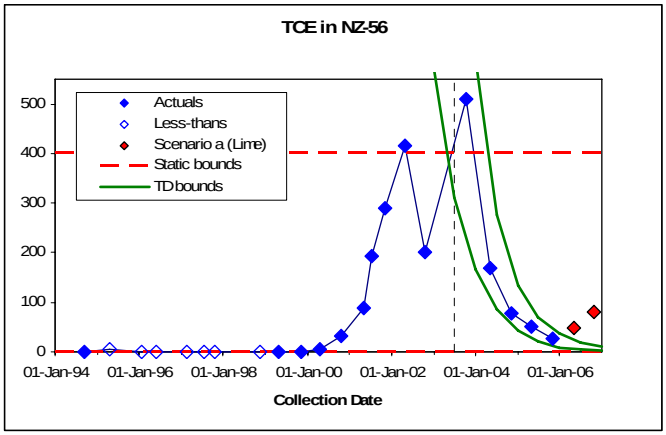
i. This involves a sudden drop to NDs in TCE in April with return to typical values in October at NZ-39 and NZ-81. DT found the drop in NZ-39 but not that in NZ-81. In NZ-81the background data are quite scattered, although with a more judicious selection of background data it would have been identified.

j. This involves a clerical error in the data, with all TCE values for the third week of October being quite low. DT found some but not all of these; again, the selection of background data is a critical issue.

<u>On background data selection and ND treatment</u>

In several instances the result obtained with DT would have been different with a different selection of background data and/or treatment of NDs. A few examples of this phenomenon are shown here.

The first is TCE in NZ-56, involved in scenario a. The data are shown in the accompanying plot. NDs are plotted at their RLs using open symbols; actual values are plotted using solid symbols. After a period of very low and ND values, a slug of TCE passed through beginning in late 2000. The highest peak in that slug occurred with the October 2003 sampling, with value ~500. Measurements have been decreasing since then. Scenario a involves a new contaminant source affecting this well; in that scenario the latest values reverse the decreasing trend.

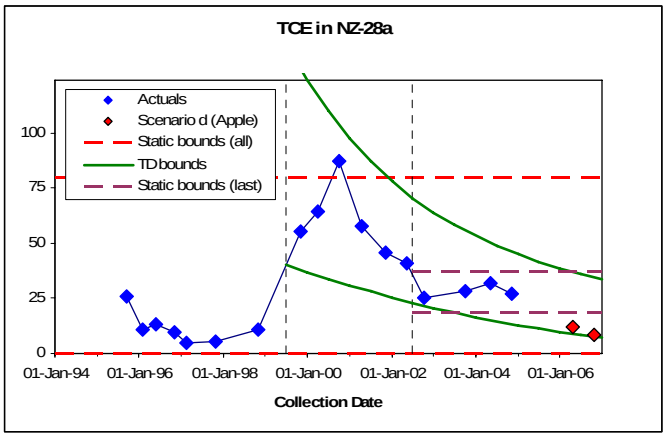


When all historical data are used as background data, the trend test gives a statistically significant overall <u>increasing</u> trend; since DT does not allow for increasing trends, it fits static prediction

bounds, which ignore the data patterns. These are shown in dashed red lines on the plot. (The lower bound, computed as -215.6, is shown as 0.0.) These static bounds are very wide, and would not be able to detect an increase due to a new source until there was substantial impact from that source.

On the other hand, if one started the background data at the highest peak (October 2003), there would be a statistically significant decreasing trend, so time-dependent prediction bounds would be used. These are shown in solid green lines. The scenario a values are above these bounds, and would be flagged.

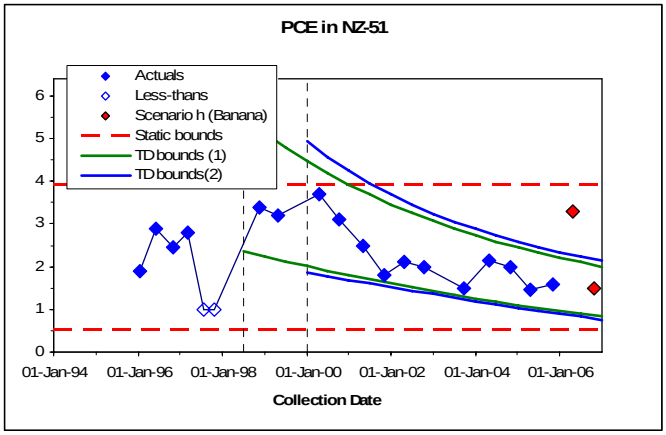
Clearly, in situations such as this the results obtained by DT are very dependent on the background data selection. The selection of background data should be made with an eye toward what sorts of patterns in future data should one want to be alerted to (or not). In selecting a background period starting with Fall 2003 (to the right of the vertical dotted line), one is saying that a continuation of the descent from the high peak would NOT be newsworthy, but deviations from that should be flagged.



A similar situation occurs with TCE in NZ-28a, involved in scenario d. Here there is an early period with moderately low concentrations, followed by a peak, then followed by two years of roughly constant concentrations around 28. The dashed red lines are the static bounds obtained by using all historical data as background data. If one starts the background data in Fall 1999 one obtains the green time-dependent bounds (starting at the peak in Fall 2000 gives virtually the same bounds). But then using only the last four, roughly steady-state observations at approximately the initial level seen in 1995 again gives static bounds. The artificial anomalies are in-, in-, and low out-of-bounds respectively depending on which background data are selected. (The actual 2006 values are 42 and 43, and would have been in-, out-, and high out-of-bounds respectively, depending on which background data were selected.)

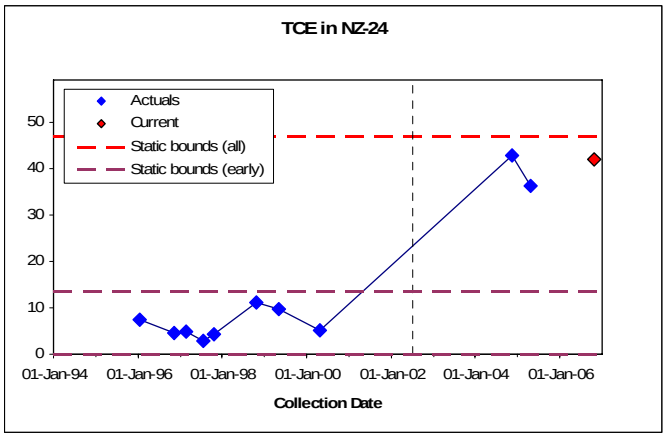
Again, the consideration that should drive the background data selection is "which historical values are representative of what is to be expected in the near future?" With TCE in NZ-28a, the period prior to 1999 is ancient history and should be ignored for the purpose of determining the prediction bounds. Whether the background data are started with the new regime in Fall 1999, the peak in Fall 2000, or the recent steady-state phase in Fall 2002 is a function of what one expects are for the future of the data, or more appropriately, what sort of patterns would one want to be alerted to in the future data.

The situation with PCE in NZ-51 is similar but more subtle; scenario h involves a spurious high value for TCE in April, returning to normal in October, with PCE tagging along. Whether DT would detect this mild anomaly depends on whether the downward trend since April 2000 was detected at the background data selection stage. Whether the background data are started in Fall 1998 or Spring 2000 makes little difference, however. A curiosity is that the anomalous value from Spring 2006 is actually quite similar to the early values in the background data used for the time-dependent decreasing bounds. Although not plotted here, starting the background data with Fall 2001 would produce static bounds, and the Spring 2006 artificial anomaly would be out-of-bounds.



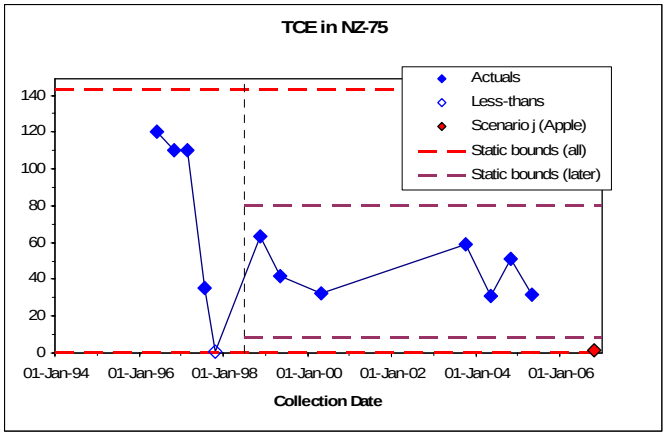
DT "found" an out-of-bounds low value for toluene in NZ-39 (plot omitted). The background data were all NDs at <1.1 (an early <11 was omitted), giving a mean of 1.1 and a standard deviation of 0.0, so the prediction bounds were (1.1 to 1.1). The value for April 2006 was the J value 0.26, which was flagged as low. Such flags can be distracting, but of little consequence since the user can tell immediately from the plot what the situation is.

There will inevitably be situations where determining the "correct" background dataset may be problematic. Consider, for example, the situation with TCE in NZ-24. None of these data are artificial. After an early history of low measurements (values no higher than 11), values jumped up to around 40 in late 2004. The current measurement is very consistent with the previous two, but one cannot know whether this is a new stable situation or a slug of contaminant passing through. In particular, there is a sizable gap in the data between Spring 2000 and Fall 2004.



DT demands a minimum of four background observations, and there are only two under the "new regime". If one wants DT to screen this COC/well combination, a background dataset containing at least four observations must be specified. If all background data are included, the prediction bounds are wide (dashed red lines), and attention is not called to this situation. On the other hand, in this situation one option is to exclude the <u>recent</u> data from the background data, at least until there is a data history of at least four observations that resemble steady-state observations. If one does this, the bounds (dashed plum-colored lines) are narrower, and the high current value is flagged. Since the true nature of the recent activity in this well is not known, the latter would be a reasonable course of action.

For a final example, consider TCE in NZ-75. In scenario j the TCE value suddenly becomes very low for all samples obtained during the third week in October; this might result from failing to properly adjust for a sample volume or dilution. In NZ-75 the early background data show a precipitous decrease from around 120 to ND. Beginning in Fall 1998, though, values vary between 30 and 65 with no noticeable pattern. The prediction bounds computed using all the data or only the recent data are considerably different; the latter represent our expectations for the future measurements better than the former.



In this case one might alternatively consider simply excluding the ND in Fall 1997 (and possibly also the low value in Spring 1997), in which case one might end up with time-dependent decreasing bounds. If one did so, and the data remained steady-state, eventually they will produce out-of-bounds high results. This will provide occasion to reconsider the background data selection.

As a general observation, the plots shown here include both the data actually used as background data and the other historical values that were excluded from the background. Doing so has advantages and is recommended as a future improvement to DT. In the final example above, for instance, if future values tended to decrease in time, one might eventually wish to consider changing from the steady-state bounds based on the more recent data to time-dependent decreasing bounds; having the whole historical record shown in the plots would assist one in considering such a decision.

Discussion of DT Results

As stated, DT and GeoTrans identified most of the anomalies present in the actual 2006 data (Orange dataset) and many of the artificially introduced anomalies in the other datasets. In some cases the anomalies were too small relative to the variation in background data to be found. In several scenarios GeoTrans was able to correlate anomalies spatially, outside of the software itself.

One shortcoming is that DT often flags very low values when the background data are mostly NDs and/or J values. Such situations are immediately apparent when one views the plots, however. DT did miss a few low-level increases in cases where the background data consisted of NDs and J values with varying reporting limits. GeoTrans simply used the RL itself as the substitute for each low ND value, deleting NDs with higher than typical RLs. If another ND treatment had been used, such as replacing all low-level NDs with the same arbitrary value, such as half the most common RL, it would have caught more real low-level increases, and also would have had more false alarms due to J values.

The GAFB data did reveal a major issue with DT, however, that of the selection of background data. The GeoTrans team initially simply put all of the historical data into the background

dataset, excluding only NDs with elevated RLs. This approach might mimic the treatment that a junior staff member might use, in spite of the concerns expressed by the site personnel. In several situations this resulted in DT missing artificial anomalies that would have been found with a more sophisticated selection of background data. Examples of these situations are discussed in the preceding material.

A recommendation that results is the following. Summit should consider including a discussion of these issues, and perhaps a tutorial using these or similar examples, in the Users' Guide and related materials. There has been some debate within the ESTCP project team about the level of knowledge and experience needed to successfully use DT. It may well be that the background data selection step requires more care and familiarity with the issues than does the routine screening of a current dataset. That initial background selection would take place when DT is used for the first time at a given facility. This recommendation presupposes that there is a mechanism for documenting and communicating the background data selection decisions (and possibly other decisions as well) between successive uses of DT to screen new datasets.

In a similar report on the use of DT with the Camp Allen data, the issue of updating the historical/ background dataset following the evaluation of one event and prior to screening the next event was raised. EnviroStat again recommends automatically adding all current data to the historical dataset following each event. Data that are unremarkable would be automatically added to the background data to be used in creating the PL bounds for the next event ONLY until there are eight background observations, however, in order to avoid masking slow trends in the data. Otherwise, and in any case where data in the current dataset have been flagged as potentially anomalous, the user should have the option of including the values in the background data or not. Decisions about updating the background dataset would require the additional expertise and care needed for the initial background dataset selection, of course. Again, implementing such a functionality will require creating some method of communicating decisions made while evaluating one event to the evaluation of subsequent events.

Concerning the PL Confidence Levels

Finally, a *caveat* regarding the confidence level used in the PL bounds is repeated. For GAFB these bounds are nominally 95% two-sided prediction limits. This is appropriate for situations such as that at GAFB. Such situations exist where there is known contamination and out-of-bounds findings are used only by the facility as a source of information on the progress of a remediation program, or for similar internal information purposes.

Such an application is rather different from the use of PL bounds in a formal monitoring program, such as groundwater detection monitoring conducted under RCRA (40 CFR Parts 264, 265, or 258) or similar state regulations. In those settings one uses background data to set upper prediction limits. An out-of-bounds value, by regulation and/or by permit, will trigger sometimes rather costly regulatory activity. In such applications the setting of the confidence level requires more sophistication. An extended discussion of this issue is beyond the scope of this report, but see "Ground-Water Monitoring Statistics Update: Part I: Progress Since 1988" (C.B. Davis and R.J. McNichols, *Ground Water Monitoring and Remediation* 1994, pp. 148-159) and articles referenced therein, for example.

Mass Flux and Mass Metric Results

Along the eastern portion of the base, where the MLU thins and contains more coarse sediments, Upper Aquifer groundwater can flow downward to the Lower Aquifer through the permeable lacustrine zone (PLZ). Site personnel are interested in quantifying the TCE mass migrating downward to the Lower Aquifer across the PLZ, and may also be interested in tracking the degree to which TCE mass is changing over time in the Upper Aquifer.

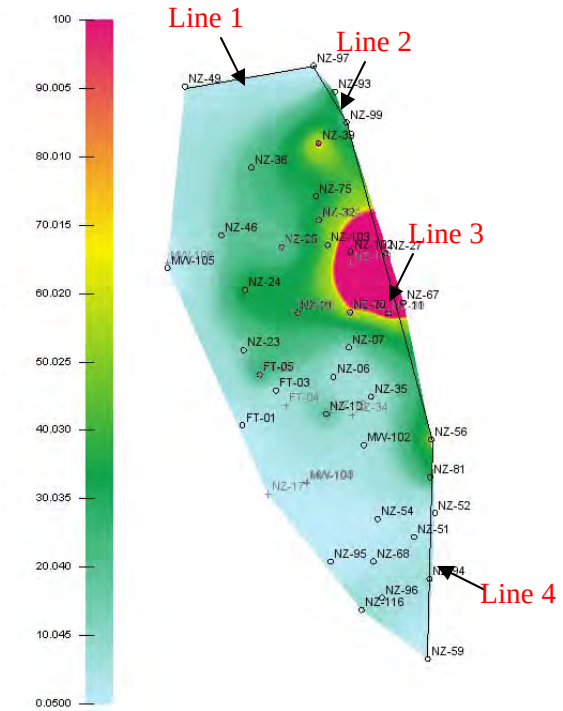
The Summit software includes a mass flux feature for tracking the flux of mass across a series of linear boundaries over time. The software also includes a mass metric feature for comparing the "relative mass" over time. It is referred to as "relative mass" because the software does not account for aquifer thickness and porosity; rather, it calculates mass per unit volume of aquifer.

Mass Flux Calculations

From water level maps in site reports, the contours of groundwater levels in the Upper Aquifer are approximately parallel to the boundary of PLZ, which implies that groundwater flow is essentially perpendicular to the boundary of the PLZ.

The software requires the user to define one or more line segments graphically. The software estimates concentrations at different points along the line segments, multiplies them by the groundwater flow rate input by the user for that line segment, and then sums up mass flux along the entire line segment. For this example, four line segments were drawn to approximate the shape of the PLZ boundary. These lines are illustrated (in black) on the figure below.

Mass Flux Boundary Lines (In Black)



The four lines are basically defined by locations of five wells:

- Line 1: NZ-49 to NZ-97
- Line 2: NZ-97 to NZ-99
- Line 3: NZ-99 to NZ-56
- Line 4: NZ-56 to NZ-59

Information regarding these lines is provided below:

Well	East Coordinate	North Coordinate	Line	Length (ft)
NZ-49	6748970	2048548		
NZ-97	6751385	2048935	1	2445.811522
NZ-99	6752002	2047870	2	1230.818427
NZ-56	6753598	2041912	3	6168.061284
NZ-59	6753529	2037788	4	4124.57719

Once the lines are defined, the user must enter the groundwater flow rate for each line segment. This was calculated with the following equation:

$$Q = vAn$$

where

v = groundwater velocity in ft/d (equal to Ki/n)
A = area through which flow occurs (segment length times saturated thickness)
n = porosity
K = hydraulic conductivity
i = hydraulic gradient

According to the 2005 Hydrogeologic Conceptual Site Model Report prepared by MWH, the groundwater flow direction in the Upper Aquifer is typically toward the northeast, with magnitude of hydraulic gradient ranging from 0.001 to 0.055. Groundwater velocity in the Upper Aquifer ranges from 0.2 to 1.0 ft/day, and a representative saturated thickness of the Upper Aquifer is 50 ft. Porosity was assigned as 0.3 by GeoTrans. The software does not account for effects of retardation and/or decay of TCE.

Mass flux calculations were performed based on a groundwater velocity of 1.0 ft/d. The calculation of flow rate for the case with groundwater velocity of 1.0 ft/d is presented below.

Flow Rates Based on GW Velocity = 1.0 ft/day

Line	Velocity (ft/d)	Porosity	Flow rate (ft^3/d)	Flow rate (liter/d)
1	1	0.3	36687.17283	1038865.045
2	1	0.3	18462.27641	522793.4487
3	1	0.3	92520.91926	2619900.677
4	1	0.3	61868.65785	1751925.293

The software requires the flow rate in units consistent with the concentrations, which in this example are µg/l. Thus, the flow rate is input as liters per day. The mass flux results over time for this scenario (i.e., groundwater velocity of 1.0 ft/d), based on the same time-varying data utilized for the spatio-temporal analysis, are provided below.

Mass Flux Rates over Time (GW Velocity = 1.0 ft/day)

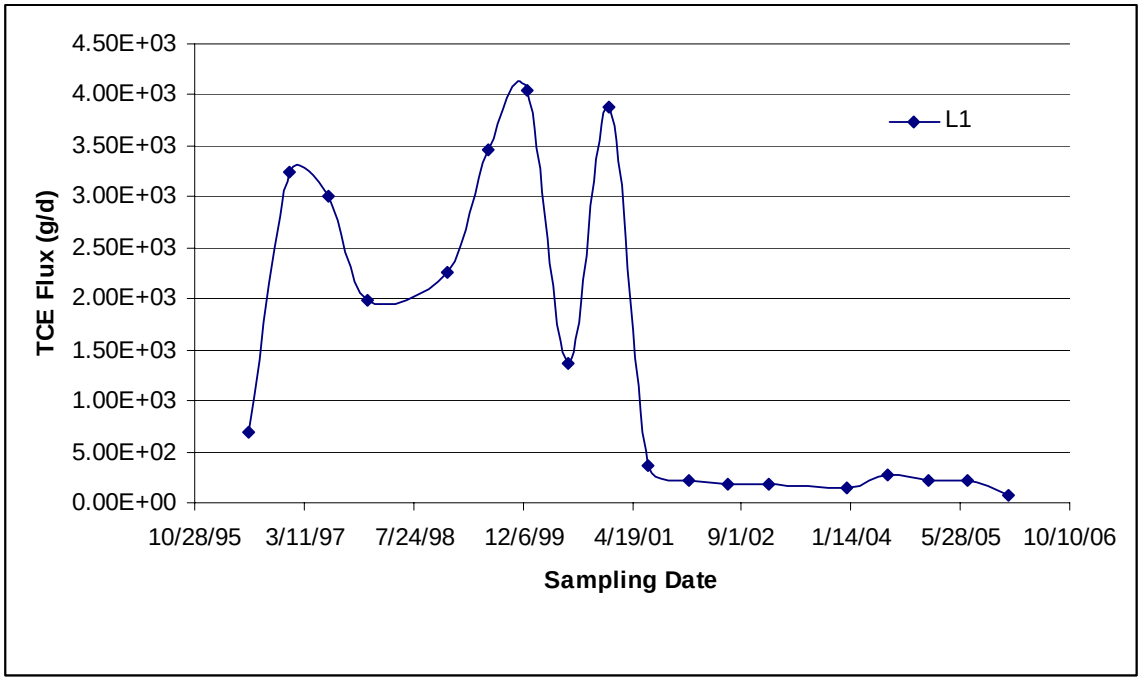
Date	Line 1 TCE flux (kg/d)	Line 2 TCE flux (kg/d)	Line 3 TCE flux (kg/d)	Line 4 TCE flux (kg/d)
6/30/1996	0.7000	0.5902	27.1488	1.6316
12/30/1996	3.2463	1.0450	42.0180	2.6568
6/30/1997	3.0109	1.0237	67.9764	2.4798
12/30/1997	1.9853	0.7006	25.1965	1.5652
12/30/1998	2.2557	1.0712	128.2250	0.3096
6/30/1999	3.4600	2.1146	83.2040	0.4338
12/30/1999	4.0521	1.9791	106.7071	1.2326
6/30/2000	1.3665	1.3043	79.4417	0.4088
12/30/2000	3.8869	1.8095	147.7453	1.8022
6/30/2001	0.3569	0.0524	94.3221	3.7328
12/30/2001	0.2197	0.1536	128.9569	1.9446
6/30/2002	0.1817	0.3142	110.9247	5.0886
12/30/2002	0.1909	0.2954	92.9866	4.3980
12/30/2003	0.1458	0.4494	120.1754	5.8329
6/30/2004	0.2649	0.5240	60.5555	4.1424
12/30/2004	0.2167	0.4803	78.7893	2.4767
6/30/2005	0.2108	0.5266	50.3532	2.0892
12/30/2005	0.0657	0.4868	48.6277	2.1710
Average mass flux (kg/d)	1.4343	0.8289	82.9641	2.4665
Mass flux percentage	1.63%	0.95%	94.60%	2.82%

Based on these results, the total average TCE flux across PLZ over these time periods (i.e., 06/1996 ~ 12/2005), with groundwater velocity of 1.0 ft/d is estimated to be 87.7 kg/day, which is calculated by adding the results for the four segments. Also, significantly more (94.6%) of the mass flux occurs along line 3 than the others, mostly because of higher concentrations and also because the line is longer.

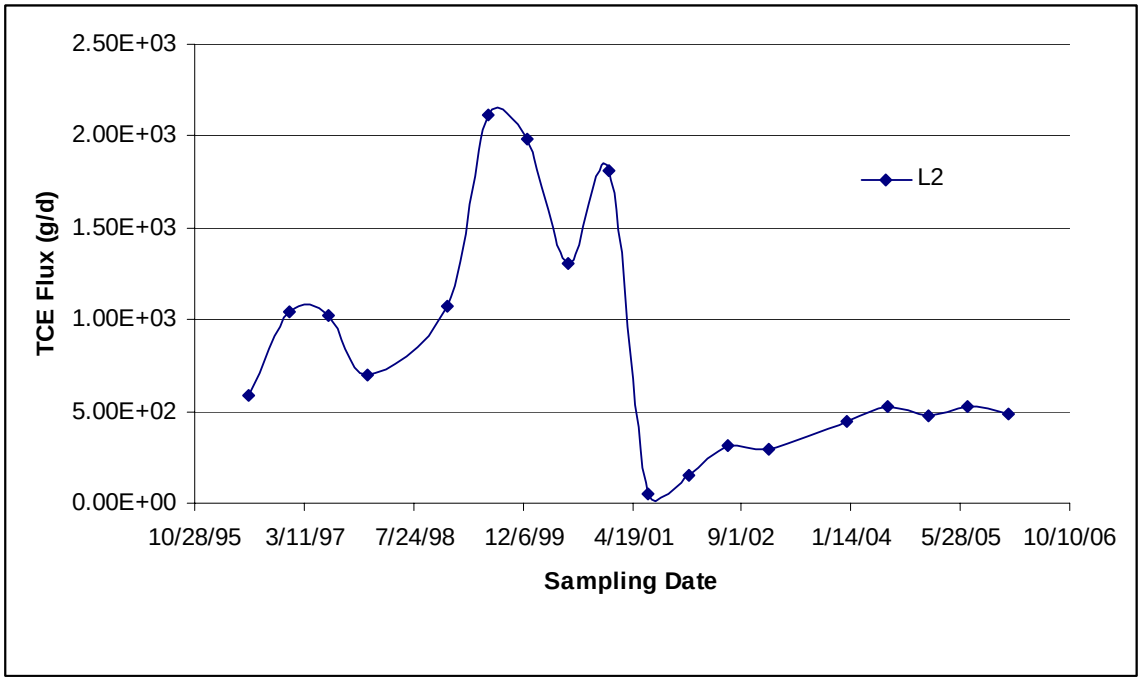
Obviously, different values will be calculated if different values for groundwater velocity and/or porosity are assumed. However, the relative mass fluxes between sampling events will remain consistent (i.e., they will be scaled consistently).

The time series plots of the mass flux across each of the four line segments (L1, L2, L3, and L4), in grams per day, are presented below:

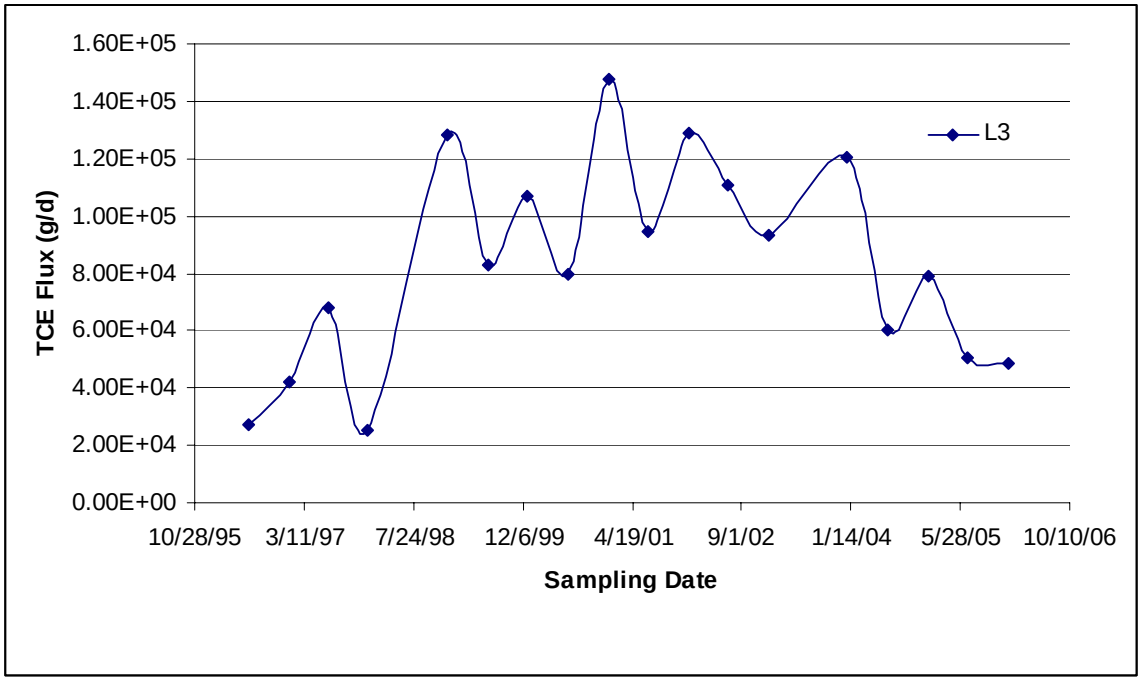
Mass Flux Rates over Time, Line 1 (GW Velocity = 1.0 ft/day)



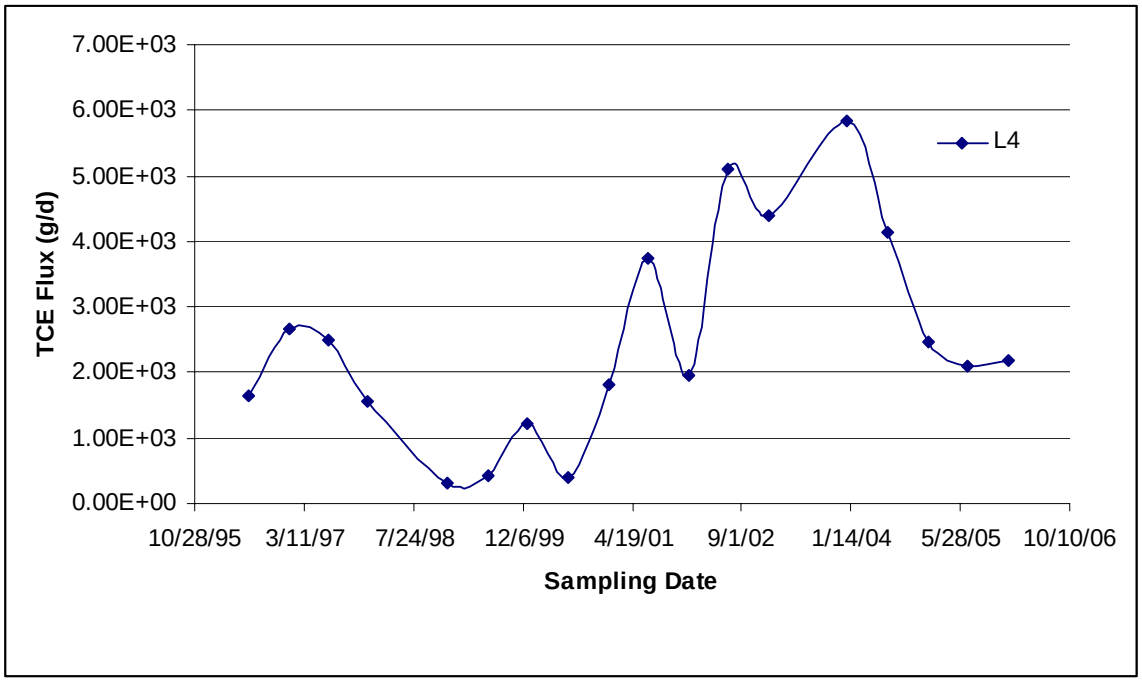
Mass Flux Rates over Time, Line 2 (GW Velocity = 1.0 ft/day)



Mass Flux Rates over Time, Line 3 (GW Velocity = 1.0 ft/day)



Mass Flux Rates over Time, Line 4 (GW Velocity = 1.0 ft/day)

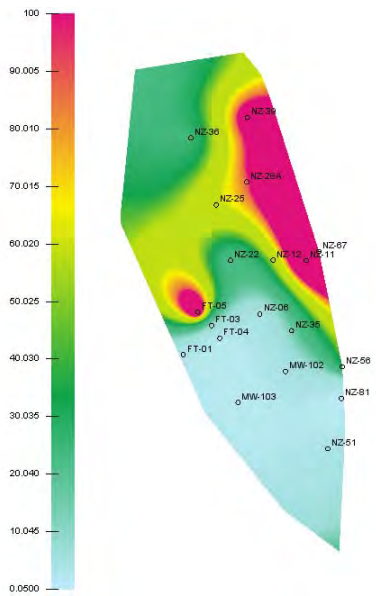


For Line 1 and Line 2, the results suggest much higher mass flux before 2001 compared to after 2001, and for Line 4 somewhat lower mass flux before 2001 compared with later. For all the lines the mass flux is relatively stable or decreasing after 2002.

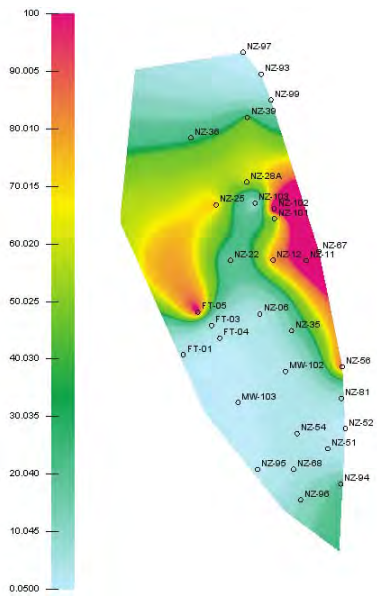
Discussion Regarding Mass Flux

It is interesting to find that the mass flux for Line 1 and Line 2 changed significantly over a short period from 12/2000 to 6/2001. The comparison of interpolated plume visualizations generated by the software for the two sampling events is provided below.

Interpolated Plume Map, 12/2000



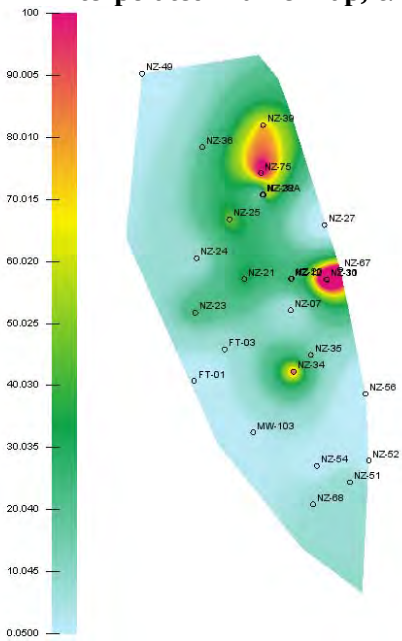
Interpolated Plume Map, 6/2001



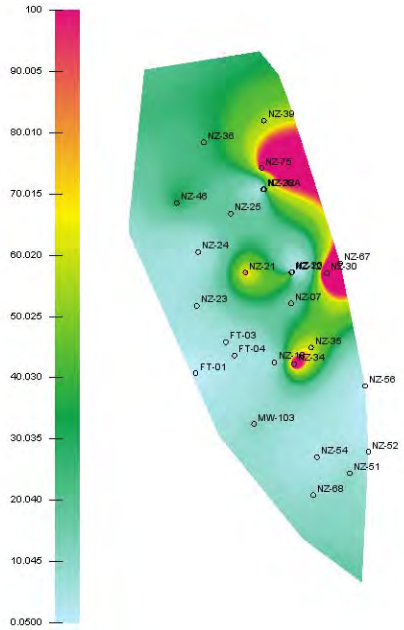
Focusing on the extreme northern portion of these figures (i.e., the vicinity of Lines 1 and 2) it is noteworthy that the interpolated concentrations in 6/2001 (right figure) are significantly lower due to the presence of three new monitoring wells (NZ-97, NZ-93, NZ-99). These wells were not present in 12/2000 or in other previous events. These three additional wells give more accurate information about the actual TCE concentrations in their vicinity. In other words, the interpolated concentrations in 12/2000 are biased due to fewer sampling locations in that area. This points out a problem with this mass flux feature, which is that different events can use different selections of wells, which can bias the calculations. This also brings about a question as to whether the data from newly installed wells should be "brought backwards" to previous sampling events, allowing comparisons to be based on similar distributions of sampling locations which would presumably result in more reliable estimation of mass flux changes over time.

Another interesting case involves wells that are sampled in one period but not a subsequent period. The upper left-hand corner of the following figures, in the vicinity of well NZ-49, provides a good example to illustrate this point.

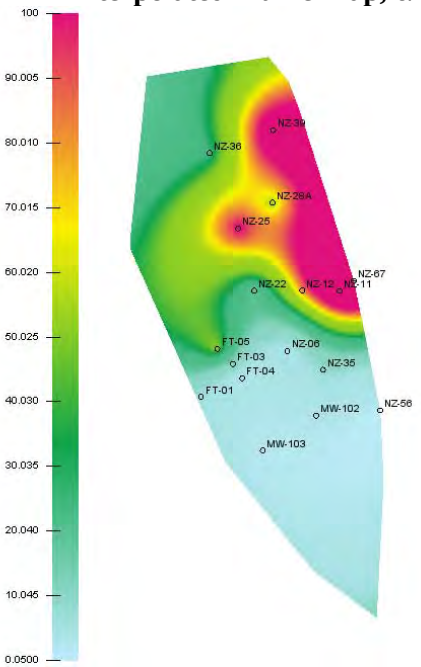
Interpolated Plume Map, 6/1996



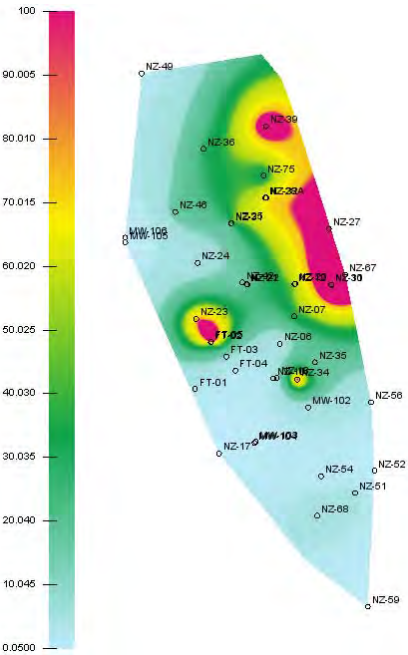
Interpolated Plume Map, 6/1997



Interpolated Plume Map, 6/1999



Interpolated Plume Map, 6/2000



In 6/1996, NZ-49 was sampled, and it appears to be clean. From 1997-1999, NZ-49 was not sampled, but in 6/2000 it was again sampled and was found to be still clean. In this case, it makes perfect sense to assume NZ-49 always had low concentrations in between the 1996 and 2000 samples. However, the interpolations for the events between these two samples are based only on spatial interpolation for wells sampled in those events, and thus there is extrapolation in the area near NZ-49 for those events that estimate higher concentrations. A different result would have occurred if values for NZ-49 had been assigned for events where it was not sampled, based on

temporal interpolation of the 1996 and 2000 results at that well. At this point the software does not accommodate this.

This discussion leads to an argument that the analyst should consider adding "estimated values" for sampling locations that are not actually sampled in specific events, so that the interpolated plumes used for the mass flux estimates in each period are based on a consistent set of locations. These estimated values would be based on actual data at those locations for events where they were sampled, plus the underlying conceptual model of the site, such as the general groundwater flow and plume transport, whether new sources of contaminant are expected, impact of remedial actions, etc. Ambiguities can still arise, however; for example, it could have happened that the TCE values in NZ-49 might have been different in 1996 and 2000, so that one would know that there was a change during that period, but not know how to fill in the missing values "correctly".

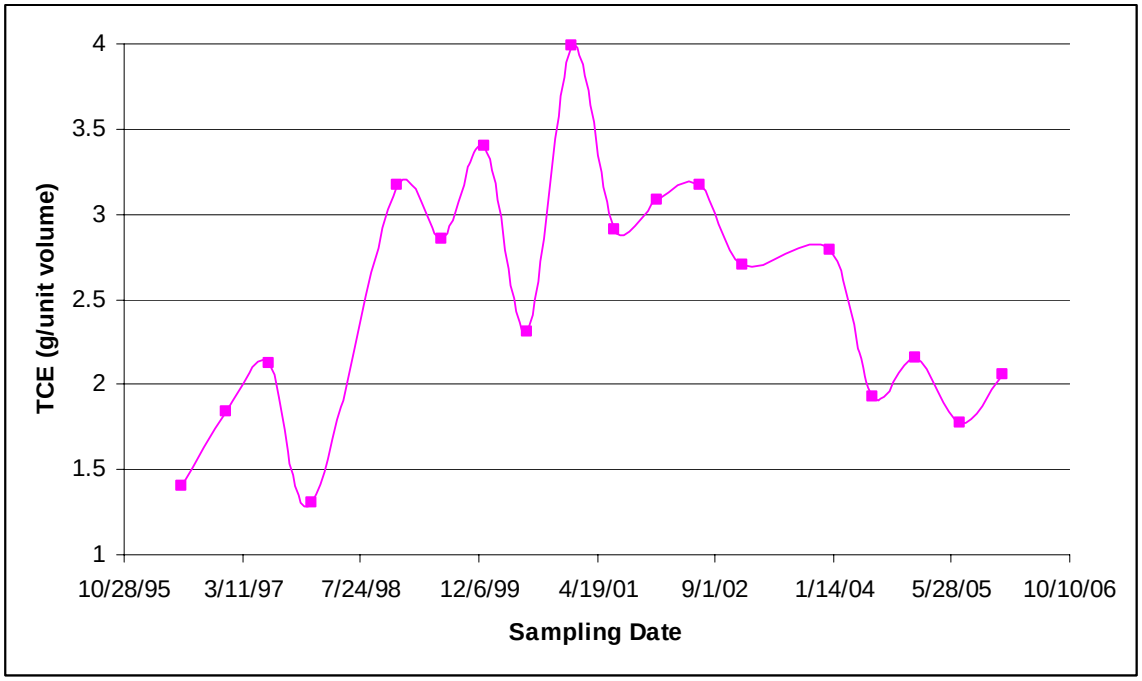
Mass Metric Results

The mass metric feature of Summit monitoring tools is designed to assist the analyst in comparing the relative mass in the plume for different sampling events. The interpolated concentration value at every cell in the plume map for that COC at that time period is summed, and each cell is assumed to represent the same volume (i.e., incorporating vertical extent and porosity) as every other cell. Thus, the results are actually given in mass per unit volume, rather than absolute mass.

The results for the mass metric, based on the same time-varying data utilized for the spatio-temporal analysis, are presented in the below table:

Date	TCE (g)	# of Sample
6/30/1996	1.405695377	27
12/30/1996	1.843573312	25
6/30/1997	2.121652894	27
12/30/1997	1.310147807	31
12/30/1998	3.173024216	38
6/30/1999	2.857299804	30
12/30/1999	3.395901221	17
6/30/2000	2.307259626	43
12/30/2000	3.989273652	19
6/30/2001	2.907015711	31
12/30/2001	3.079288147	39
6/30/2002	3.172984274	44
12/30/2002	2.701745975	28
12/30/2003	2.789606450	50
6/30/2004	1.926729518	48
12/30/2004	2.153241681	52
6/30/2005	1.776017540	39
12/30/2005	2.058060098	35

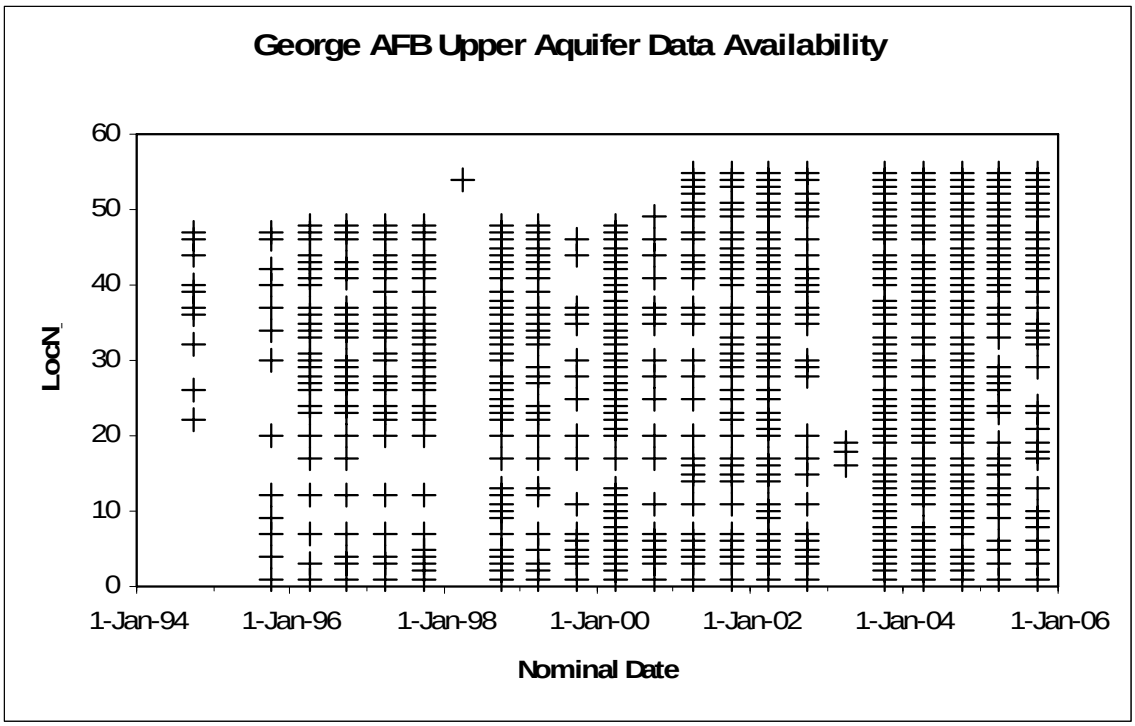
Mass Metric versus Time



It can be seen from the above mass metric plot that the mass of TCE per unit aquifer volume varies over time – it grows steadily from 1995 to 2001, when the TCE mass reaches its maximum of nearly 4 g/unit volume. After 2001 the mass is decreasing constantly; this may be caused by the implementation of the aquifer remediation. Additionally, the data for 12/30/2000 has a relatively high mass. Although this is partly explained by higher concentrations measured at some points in 12/30/2000, an examination of the sampling history reveals that there are only 19 data points for that event, compared to 43 for 6/30/2000 and 31 for 6/30/2001. Fewer sampling locations result in lower accuracy of the interpolation – in other words, the insufficiency of information causes some interpolation bias (in this case the interpolated values seem to be overestimated). The extent of the zone with highest concentrations is greater for the 12/30/2000 event partly because of lack of sampling at some wells that were sampled in the event prior and/or the event subsequent. This helps explain why the calculated mass metric has a particularly high value for 12/30/2000. It again suggests that it may be important to have similar data distribution per event for making these types of comparisons over time, based on some rule(s) for assigning data values at locations not sampled in specific events.

Discussion Regarding Mass Flux With Incomplete Data Filled in Manually

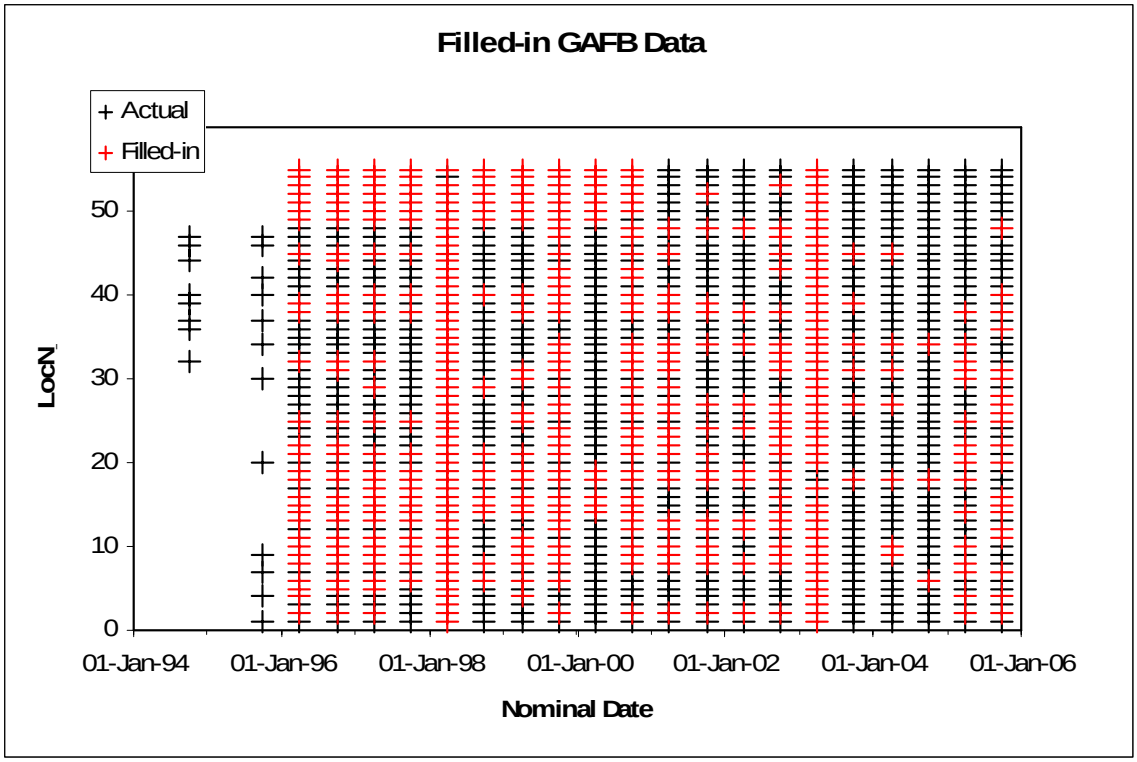
To assess the difference in the mass calculations that might occur if each dataset has similar data distribution, EnviroStat created an additional dataset where data gaps were eliminated by assigning values based on temporal interpolation. The schematic below shows the data availability over time for TCE at the former GAFB. The dates are those assigned to the spatiotemporal dataset. There are a few events with little data, but for those nearly all wells are missing, and for the others the majority of the wells are present.



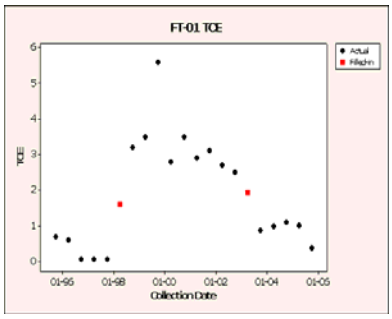
The table below indentifies the actual well associated with each location represented in the figure presented above.

Translation of "LocN" to SiteID									
LocN	SiteID	LocN	SiteID	LocN	SiteID	LocN	SiteID	LocN	SiteID
1	FT-01	12	NZ-07	23	NZ-20	34	NZ-34	45	NZ-59
2	FT-02	13	NZ-10	24	NZ-21	35	NZ-35	46	NZ-67
3	FT-03	14	NZ-101	25	NZ-22	36	NZ-36	47	NZ-68
4	FT-04	15	NZ-102	26	NZ-23	37	NZ-39	48	NZ-75
5	FT-05	16	NZ-103	27	NZ-24	38	NZ-42	49	NZ-81
6	MW-102	17	NZ-11	28	NZ-25	39	NZ-46	50	NZ-93
7	MW-103	18	NZ-111	29	NZ-27	40	NZ-49	51	NZ-94
8	MW-104	19	NZ-116	30	NZ-28A	41	NZ-51	52	NZ-95
9	MW-105	20	NZ-12	31	NZ-30	42	NZ-52	53	NZ-96
10	MW-106	21	NZ-17	32	NZ-31	43	NZ-54	54	NZ-97
11	NZ-06	22	NZ-18	33	NZ-32	44	NZ-56	55	NZ-99

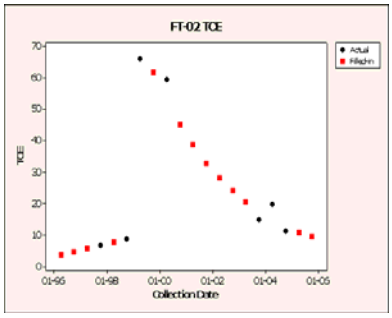
For 1996-2005 there are 625 actual values for 55 wells and 20 events. A completely filled-in data set would have 1100 values (55 wells x 20 events) and therefore 475 values would need to be supplied to have a "complete" dataset. About 100 of these occur in two events, those of April 1998 and April 2003. Others dominate the early data (the wells were added later), with SiteIDs NZ-81 and higher. The proportions of filled-in data range from 9% for NZ-39 up to 75% for several wells. Well NZ-111 with fewer than four actual values is omitted.



It would be difficult to imagine using the filled-in data as a basis for any sort of temporal optimization, since overall 43% of the values are artificial. Since the artificial values are based on interpolation or extrapolation, they are likely to exhibit less variability than the actual data would have, had they been obtained.

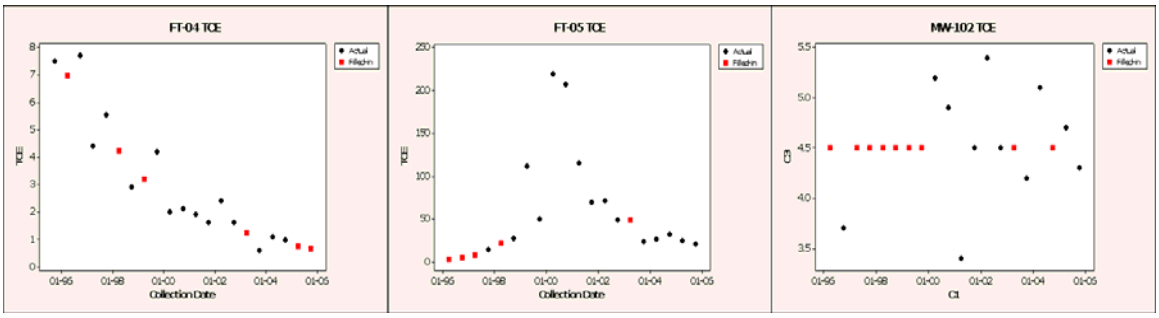


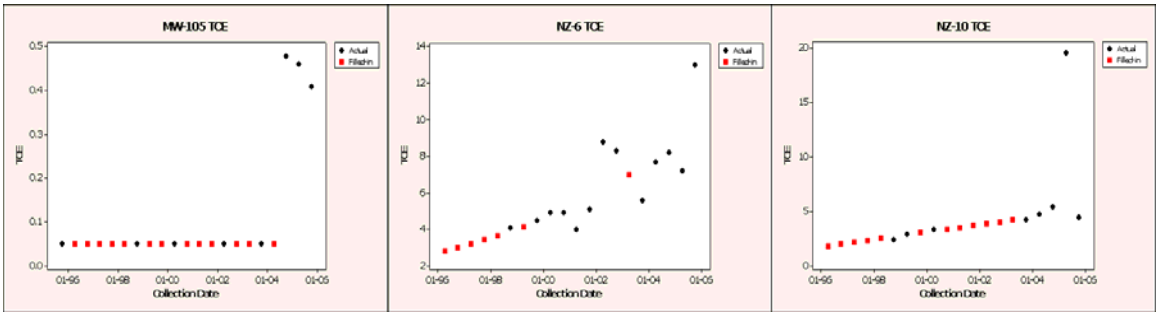
Here are a few examples of data histories with filled-in data, to illustrate the challenges involved. NDs are assigned the value 0.05. In the first, FT-01, the slug of TCE started somewhere during the data gap, so the values before and after were averaged. The other filled-in value is from a straight-line regression fit to data starting at the peak; no data transformation was used.

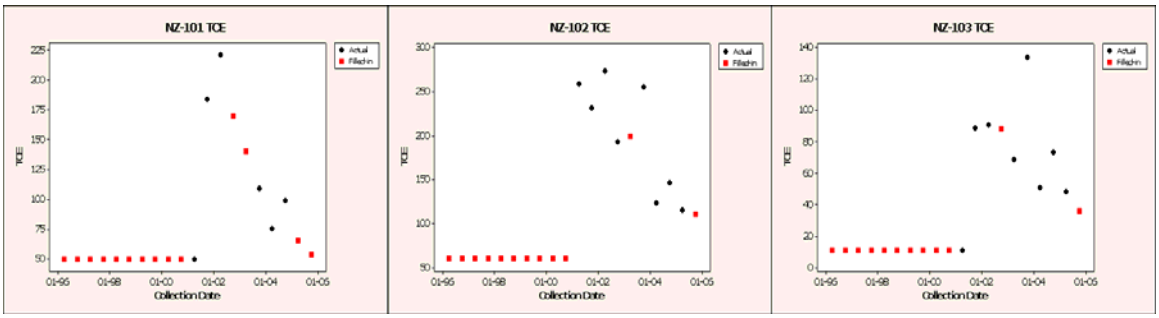


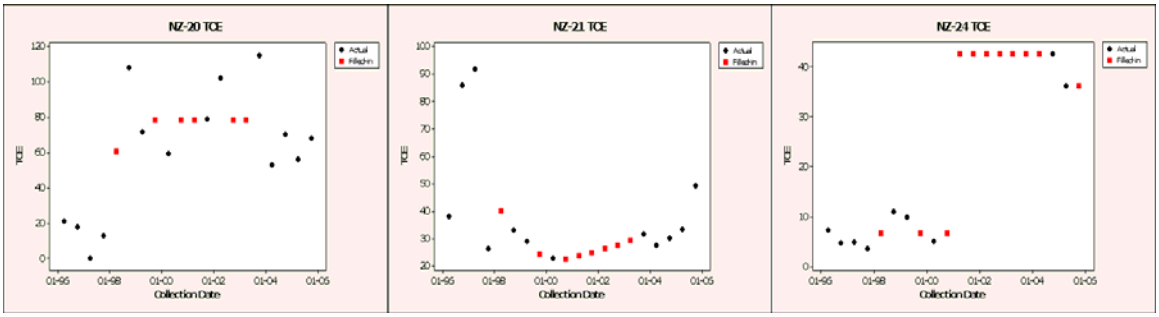
In FT-02 there are large gaps, with data missing before and after the available data. A linear fit was used at the beginning, but log-linear after since the linear fit would predict virtually zero for the last value. There are a lot of unverifiable assumptions in this plot!

With no statistically significant slope and no patterns, one approach is to simply fill in using the mean concentration value for the well. Twelve more examples follow, chosen to represent a range of challenges.









EnviroStat generally did not extrapolate prior to the first actual value, but just filled in back in time using the first actual value. In some cases where the first actual value was above ND, this surely misrepresents the actual beginning of elevated concentrations, but one cannot know what the reality might have been. One exception is with wells NZ-101 and NZ-102; EnviroStat applied the pattern from NZ-101 (which had an earlier data value with low concentration) to adjacent well NZ-102, and thus filled in the initial values for both wells using low values.

NZ-24 is also problematic, in that the early data are all low values (but not NDs), then there is a gap of several years, and the last two actual values are around 40 µg/l. Surely the concentrations were increasing during that gap, but the details are not knowable.

The most common interpolation technique was local log-linear regression; that is, selecting a subset of the data that seem to be following the same trend and doing a linear least-squares fitting on log scale. This avoids fitting negative values. On the other hand, one must take care with the NDs, as the rather low "graphing value" of 0.05 used for NDs is usually an outlier on log scale. (It works fine for its intended purpose on the original data scale.)

Other techniques used are simple linear regression, simply averaging the data values on either side of a one- or two-event gap (weighted average in the latter case), and of course carrying the earliest value backwards or the latest value forwards. Carrying values backwards is of course not really possible for actual monitoring, since (for example) the value for April 2008 is not known when one is attempting to find a fill-in value for October 2007.

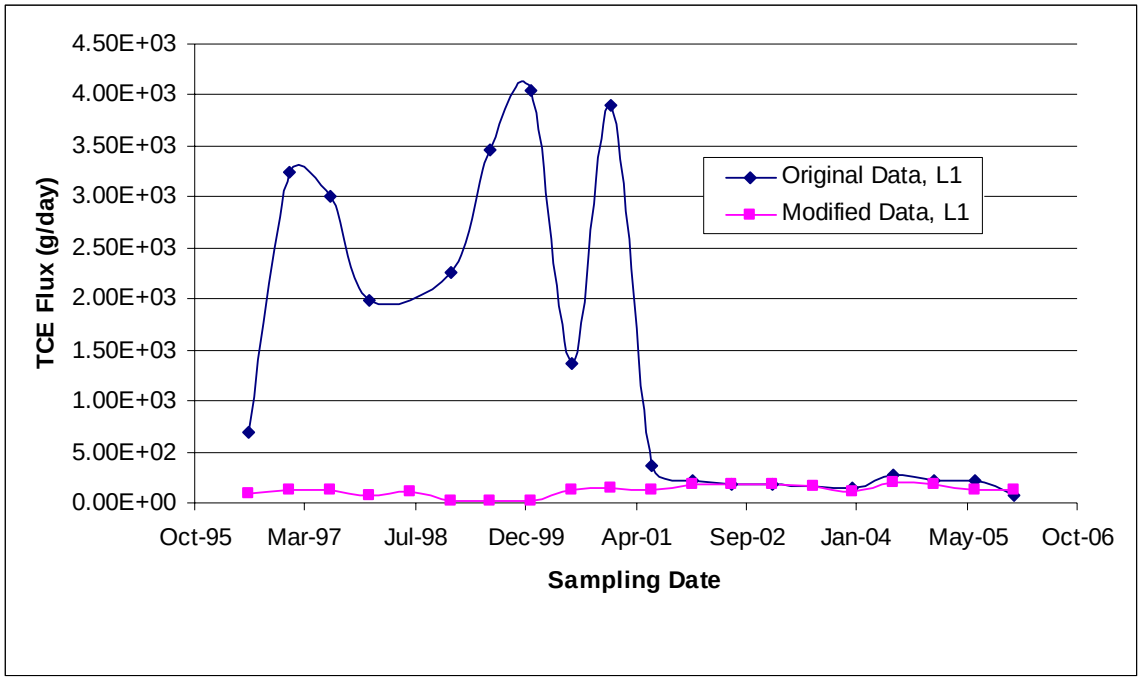
GeoTrans then utilized this additional dataset provided by EnviroStat for the mass calculations. This allowed for comparison of results for the dataset with incomplete data versus the dataset with incomplete data filled in. All other input parameters (e.g., groundwater velocity) were the same.

Mass Flux Rates over Time with "Filled-in" Data (GW Velocity = 1.0 ft/day)

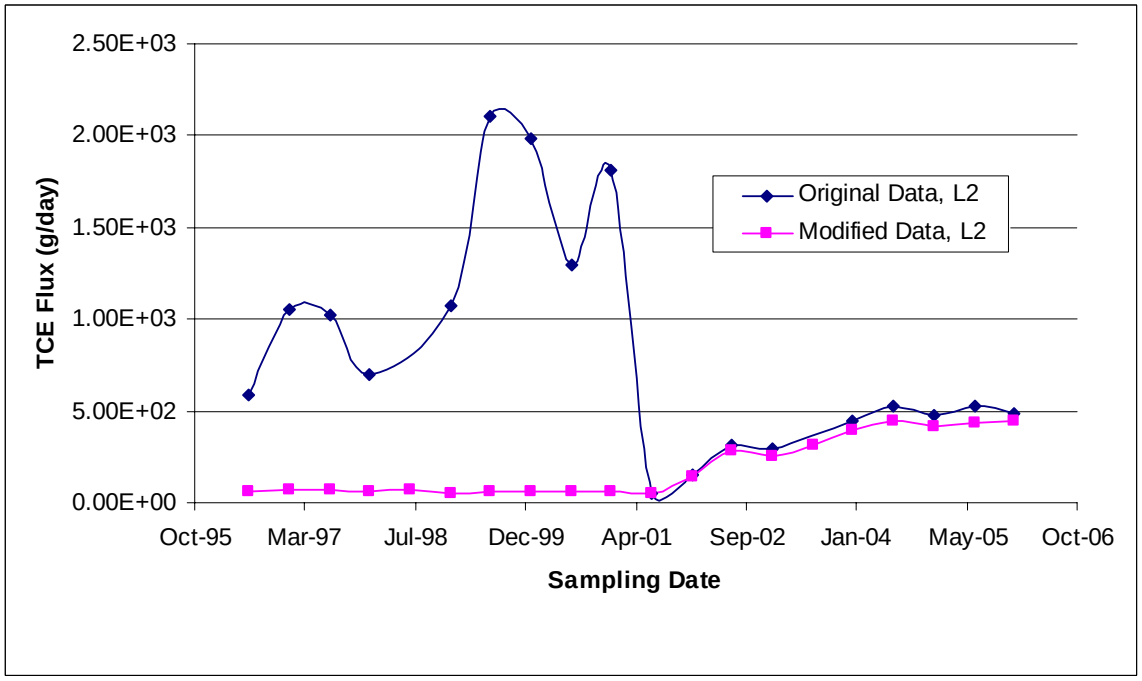
Date	Line 1 TCE flux (µg/d)	Line 2 TCE flux (µg/d)	Line 3 TCE flux (µg/d)	Line 4 TCE flux (µg/d)
6/30/1996	8.63E+07	5.59E+07	1.59E+10	1.43E+09
12/30/1996	1.19E+08	6.81E+07	2.63E+10	1.68E+09
6/30/1997	1.19E+08	6.82E+07	3.17E+10	1.75E+09
12/30/1997	7.40E+07	6.21E+07	1.90E+10	1.35E+09
6/30/1998	1.15E+08	7.36E+07	3.62E+10	1.18E+09
12/30/1998	1.94E+07	5.55E+07	4.13E+10	1.08E+09
6/30/1999	2.26E+07	6.12E+07	5.09E+10	1.15E+09
12/30/1999	2.29E+07	6.23E+07	5.24E+10	1.05E+09
6/30/2000	1.26E+08	6.47E+07	5.12E+10	1.33E+09
12/30/2000	1.46E+08	5.91E+07	6.77E+10	1.50E+09
6/30/2001	1.23E+08	5.49E+07	7.74E+10	2.65E+09
12/30/2001	1.85E+08	1.45E+08	9.14E+10	3.84E+09
6/30/2002	1.76E+08	2.82E+08	8.75E+10	1.07E+10
12/30/2002	1.74E+08	2.52E+08	7.47E+10	6.26E+09
6/30/2003	1.55E+08	3.16E+08	7.72E+10	1.02E+10
12/30/2003	1.13E+08	3.94E+08	9.45E+10	1.02E+10
6/30/2004	1.92E+08	4.42E+08	5.52E+10	4.55E+09
12/30/2004	1.78E+08	4.16E+08	6.48E+10	3.48E+09
6/30/2005	1.35E+08	4.32E+08	4.45E+10	2.73E+09
12/30/2005	1.32E+08	4.45E+08	4.44E+10	2.51E+09
Average mass flux	2.41E+09	3.81E+09	1.10E+12	7.06E+10
Mass flux percentage	0.21%	0.32%	93.5%	6.00%

The figures below compare the results for each mass flux line for the original data versus the modified data.

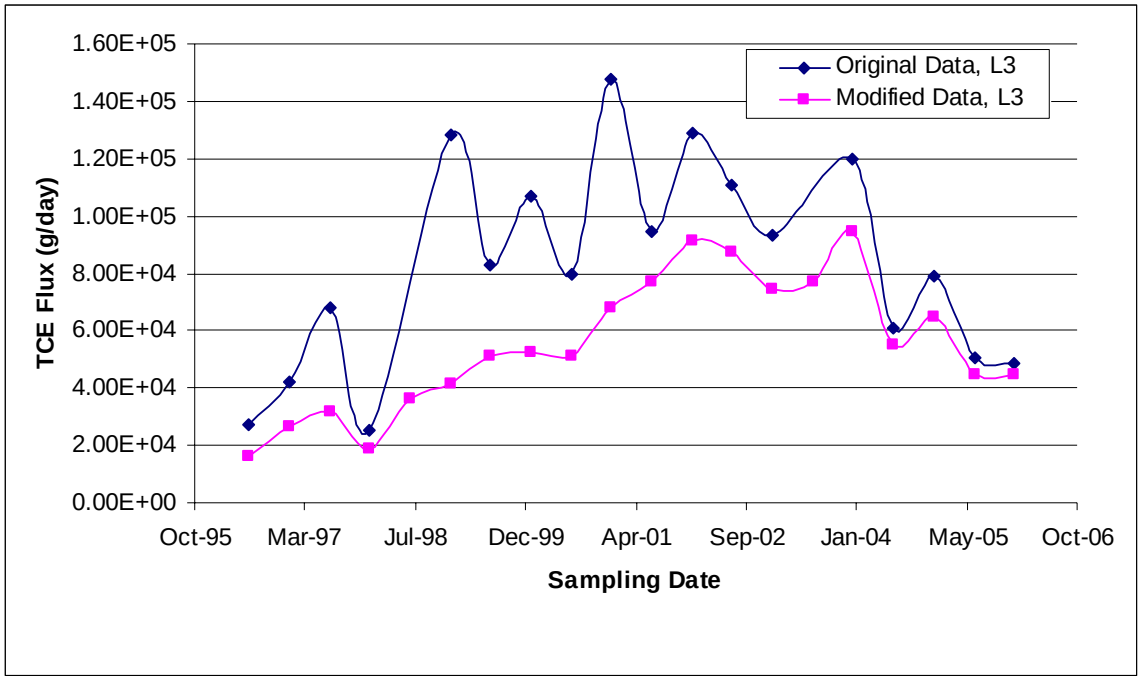
Mass flux across Line 1 (L1), Original Data versus Modified Data



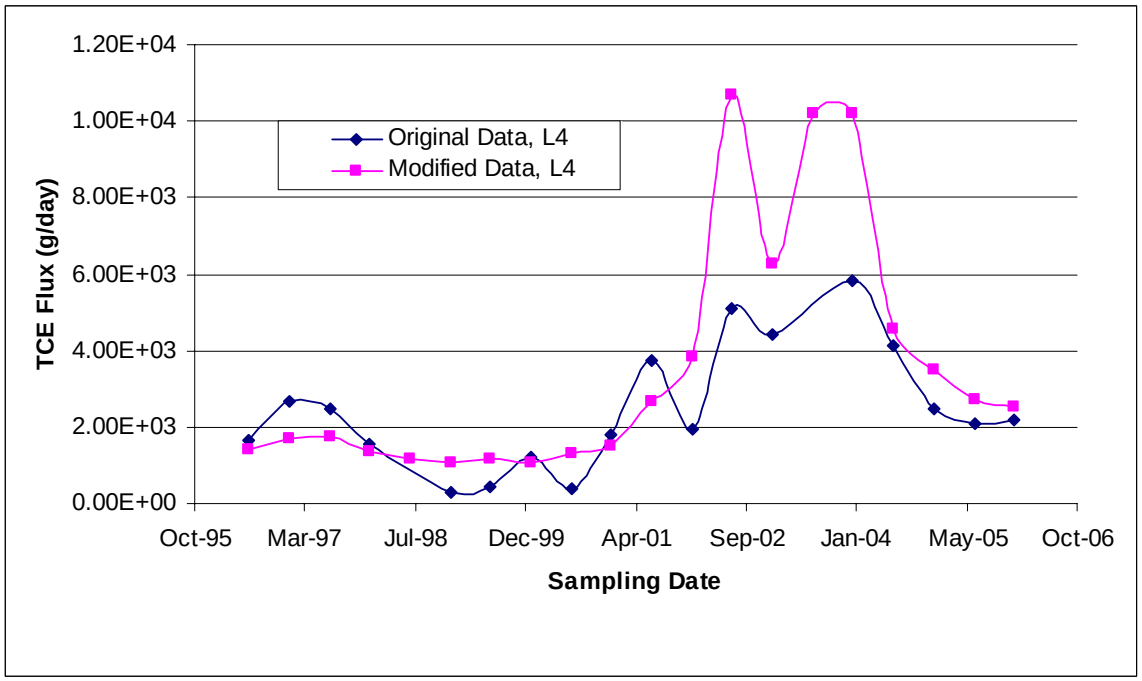
Mass flux across Line 2 (L2), Original Data versus Modified Data



Mass flux across Line 3 (L3), Original Data versus Modified Data



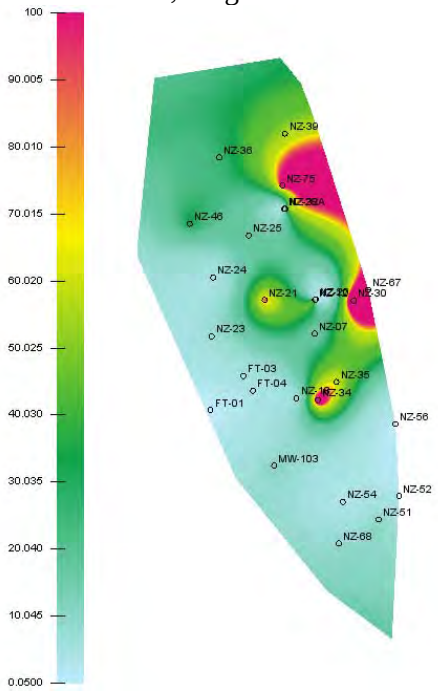
Mass flux across Line 4 (L4), Original Data versus Modified Data



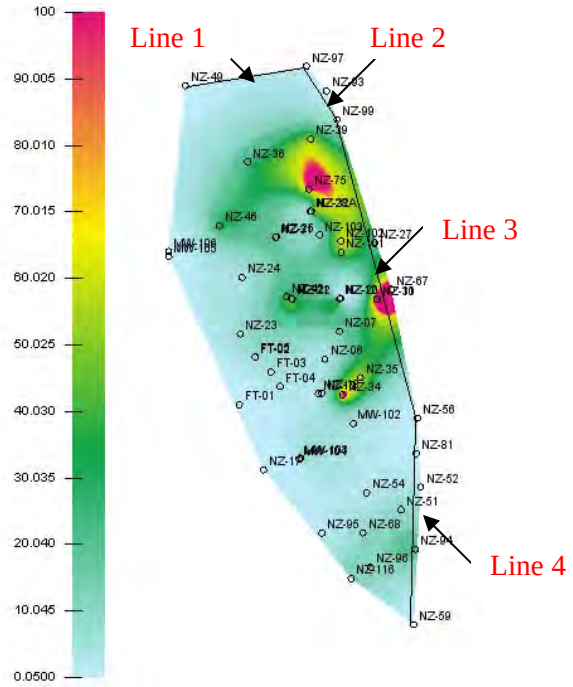
The above time-series plots of mass flux for each line illustrate that the mass flux results for modified data (i.e., filled-in data) are generally more stable than for the original data; that is, the mass flux trends look smoother without abrupt changes between adjacent events. Moreover, the estimated mass flux for original data is generally greater than for the modified data, except for Line 4, which has higher mass flux with modified data (possibly caused by NZ-94 which has relatively higher concentrations). It is clear by comparing the plume maps provided below for

selected sampling events that the original data result in overestimated concentrations in some areas.

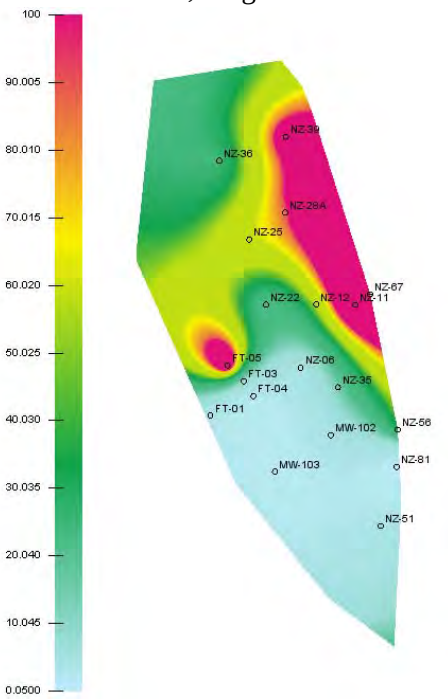
6/1997, Original data



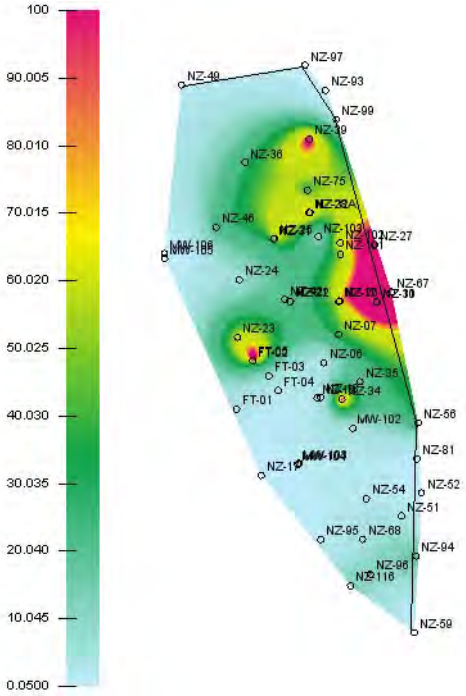
6/1997, Modified data



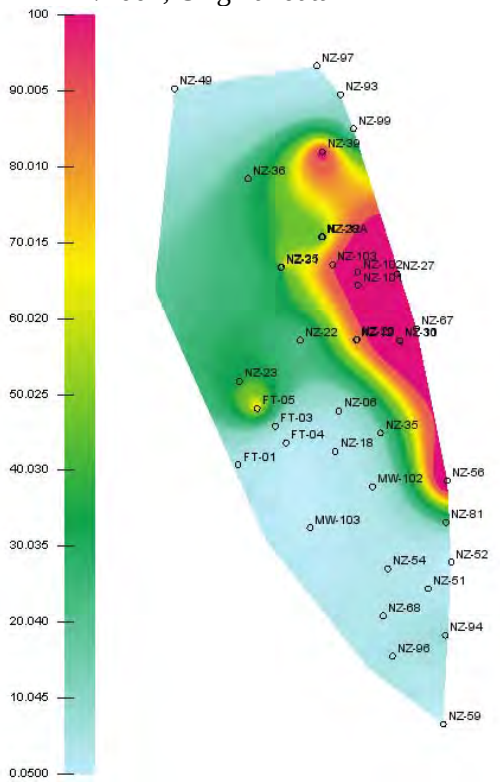
12/2000, Original data



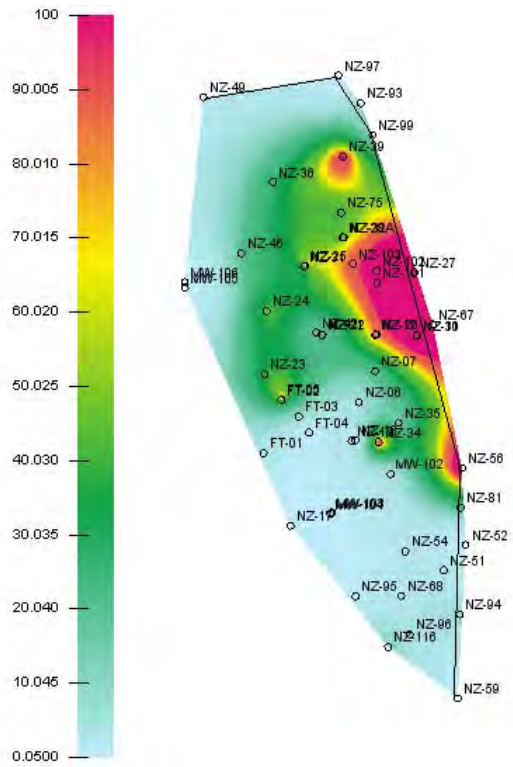
12/2000, Modified data



12/2001, Original data



12/2001, Modified data

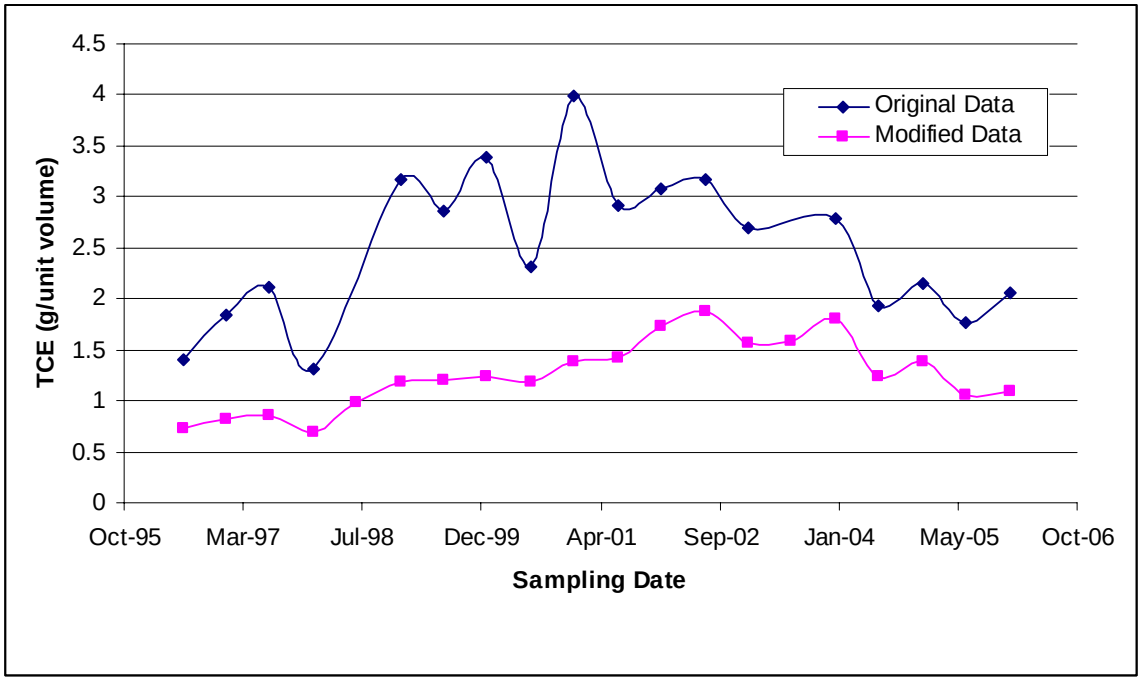


By filling in the missing data that overestimation is constrained (i.e., lower values were assigned using temporal interpolation/extrapolation manually versus higher values interpolated or extrapolated spatially by the software within each event). For example, as discussed previously, NZ-49 and NZ-97 (in the northern part of the site along Line 1) were problematic using the actual data. For NZ-49, there were some early data with low values, but other events with no sample at the well. At NZ-97, there were results from later events with low values, but no data from early events. In each case, using only the actual data, the software assigned higher values in those areas for events where samples were not taken, based on spatial extrapolation of other locations sampled in those periods. However, when data were filled in manually for events where samples were not taken, lower values were assigned. This additional information prevented the software from overestimating the values in those areas.

Another example is the vicinity of NZ-101 and NZ-102, located in the eastern part of the site along Line 3. By assigning values manually to complete the dataset, much lower concentrations were assigned at these wells for events where they were not sampled versus the values estimated by the software using spatial interpolation within each event. As a result, the area of highest concentrations is interpreted to be much smaller for the modified dataset versus the actual dataset, resulting in reduced mass flux estimates along Line 3 over time.

The results for the mass metric, based on the same time-varying data utilized for the spatio-temporal analysis but with incomplete data filled in, are compared to the results for the actual dataset on the figure below.

Mass Metric versus Time, Original Data versus Modified Data



Again, the mass metric plot of modified data looks steadier and its TCE mass is lower than original data, due to the reasons aforementioned.

Model Validation

The original 2006 dataset prepared by EnviroStat for validation indicates that some of the wells were sampled semi-annually, which means there are two 2006 concentration values for those wells for validation. For model validation purposes (for both the spatial and spatio-temporal analyses), only one sampling event is required for input data. Because no abrupt changes in concentrations were observed for wells with two sampling events during 2006, their concentrations were averaged. The modified reserved 2006 sampling data therefore includes 55 sampling locations with one value at each location.

The following general procedures were applied for performing model validation:

- Create a full dataset which includes all reserved sampling data; make sure the locations (i.e., well ID) are identical to historical ones. For example, if 55 wells were sampled in 2005 (historical event), the current 2006 dataset should contain the exact same wells. For wells that were not sampled in 2006, the latest available data value was assigned to it so that the computational domain remains the same.

- For wells that are recommended to be shut off by the Optimizer, set both their Max/Min sampling frequency to be "off"; for the remaining wells, set both their Max/Min sampling frequency to be "on". This way no sampling optimization will be performed by the software since the sampling frequency for each individual well has been predetermined.

- Run Optimizer with the above settings with only one generation, and display the interpolated concentrations at removed well locations.

- Compare the interpolated values against actual values, and evaluate the loss of information due to the removal of those wells.

The reason we need to use Optimizer in this process is that Model Builder does not extrapolate concentration values at the locations outside of the convex hull of the data points, nor does it interpolate a concentration value at a specific (i.e., removed) location. Therefore, if points located at the plume edge are removed, the area surrounding the removed point is not interpolated and appears to be "missing" in Model Builder. Thus, it is hard to compare the concentration values at removed locations to the actual values purely based on plume maps using Model Builder. This is a non-standard use of Model Builder, of course, that would not arise in routine application of the Summit Tools.

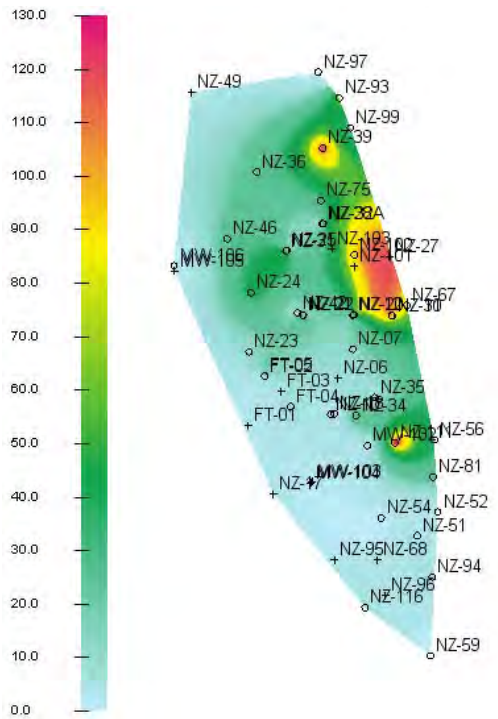
Model Validation for Spatial Analysis

Plan 97 (14 of 55 wells removed) and Plan 14 (25 of 55 wells removed) from the Spatial Analysis results (Dataset A, cutoff = 25 µg/l) were used for the model validation analysis. The figure below illustrates the plume maps generated for all sampling locations versus the locations recommended by Plan 97. The posted values for plan 97 are the interpolated values at the removed wells.

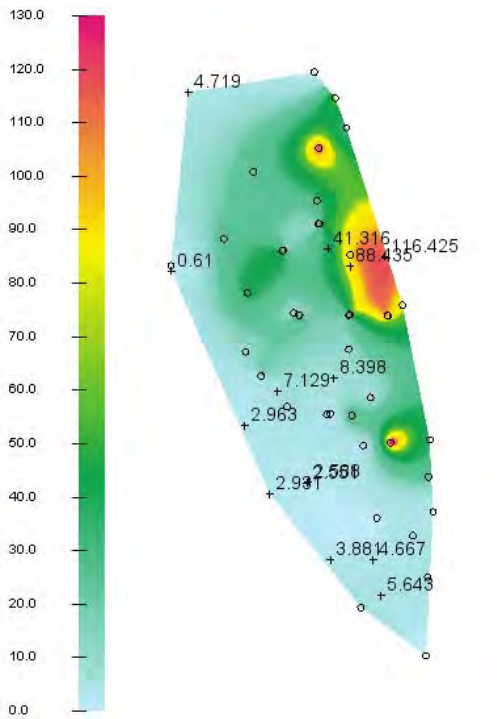
{this gap is intentional}

All Sampling Locations Versus Plan 97, Validation Data

All sampling locations



Plan 97 with interpolated values posted



Visually, the figures seem reasonably similar, validating that there is little loss of information caused the removal of the wells. The table below presents a comparison of the interpolated concentrations at the removed locations versus the actual measured concentrations at those locations in the validation data

Interpolated Concentrations versus Actual Concentrations at Removed Wells
Optimal Plan 97

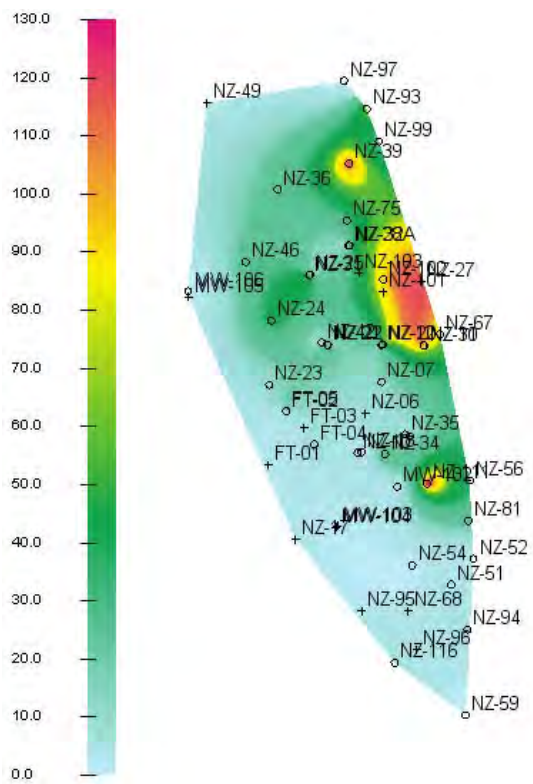
Removed Wells	Interpolated TCE Concentrations (µg/l)	Measured TCE Concentrations (µg/l)	Absolute Deviation (µg/l)
FT-01	2.963	0.39	2.573
FT-03	7.129	7.15	0.021
MW-103	2.558	0.82	1.738
MW-104	2.531	1.9	0.631
MW-105	0.610	0.6	0.01
NZ-06	8.398	8.5	0.102
NZ-101	88.435	48	40.435
NZ-103	41.316	37.5	3.816
NZ-17	2.931	0.5	2.431
NZ-27	116.425	135	18.575
NZ-49	4.719	2.45	2.269
NZ-68	4.667	6	1.333
NZ-95	3.881	1.06	2.821
NZ-96	5.643	4.1	1.543

Our analyst concluded that there were no major concerns with the interpolation with wells removed versus the actual data, but that would ultimately be up to site personnel. Furthermore, as would be expected, locations with high concentrations tend to have higher errors. Well NZ-27, for example, has a measurement of 135 µg/l, whereas its interpolated value is 116.425 µg/l. Well NZ-101 has a measurement of 48 µg/l, whereas its interpolated value is 88.435 µg/l. This finding is consistent with the concentration uncertainty analysis, which indicated that higher uncertainties are associated with areas where measured concentrations are relatively high.

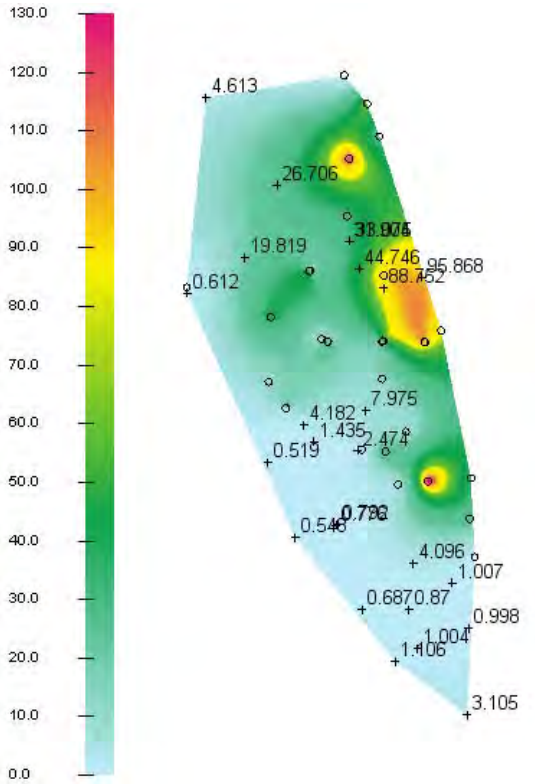
The figure below illustrates the plume maps generated for all sampling locations versus the locations recommended by Plan 14. The posted values for plan 14 are the interpolated values at the removed wells.

All Sampling Locations Versus Plan 14, Validation Data

All sampling locations



Plan 14 with interpolated values posted



Visually, the figures seem reasonably similar, though not as close a match as for Plan 97. This is because more wells are removed for Plan 14 versus Plan 97. Thus the validation results are consistent with the spatial analysis results, in that both Plan 97 and Plan 14 are reasonable representations, but Plan 14 has more error.

The table below presents a comparison of the interpolated concentrations at the removed locations versus the actual measured concentrations at those locations in the validation data.

Interpolated Concentrations Versus Actual Concentrations at Removed Wells
Optimal Plan 14

Removed Wells	Interpolated TCE Concentrations (µg/l)	Measured TCE Concentrations (µg/l)	Absolute Deviation (µg/l)
FT-01	0.519	0.39	0.129
FT-03	4.182	7.15	2.968
FT-04	1.435	8.7	7.265
MW-103	0.792	0.82	0.028
MW-104	0.776	1.9	1.124
MW-105	0.612	0.6	0.012
NZ-06	7.975	8.5	0.525
NZ-10	2.474	4.8	2.326
NZ-101	88.752	48	40.752
NZ-103	44.746	37.5	7.246
NZ-116	1.106	7	5.894
NZ-17	0.548	0.5	0.048
NZ-27	95.868	135	39.132
NZ-28A	33.975	12	21.975
NZ-32	31.004	40.5	9.496
NZ-36	26.706	25	1.706
NZ-46	19.819	25	5.181
NZ-49	4.613	2.45	2.163
NZ-51	1.007	5.55	4.543
NZ-54	4.096	0.9	3.196
NZ-59	3.105	0.5	2.605
NZ-68	0.870	6	5.13
NZ-94	0.998	5.85	4.852
NZ-95	0.687	1.06	0.373
NZ-96	1.004	4.1	3.096

The deviations for Plan 14 are generally larger than those for Plan 97 because there is more interpolation (since more wells are removed).

Overall, these results do validate the spatial optimization because there is no significant loss of data in the validation dataset, which was also the case in the base optimization results.

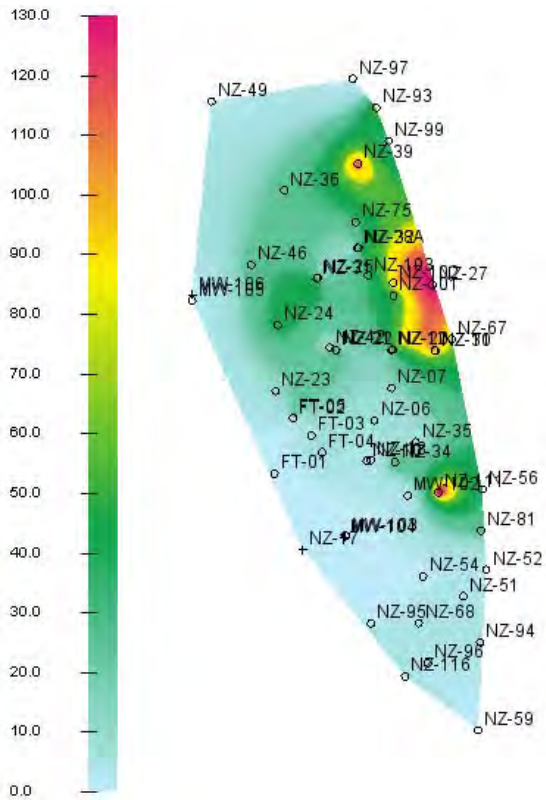
Model Validation for Spatio-Temporal Analysis

The validation process is essentially identical to that used for the spatial analysis validation, except that the Max/Min sampling frequencies are set to be "Off" at wells recommended for removal and also at wells removed recommended to be sampled less than annually. All other wells are forced to be on.

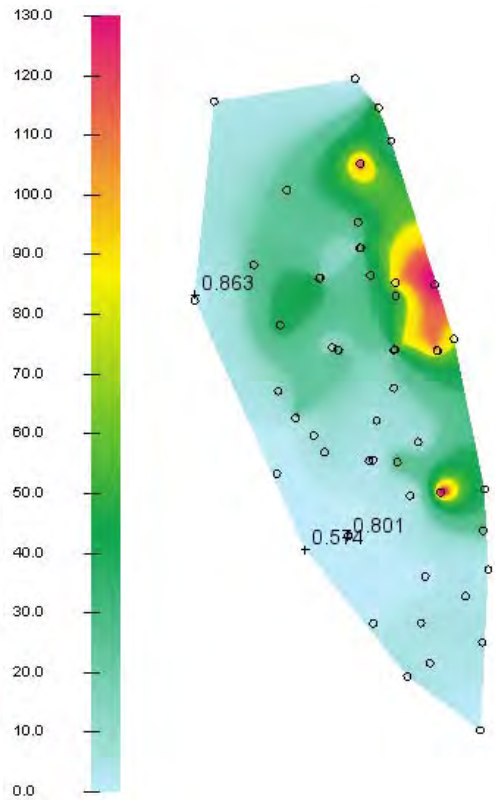
Plan 44 (lower error) and Plan 26 (higher error) from the Spatio-Temporal Analysis results (Dataset A, cutoff = 25 µg/l) were used for the model validation analysis. The figure below illustrates the plume maps generated for all sampling locations versus the locations recommended by Plan 44. The posted values for plan 44 are the interpolated values at the removed wells.

All Sampling Locations Versus Plan 44, Validation Data

All sampling locations



Plan 44 with interpolated values posted



Visually, the figures seem reasonably similar, validating that there is little loss of information caused by the removal of the wells. Note that fewer wells are removed relative to the spatial analysis results. The table below presents a comparison of the interpolated concentrations at the removed locations versus the actual measured concentrations at those locations in the validation data.

Interpolated Concentrations Versus Actual Concentrations at Removed Wells
Optimal Plan 44

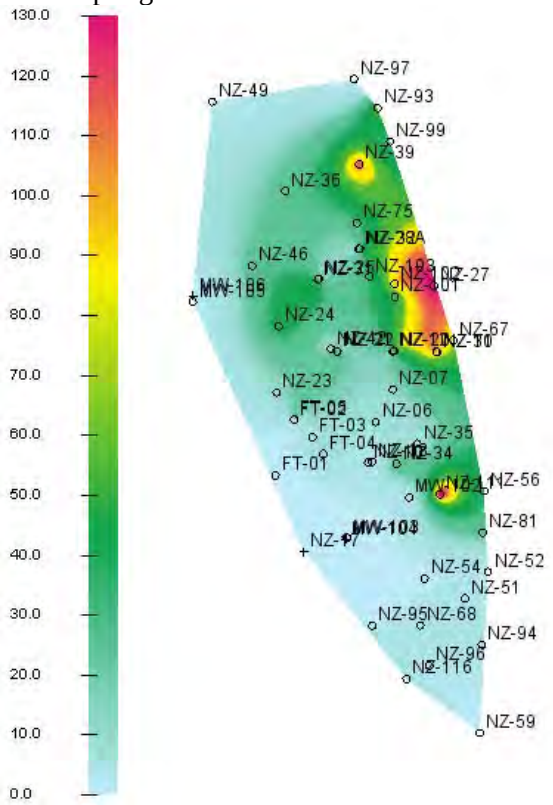
Removed/Non-sampled Wells	Interpolated TCE Concentrations (µg/l)	Measured TCE Concentrations (µg/l)	Absolute Deviation (µg/l)
MW-104	0.801	1.9	1.099
MW-106	0.863	0.5	0.363
NZ-17	0.574	0.5	0.074

There is very little information lost by removing these locations.

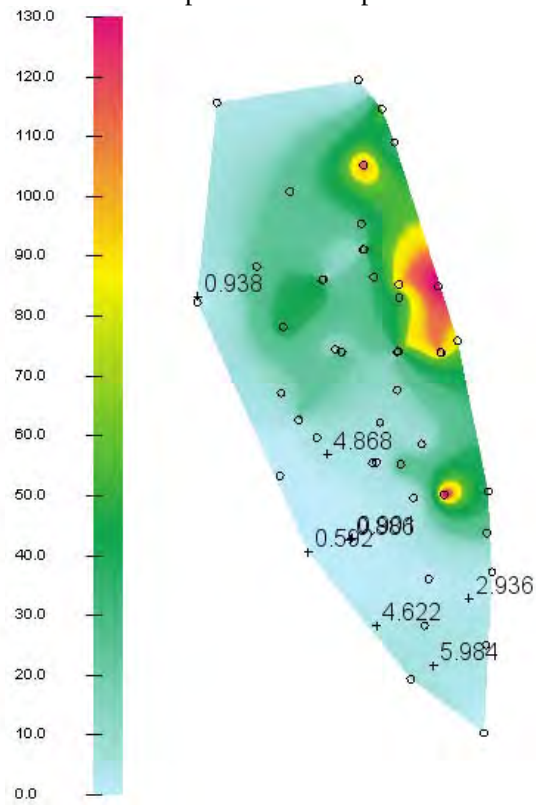
The figure below illustrates the plume maps generated for all sampling locations versus the locations recommended by Plan 26. The posted values for Plan 26 are the interpolated values at the removed wells.

All Sampling Locations Versus Plan 26, Validation Data

All sampling locations



Plan 26 with interpolated values posted



Again, the results indicate little loss of information due the reduced sampling. The table below presents a comparison of the interpolated concentrations at the removed locations versus the actual measured concentrations at those locations in the validation data.

Interpolated Concentrations Versus Actual Concentrations at Removed Wells
Optimal Plan 26

Removed/Non-sampled Wells	Interpolated TCE Concentrations (µg/l)	Measured TCE Concentrations (µg/l)	Absolute Deviation (µg/l)
FT-04	4.868	8.7	3.832
MW-103	0.906	0.82	0.086
MW-104	0.881	1.9	1.019
MW-106	0.938	0.5	0.438
NZ-17	0.502	0.5	0.002
NZ-51	2.936	5.55	2.614
NZ-95	4.662	1.06	3.602
NZ-96	5.984	4.1	1.884

The errors due to the reduced sampling are very minor. Compared to errors associated with spatial optimizations, smaller deviations are found with spatio-temporal results because fewer wells are recommended to be turned off and more information is maintained. Also note that the

way this validation was performed is conservative because all wells recommended for less-than-annual sampling were assumed to not be sampled for purposes of validation.

LTMO Follow-up: George Air Force Base Site

As described in the Technical Demonstration Plan, one site (GAFB) was elected for follow-up to track outcomes of the recommendations through additional statistical validations similar to those performed during the previous phase of the project. This follow-up consisted of three components:

- Additional validation of spatial and spatio-temporal optimization results
- Data tracker analysis
- Mass Flux and Mass Metric calculations

These follow-up evaluations were performed using sampling data collected from the Spring 2007 and Fall 2007 sampling events, which were collected under the previous sampling plan and subsequent to the original LTMO analysis for the site. Each component of the follow-up analysis is described below.

Additional Validation of Spatial and Spatio-Temporal Optimization Results

The initial LTMO analysis included validation of spatial and spatio-temporal optimization results using reserved data from two sampling events (Spring 2006 and Fall 2006). For this follow-up, additional validation was performed using sampling data from the Spring 2007 and Fall 2007 sampling events.

The dataset prepared by EnviroStat for this follow-up validation indicates that only four of the 55 monitoring wells (NZ-10, NZ-51, NZ-68, NZ-95) were sampled both in April and in October of 2007. A few of the wells in the initial optimization dataset were not sampled in 2007 but were sampled in 2006 (MW-104, MW-105, NZ-17, NZ-18, NZ-33, NZ-34, NZ-54, NZ-59, NZ-75, and NZ-97). To maintain a consistent computational domain, 2006 values were assigned those wells where no 2007 data were collected. Additionally, NZ-33 and NZ-43 were not sampled in recent years but were sampled in 2007. To maintain a consistent computational domain, they were therefore removed from the 2007 dataset for the purpose of model validation.

For model validation purposes (for both the spatial and spatio-temporal analyses), we combined the Spring 2007 and Fall 2007 sampling results into one event. Because no abrupt changes in concentrations were observed for wells sampled in both Spring 2007 and Fall 2007, their concentrations were averaged. The sampling date for the follow-up dataset was also normalized to 10/01/2007, and includes 55 sampling locations with one value at each location. TCE is the only COC evaluated, consistent with the original spatial and spatio-temporal optimization evaluation.

The following general procedures were applied for performing model validation:

- Create a follow-up sampling dataset as described above

- For spatial optimization validation, for wells that were recommended to be shut off by the Optimizer during the original LTMO evaluation, set both their Max/Min sampling frequency to be "off". For the remaining wells, set both their Max/Min sampling frequency to be "on". By doing so, no sampling optimization will be performed by the software since the sampling frequency for each individual well has been predetermined. For spatio-temporal optimization validation, the process is identical except that the Max/Min sampling frequencies are set to be "Off" at wells recommended for removal and also at wells whose recommended sampling frequency is less than two years. All the remaining wells are forced to be on in the software.

- Run Optimizer with the above settings for only one generation, and display the interpolated concentrations at removed well locations.

- Compare the interpolated values against actual values, and evaluate the loss of information due to the removal of those wells.

The reason we need to use Optimizer in this process is that Model Builder does not extrapolate concentration values at the locations outside of the convex hull of the data points, nor does it interpolate a concentration value at a specific (i.e., removed) location. Therefore, if points located at the plume edge are removed, the area surrounding the removed point is not interpolated and appears to be "missing" in Model Builder. Thus, it is hard to compare the concentration values at removed locations to the actual values purely based on plume maps using Model Builder. This is a non-standard use of Model Builder, of course, that would not arise in routine application of the Summit Tools.

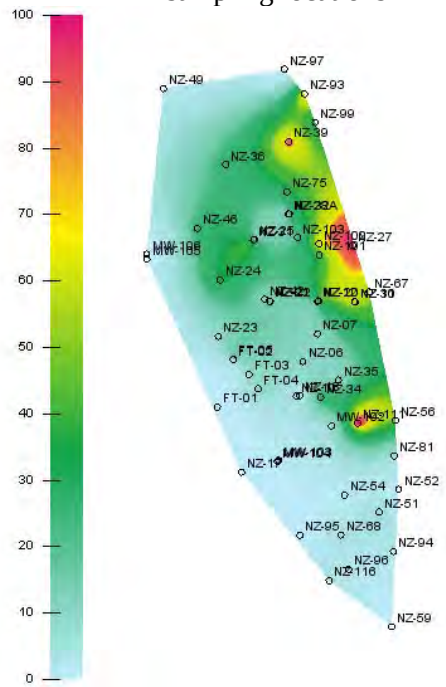
Follow-Up Validation for Spatial Analysis

Plan 97 (14 of 55 wells removed) and Plan 14 (25 of 55 wells removed) from the original Spatial Analysis results (Dataset A, cutoff = 25 µg/l) were used for the model validation analysis. The figure below illustrates the plume maps generated for all sampling locations versus the interpolated plume for Plan 97 (14 wells removed). The posted values for plan 97 are the interpolated values at the removed wells.

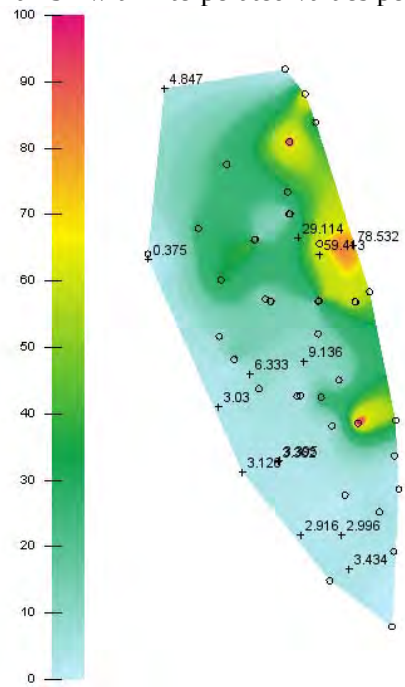
{this gap is intentional}

All Sampling Locations Versus Plan 97, Follow-up Dataset

All sampling locations



Plan 97 with interpolated values posted



Visually, the figures seem reasonably similar, validating that there is little loss of information caused the removal of the 14 wells. The table below presents a comparison of the interpolated concentrations at the removed locations versus the actual concentrations in the validation data

Interpolated Concentrations versus Actual Concentrations at Removed Wells
Optimal Plan 97, Follow-Up Dataset

Removed Wells	Interpolated TCE Concentrations (µg/l)	Measured TCE Concentrations (µg/l)	Absolute Deviation (µg/l)
FT-01	3.03	0.42	2.61
FT-03	6.333	8.6	2.27
MW-103	3.395	1.6	1.80
MW-104	3.352	1.9	1.45
MW-105	0.375	0.6	0.23
NZ-06	9.136	11	1.86
NZ-101	59.413	40	19.41
NZ-103	29.114	23	6.11
NZ-17	3.126	0.5	2.63
NZ-27	78.532	99	20.47
NZ-49	4.847	2.3	2.55
NZ-68	2.996	4.8	1.80
NZ-95	2.916	0.65	2.27
NZ-96	3.434	2.8	0.63

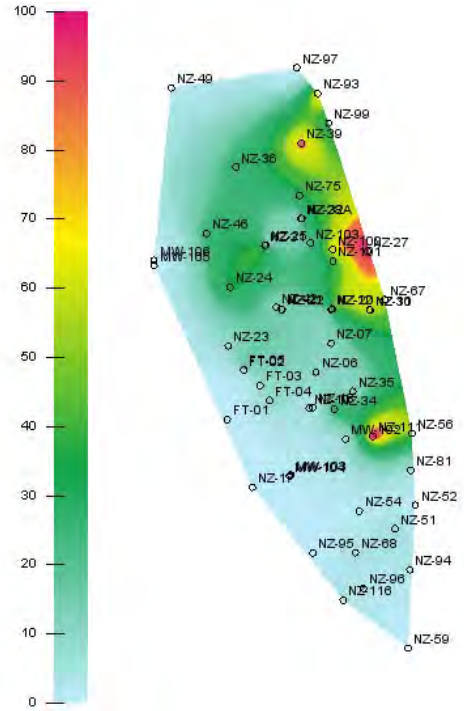
Our analyst concluded that there were no major concerns with the interpolation with wells removed versus the actual data, but that would ultimately be up to site personnel. Furthermore, as would be expected, locations with high concentrations tend to have higher errors. Well NZ-27,

for example, has a measurement of 99 µg/l, whereas its interpolated value is 78.532 µg/l. Well NZ-101 has a measurement of 40 µg/l, whereas its interpolated value is 59.413 µg/l.

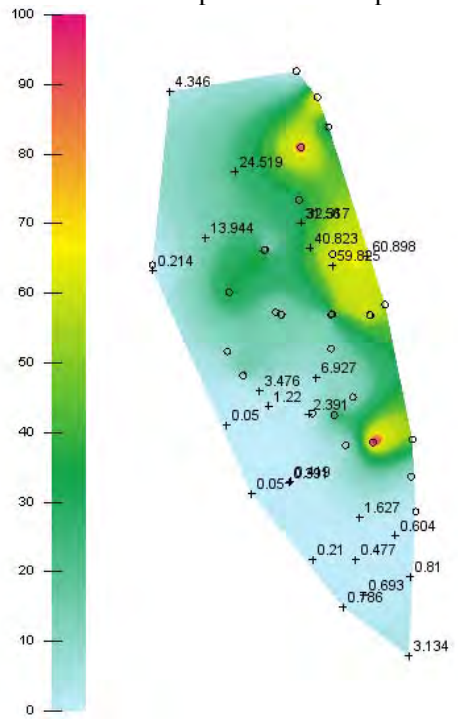
The figure below illustrates the plume maps generated for all sampling locations versus the interpolated plume for Plan 14 (25 wells removed). The posted values for plan 14 are the interpolated values at the removed wells.

All Sampling Locations Versus Plan 14, Follow-Up Dataset

All sampling locations



Plan 14 with interpolated values posted



Visually, the figures seem reasonably similar, though not as close a match as for Plan 97 in that the concentrations in high concentration area were interpolated lower for Plan 14. This is because more wells are removed for Plan 14 versus Plan 97. Thus the validation results are consistent with the spatial analysis results, in that both Plan 97 and Plan 14 are reasonable representations, but Plan 14 has more error.

The table below presents a comparison of the interpolated concentrations at the removed locations versus the actual measured concentrations at those locations in the validation data.

{this gap is intentional}

Interpolated Concentrations Versus Actual Concentrations at Removed Wells
Optimal Plan 14, Follow-Up Dataset

Removed Wells	Interpolated TCE Concentrations (µg/l)	Measured TCE Concentrations (µg/l)	Absolute Deviation (µg/l)
FT-01	0.05	0.42	0.37
FT-03	3.476	8.6	5.124
FT-04	1.22	9.4	8.18
MW-103	0.419	1.6	1.181
MW-104	0.391	1.9	1.509
MW-105	0.214	0.6	0.386
NZ-06	6.927	11	4.073
NZ-10	2.391	5.25	2.859
NZ-101	59.825	40	19.825
NZ-103	40.823	23	17.823
NZ-116	0.786	4.3	3.514
NZ-17	0.05	0.5	0.45
NZ-27	60.898	99	38.102
NZ-28A	31.317	13	18.317
NZ-32	32.56	26	6.56
NZ-36	24.519	26	1.481
NZ-46	13.944	26	12.056
NZ-49	4.346	2.3	2.046
NZ-51	0.604	5.5	4.896
NZ-54	1.627	0.9	0.727
NZ-59	3.134	0.5	2.634
NZ-68	0.477	4.8	4.323
NZ-94	0.81	2.9	2.09
NZ-95	0.21	0.65	0.44
NZ-96	0.693	2.8	2.107

The deviations for Plan 14 are generally larger than those for Plan 97 because there is more interpolation (since more wells are removed).

Overall, these results do validate the spatial optimization because there is no significant loss of data in the validation dataset, which was also the case in the base optimization results. It would ultimately be up to site personnel to determine if the loss of accuracy in either Plan 97 or 14 is acceptable.

{this gap is intentional}

Follow-Up Validation for Spatio-Temporal Analysis

The table below lists the recommended sampling frequency for these plans and the maximum error for each plan.

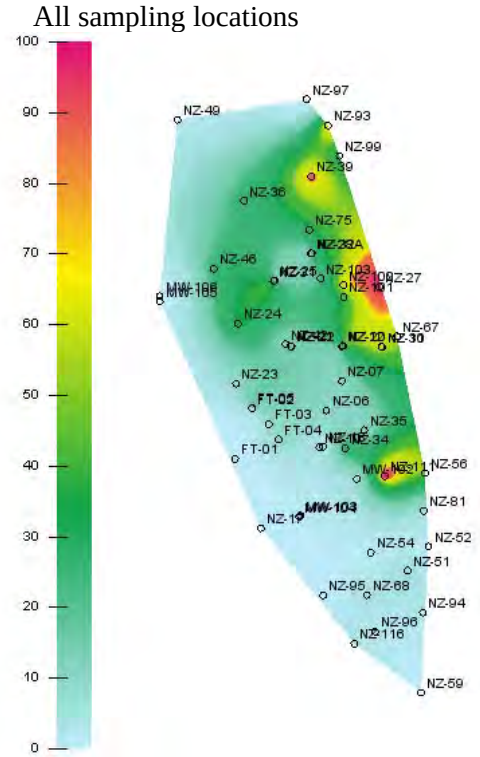
Original Spatio-Temporal Analysis Recommendations (Dataset A, Cutoff = 25 µg/l)

Plan #	Recommended Sampling Frequency						Max TCE Error
	Semi-Annually	Annually	Every 2 Years	Every 3 Years	Every 4 Years	Off	
44	All the remaining wells					MW-104 MW-106 NZ-17	0.4506
26	All the remaining wells		NZ-51	NZ-96	MW-103	FT-04 MW-104 MW-106 NZ-17 NZ-95	1.388

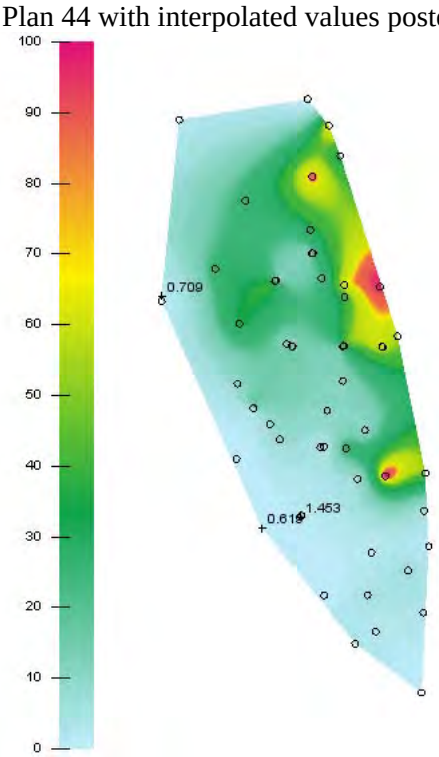
Plan 44 (lower error) and Plan 26 (higher error) from the original Spatio-Temporal Analysis results (Dataset A, cutoff = 25 µg/l) were used for the follow-up analysis. The figure below illustrates the plume maps generated for all sampling locations versus the interpolated plumes for Plan 44. The posted values for plan 44 are the interpolated values at the removed wells.

All Sampling Locations Versus Plan 44, Follow-Up Dataset

All sampling locations



Plan 44 with interpolated values posted



Visually, the figures look very similar, validating that there is little loss of information caused by the removal of the wells. Note that fewer wells are removed relative to the spatial analysis results. The table below presents a comparison of the interpolated concentrations at the removed locations versus the actual measured concentrations at those locations in the validation data.

Interpolated Concentrations Versus Actual Concentrations at Removed Wells
Optimal Plan 44, Follow-Up Dataset

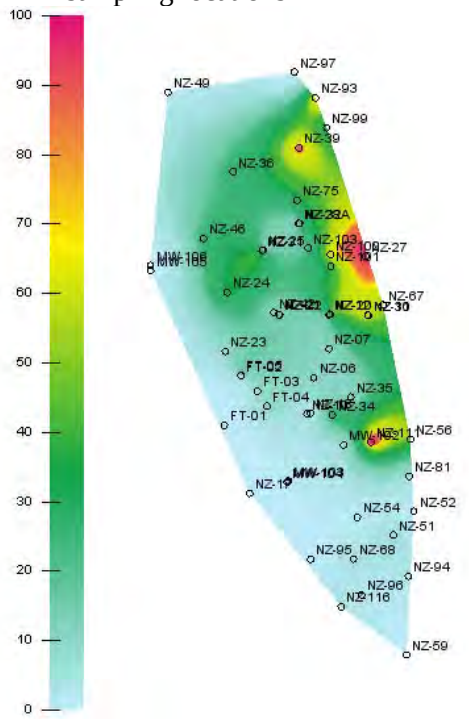
Removed/Non-sampled Wells	Interpolated TCE Concentrations (µg/l)	Measured TCE Concentrations (µg/l)	Absolute Deviation (µg/l)
MW-104	1.453	1.9	0.447
MW-106	0.709	0.05	0.659
NZ-17	0.619	0.5	0.119

There is very little information lost by removing these locations.

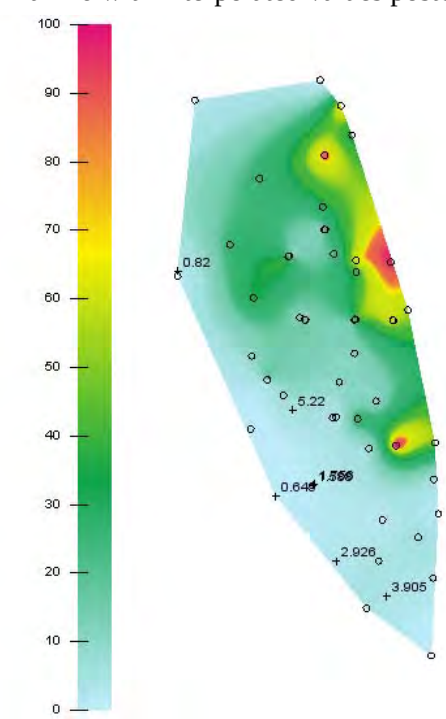
The figure below illustrates the plume maps generated for all sampling locations versus the interpolated plume for Plan 26. The posted values for Plan 26 are the interpolated values at the removed wells.

All Sampling Locations Versus Plan 26, Follow-Up Dataset

All sampling locations



Plan 26 with interpolated values posted



Again, the results indicate little loss of information due the reduced sampling. The table below presents a comparison of the interpolated concentrations at the removed locations versus the actual concentrations at those locations in the validation data.

Interpolated Concentrations Versus Actual Concentrations at Removed Wells
Optimal Plan 26, Follow-Up Dataset

Removed/Non-sampled Wells	Interpolated TCE Concentrations (µg/l)	Measured TCE Concentrations (µg/l)	Absolute Deviation (µg/l)
FT-04	5.22	9.4	4.18
MW-103	1.756	1.6	0.156
MW-104	1.588	1.9	0.312
MW-106	0.82	0.05	0.77
NZ-17	0.649	0.5	0.149
NZ-95	2.926	0.65	2.276
NZ-96	3.905	2.8	1.105

Again, the errors due to the reduced sampling are very minor. Compared to errors associated with spatial optimizations, smaller deviations are found with spatio-temporal results because fewer wells are recommended to be turned off and more information is maintained. Also note that the way this validation was performed is conservative because all wells recommended for less-than-biennial sampling were assumed to not be sampled for purposes of validation.

Data Tracker Results

Data Tracker (DT) indicates if sampling results in a "current" dataset are "in-bounds" or "out-of-bounds" based on statistical characterization of values in a "background dataset". In the original analysis, the current dataset was Spring 2006 and Fall 2006, and the background data were from 1994 to 2005. To perform DT on the follow-up dataset (2007 data) a decision needs to be made by the user whether or not to integrate the 2006 data into the background dataset. We tried both approaches. Furthermore, as part of the original DT evaluation, EnviroStat indicated that an alternative treatment of ND values might have led to improved results. Specifically, EnviroStat recommended that one "graphing value" be assigned for ND and J values (either an arbitrary low value or half of the most common reporting limit (RL)), versus the approach in the original DT evaluation that utilized the RL for each sample for the NDs. This can be problematic if the RL is elevated for a specific sample due to sample dilution in the laboratory, for example. Again, we utilized both approaches. Therefore, six DT scenarios were considered:

1. 2006 data not included in the background dataset

    A. replacing NDs with the RL (as in original DT analysis)
    B. replacing NDs with graphing value of 0.05 µg/l
    C. replacing NDs with ½ of the most common RL of each CoC

2. 2006 data included in the background dataset

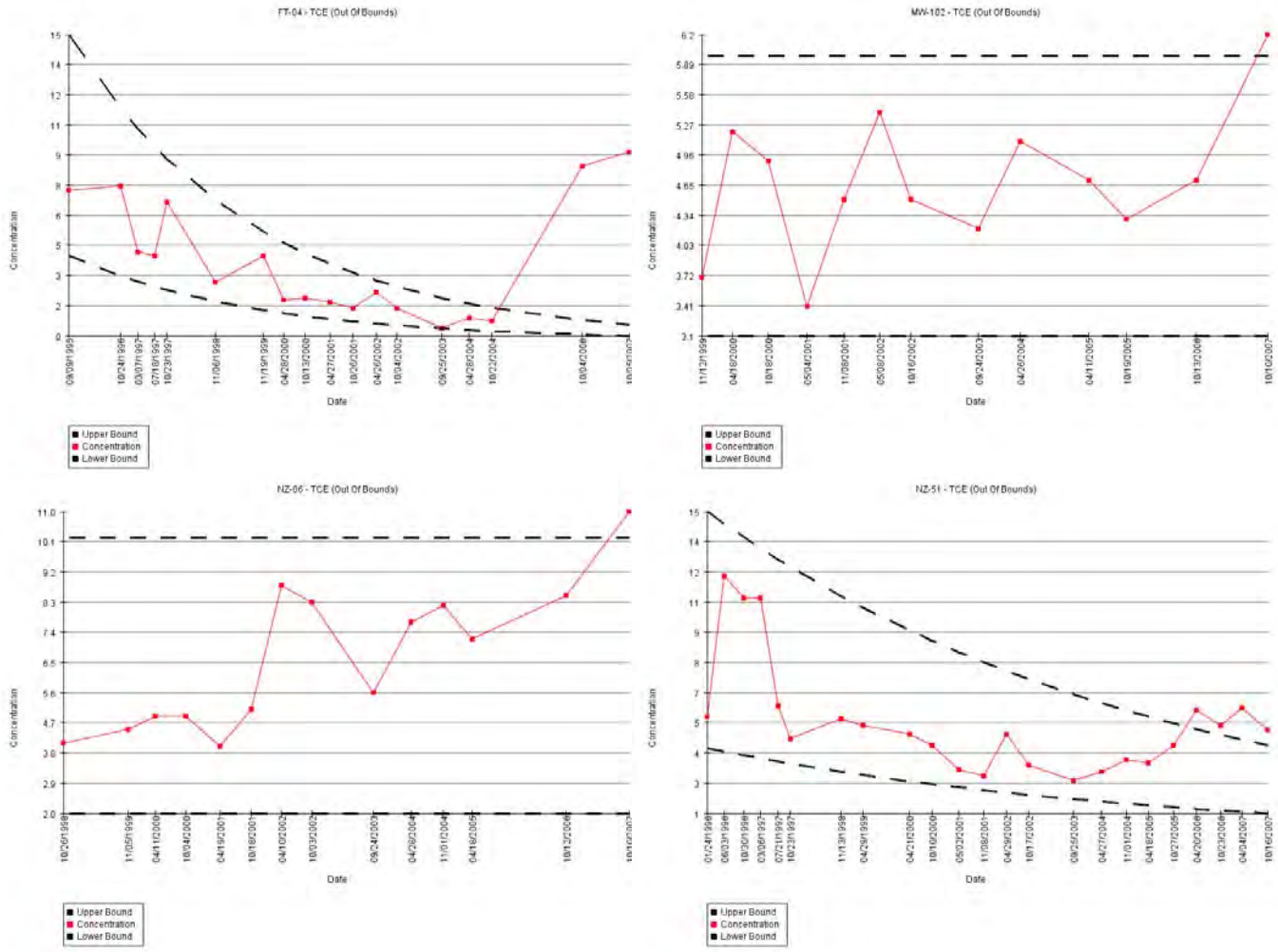
    A. replacing NDs with the RL (as in original DT analysis)
    B. replacing NDs with graphing value of 0.05 µg/l
    C. replacing NDs with ½ of the most common RL of each CoC

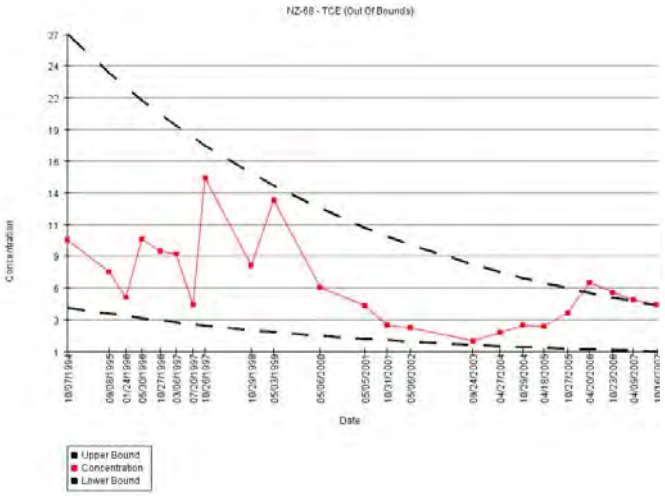
For Scenario C, the most common RL for each CoC is as follows: TCE – 0.5 µg/l, c12DCE – 1.2 µg/l l, PCE – 1.4 µg/l, Benzene – 0.4 µg/l, Toluene – 1.1 µg/l.

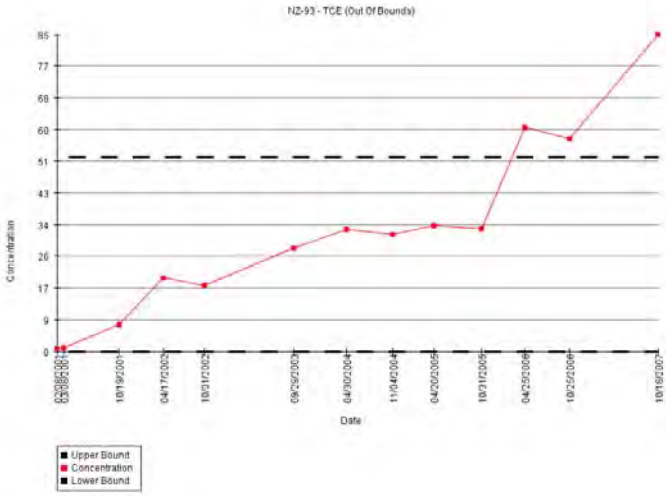
The figures presented below illustrate the concentration histories for wells where the values in the 2007 sampling data are out-of-bounds, for each of the DT scenarios.

*Scenario 1A (2006 data not included in background, ND values set to RL of each sample)***:**
*note: 2006 data, while not in background for calculating bounds, are included on the plots

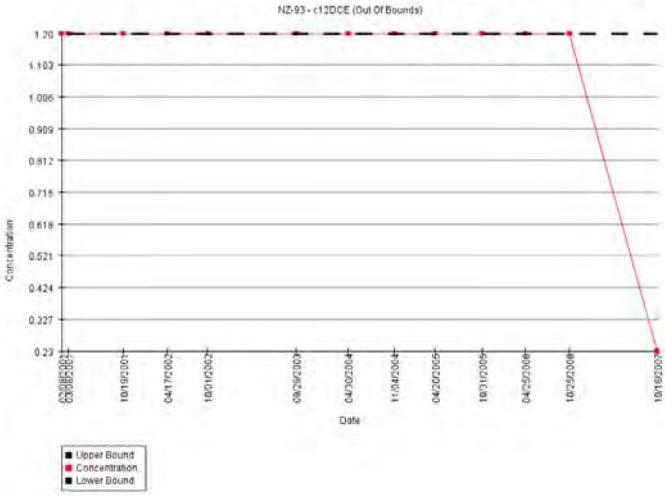
TCE: FT-04, MW-102, NZ-06, NZ-51, NZ-68, NZ-93



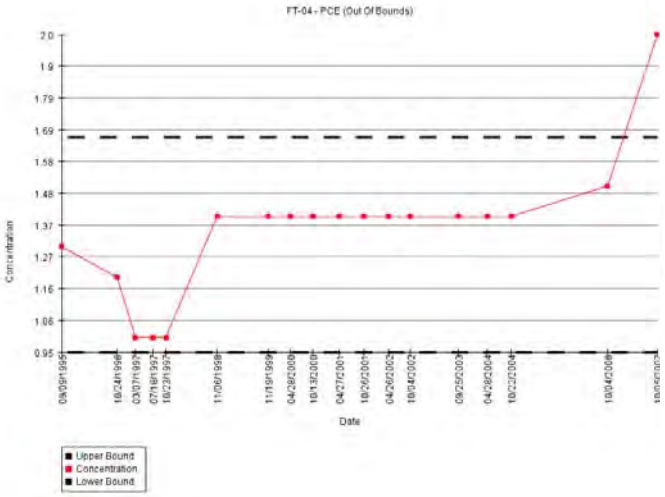


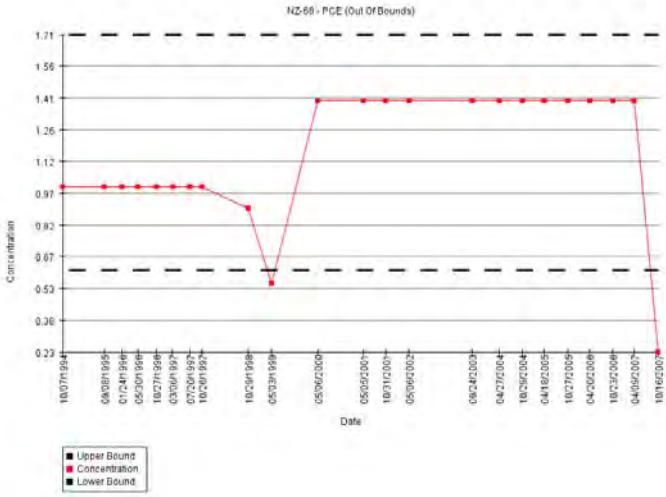


c12DCE: NZ-93

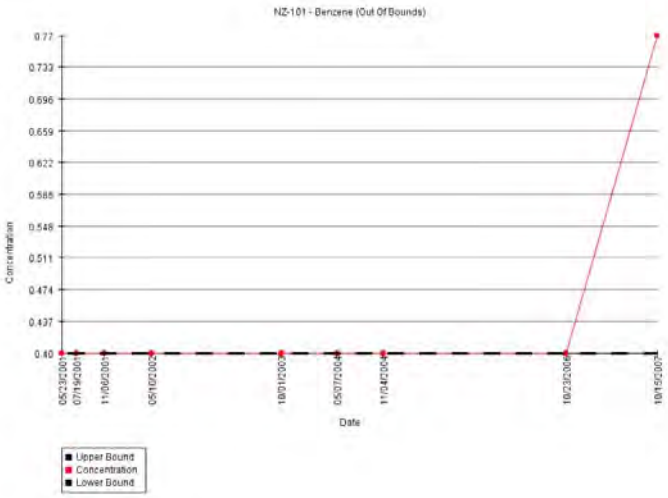


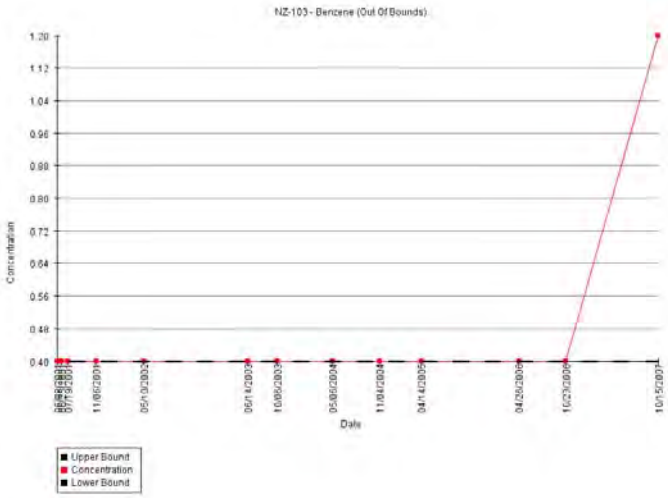
PCE: FT-04, NZ-68

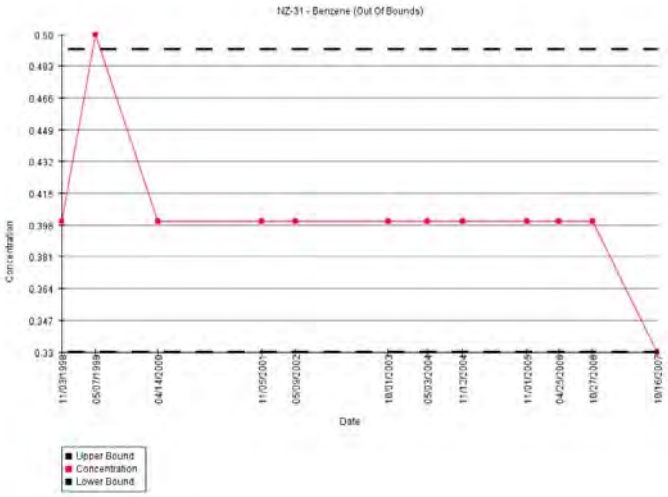




Benzene: NZ-101, NZ-103, NZ-31



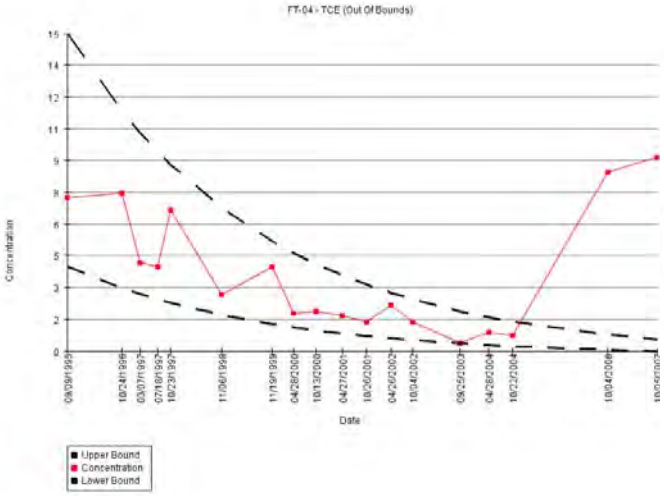


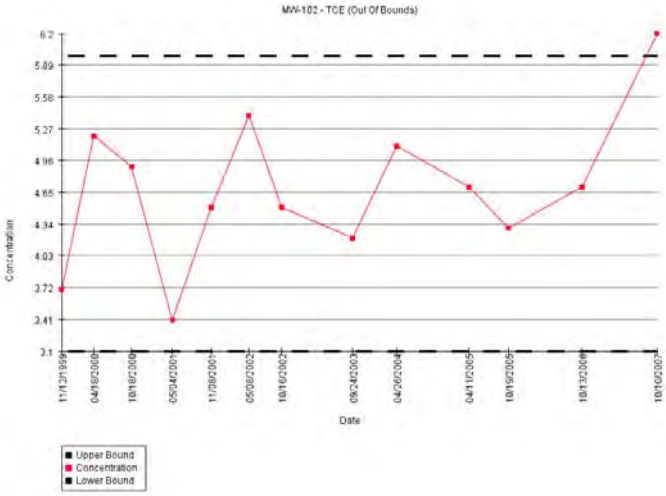


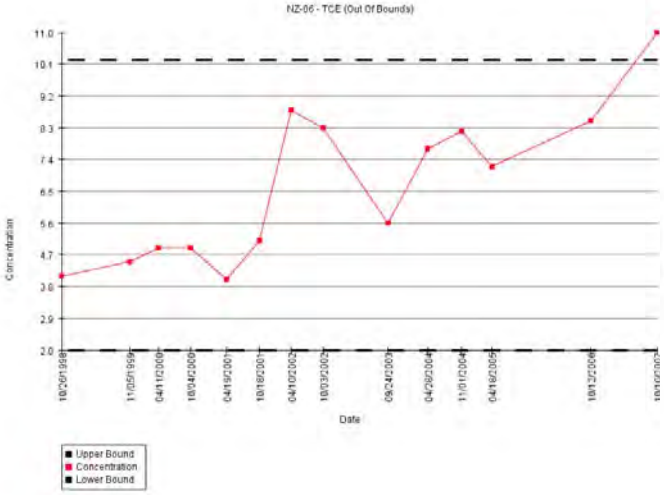
{this gap is intentional}

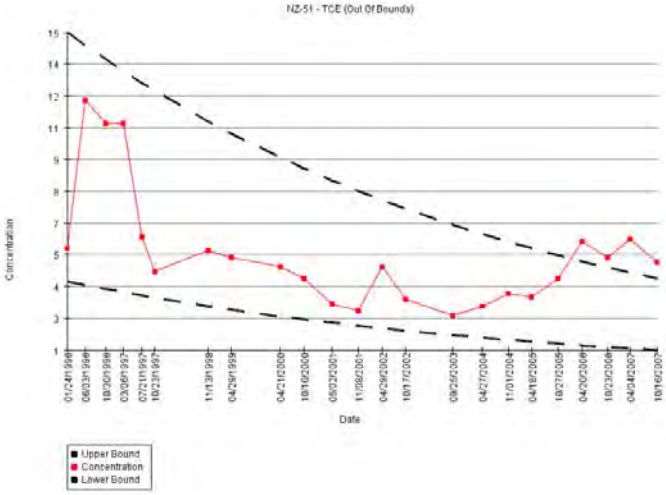
*Scenario 1B (2006 data not included in background, ND values set to 0.05 µg/l)***:**
*note: 2006 data, while not in background for calculating bounds, are included on the plots

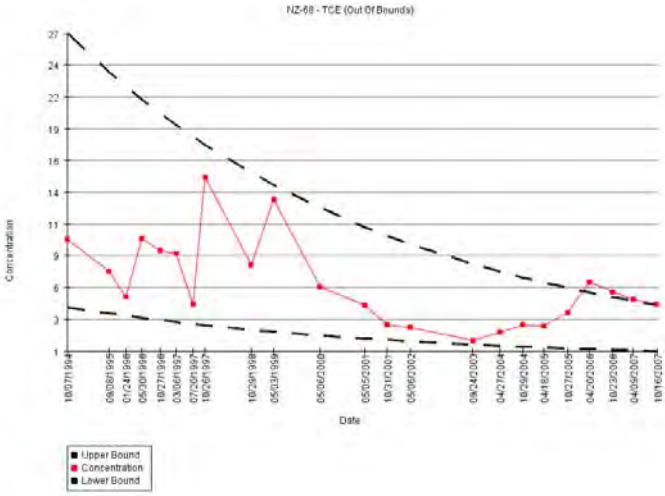
TCE: FT-04, MW-102, NZ-06, NZ-51, NZ-68, NZ-93

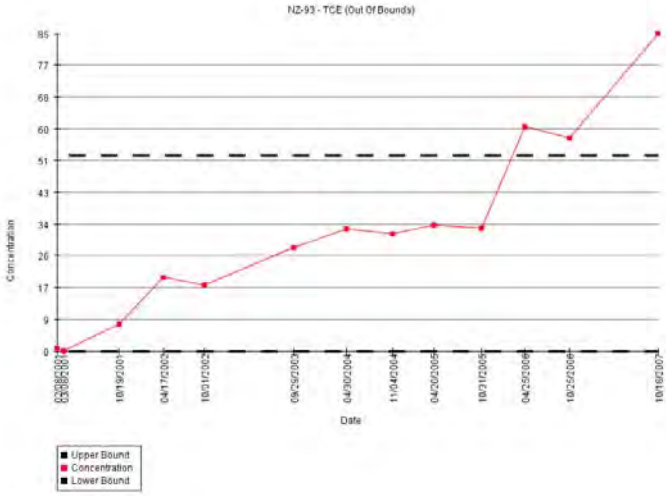




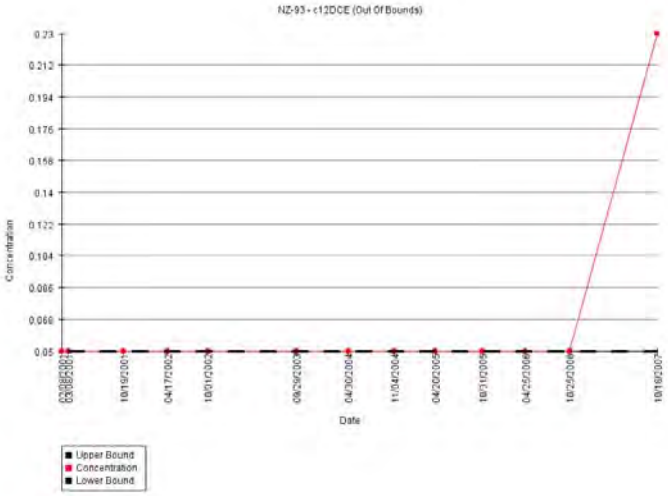




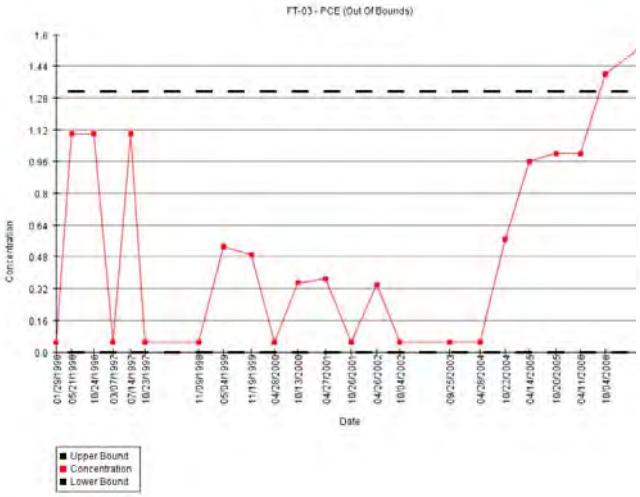


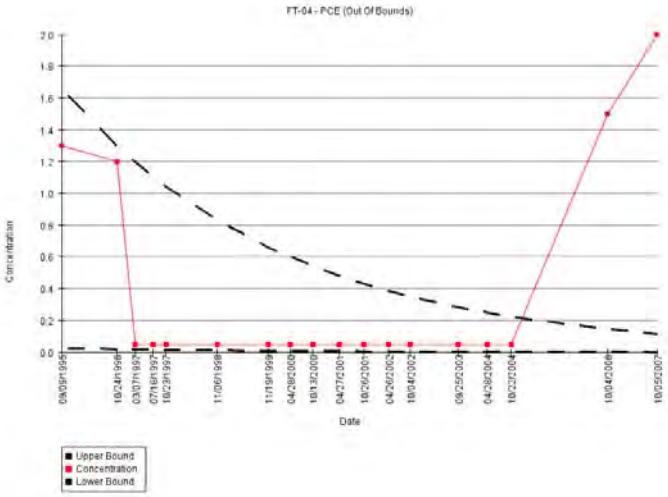


c12DCE: NZ-93

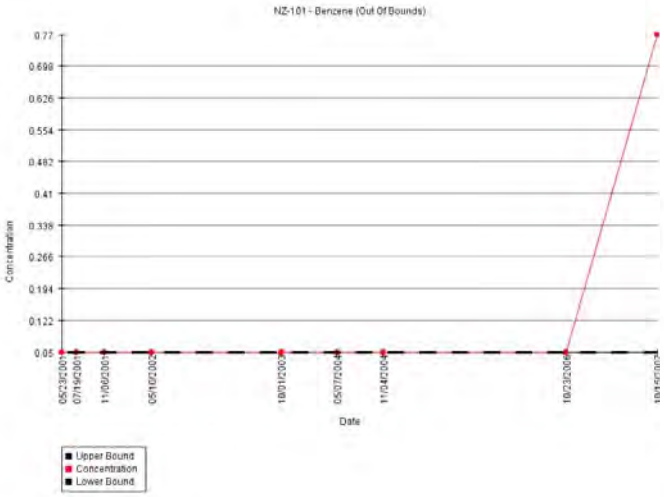


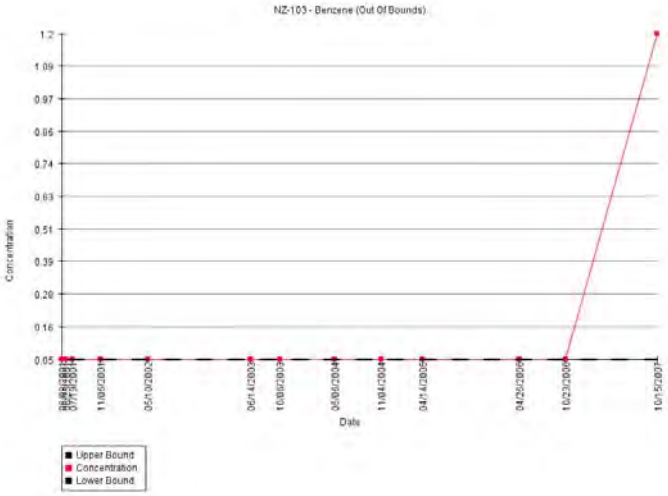
PCE: FT-03, FT-04

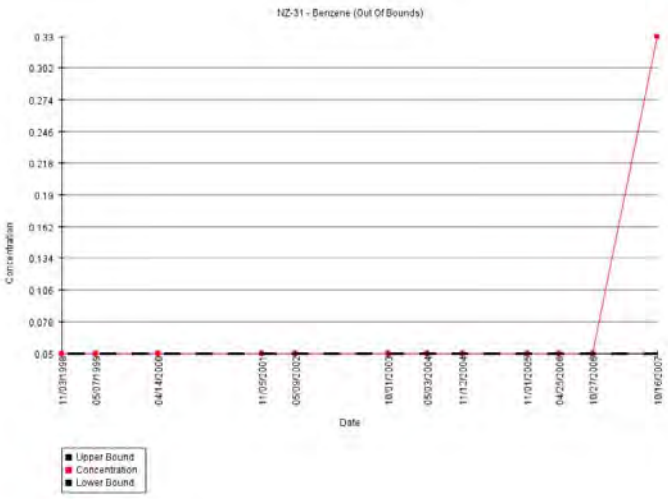




Benzene: NZ-101, NZ-103, NZ-31



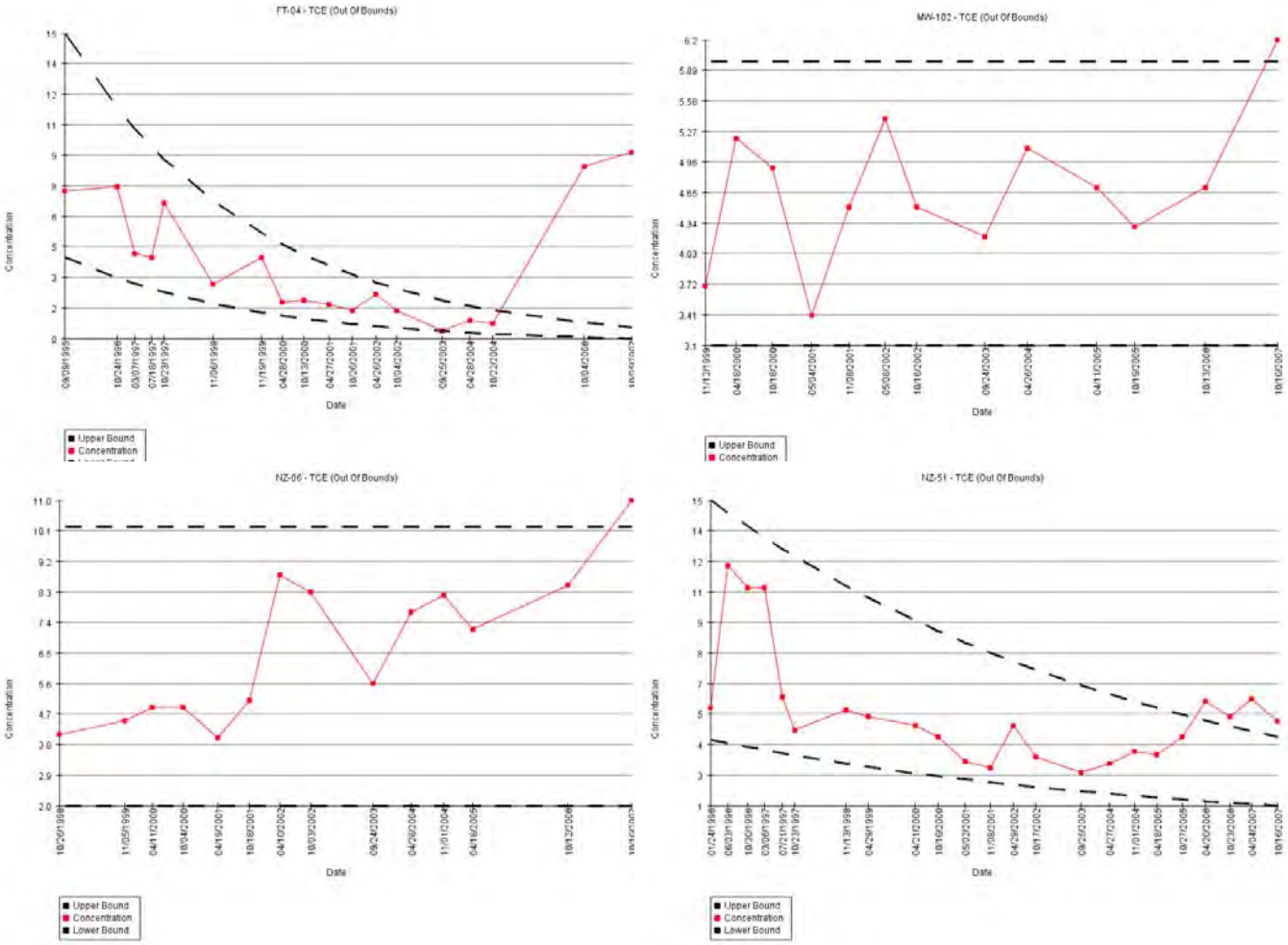


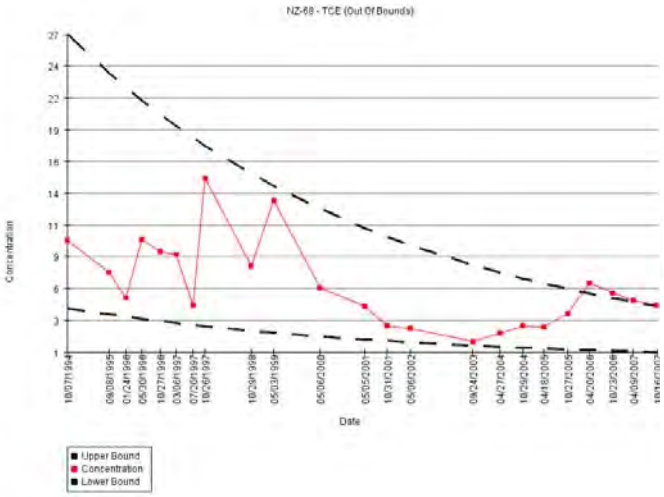


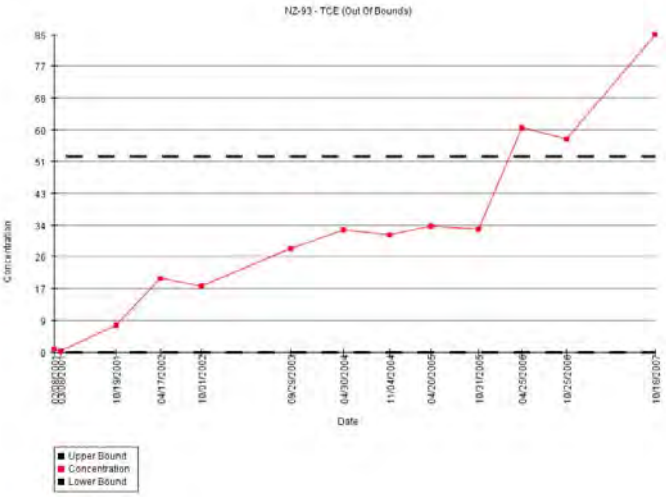
*Scenario 1C (2006 data not included in background, ND values set to ½ of the most common RL of each CoC)***:**

*note: 2006 data, while not in background for calculating bounds, are included on the plots

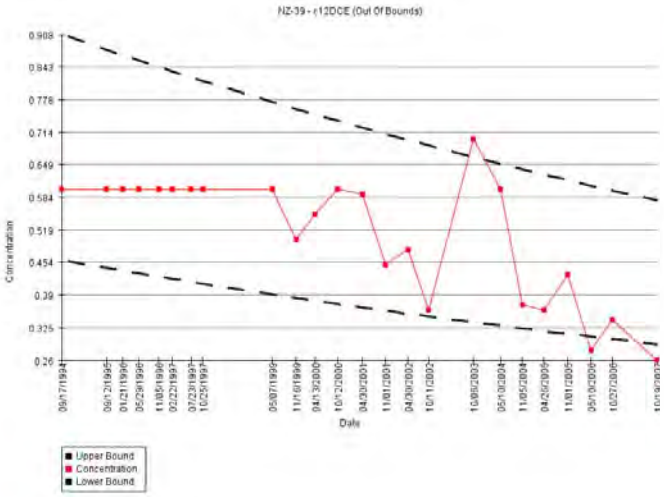
TCE: FT-04, MW-102, NZ-06, NZ-51, NZ-68, NZ-93

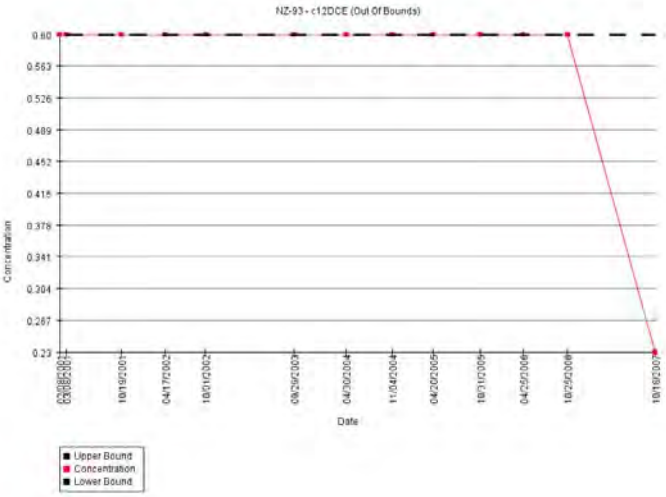




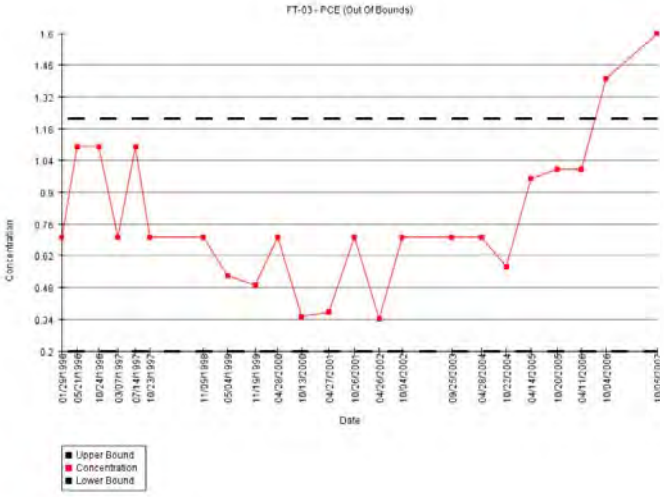


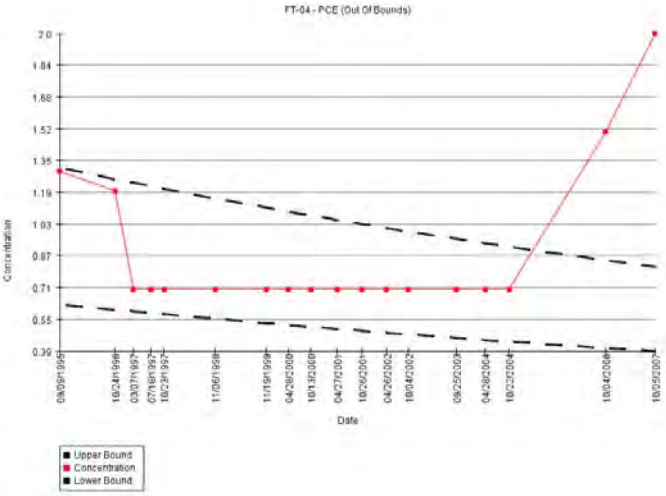
c12DCE: NZ-39, NZ-93

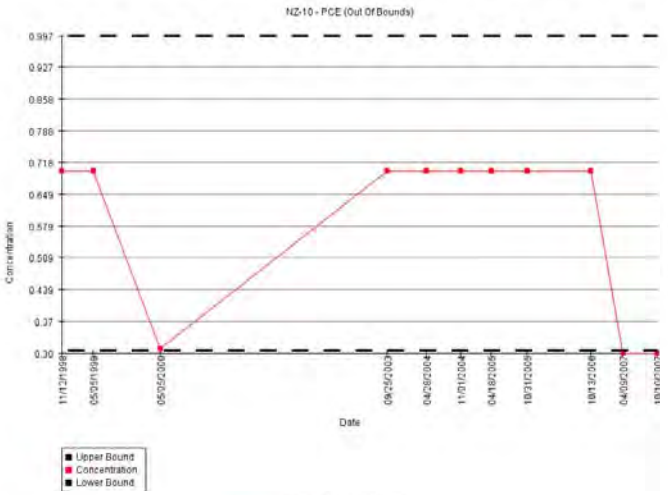


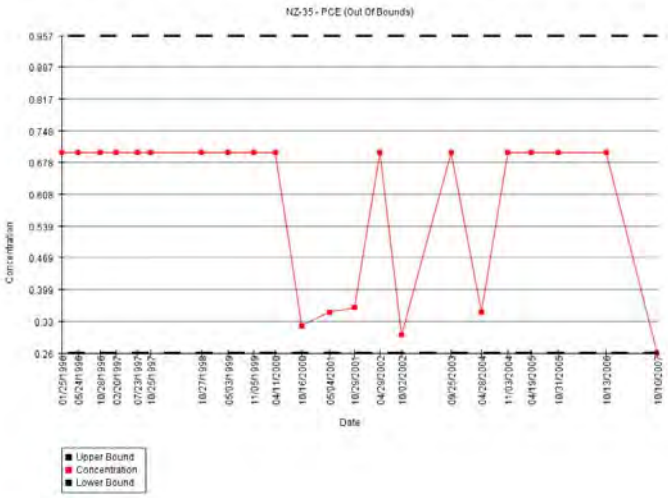


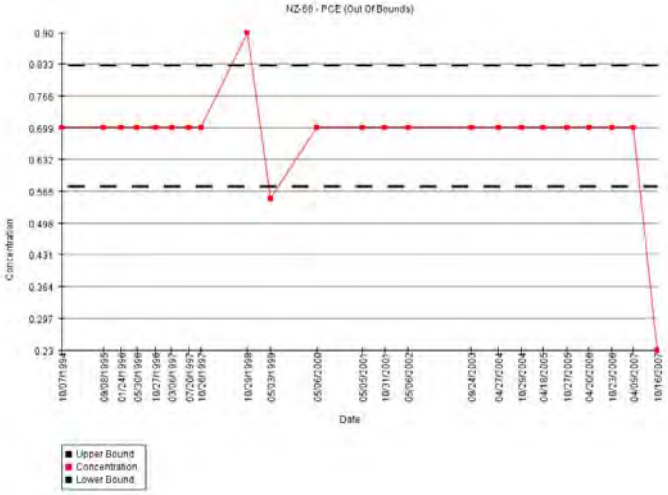
PCE: FT-03, FT-04, NZ-10, NZ-35, NZ-68



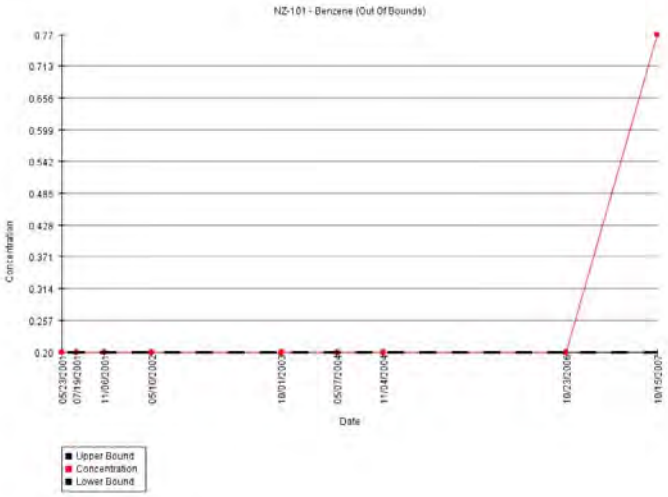


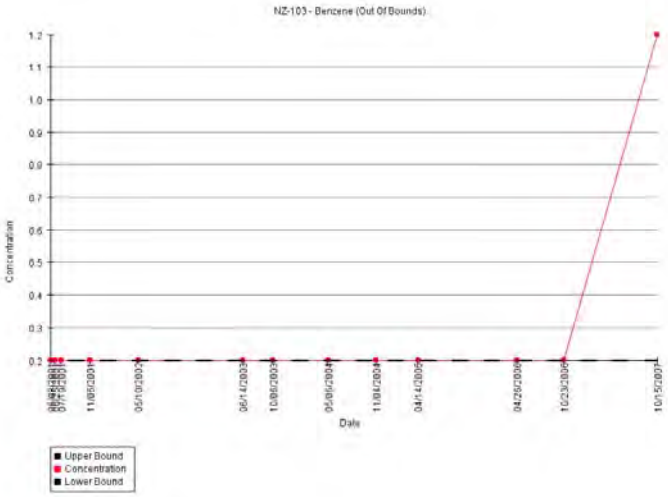


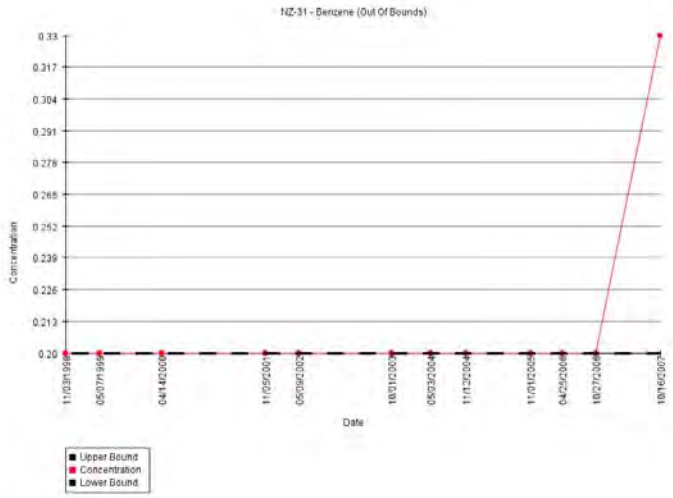




Benzene: NZ-101, NZ-103, NZ-31

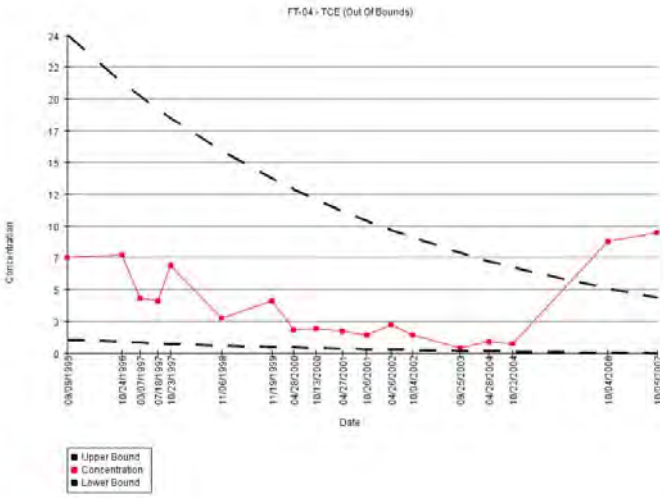


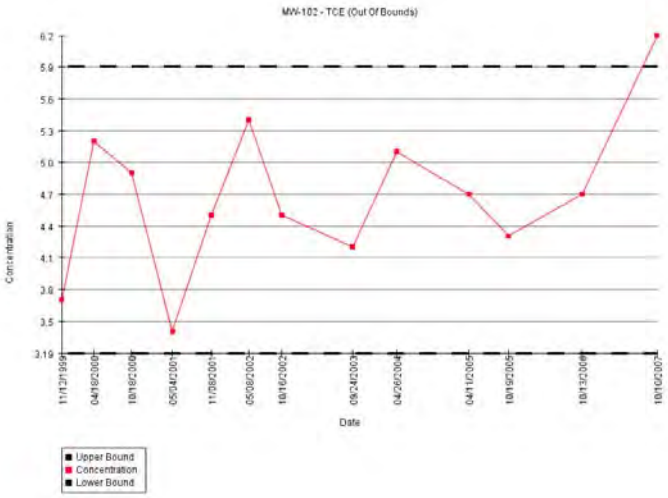


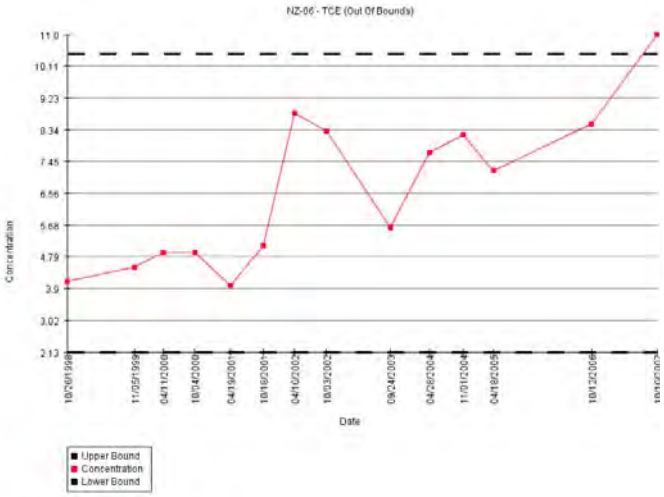


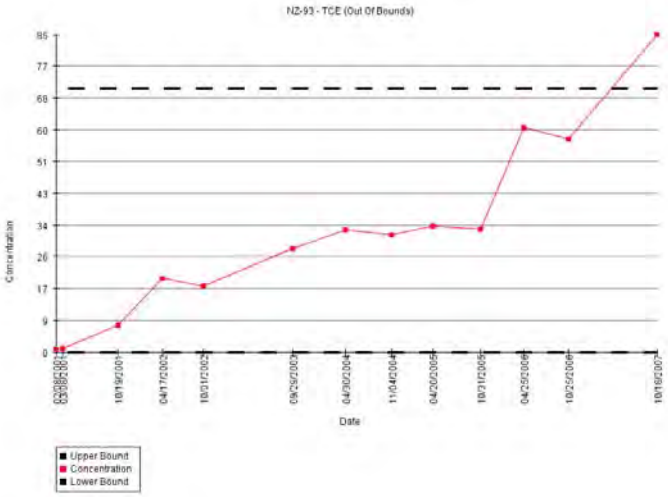
Scenario 2A (2006 data included in background, ND values set to RL of each sample):

TCE: FT-04, MW-102, NZ-06, NZ-93

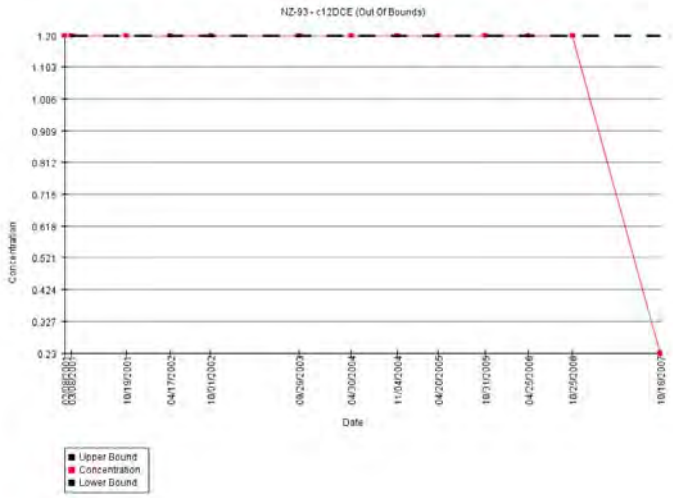




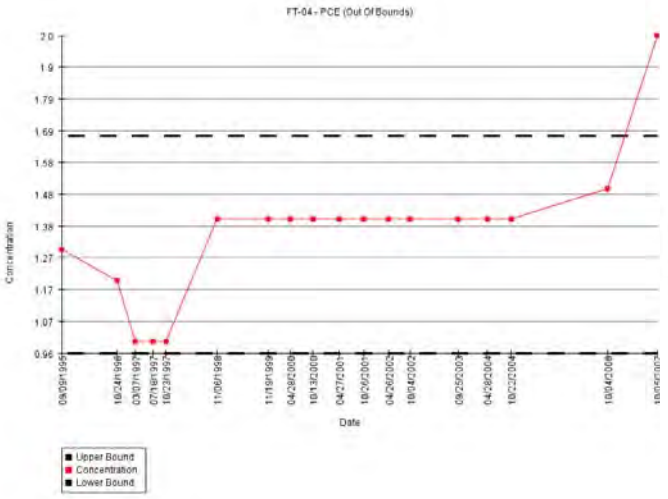


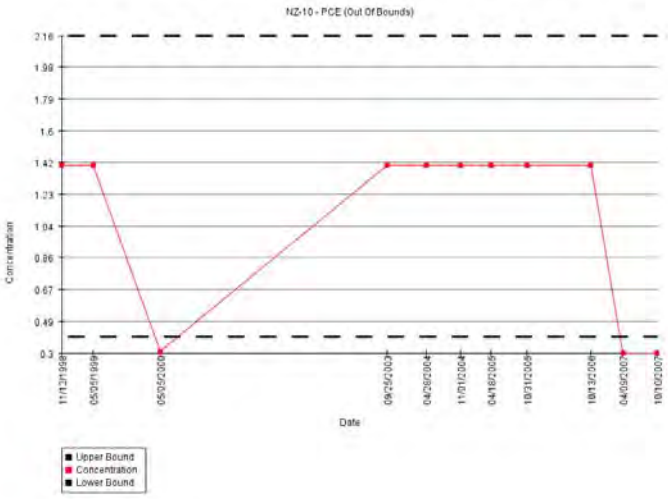


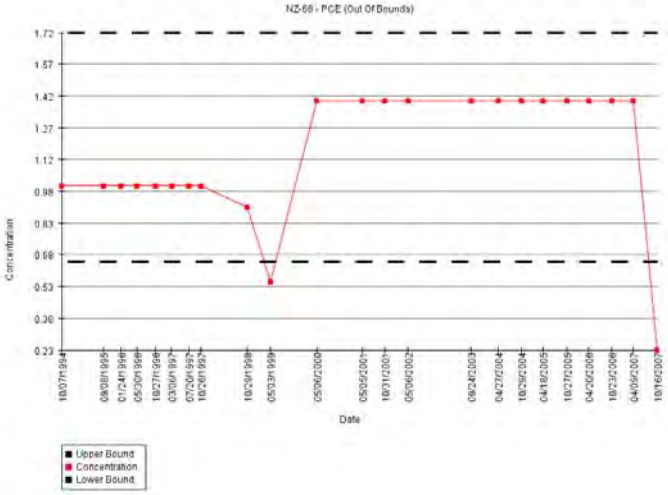
c12DCE: NZ-93



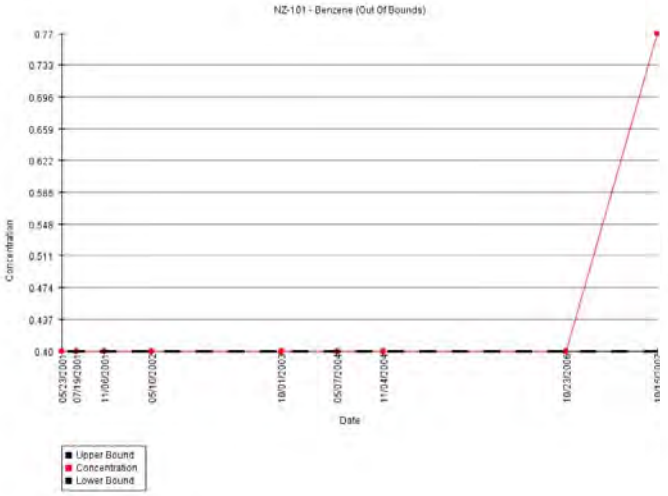
PCE: FT-04, NZ-10, NZ-68

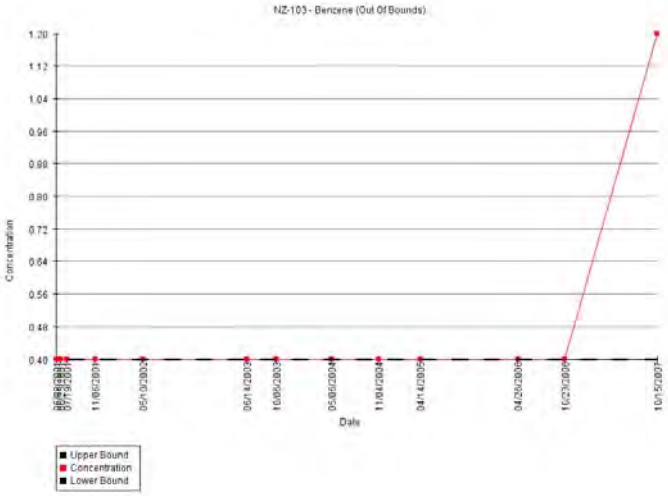


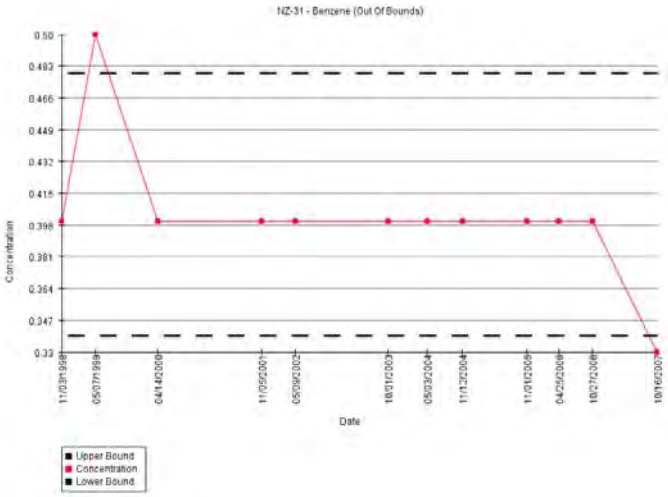




Benzene: NZ-101, NZ-103, NZ-31

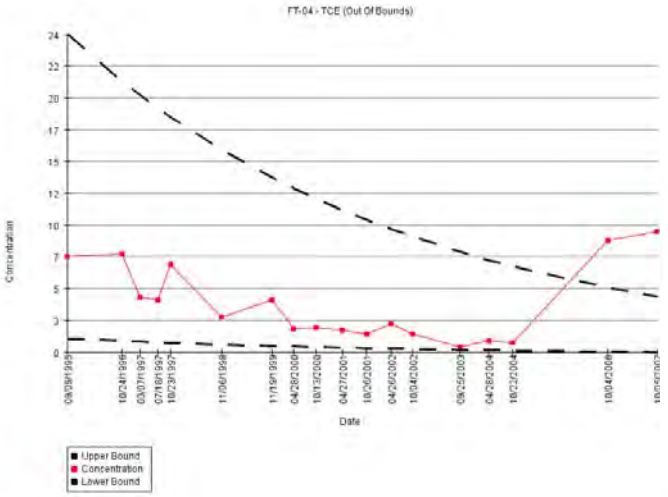


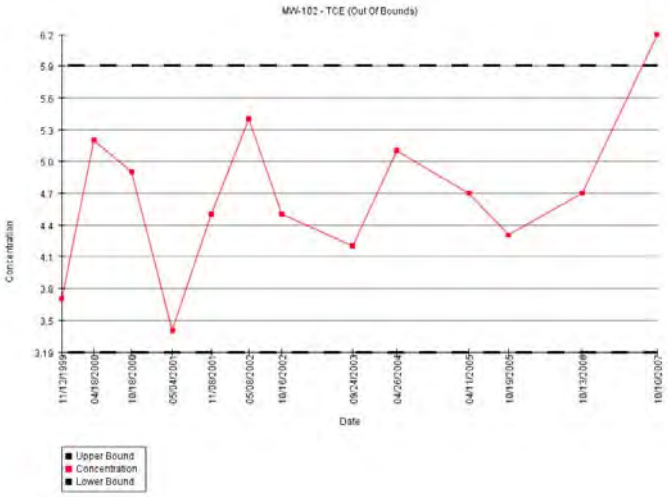


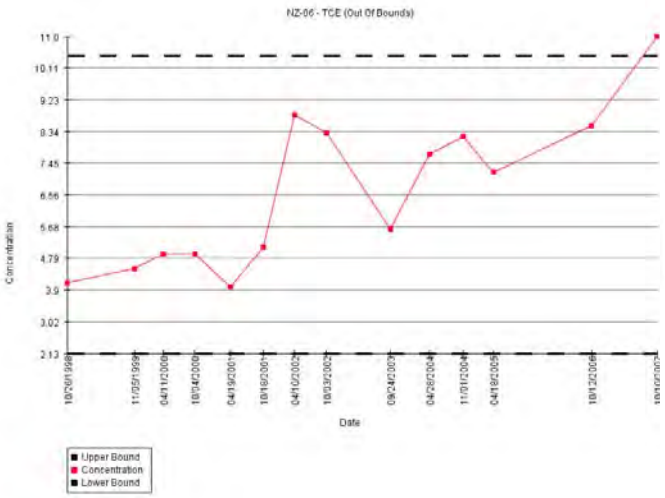


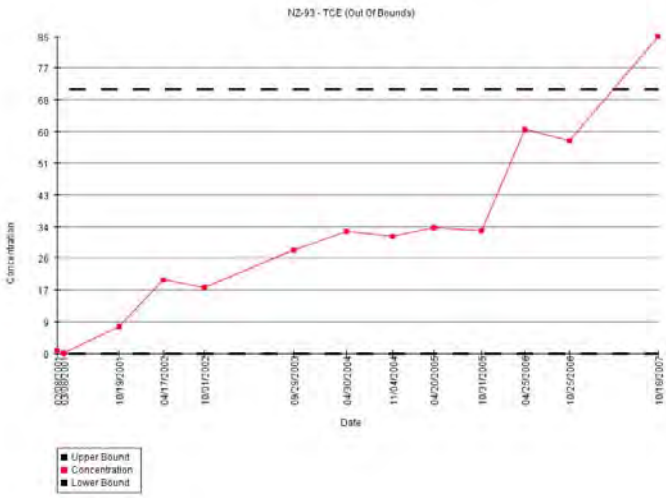
Scenario 2B (2006 data included in background, ND values set to 0.05 µg/l):

TCE: FT-04, MW-102, NZ-06, NZ-93

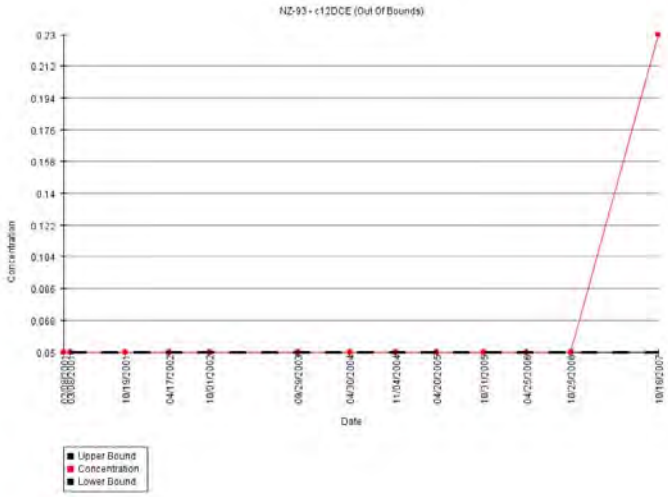




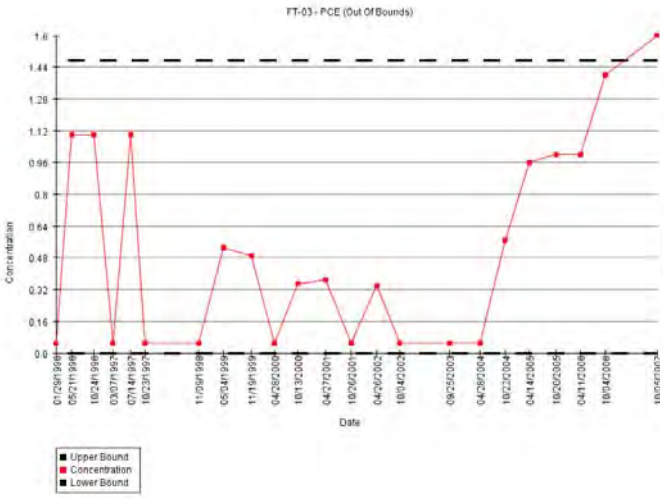


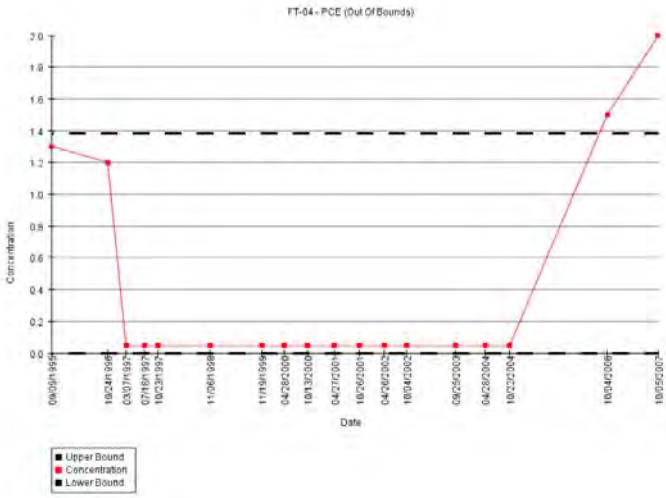


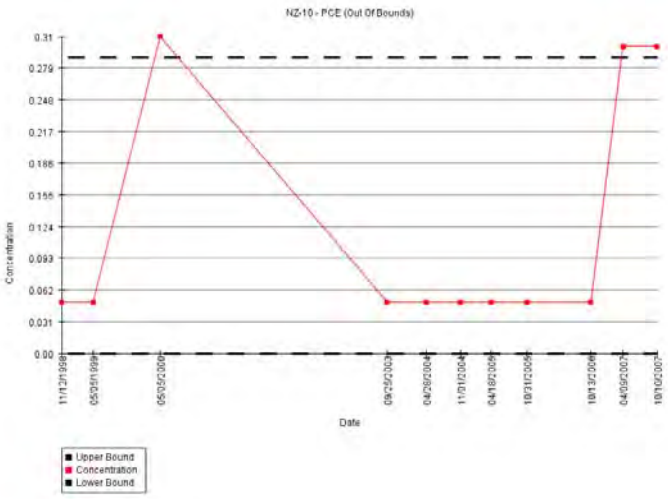
c12DCE: NZ-93



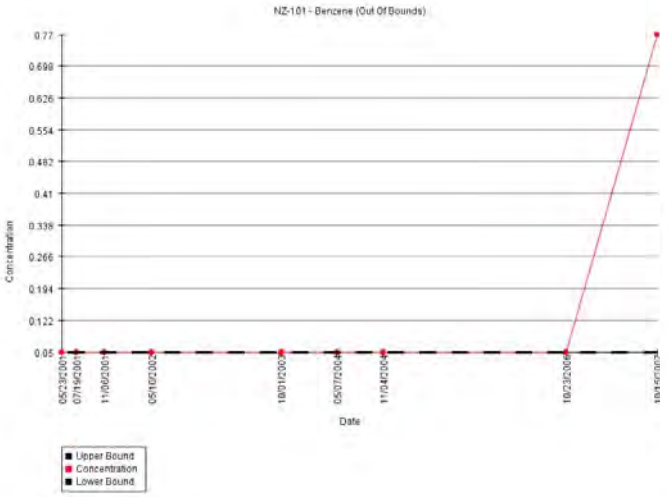
PCE: FT-03, FT-04, NZ-10

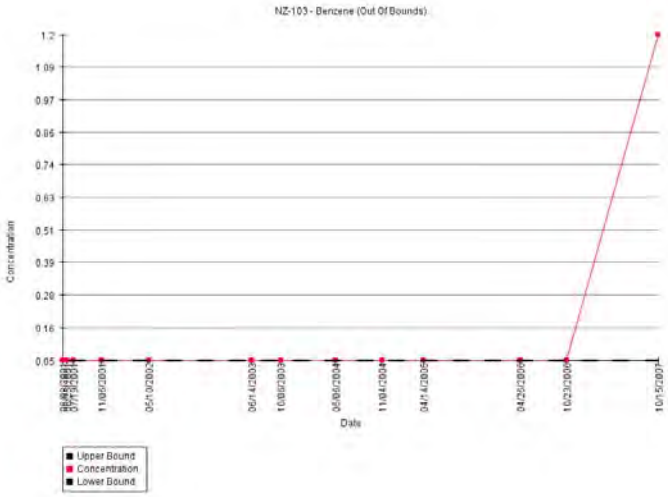


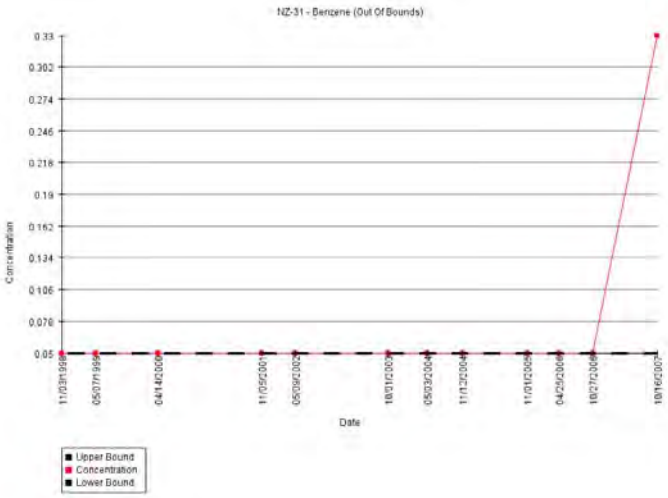




Benzene: NZ-101, NZ-103, NZ-31

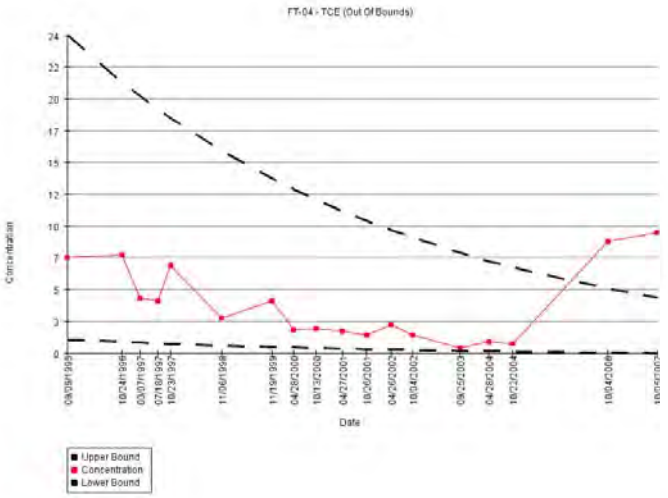


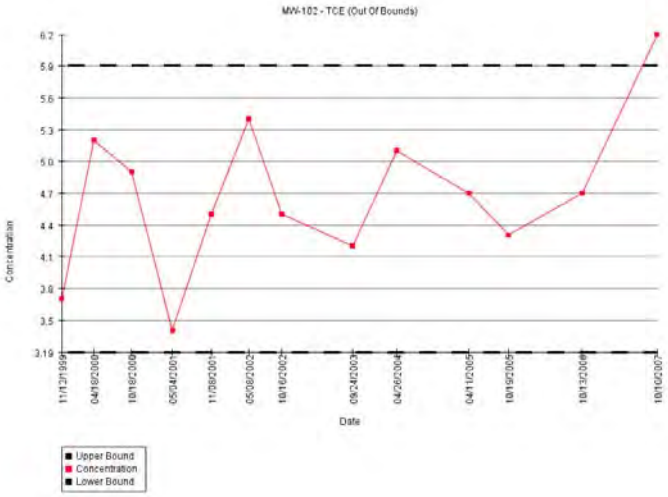


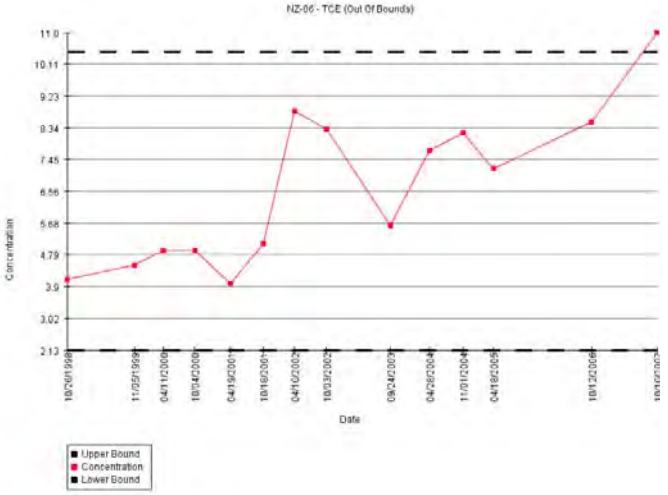


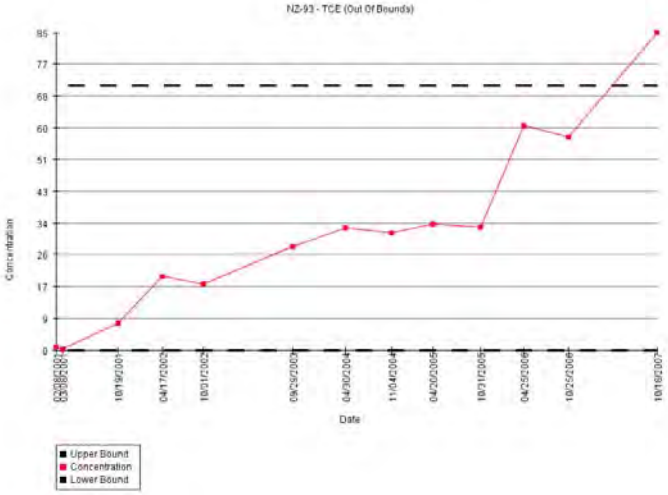
Scenario 2C (2006 data included in background, ND values set to ½ of the most common RL of each CoC):

TCE: FT-04, MW-102, NZ-06, NZ-93

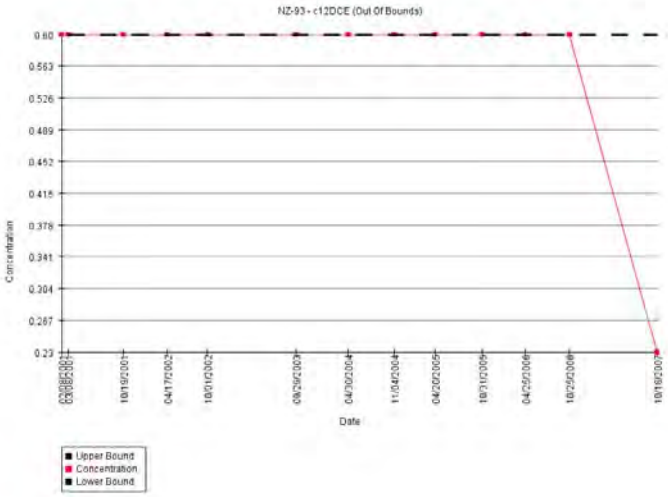




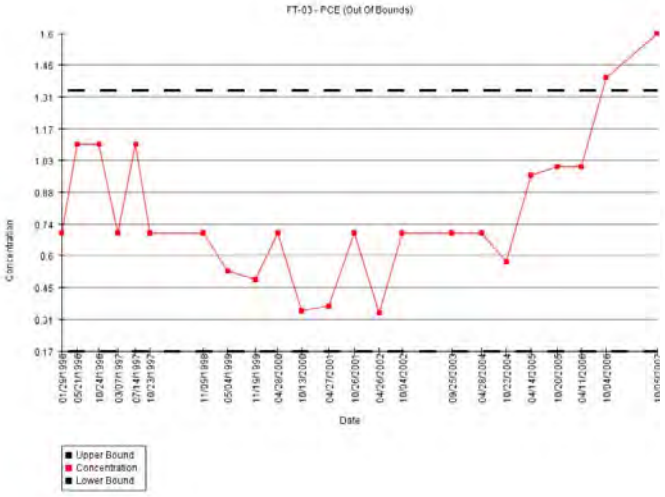


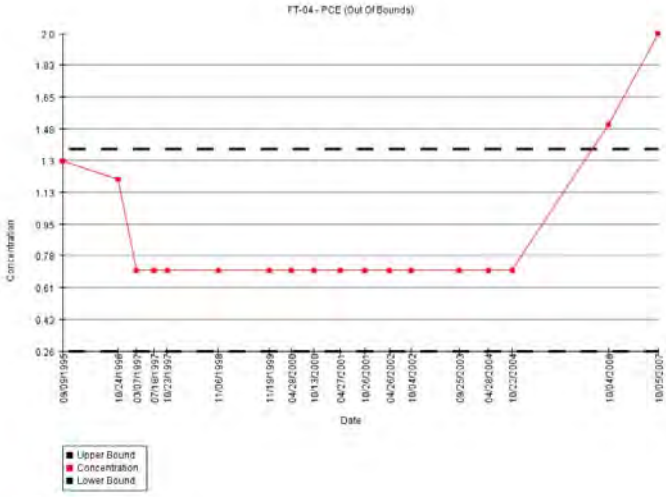


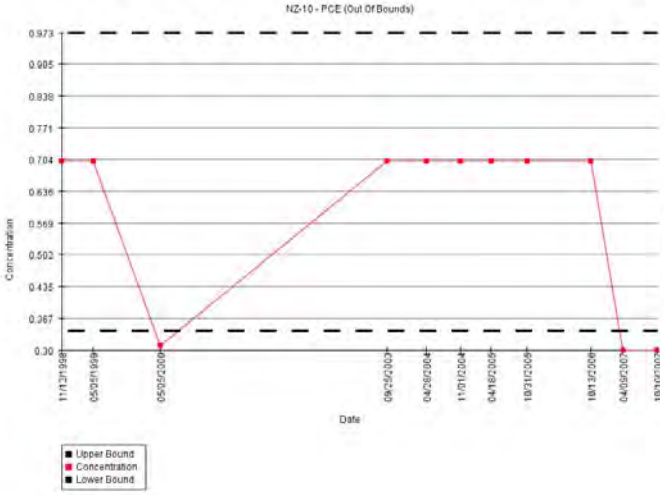
c12DCE: NZ-93

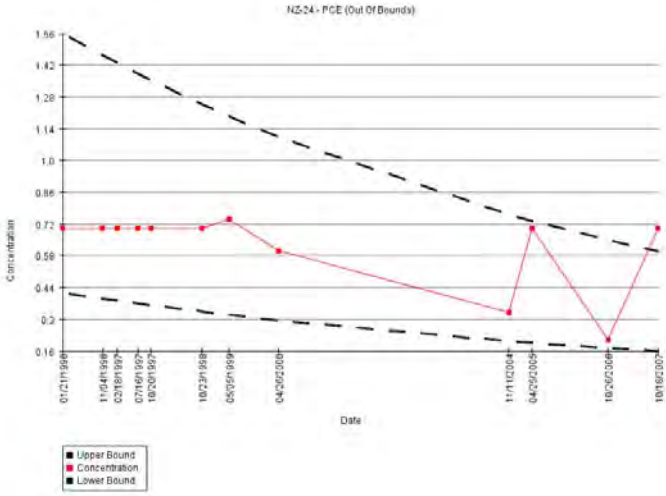


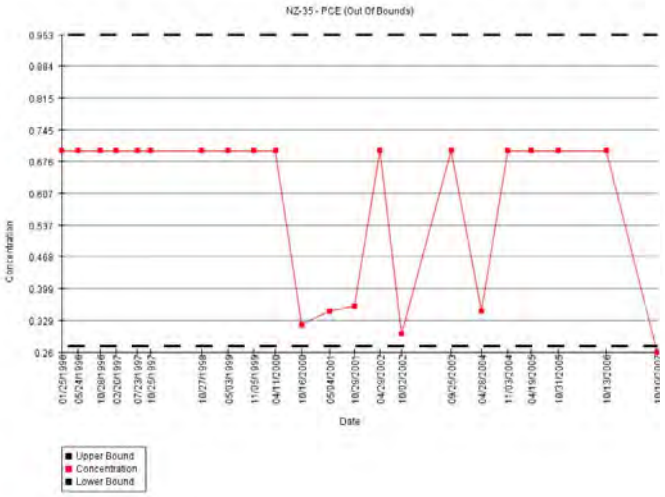
PCE: FT-03, FT-04, NZ-10, NZ-24, NZ-35, NZ-68

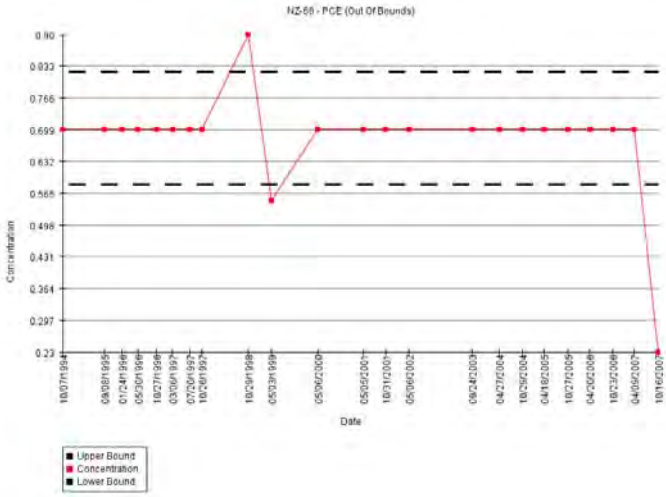




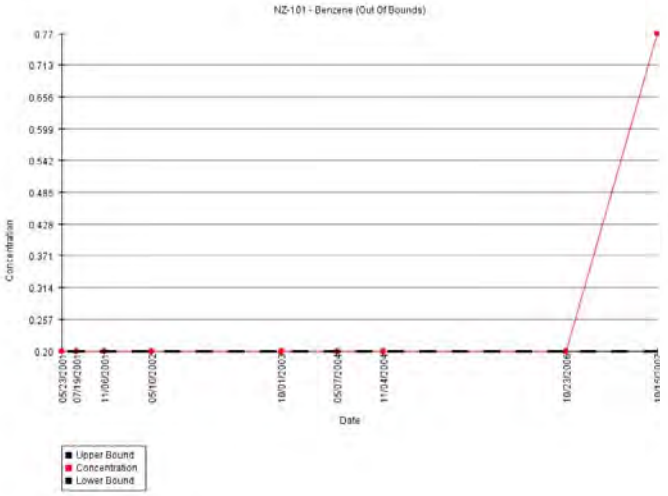


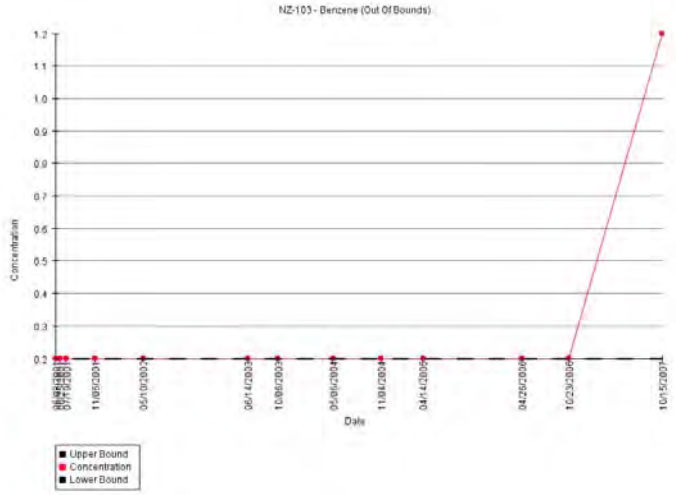


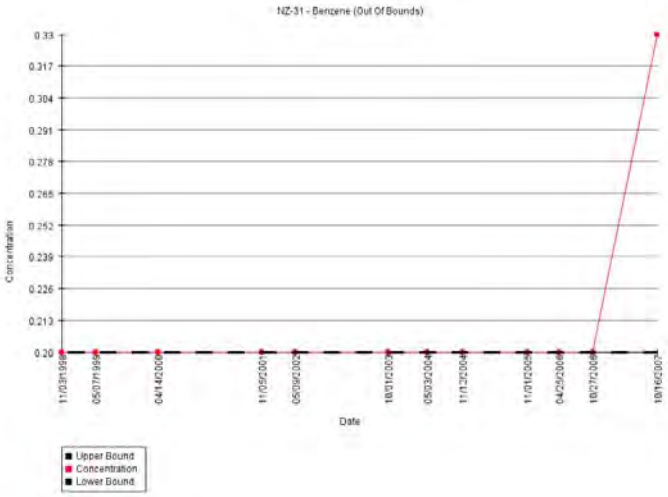




Benzene: NZ-101, NZ-103, NZ-31

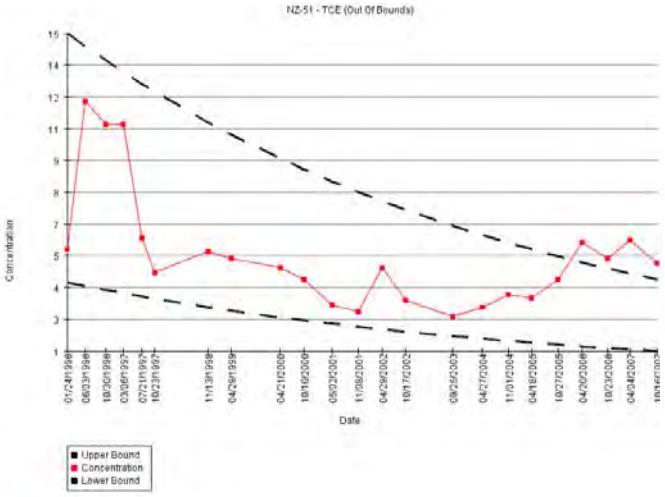




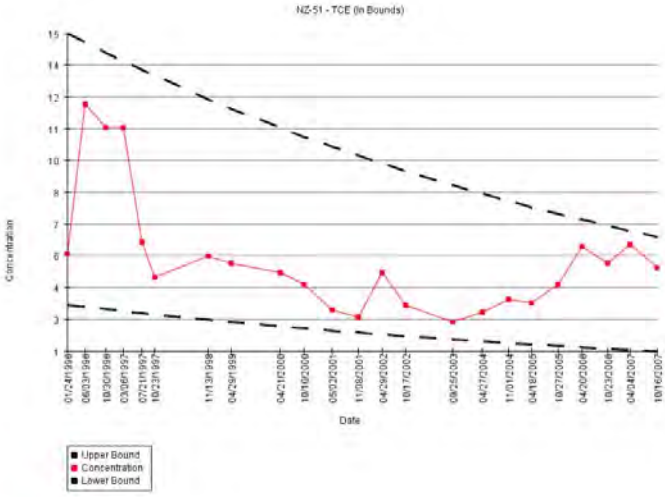


It can be concluded from the above results that adding the 2006 data into the historical data (Scenario 2A and Scenario 2B) resulted in a reduction in the number of out-of-bounds TCE values detected by DT, since 2006 data provide more information for the statistical bound calculation, resulting in broader determination of bounds. For example, TCE in NZ-51 was detected as out-of-bounds for Scenario 1B (2006 data not included in the historical data), but not in Scenario 2B (2006 data included in the historical data), as illustrated below:

Scenario 1B

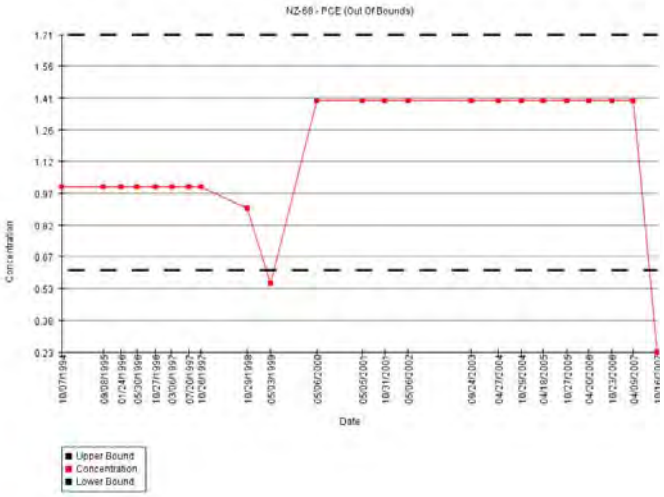


Scenario 2B

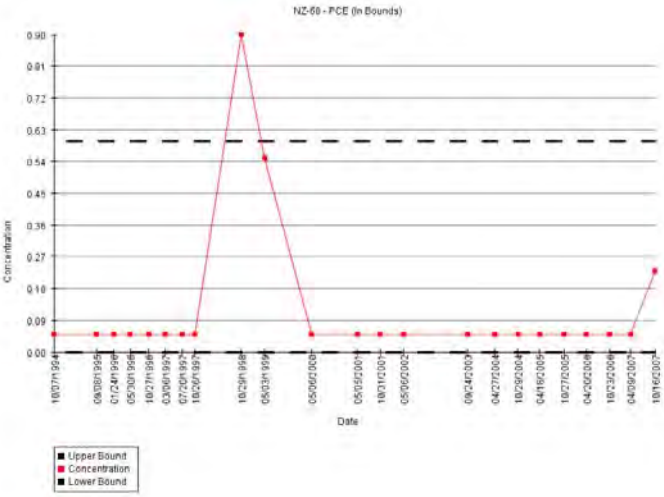


Moreover, substituting ND values with the graphing value 0.05 resulted in differences in the out-of-bounds values, particularly for PCE for which concentrations are low. This is because by assigning 0.05 to ND values which are typically greater than 0.05, the historical statistical bound is calculated lower, resulting in a narrower bound. For instance, as illustrated below, the lower bound of PCE in NZ-68 in Scenario 1A (replacing the ND values with the reporting limits) is approximately 0.61, and the 2007 value of 0.23 was detected out-of-bounds; however, its lower bound in Scenario 1B (replacing the ND values with the graphing value 0.05) is close to zero, therefore making the 2007 data an in-bound value.

Scenario 1A



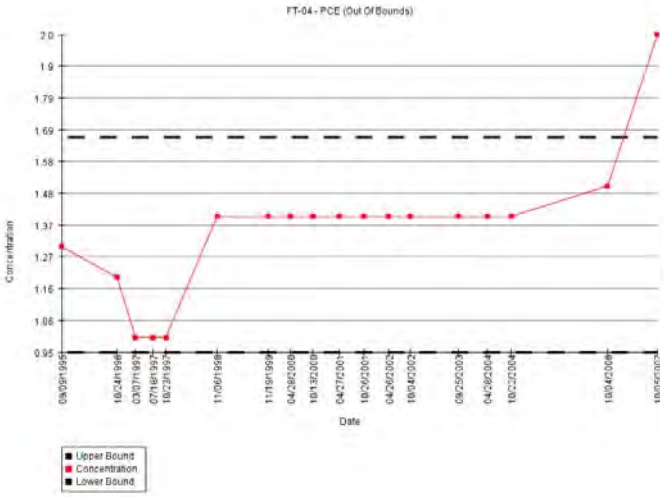
Scenario 1B



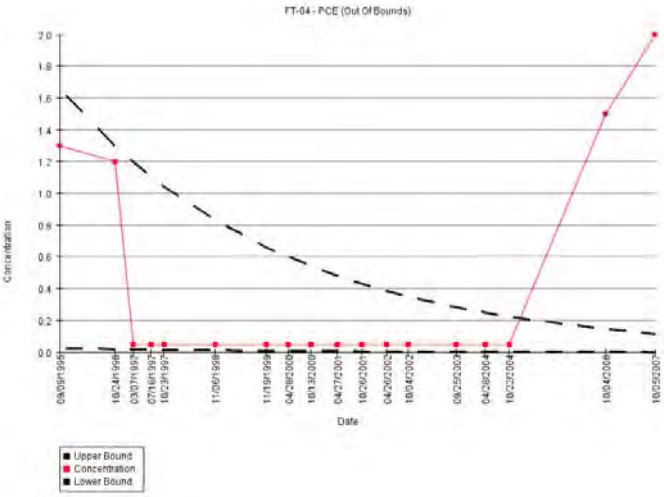
In addition, substituting ND values with the most common RL of each CoC (Scenario 1C and Scenario 2C) resulted in additional "out-of-bounds" values for PCE and c12DCE, which are mostly low concentrations. Those out-of-bounds items may not be of particular interest to the Site personnel due to their historically consistent low concentrations.

Another example of how using a consistent graphing value for NDs is PCE at well FT-04, illustrated below. When the RL is used for each sample (Scenario 1A) the software uses a static bound. However, when the graphing value of 0.05 µg/l is used to replace NDs (Scenario 1B), the software uses a time-dependant bound. In both cases the value from 2007 is out-of-bounds. However, the value from 2006 (which was not used in the background data for these scenarios) was out-of-bounds in Scenario 1B but not in Scenario 1A.

Scenario 1A

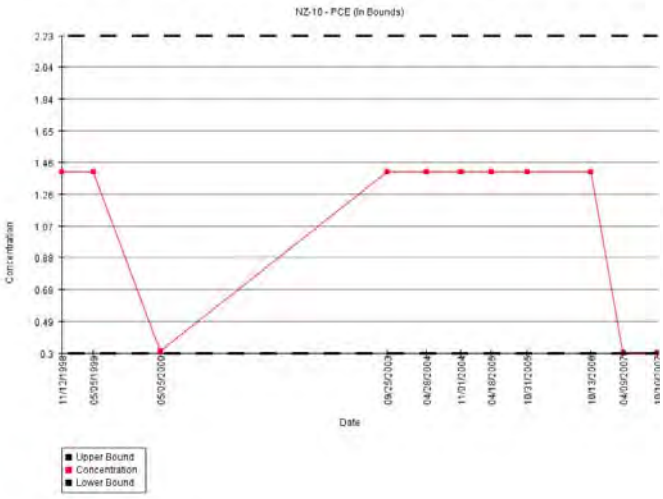


Scenario 1B

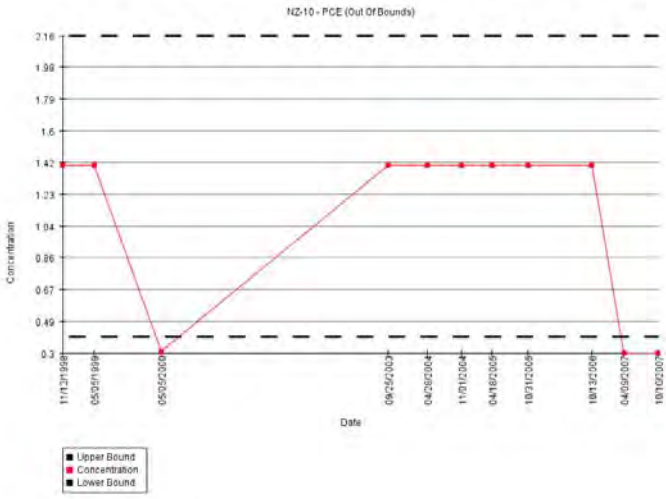


Another interesting observation, illustrated below, is PCE in NZ-10. By comparing the PCE historical plots in NZ-10 for Scenario 1A (without 2006 included in the background data) and Scenario 2A (with 2006 data included in the background data), it can been seen that adding the 2006 data (which was a single ND value with an RL of 1.4) to the background data increased the lower bound calculated by the software from 0.3 to about 0.46, thus causing the 2007 data (which are very low values) to be out-bounds. It is not clear to our software user why adding the non-detect value with an RL value of 1.4 in 2006 to the background data in Scenario 2A would cause the lower bound to be calculated higher.

Scenario 1A



Scenario 2A



As EnviroStat identified and pointed out regarding the original DT analysis, DT often flags very low values in the current data as out-of-bounds when the background data are mostly J/NDs. This is one area of improvement that Summit may consider working on to enhance the performance of DT. It is possible for a background dataset consisting entirely of "J" and "ND" values to appear to have a trend because of slight variations in RLs or J values, which is generally not something the user would consider to be a trend.

One noticeable observation regarding the DT results that merits further attention involves a steady and distinct increase of TCE concentrations since 2004 in well NZ-93, which is located in the northeast portion of the site, to values greater than 80 µg/l. These increases suggest potential new sources of TCE contamination and/or TCE plume migration toward the northeastern corner of the site.

A list of the wells that were out-of-bounds for each scenario is provided below:

COC / Scenario	TCE	c12DCE	PCE	Benzene	Toluene
1A	FT-04, MW-102, NZ-06, NZ-51, NZ-68, NZ-93	NZ-93	FT-04, NZ-68	NZ-101, NZ-103, NZ-31	None
1B	FT-04, MW-102, NZ-06, NZ-51, NZ-68, NZ-93	NZ-93	FT-03, FT-04	NZ-101, NZ-103, NZ-31	None
1C	FT-04, MW-102, NZ-06, NZ-51, NZ-68, NZ-93	NZ-93, NZ-39	FT-03, FT-04, NZ-10, NZ-35, NZ-68	NZ-101, NZ-103, NZ-31	None
2A	FT-04, MW-102, NZ-06, NZ-93	NZ-93	FT-04, NZ-10, NZ-68	NZ-101, NZ-103, NZ-31	None
2B	FT-04, MW-102, NZ-06, NZ-93	NZ-93	FT-03, FT-04, NZ-10	NZ-101, NZ-103, NZ-31	None
2C	FT-04, MW-102, NZ-06, NZ-93	NZ-93	FT-03, FT-04, NZ-10, NZ-35, NZ-68, NZ-24	NZ-101, NZ-103, NZ-31	None

During this follow-up analysis with DT, no out-of-bounds values were interpreted by our analyst as “bad data”. Rather, some of the out-of-bounds values appear to be continuations of increasing trends, while others are very small values where the current data are either just above or just below the software-calculated bounds but do not appear to be a concern. Our analyst also agrees with EnviroStat that using a consistent small value for NDs provides for an improved analysis. The GeoTrans analyst concluded there was not a compelling reason to not update the previous background data with the 2006 data prior to using DT on the newer 2007 data. However, EnviroStat points out that with continual updating of background there may be cases with a *slowly increasing trend* where each individual observation would be “in-bounds” because the bounds widen with each additional observation that is added to the background. Eventually, of course, this should result in a statistically significant increasing trend, but DT does not currently indicate which wells have increasing trends, it simply falgs “out-of-bounds” values. This leads to a basic question “What should DT bring to our attention?”. If the goal is for DT to be sentive to increasing trends (i.e., indicate the new values are “out-of-bounds”, then not updating the background values would be preferred. However, that approach would ultimately require periodic updates of the background data.

Mass Flux and Mass Metric Results

Along the eastern portion of the GAFB Site, where the MLU thins and contains more coarse sediments, Upper Aquifer groundwater can flow downward to the Lower Aquifer through the permeable lacustrine zone (PLZ). Site personnel are interested in quantifying the TCE mass migrating downward to the Lower Aquifer across the PLZ, and may also be interested in tracking the degree to which TCE mass is changing over time in the Upper Aquifer.

The Summit software includes a mass flux feature for tracking the flux of mass across a series of linear boundaries over time. The software also includes a mass metric feature for comparing the "relative mass" over time. It is referred to as "relative mass" because the software does not account for aquifer thickness and porosity; rather, it calculates mass per unit volume of aquifer. The original analysis for this site included evaluation of both the mass flux and mass metric features within the software using data through 2006. For this follow-up analysis, the same features are applied using the 2007 data in addition to the previous sampling events.

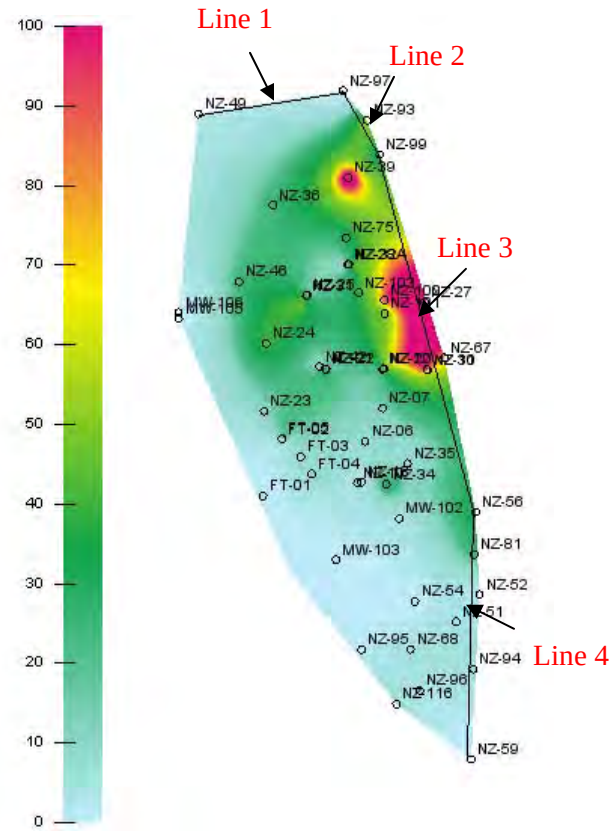
Mass Flux Calculations

From water level maps in site reports, the contours of groundwater levels in the Upper Aquifer are approximately parallel to the boundary of PLZ, which implies that groundwater flow is essentially perpendicular to the boundary of the PLZ.

The software requires the user to define one or more line segments graphically. The software estimates concentrations at different points along the line segments, multiplies them by the groundwater flow rate input by the user for that line segment, and then sums up mass flux along the entire line segment. For this example, four line segments were drawn to approximate the shape of the PLZ boundary. These lines are illustrated (in black) on the figure below.

{this gap is intentional}

Mass Flux Boundary Lines (In Black)



The four lines are basically defined by locations of five wells:

- Line 1: NZ-49 to NZ-97
- Line 2: NZ-97 to NZ-99
- Line 3: NZ-99 to NZ-56
- Line 4: NZ-56 to NZ-59

Information regarding these lines is provided below:

Well	East Coordinate	North Coordinate	Line	Length (ft)
NZ-49	6748970	2048548		
NZ-97	6751385	2048935	1	2445.811522
NZ-99	6752002	2047870	2	1230.818427
NZ-56	6753598	2041912	3	6168.061284
NZ-59	6753529	2037788	4	4124.57719

Once the lines are defined, the user must enter the groundwater flow rate for each line segment. This was calculated with the following equation:

$$Q = vAn$$

where

v = groundwater velocity in ft/d (equal to Ki/n)
A = area through which flow occurs (segment length times saturated thickness)
n = porosity
K = hydraulic conductivity
i = hydraulic gradient

According to the 2005 Hydrogeologic Conceptual Site Model Report prepared by MWH, the groundwater flow direction in the Upper Aquifer is typically toward the northeast, with magnitude of hydraulic gradient ranging from 0.001 to 0.055. Groundwater velocity in the Upper Aquifer ranges from 0.2 to 1.0 ft/day, and a representative saturated thickness of the Upper Aquifer is 50 ft. Porosity was assigned as 0.3 by GeoTrans. The software does not account for effects of retardation and/or decay of TCE.

Mass flux calculations were performed based on a groundwater velocity of 1.0 ft/d. The calculation of flow rate for the case with groundwater velocity of 1.0 ft/d is presented below.

Flow Rates Based on GW Velocity = 1.0 ft/day

Line	Velocity (ft/d)	Porosity	Flow rate (ft^3/d)	Flow rate (liter/d)
1	1	0.3	36687.17283	1038865.045
2	1	0.3	18462.27641	522793.4487
3	1	0.3	92520.91926	2619900.677
4	1	0.3	61868.65785	1751925.293

The software requires the flow rate in units consistent with the concentrations, which in this example are µg/l. Thus, the flow rate is input as liters per day. The mass flux results over time for this scenario (i.e., groundwater velocity of 1.0 ft/d), based on the same time-varying data utilized for the spatio-temporal analysis and through 2007, are provided below.

Mass Flux Rates over Time (GW Velocity = 1.0 ft/day)

Date	Line 1 TCE flux (kg/d)	Line 2 TCE flux (kg/d)	Line 3 TCE flux (kg/d)	Line 4 TCE flux (kg/d)
6/30/1996	0.55	0.54	22.96	1.35
12/30/1996	2.59	0.93	34.16	2.50
6/30/1997	2.40	0.92	54.39	2.07
12/30/1997	1.58	0.63	20.54	1.30
12/30/1998	1.79	1.05	105.42	0.28
6/30/1999	2.75	1.96	68.56	0.39
12/30/1999	3.22	1.81	85.59	1.01
6/30/2000	1.08	1.19	63.53	0.44
12/30/2000	3.08	1.69	117.29	1.88
6/30/2001	0.29	0.06	76.25	4.48
12/30/2001	0.18	0.15	108.22	5.26
6/30/2002	0.15	0.29	100.03	12.01
12/30/2002	0.15	0.27	79.30	8.53
12/30/2003	0.12	0.41	105.85	13.11
6/30/2004	0.21	0.47	52.38	6.35
12/30/2004	0.18	0.44	65.51	3.81

Date	Line 1 TCE flux (kg/d)	Line 2 TCE flux (kg/d)	Line 3 TCE flux (kg/d)	Line 4 TCE flux (kg/d)
6/30/2005	0.17	0.47	42.20	2.89
12/30/2005	0.05	0.45	40.77	2.61
6/30/2006	0.15	0.66	43.48	2.30
12/30/2006	0.25	0.64	39.44	2.58
12/30/2007	1.74	1.19	31.76	1.82
Average mass flux (kg/d)	1.08	0.77	64.65	3.67
Mass flux percentage	1.54%	1.10%	92.14%	5.22%

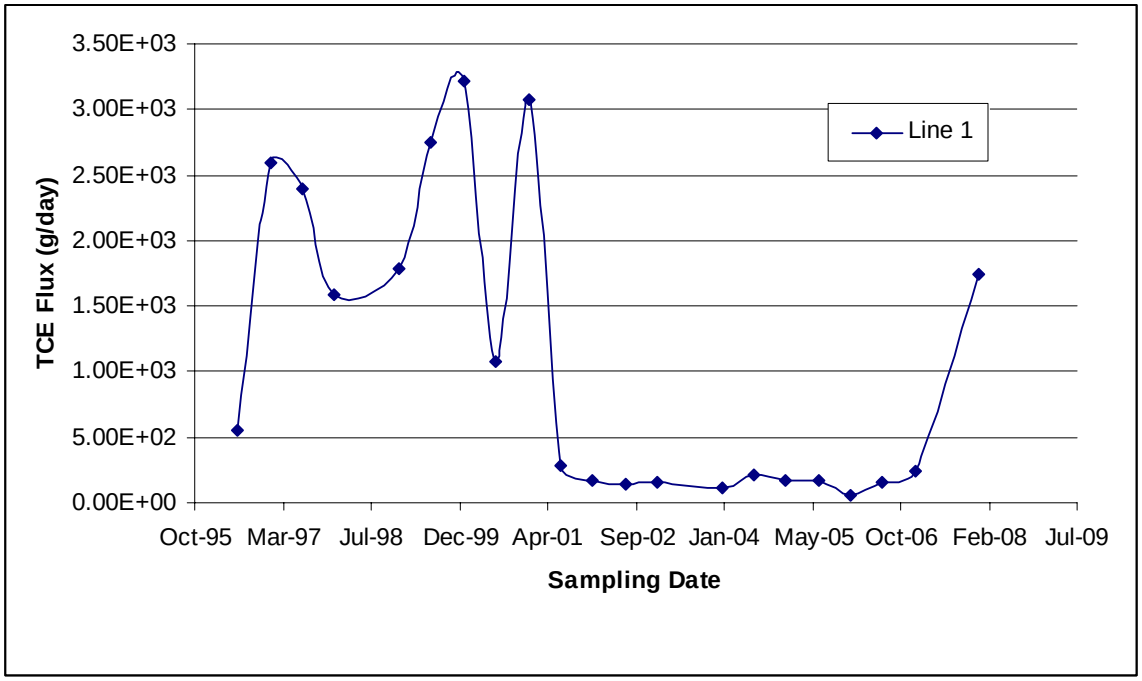
Based on these results, the total average TCE flux across PLZ over these time periods (i.e., 06/1996 ~ 12/2007), with groundwater velocity of 1.0 ft/d is estimated to be 70.17 kg/day, which is calculated by adding the results for the four segments. Also, significantly more (92.14%) of the mass flux occurs along line 3 than the others, mostly because of higher concentrations and also because the line is longer. The mass flux percentage of each line segment is similar to the previous report, except that Line 4 increased from 2.82% to 5.22%. GeoTrans notes that the software requires the lines to be defined manually in a graphical approach, and they were defined slightly differently in the follow-up analysis versus the original analysis. It is not possible to assign exactly the same line each time because coordinates cannot be entered. This is the likely explanation for the difference in mass flux percentage for Line 4.

Obviously, different values will be calculated if different values for groundwater velocity and/or porosity are assumed. However, the relative mass fluxes between sampling events will remain consistent (i.e., they will be scaled consistently).

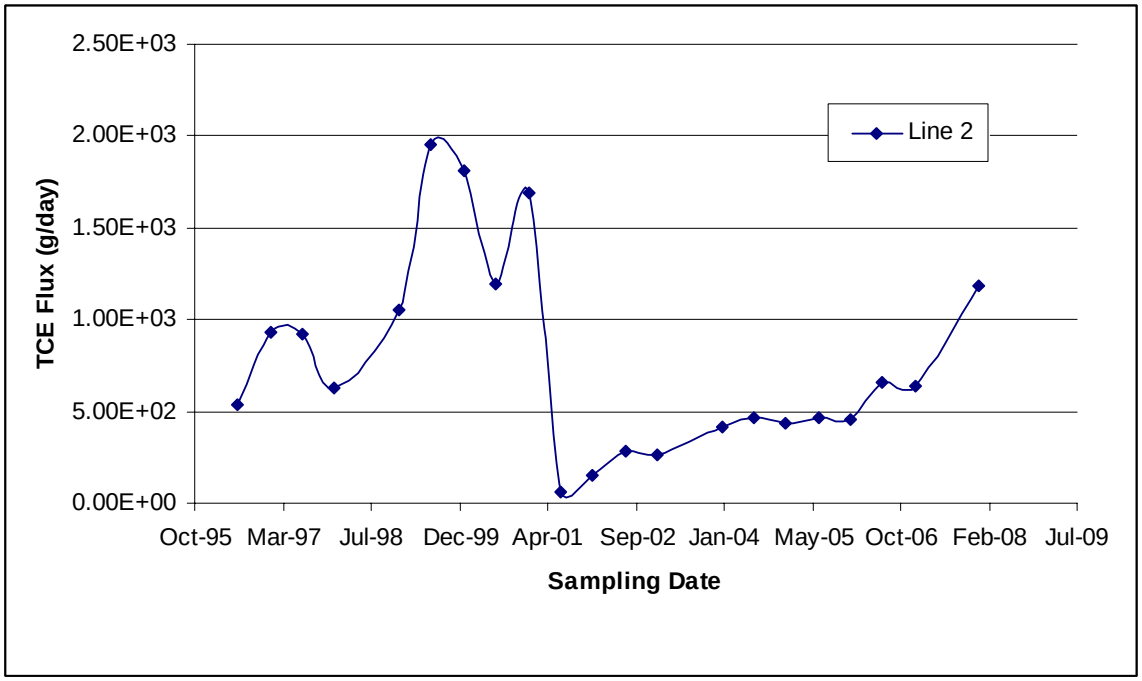
The time series plots of the mass flux across each of the four line segments (L1, L2, L3, and L4), in grams per day, are presented below:

{this gap is intentional}

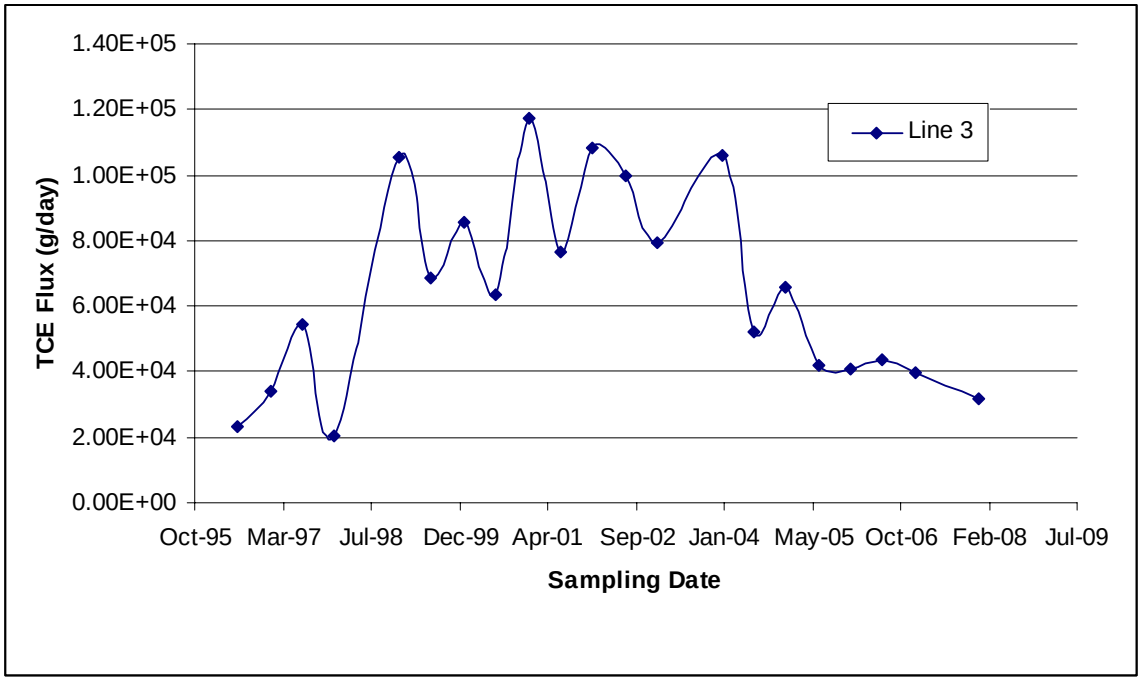
Mass Flux Rates over Time, Line 1 (GW Velocity = 1.0 ft/day)



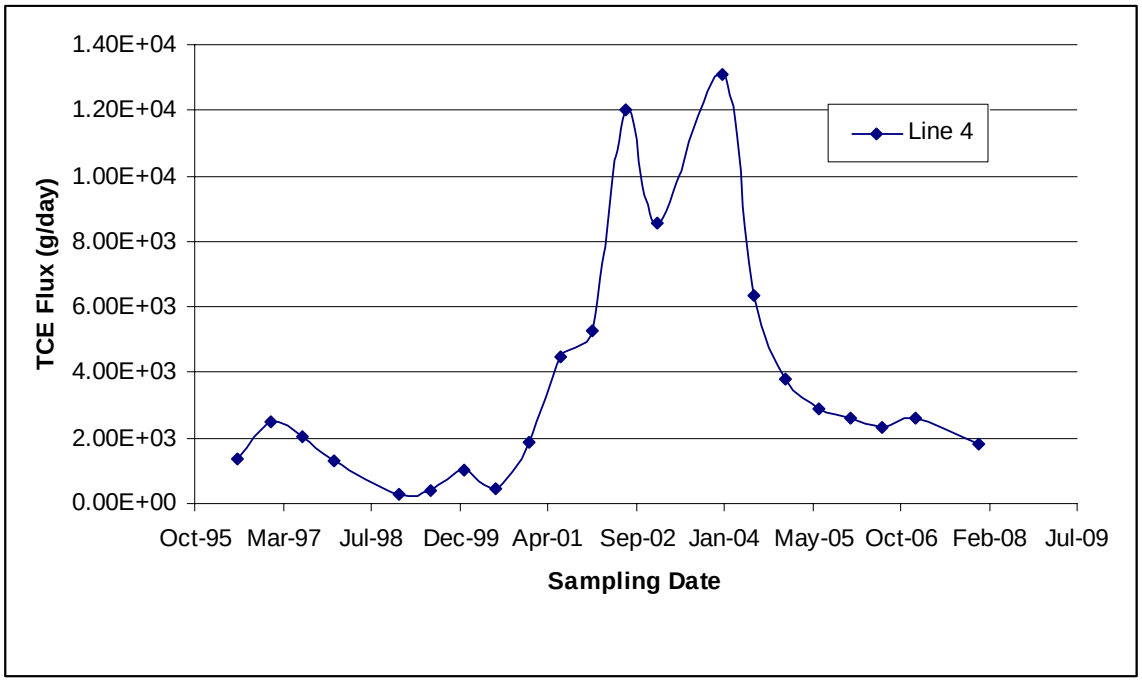
Mass Flux Rates over Time, Line 2 (GW Velocity = 1.0 ft/day)



Mass Flux Rates over Time, Line 3 (GW Velocity = 1.0 ft/day)



Mass Flux Rates over Time, Line 4 (GW Velocity = 1.0 ft/day)

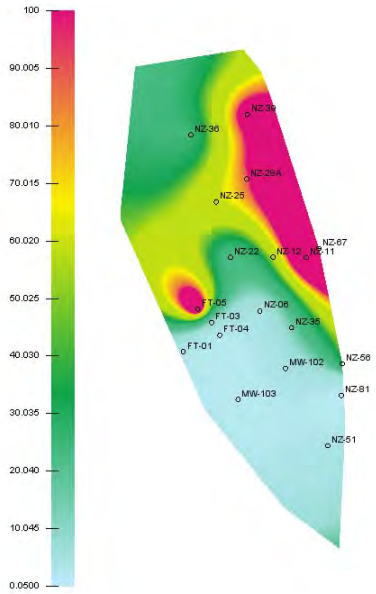


For Line 1 and Line 2, the results suggest much higher mass flux before 2001 compared to after 2001, and for Line 4 somewhat lower mass flux before 2001 compared with later. Moreover, for Line 1 and Line 2, there is a noticeable increase in TCE flux from 2005 through 2008, suggesting that plume might be migrating toward the northeastern corner of the Site.

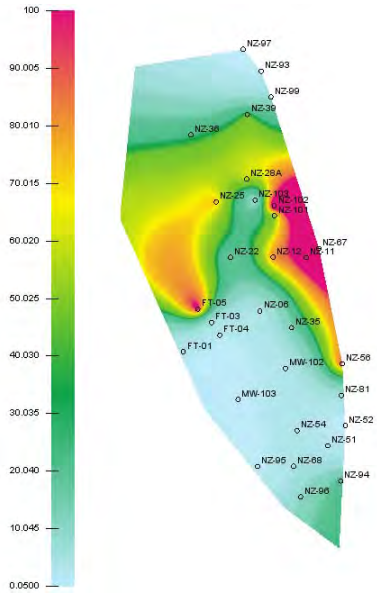
Discussion Regarding Mass Flux

It is interesting to find that the mass flux for Line 1 and Line 2 changed significantly over a short period from 12/2000 to 6/2001. The comparison of interpolated plume visualizations generated by the software for the two sampling events is provided below.

Interpolated Plume Map, 12/2000



Interpolated Plume Map, 6/2001

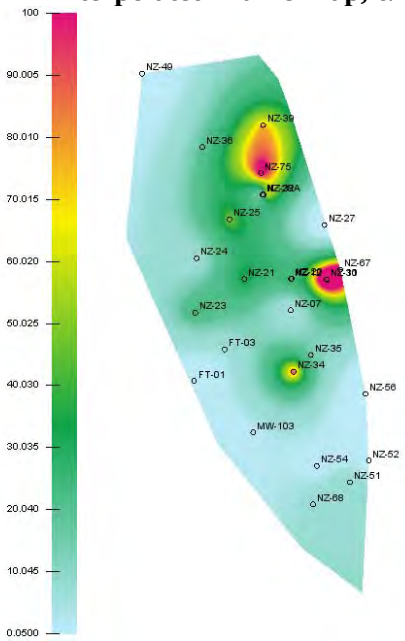


Focusing on the extreme northern portion of these figures (i.e., the vicinity of Lines 1 and 2) it is noteworthy that the interpolated concentrations in 6/2001 (right figure) are significantly lower due to the presence of three new monitoring wells (NZ-97, NZ-93, NZ-99). These wells were not present in 12/2000 or in other previous events. These three additional wells give more accurate information about the actual TCE concentrations in their vicinity. In other words, the interpolated concentrations in 12/2000 are biased due to fewer sampling locations in that area. This points out a problem with this mass flux feature, which is that different events can use different selections of wells, which can bias the calculations. This also brings about a question as to whether the data from newly installed wells should be "brought backwards" to previous sampling events, allowing comparisons to be based on similar distributions of sampling locations which would presumably result in more reliable estimation of mass flux changes over time.

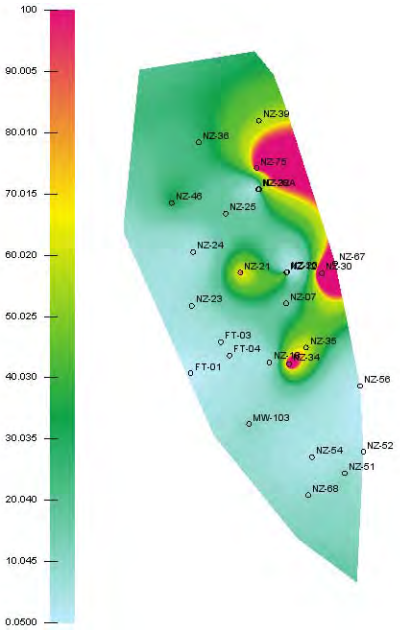
Another interesting case involves wells that are sampled in one period but not a subsequent period. The upper left-hand corner of the following figures, in the vicinity of well NZ-49, provides a good example to illustrate this point.

{this gap is intentional}

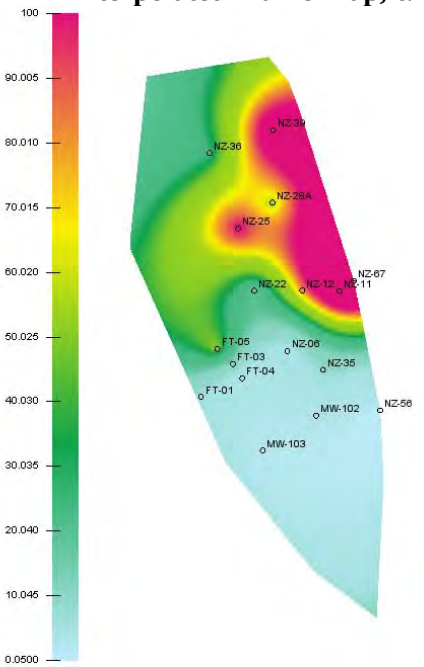
Interpolated Plume Map, 6/1996



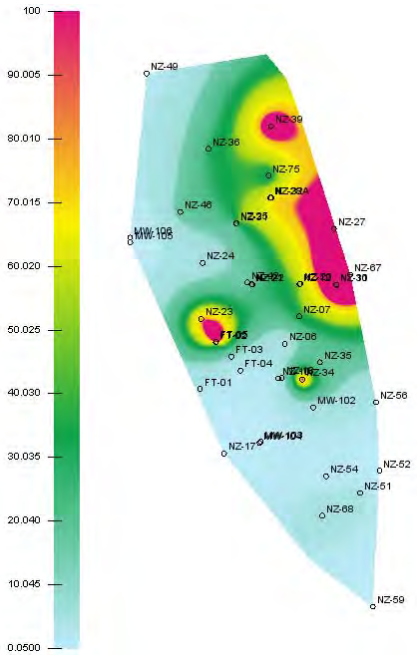
Interpolated Plume Map, 6/1997



Interpolated Plume Map, 6/1999



Interpolated Plume Map, 6/2000



In 6/1996, NZ-49 was sampled, and it appears to be clean. From 1997-1999, NZ-49 was not sampled, but in 6/2000 it was again sampled and was found to be still clean. In this case, it makes perfect sense to assume NZ-49 always had low concentrations in between the 1996 and 2000 samples. However, the interpolations for the events between these two samples are based only on spatial interpolation for wells sampled in those events, and thus there is extrapolation in the area near NZ-49 for those events that estimate higher concentrations. A different result would have occurred if values for NZ-49 had been assigned for events where it was not sampled, based on

temporal interpolation of the 1996 and 2000 results at that well. At this point the software does not accommodate this.

This discussion leads to an argument that the analyst should consider adding "estimated values" for sampling locations that are not actually sampled in specific events, so that the interpolated plumes used for the mass flux estimates in each period are based on a consistent set of locations. These estimated values would be based on actual data at those locations for events where they were sampled, plus the underlying conceptual model of the site, such as the general groundwater flow and plume transport, whether new sources of contaminant are expected, impact of remedial actions, etc. Ambiguities can still arise, however; for example, it could have happened that the TCE values in NZ-49 might have been different in 1996 and 2000, so that one would know that there was a change during that period, but not know how to fill in the missing values "correctly".

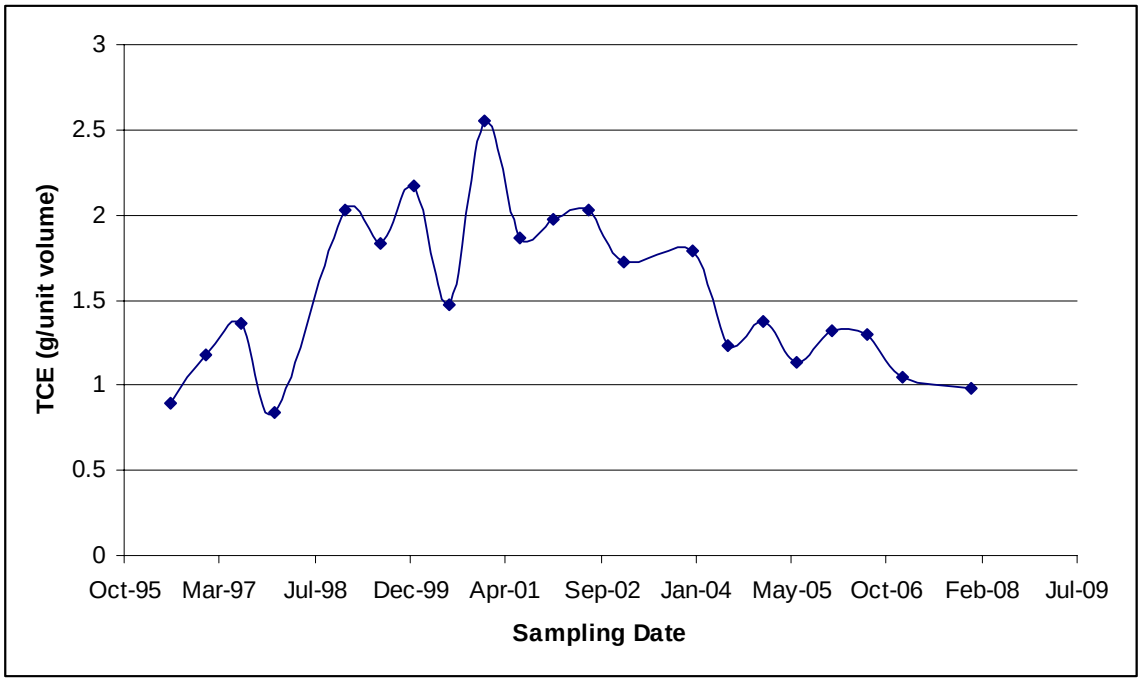
Mass Metric Results

The mass metric feature of Summit monitoring tools is designed to assist the analyst in comparing the relative mass in the plume for different sampling events. The interpolated concentration value at every cell in the plume map for that COC at that time period is summed, and each cell is assumed to represent the same volume (i.e., incorporating vertical extent and porosity) as every other cell. Thus, the results are actually given in mass per unit aquifer volume, rather than absolute mass.

The results for the mass metric, based on the same time-varying data utilized for the spatio-temporal analysis and through 2007, are presented in the below table:

Date	TCE (g)	# of Samples
6/30/1996	0.899803	27
12/30/1996	1.180228	25
6/30/1997	1.358198	27
12/30/1997	0.838559	31
12/30/1998	2.030911	38
6/30/1999	1.828882	30
12/30/1999	2.173369	17
6/30/2000	1.476845	43
12/30/2000	2.553259	19
6/30/2001	1.86038	31
12/30/2001	1.970939	39
6/30/2002	2.030898	44
12/30/2002	1.728991	28
12/30/2003	1.78555	50
6/30/2004	1.233087	48
12/30/2004	1.37812	52
6/30/2005	1.136666	39
12/30/2005	1.317143	35
6/30/2006	1.302859	32
12/30/2006	1.051141	52
12/30/2007	0.979951	47

Mass Metric versus Time

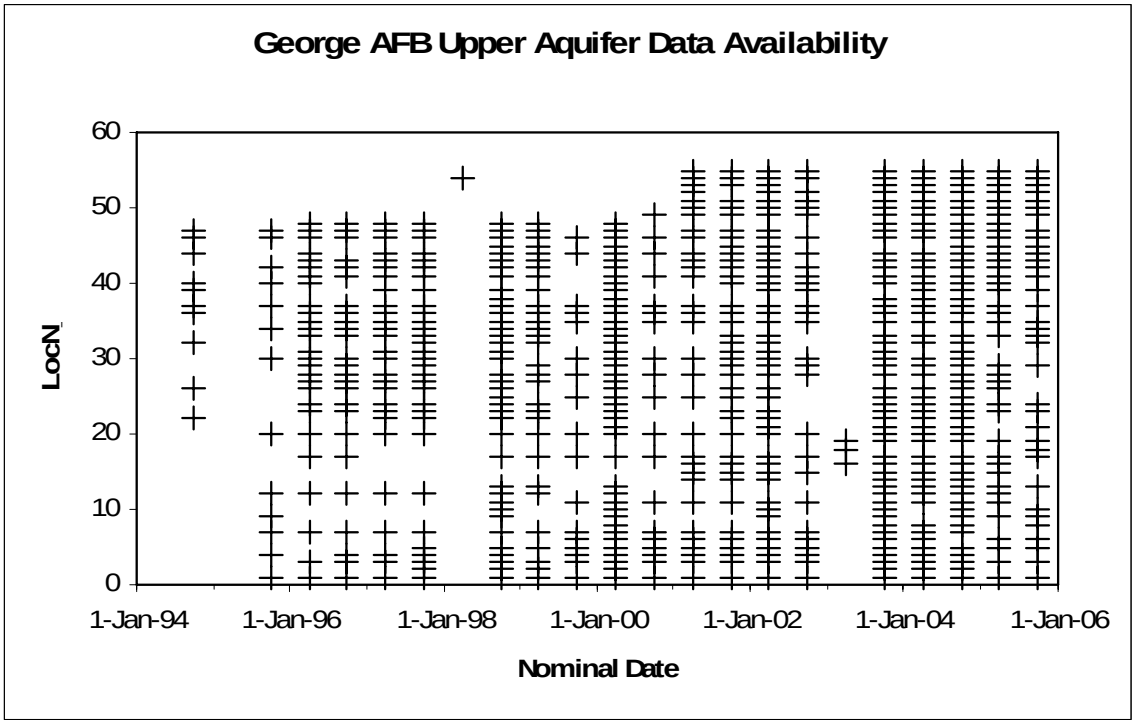


It can be seen from the above mass metric plot that the mass of TCE per unit aquifer volume varies over time – it grows steadily from 1995 to 2001, when the TCE mass reaches its maximum of nearly 2.5 g/unit volume. After 2001 the mass is decreasing constantly; this may be caused by the implementation of the aquifer remediation. Additionally, the dataset for 12/30/2000 has a relatively high mass. Although this is partly explained by higher concentrations measured at some points in 12/30/2000, an examination of the sampling history reveals that there are only 19 data points for that event, compared to 43 for 6/30/2000 and 31 for 6/30/2001. Fewer sampling locations result in lower accuracy of the interpolation – in other words, the insufficiency of information causes some interpolation bias (in this case the interpolated values seem to be overestimated). The extent of the zone with highest concentrations is greater for the 12/30/2000 event partly because of lack of sampling at some wells that were sampled in the event prior and/or the event subsequent. This helps explain why the calculated mass metric has a particularly high value for 12/30/2000. It again suggests that it may be important to have similar data distribution per event for making these types of comparisons over time, based on some rule(s) for assigning data values at locations not sampled in specific events.

GeoTrans also looked at the mass flux and mass metric plots and compared them with those from the original analysis (i.e., through 2006). There is some differences in the mass flux values computed during the follow-up analysis versus the original analysis, due to the manner in which the line segments are defined (i.e., graphically within the software). GeoTrans performed an experiment where it used the same dataset and manually created the line segments, and each time slightly different mass flux results were determined due to the unrepeatable process of creating line segments graphically.

Discussion Regarding Mass Flux With Incomplete Data Filled in Manually

As was done in the original analysis, to assess the difference in the mass calculations that might occur if each dataset has similar data distribution, EnviroStat created an additional dataset where data gaps were eliminated by assigning values based on temporal interpolation. The schematic below shows the data availability over time for TCE at the former GAFB for the historical data. The dates are those assigned to the spatiotemporal dataset. For some nearly all wells are missing, and for the others the majority of the wells are present.

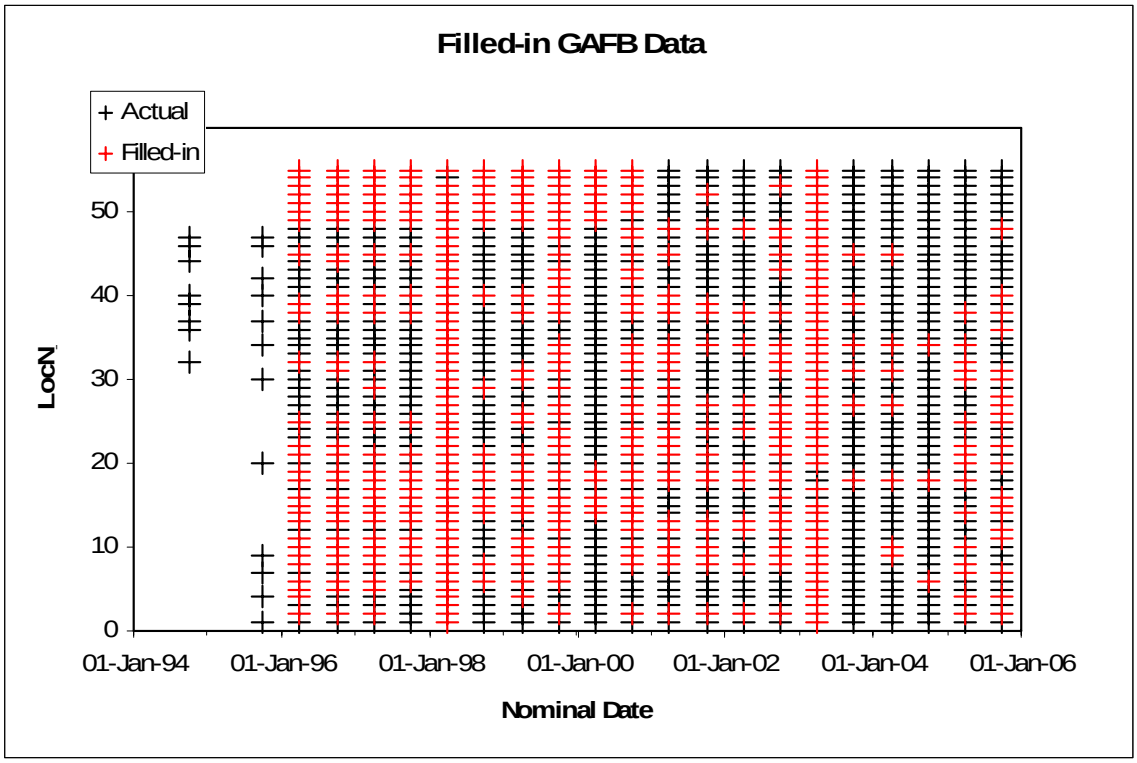


The table below indentifies the actual well associated with each location represented in the figure presented above.

Translation of "LocN" to SiteID									
LocN	SiteID	LocN	SiteID	LocN	SiteID	LocN	SiteID	LocN	SiteID
1	FT-01	12	NZ-07	23	NZ-20	34	NZ-34	45	NZ-59
2	FT-02	13	NZ-10	24	NZ-21	35	NZ-35	46	NZ-67
3	FT-03	14	NZ-101	25	NZ-22	36	NZ-36	47	NZ-68
4	FT-04	15	NZ-102	26	NZ-23	37	NZ-39	48	NZ-75
5	FT-05	16	NZ-103	27	NZ-24	38	NZ-42	49	NZ-81
6	MW-102	17	NZ-11	28	NZ-25	39	NZ-46	50	NZ-93
7	MW-103	18	NZ-111	29	NZ-27	40	NZ-49	51	NZ-94
8	MW-104	19	NZ-116	30	NZ-28A	41	NZ-51	52	NZ-95
9	MW-105	20	NZ-12	31	NZ-30	42	NZ-52	53	NZ-96
10	MW-106	21	NZ-17	32	NZ-31	43	NZ-54	54	NZ-97
11	NZ-06	22	NZ-18	33	NZ-32	44	NZ-56	55	NZ-99

For 1996-2005 there are 625 actual values for 55 wells and 20 events. A completely filled-in data set would have 1100 values (55 wells x 20 events) and therefore 475 values would need to be

supplied to have a "complete" dataset. About 100 of these occur in two events, those of April 1998 and April 2003. Others dominate the early data (the wells were added later), with SiteIDs NZ-81 and higher. The proportions of filled-in data range from 9% for NZ-39 up to 75% for several wells. Well NZ-111 with fewer than four actual values is omitted.



To ensure a consistent sampling distribution over the sampling events, some wells that are missing in the 2006 and 2007 datasets were added in by assigning the data from the very last event for those locations. The assigned values for those wells not sampled in 2006 and 2007 are listed below:

4/1/2006 add-in data

Well ID	Value Assigned
FT-02	9.4
FT-04	0.66
MW-102	4.3
MW-103	0.54
MW-106	0.05
NZ-06	13
NZ-07	10.5
NZ-10	4.4
NZ-101	53.6
NZ-102	111
NZ-11	92.3
NZ-12	20.4
NZ-18	1.83

Well ID	Value Assigned
NZ-21	49.4
NZ-22	8.95
NZ-23	9.93
NZ-24	36.1
NZ-25	18.6
NZ-35	7.8
NZ-36	28.3
NZ-42	6.63
NZ-75	40.1

10/1/2006 add-in data

Well ID	Value Assigned
MW-104	1.9
NZ-17	0.25

10/1/2007 add-in data

Well ID	Value Assigned
MW-104	1.9
MW-105	0.55
NZ-17	0.25
NZ-18	3
NZ-34	31
NZ-54	0.81
NZ-59	0.25
NZ-75	30
NZ-97	0.25

GeoTrans then utilized this additional dataset provided by EnviroStat for the historical values, plus the assigned values provided above for 2006 and 2007, to perform the mass flux and mass metric calculations (i.e., each period has a consistent number of samples).

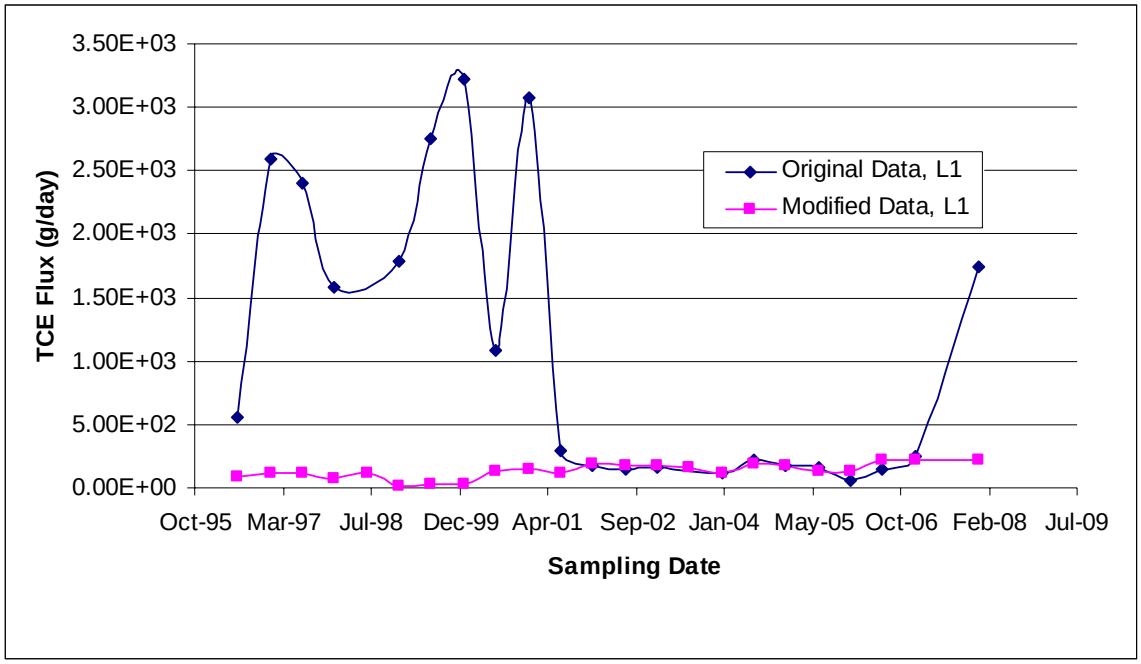
Mass Flux Rates over Time with "Filled-in" Data (GW Velocity = 1.0 ft/day)

Date	Line 1 TCE flux (kg/d)	Line 2 TCE flux (kg/d)	Line 3 TCE flux (kg/d)	Line 4 TCE flux (kg/d)
6/30/1996	0.0861	0.0600	16.0037	1.4182
12/30/1996	0.1184	0.0710	26.2298	1.6642
6/30/1997	0.1180	0.0705	31.6838	1.7311
12/30/1997	0.0737	0.0677	18.9818	1.3410
6/30/1998	0.1148	0.0765	36.3041	1.1741
12/30/1998	0.0194	0.0607	41.4241	1.0746
6/30/1999	0.0226	0.0673	50.9680	1.1334
12/30/1999	0.0229	0.0684	52.3299	1.0494
6/30/2000	0.1274	0.0708	51.0431	1.3349
12/30/2000	0.1474	0.0652	67.4468	1.5914
6/30/2001	0.1243	0.0610	77.3286	2.9643
12/30/2001	0.1858	0.1540	91.4490	4.7797

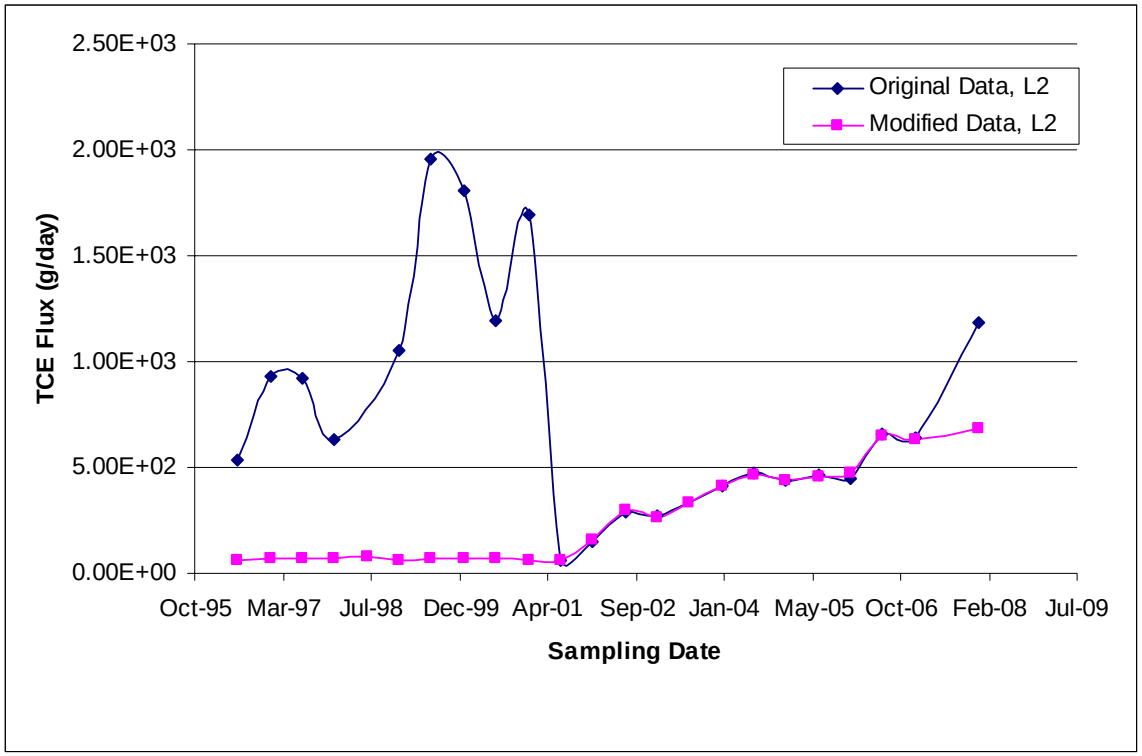
Date	Line 1 TCE flux (kg/d)	Line 2 TCE flux (kg/d)	Line 3 TCE flux (kg/d)	Line 4 TCE flux (kg/d)
6/30/2002	0.1768	0.2958	88.1702	12.2849
12/30/2002	0.1753	0.2654	74.9721	7.1424
6/30/2003	0.1562	0.3304	77.7014	11.5882
12/30/2003	0.1144	0.4133	95.0114	12.0773
6/30/2004	0.1927	0.4643	55.4062	5.2053
12/30/2004	0.1795	0.4367	64.9080	3.8037
6/30/2005	0.1366	0.4524	44.6181	2.9332
12/30/2005	0.1336	0.4694	44.5001	2.6134
6/30/2006	0.2210	0.6532	43.2688	2.4718
12/30/2006	0.2210	0.6303	39.0943	2.5631
12/30/2007	0.2161	0.6878	31.7309	1.6118
Average mass flux	0.13	0.26	53.07	3.72
Mass flux percentage	0.2345%	0.4556%	92.8051%	6.5048%

The figures below compare the results for each mass flux line for the original data versus the modified data.

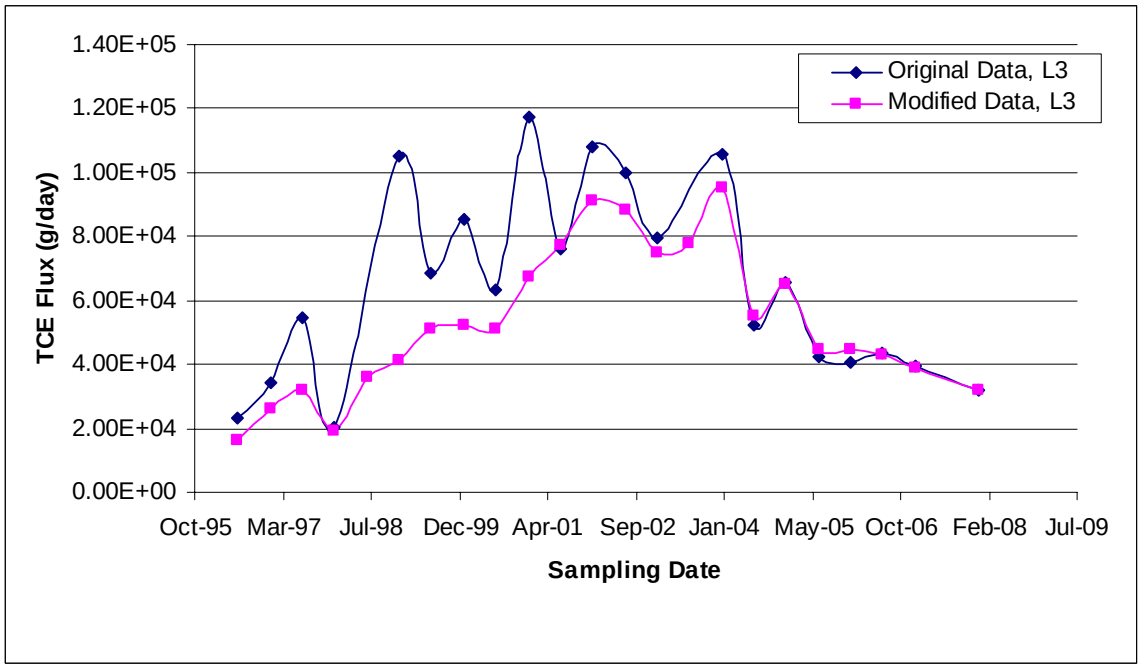
Mass flux across Line 1 (L1), Original Data versus Modified Data



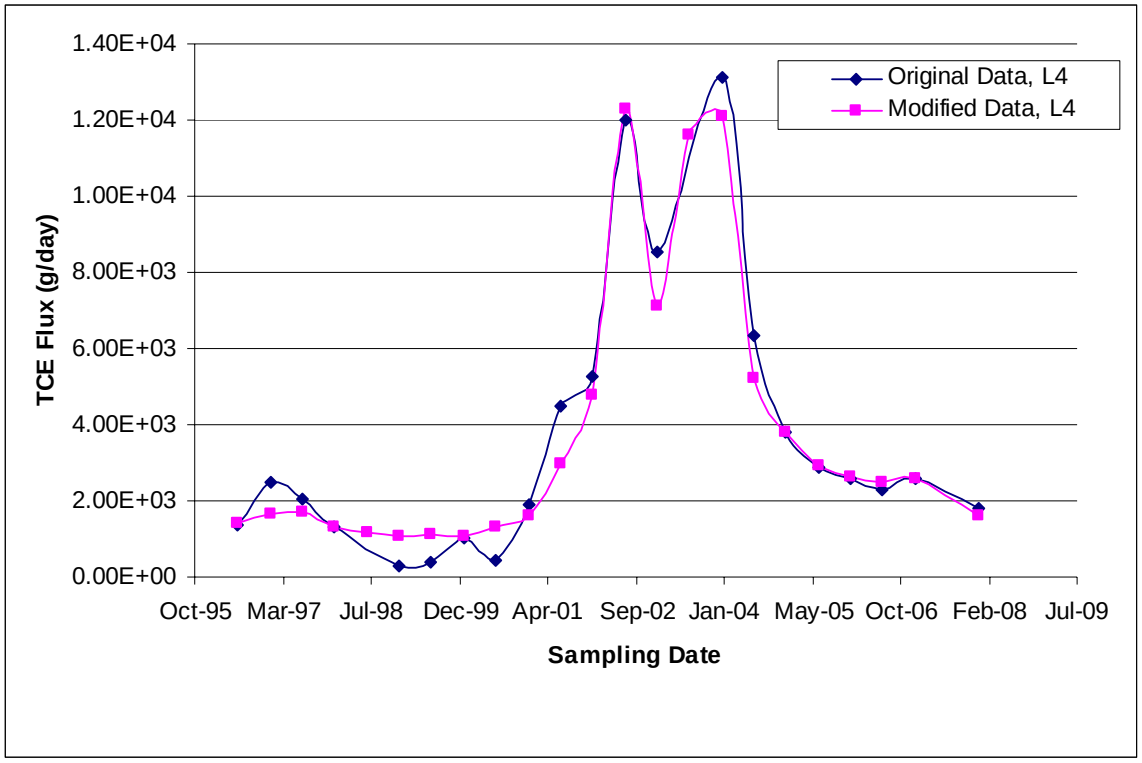
Mass flux across Line 2 (L2), Original Data versus Modified Data



Mass flux across Line 3 (L3), Original Data versus Modified Data



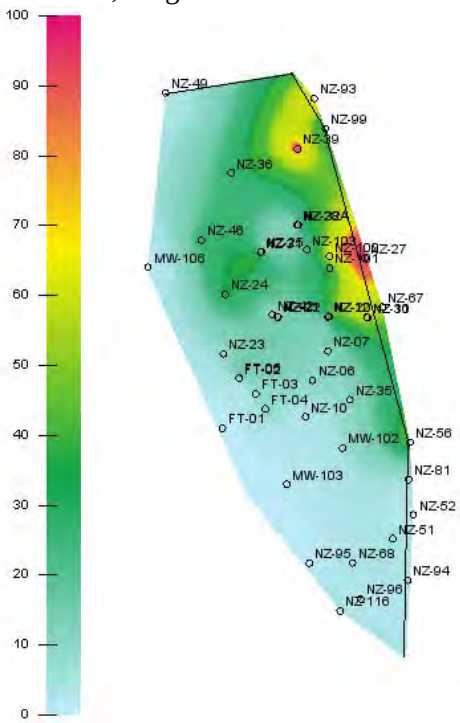
Mass flux across Line 4 (L4), Original Data versus Modified Data



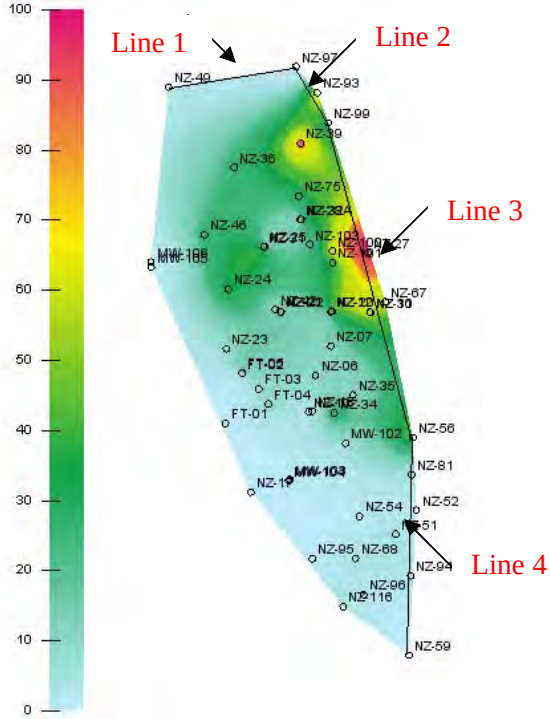
The above time-series plots of mass flux for each line illustrate that the mass flux results for modified data (i.e., filled-in data) are generally more stable than for the original data; that is, the mass flux trends look smoother without abrupt changes between adjacent events. Moreover, the estimated mass flux for original data is generally greater than for the modified data. It is clear by comparing the plume maps provided below for selected sampling events that the original data result in overestimated concentrations in some areas.

{this gap is intentional}

10/2007, Original data



10/2007, Modified data



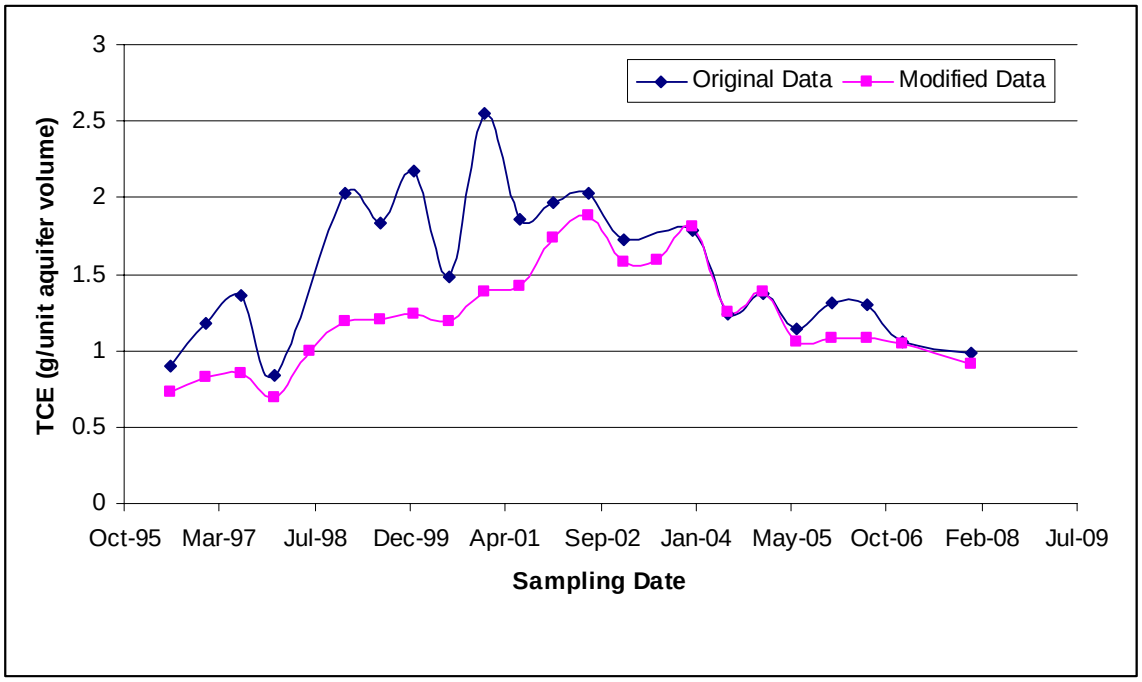
By filling in the missing data that overestimation is constrained (i.e., lower values were assigned using temporal interpolation/extrapolation manually versus higher values interpolated or extrapolated spatially by the software within each event). For example, well NZ-97 on Line 2 has historically low concentrations (0.05 µg/l), but was not sampled in 2007. The interpolated concentrations based on the original data were estimated higher than actual because of the constraint of NZ-97, thus resulting in a greater mass-flux along Line 2 than the modified data, as illustrated in the mass-flux figure above on Line 2. In this case, using only the actual data, the software assigned higher values in those areas for events where samples were not taken, based on spatial extrapolation of other locations sampled in those periods. However, when data were filled in manually for events where samples were not taken, lower values were assigned. This additional information prevented the software from overestimating the values in those areas.

The results for the mass metric, based on the same time-varying data utilized for the spatio-temporal analysis but with incomplete data filled in, are compared to the results for the actual dataset on the figure below.

Date	TCE (g)
6/30/1996	0.732834822
12/30/1996	0.821937321
6/30/1997	0.849770382
12/30/1997	0.691162667
6/30/1998	0.992593746
12/30/1998	1.189908831
6/30/1999	1.200945565
12/30/1999	1.242038928

Date	TCE (g)
6/30/2000	1.1894639
12/30/2000	1.380838196
6/30/2001	1.418434523
12/30/2001	1.735933797
6/30/2002	1.880398521
12/30/2002	1.573330692
6/30/2003	1.585405877
12/30/2003	1.809658984
6/30/2004	1.245283446
12/30/2004	1.38453728
6/30/2005	1.057709417
12/30/2005	1.086839071
6/30/2006	1.086016933
12/30/2006	1.043346684
12/30/2007	0.905839196

Mass Metric versus Time, Original Data versus Modified Data



From the figure above, the mass metric plot of modified data again looks steadier and its TCE mass is lower than original data, due to the reasons aforementioned. Moreover, the TCE mass per unit aquifer volume has been decreasing since year 2004. This finding is consistent with the result from mass-flux calculation for the eastern boundary (Line 3) of the Site, where the majority of mass flux occurs.

Suggested Approach for Utilizing our Results

The following conceptual approach for utilizing our results is suggested:

- Focus on spatial optimization results rather than spatio-temporal results, which appear to be overly conservative due to inconsistent sampling locations for different events

- Determine if eliminated well locations in any of the recommended plans (from the spatial analysis results) are reasonable and acceptable

- Frequency of sampling at remaining locations is not provided by spatial optimization results – it should be determined based on locations where changes in concentration are expected and/or are of greatest concern

- For wells not sampled in a specific event, develop rules for estimating the values at those locations for developing plume maps and/or mass calculations (e.g., latest value, moving average of latest values, etc.)

APPENDIX C:

SITE-SPECIFIC ANALYSIS – NOP SITE

Application of Summit Tools to LTMO Optimization Write-Up for ESTCP NOP Site, Nebraska

(also referred to as the "Mead Site")

Sections

Background Site Information **2**

LTMO Software Demonstrated **6**

Optimization Formulation **7**

Data Preparation **10**
Other Data Preparation Observations 14

Sampling Optimizer (Including Model Builder) **14**
Brief Overview of Functionality 14
Observations Regarding Use of Sampling Optimizer Including Model Builder 17
Computation Time for Model Builder and Optimizer 19
Spatial Analysis Results 20
Spatio-Temporal Analysis Results 35
Relative Uncertainty Analysis 37

Data Tracker Results (Reported by Charles Davis, EnviroStat) **39**
Overview and Summary 39
Using DT 39
Testing DT using the Mead data 43
DT- Discussion 48

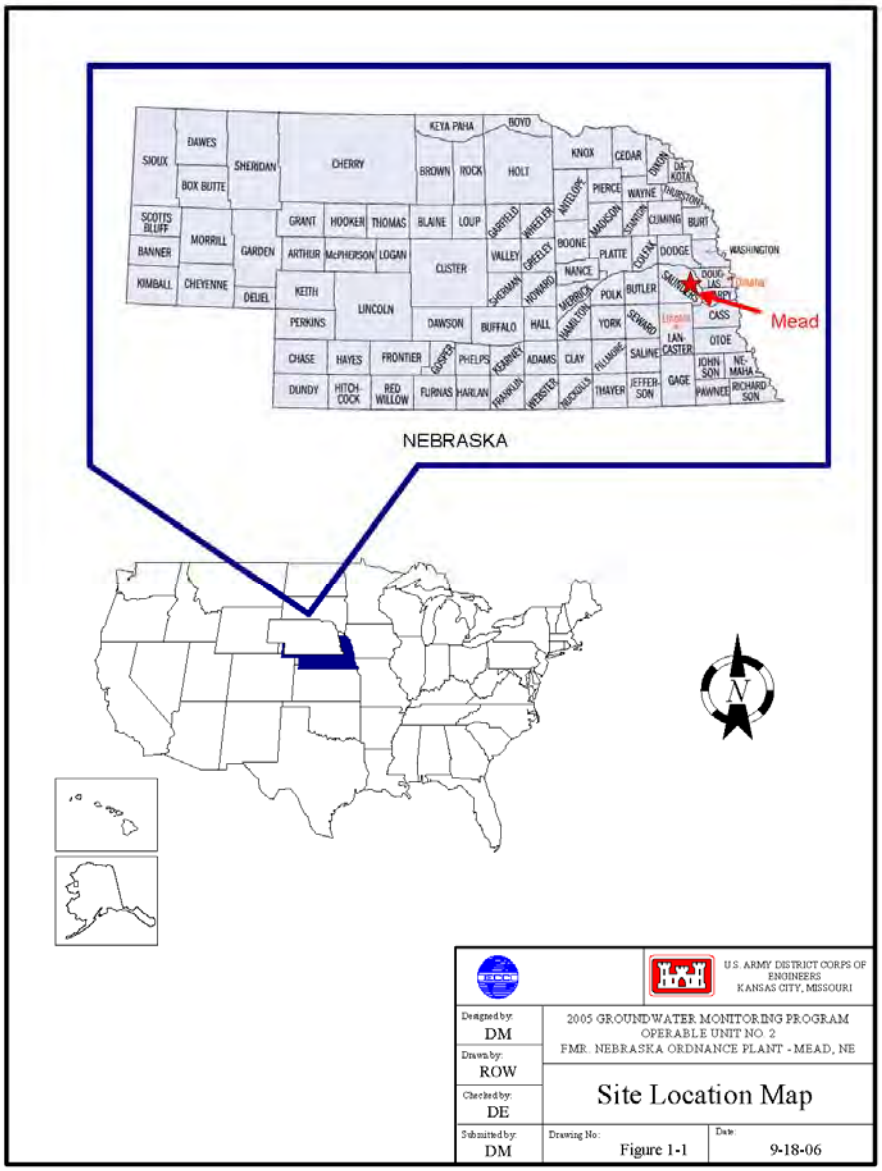
Mass Metric and Mass Flux Results **49**

Model Validation **49**
Model Validation for Spatial Analysis 50

Background Site Information

The former Nebraska Ordnance Plant (NOP) (the Site) occupies approximately 17,250 acres (27 square miles) located 0.5 miles south of the town Mead, Saunders County, Nebraska. We refer to this as the "NOP site" or the "Mead Site".

Location of Former Nebraska Ordnance Plant, Mead, NE



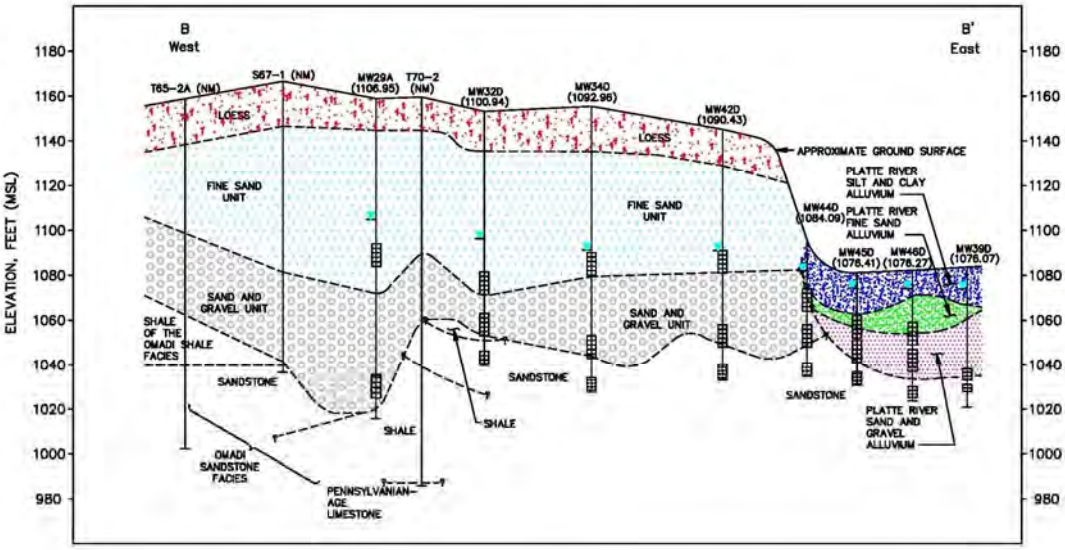
The Site is nearly flat, with a few gentle slopes. There are numerous surface water features in the vicinity of the site. During World War II and the Korean Conflict, bombs, shells, and rockets were assembled at the site. The site includes four load lines (LL1 is furthest west and LL4 is furthest east) where bombs, shells, and rockets were assembled; the Burning/Proving Grounds; a

Bomb Booster Assemble Area; an Administrative Area; an Air Force Ballistic Missile Division Technical Area; and an Atlas Missile Area. According to previous reports, wastewater from both the load line plant operations and a laundry was discharged into a series of sumps, ditches, and underground pipes. The site was placed on the U.S. Environmental Protection Agency (EPA) National Priorities List (NPL) of Superfund sites in August 1990 because contamination was identified in the groundwater and the soils at the site, and the releases of contamination from this site are considered to be a potential threat to public health, welfare, and the environment.

The former NOP site is located in the Todd Valley, an abandoned alluvial valley of the ancestral Platte River. The thickness of the unconsolidated material above bedrock in the Todd Valley at the site ranges from approximately 81-157 feet. The unconsolidated material consists of topsoil, loess, sand, and gravel of Pleistocene age. The uppermost bedrock unit is the Omadi Shale in the northwest and the Omadi Sandstone in the southeast portions of the site.

Three aquifers are present at the site: the Omadi Sandstone aquifer, the Todd Valley aquifer, and the Platte River alluvial aquifer. The Todd Valley aquifer is the first aquifer beneath the site. Towards the Platte River (i.e., towards the east) it grades horizontally into the Platte River alluvial aquifer. The Omadi Sandstone underlies these aquifers, and is part of the bedrock. In places, the Omadi Shale aquitard separates the deeper Omadi Sandstone aquifer from the overlying aquifer(s). Where the Omadi Shale is absent, the Todd Valley aquifer and the Platte River alluvial aquifer are in hydraulic communication with the Omadi Sandstone and behave as a single aquifer without hydraulic barriers. The Pennsylvania Shale aquitard underlies the Omadi Sandstone aquifer.

Aquifers and Aquitards at the Mead Site



Monitoring well locations at the Site were established based on regional groundwater flow (generally towards the south and southeast). The water-bearing portions of the unconsolidated material in the Todd Valley are divided into an upper fine sand unit (12-17 feet thick) and a lower sand and gravel unit (17.5-72 feet thick). The upper sand unit is overlain by 4-23 feet of Peoria Loess. The unconsolidated material in the Platte River Valley (i.e., in the immediate vicinity of

the Platte River) ranges in the thickness from 39 to 49 feet. Overbank silts and clays ranging from 10-17 feet thick overlie the Platte River alluvial sands and gravels.

The water table surface of the Todd Valley slopes toward the south-southeast with depths to groundwater in the Todd Valley ranging from 6.6 feet to 58.0 feet. A local zone of groundwater discharge is located along the western side of the Platte River floodplain in the southeastern portion of the Site. East of Johnson Creek, the water table surface of the Platte River alluvial aquifer slopes to the south, paralleling the Platte River Valley with depths to groundwater in the Platte Valley ranging from 0.0-10.2 feet.

The following VOCs and explosive compounds were identified at the site (primary COCs are indicated with a "*"):

VOCs:

- Trichloroethene (TCE)*
- Methylene chloride;
- 1,2-dichloropropane; and

Explosive compounds:

- Hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX)*
- 1,3,5-trinitrobenzene (TNB)
- 2,4,6- trinitrotoluene (TNT)
- 2,4-dinitrotoluene (2,4-DNT); and

The site generally distinguishes the plumes based on TCE and RDX. The four plumes of groundwater contamination identified at the Site are:

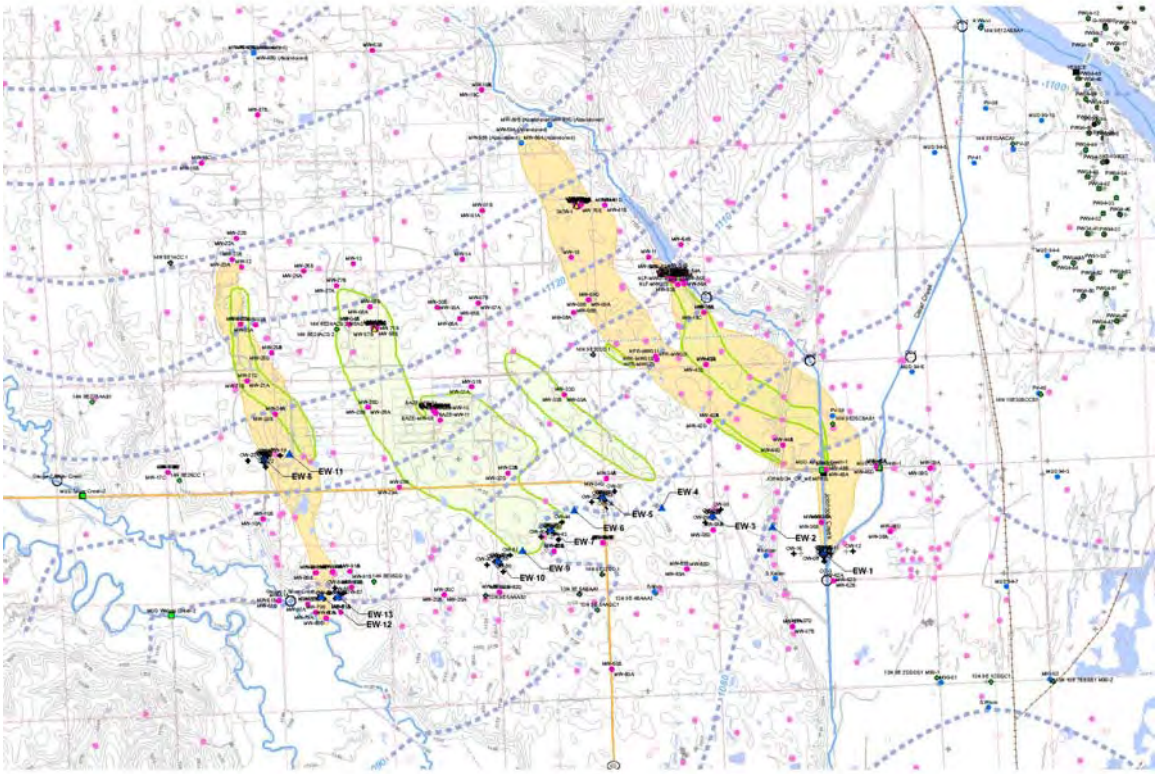
- TCE plume with the suspected source from the Atlas Missile Area, which is north of the eastern load lines (LL3 and LL4);
- TCE plume with the suspected source from Load Line 1 (LL1);
- RDX plumes with the suspected sources from LL1, LL2, LL3, and LL4.

According to site reports, the migration of these contaminant plumes is dictated primarily by the southeastward direction of the groundwater flow. The TCE and RDX plumes overlap in two areas: LL1 and LL4. The overlap at LL4 is due to migration of TCE from the Atlas Missile Area. Higher groundwater contamination is found in the upper fine sand units than in the sand and gravel units below. Generally, lower contamination is found in the deepest of the three aquifers (the Omadi Sandstone aquifer).

{this gap is intentional}

The figure below illustrates the four plumes described above. LL1 is furthest to the left and LL4 is furthest to the right.

TCE (orange) and RDX (green) Plume Map at the Mead Site



Groundwater extraction and treatment has been operated and improved since 2002 to prevent further migration to the south and east. The system consists of a network of thirteen extraction wells generally located at the downgradient portion of each plume (except two recovery wells in the middle portion of the LL1 plume, which have their own treatment plant). Two groundwater circulation wells are also in operation to mitigate contamination hot spots (LL2 and LL4). Alternate water supply is provided to residents whose domestic wells are contaminated from military activities at the site.

Excluding groundwater extraction wells associated with the groundwater remedy, there are generally three types of wells at the Site: residential, irrigation, and monitoring. According to the 2005 Groundwater Monitoring Program Annual Report, the groundwater Monitoring Well (MW) network for the Site includes 170 MWs (69 shallow wells, 66 intermediate wells, and 35 deep wells). The depths of the shallow MWs at the Site range from 17.4 feet below ground surface (bgs) to 98.5 ft bgs in the Todd Valley, and from 26.5 feet bgs to 31.8 feet bgs in the Platte River alluvial valley. Intermediate MWs range in depth from 49.3 feet bgs to 156 feet bgs in the Todd Valley, and from 38.5 feet bgs to 50.9 feet bgs in the Platte River alluvial valley. Deep MWs at the Site range in depth from 77.1 feet bgs to 168.8 feet bgs in the Todd Valley, and from 48.5 feet bgs in the Platte River alluvial valley.

The groundwater sampling is conducted semi-annually with additional sampling occurring sporadically in some years. Site-wide groundwater monitoring started in 1992. No groundwater

sampling was conducted between the June 1993 and September 1994 events, or between the December 2000 and January 2003 events. An additional 70 monitoring wells were installed in multi-depth clusters in late 2006 along the eastern edge and the southern edge of the site (i.e., the downgradient edge of the site) with the first sampling scheduled in January 2007. In addition, monitoring at numerous water supply wells across the site has generally been performed quarterly to semi-annually since December 1993.

LTMO Software Demonstrated

The goal of this project is to demonstrate and validate the use of the Sampling Optimizer tools (Summit Envirosolutions, Inc.) to reduce cost and improve effectiveness of long term monitoring at DoD sites. Two major modules comprise the Summit software: Sampling Optimizer and Data Tracker.

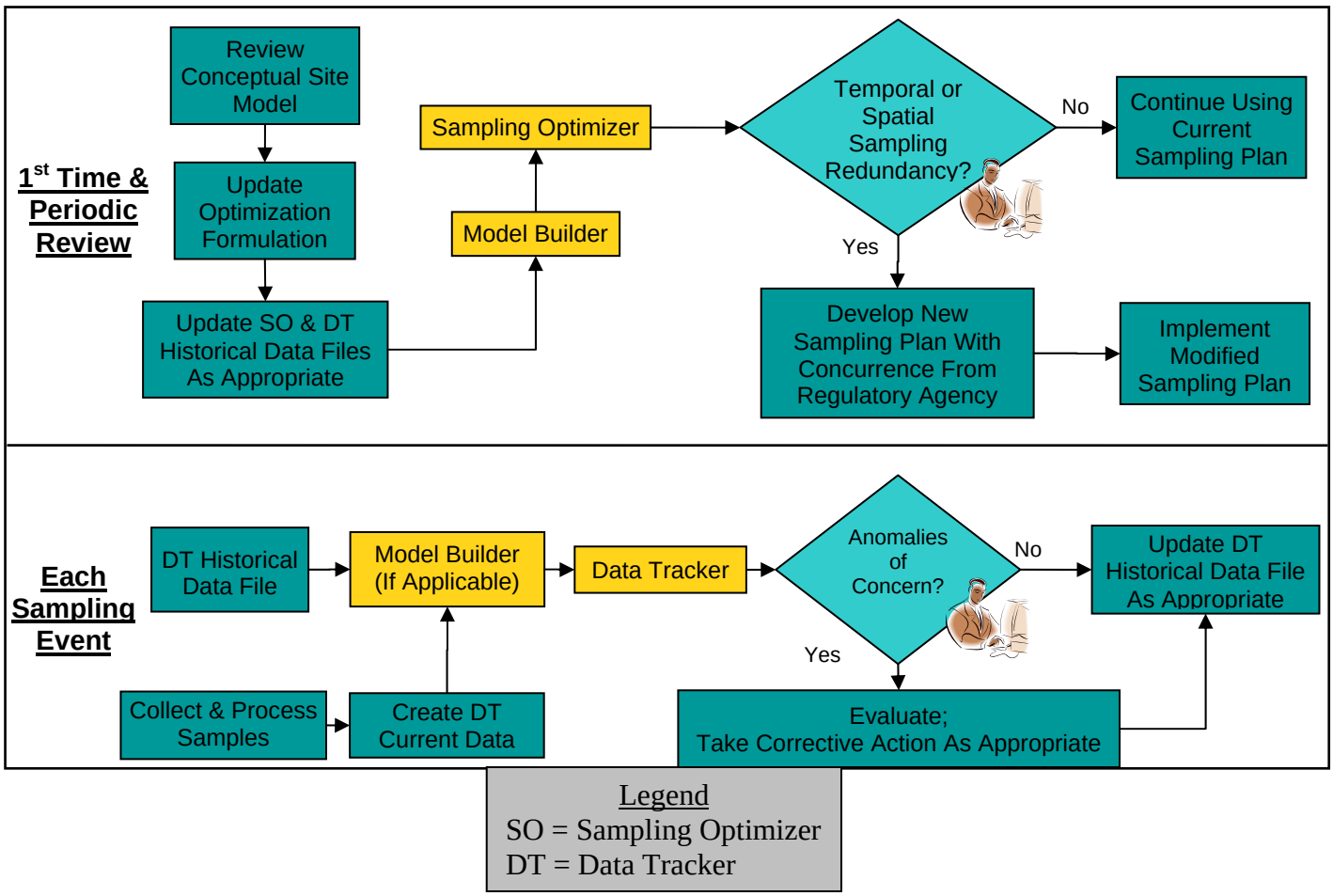
- Sampling Optimizer identifies redundant sampling locations and/or frequencies in historical data. This module identifies redundant sampling locations and/or frequencies using a multi-objective genetic algorithm to obtain monitoring designs that represent optimal tradeoffs among two or more monitoring objectives, such as minimizing the number of samples and minimizing the concentration error (i.e., error typically increases as number of wells decreases, resulting in a tradeoff).

- Data Tracker allows current monitoring data to be reviewed against historical data to identify cases where current data deviate from expectations that are based on the historical values and patterns.

Model Builder is an additional component within the software that is utilized by the Sampling Optimizer and, in some cases, by Data Tracker. Model Builder has two sections: one for model fitting, visualization, and analysis (with kriging or inverse distance weighting), and another for visualizing relative uncertainty. A general flowchart of the software modules is presented below.

{this gap is intentional}

General Flowchart of Software Modules



Optimization Formulation

A site visit was conducted by the ESTCP project team on January 29, 2007, and a meeting with site personnel followed the next day. A subsequent site tour was conducted by Dave Becker from the US Army Corps of Engineers (part of the ESTCP project team) on February 5, 2008 as part of a meeting associated with a different LTMO project. Based on these two meetings, LTMO objectives and constraints of the Mead site include:

- The plumes of TCE and RDX associated with LL1 were originally not going to be included as part of the ESCTP project evaluation (due to recent changes to the remedy two new extraction wells started operation in 2006), but the other LTMO project wants to consider LL1 so we will include it to enhance comparability.

- The site personnel are more interested in the accuracy of the concentrations at leading edge of the plumes and less interested in the concentration accuracy in the interior of the plumes.

- The site personnel are interested in tracking plume containment, if possible, and also have interest in estimating contaminant mass and size of plume footprint.

- The site personnel are interested in performing optimization analysis for all three layers (shallow, intermediate, and deep), and the different vertical intervals will be analyzed separately (i.e., a 2.5 dimensional analysis).

- TCE and RDX will be used for Model Builder and Sampling Optimizer, and all seven COCs will be used for Data Tracker.

- There are some "sentinel wells" and new perimeter wells that will be defined as wells that cannot be entirely eliminated from the monitoring program, but their sampling frequency should be evaluated using Sampling Optimizer. These wells should also be checked for continued "non-detects" with Data Tracker.

- For the plumes that will be optimized, monitoring wells north of the extraction wells are intended for early-warning because the main treatment plant is designed for very low influent concentrations. These wells should not be entirely eliminated from the monitoring program, but their sampling frequencies should be evaluated using Sampling Optimizer.

- Water supply wells may be used for Data Tracker but will not be used for Model Builder or Sampling Optimizer (these wells were ultimately omitted from the DT evaluation).

- The "baseline" data for the ESTCP project will be data obtained through (i.e., including) Spring 2007, which allows many new wells first sampled in Spring 2007 to be included in the baseline spatial analysis. The September 2007 sampling event, and possibly the Spring 2008 sampling event, will be reserved for the "validation" analysis.

- Both the GTS and Summit tools will be applied to the Mead site by Dave Becker (external to this ESTCP project).

- MWs east of the plume must have at least annual sampling frequency, but may not be the same group of wells every year.

As stated in the 1997 EPA Superfund Record of Decision (ROD), the overall OU2 (i.e., groundwater) RAOs are

- Minimize the potential for ingestion of contaminated groundwater, or reduce concentrations to acceptable health-base levels

- Minimize the potential for dermal exposure to contaminated groundwater, or reduce concentrations to acceptable health-based levels

- Minimize the potential for inhalation of chemicals released during the use of contaminated groundwater, or reduce concentrations to acceptable health-based levels

The objectives of the site-wide groundwater monitoring program are as follows:

- To monitor and evaluate potential changes in concentrations of COCs defined in the OU2 ROD and the migration of contaminated groundwater;

- To provide data to evaluate whether the contaminant plume(s) are being contained by the groundwater extraction network; and

- To evaluate the performance of the remedy relative to the requirements outlined in the ROD signed in 1997.

The groundwater target cleanup goals for COCs detected in groundwater at the Site are outlined in the below table.

Chemical of Concern	Target Cleanup Goal (µg/l)
VOCs	
Trichloroethene (TCE) Methylene chloride(MCl) 1,2-dichloropropane (DCP)	5 5 5
Explosive Compounds	
Hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) 1,3,5-trinitrobenzene (TNB) 2,4,6- trinitrotoluene (TNT) 2,4-dinitrotoluene (2,4-DNT)	2 0.778 2 1.24

Source: Woodward Clyde, Mead ROD, 1996

The Optimizer provides "tradeoff curves" associated with two competing objectives:

- Minimize the number of sampling points and (optionally) the sampling frequency at retained sampling points; and

- Minimize the maximum concentration errors (emphasizing the importance of errors near the plume boundary) that result from removing specific sampling locations.

The error function built into the objective function in the software accounts for the different significance of errors near the plume boundary versus the plume interior, placing greater emphasis on the significance of errors near the plume boundary. The user defines a cutoff concentration between low values (i.e., plume boundary) and high values (i.e., plume interior). For locations where the actual value is below the cutoff, the error is calculated as the difference between actual and estimated concentration, divided by the "acceptable error for low concentrations" for the specific parameter. The "acceptable error for low concentrations" is frequently assigned as the MCL. For locations where the actual value is above the cutoff, the error is calculated as the difference between actual and estimated concentration, divided by a specified percentage of the actual value. Details are provided in the software documentation. This function scales the difference between observed and estimated values by a different amount

for different contaminants in the plume boundary area (e.g., by the MCL for each contaminant), and also diminishes the importance of errors in high concentration regions.

The following acceptable error levels and cutoff values are used for monitoring optimizations:

	TCE (µg/l)	**RDX (µg/l)**
Acceptable error level for low concentrations	5	2
Cutoff value between low and high concentrations	25	10
Acceptable percentage error for high concentrations*	20%	20%

**percentage chosen such that there is continuity of the error function at the cutoff value*

The "Plume Interior" and "Plume Boundary" is defined by the cutoff values between low and high concentrations as listed in the above table; that is concentrations above the cutoff value are considered to be plume interior, on the other hand concentrations below the cutoff are assigned to plume boundary. A new feature of Model Builder added by Summit prior to the evaluation of this site is that different objective function parameters can be specified separately for user-defined "interior wells" and "exterior wells"; this modification can be used to explicitly categorize wells in the software as either plume "interior well" or "exterior well".

Site personnel defined specific wells as "non-removable" for the purposes of sentinel and early-warning. A list of 73 non-removable wells is presented below:

List of Wells That Cannot Be Removed

MW-100A	MW-108A	MW-115D	MW-82D
MW-100B	MW-108B	MW-116A	MW-87A
MW-100D	MW-108D	MW-116B	MW-87B
MW-101A	MW-10A	MW-116D	MW-87D
MW-101B	MW-10B	MW-20A	MW-88A
MW-101D	MW-110A	MW-20B	MW-88B
MW-102A	MW-110B	MW-20C	MW-88D
MW-102B	MW-110D	MW-41A	MW-95A
MW-102D	MW-112A	MW-41B	MW-95B
MW-103A	MW-112B	MW-41D	MW-95D
MW-103B	MW-113A	MW-46A	MW-96A
MW-103D	MW-113B	MW-46B	MW-96B
MW-106A	MW-113D	MW-46D	MW-96D
MW-106B	MW-114A	MW-62A	MW-97A
MW-106D	MW-114B	MW-62B	MW-97B
MW-107A	MW-114D	MW-62D	MW-97D
MW-107B	MW-115A	MW-82A	MW-98A
MW-107D	MW-115B	MW-82B	MW-98B
			MW-98D

Data Preparation

Data were received from site personnel in Microsoft EXCEL files containing the following information:

- Well details including ID, northing and easting, measuring point elevation and top and bottom screen depths relative to that measuring point;

- Historical water level data for June 1992-April 2008; and

- Chemical data for August 1992-April 2008.

Chemical data were for the following constituents:

- TCE – Trichloroethylene
- DCP – 1,2-Dichloropropane
- MCl – Methylene Chloride
- RDX - Cyclotrimethylene trinitramine
- DNT – Dinitrotoluene
- TNB – Trinitrobenzene
- TNT – Trinitrotoluene

The first three are VOCs and the remaining four are explosives. The primary Constituents of Concern (COCs), which are the focus of the groundwater monitoring optimization effort, are TCE and RDX. The remaining COCs are tracked, but not involved in the optimization effort.

Data from both monitoring wells and water supply wells were given to the team. Water supply wells are located throughout the site itself and on the periphery; the latter include a number of domestic water supply wells. In the course of preparing data for use in the project the decision was made to exclude the water supply well data. The reasons for that decision involve data comparability and the completeness of data and well construction information (e.g., geographical coordinates and screen depth). For the spatial optimization and data tracker demonstrations the number of monitoring wells was judged to be adequate.

Data were supplied for three strata: shallow, intermediate, and deep. Several other reports were also supplied. The data files were complete and consistent; few problems such as differing well IDs in different documents, missing coordinates, or missing screen depths, etc. were found, and those problems occurred only in wells with only one historical sampling event present. The COC data were in a standard format (one row per constituent per well per event). EnviroStat used this data file to prepare CSV files to be used as input to the Summit Monitoring Tools as well as other information displays for use by the project team.

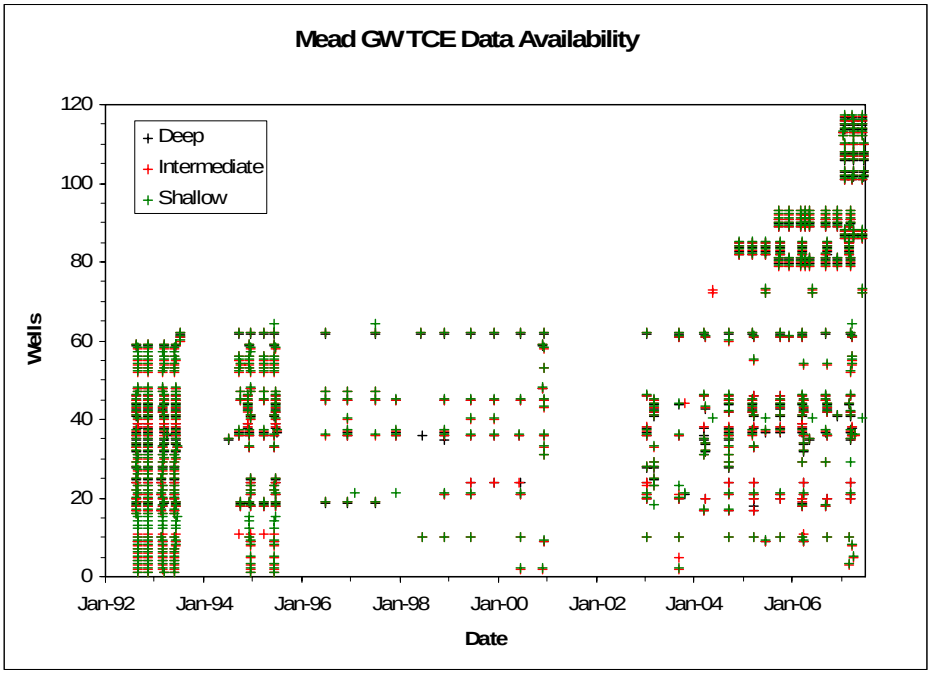
First, a master EXCEL file containing several worksheets was created. The COCs were arranged into columns keyed to a common well and sampling date. Separate worksheets were created for the “historical” data (1992 – June 2007) and the “current” data (July 2007 – April 2008). The “current” data were reserved for the “validation” analysis as well as “current data” for the DT analysis. Worksheets were created for the well location and depth information and for the water level histories.

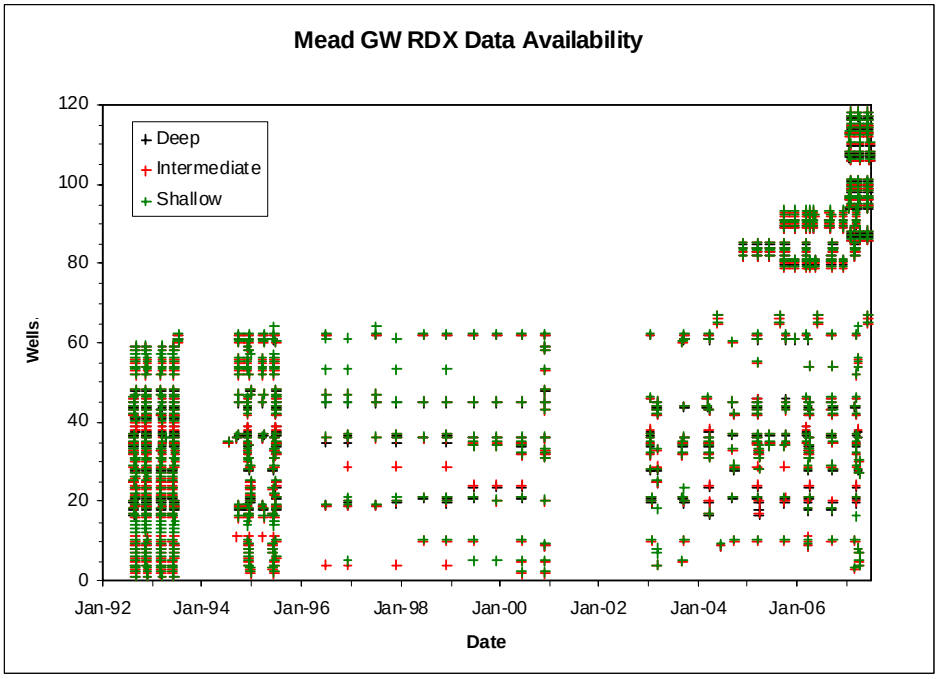
Summaries of data availability were prepared and included in the master file, as were Time Series Plots (TSPlots) of TCE and RDX data at wells with “interesting” histories. TSPlots of water level and plots of well screen elevations were included as well. The master data file, as well as other files, contains columns designating the stratum for each well, since the well ID system is somewhat inconsistent and non-intuitive in this regard, and also assigns each well either to Load

Line 1 (LL1) or the rest of the site (Main), as there had been some discussion about possibly considering the two portions of the site separately.

Data were reviewed for inclusion or exclusion with regard to use for by Sampling Optimizer (SO). The issue here was that many wells had been sampled frequently during the mid-1990s, then perhaps again in 2000, and then only sporadically, as shown in the following two plots.

.





Some wells had latest sampling dates as early as 1993. Ultimately all wells with northing and easting coordinates were included in the dataset used for SO; the wells with no recent data had almost entirely nondetect (ND) data when sampled, with a few low-level exceptions, the highest being a measurement of 18 µg/L for TCE.A version of the master file excluding the current COC and water level data was supplied to GeoTrans and the rest of the project team.

The master file was then used to prepare a CSV file of historical data for SO. The latest available data value from each well was used. Some further minor data massaging was needed at this step. Field duplicate values were averaged with regular sample values, with some adjustment where one value was a nondetect (ND) and one not or where the two values were NDs with different reporting limits. NDs with typical reporting limits were replaced by a low "graphing value" (0.05 µg/L for TCE and 0.005 µg/L for RDX). NDs with elevated reporting limits (RLs) were omitted. The three wells where the latest sampling date was prior to 2004 were omitted from all historical datasets. Only TCE and RDX were included in the SO CSV file.

After careful examination of data patterns the project team decided not to attempt to perform a spatiotemporal optimization (SO-st). The reason is the irregularity and sparseness of the recent data, as shown in the two plots on the preceding page. SO-st implicitly assumes that the datasets for successive events are comparable. To achieve that, however, one would have to interpolate artificial data to a great extent within individual COC/well combinations. This was done with the GAFB data for the purposes of using DT for mass metric tracking, but the Mead data are much more incomplete than those of GAFB. The inclusion of such a large amount of interpolated (essentially manufactured) data would be inappropriate for SO-st, since the interpolated data could not be reliably known to present the actual temporal characteristics of actual data. Similarly, the project team declined to use DT in the mass-metric mode, since that would similarly need large amounts of interpolated data to provide comparable datasets for successive events.

For DT the data preparation effort involved first identifying an appropriate treatment for the NDs. The Reporting Limits (RLs) varied considerably in the data, even for the primary COCs TCE and RDX. For the secondary COCs (DCP, MCl, DNT, TNB, and TNT) there were also NDs with elevated RLs, which occurred when samples were diluted because of high concentrations of their associated primary COCs. In order to accomplish this, the distributions of RLs was evaluated for VOCs and for explosives; these are described elsewhere. Common values were assigned as substitutes for the ND data: 0.5 µg/L for VOCs and 0.1 µg/L for explosives, which are round numbers close to the medians of the RLs that were reported.

Low NDs were replaced with those common values in most cases. NDs with higher RLs were generally excluded from the datasets, except when that would leave fewer than the four background data values required by DT. In those cases half the RL was generally used as a substitute value. Similarly, for many COC/well combinations there were only three rather than the requisite four background observations. In these cases an artificial data value and date were supplied. There were generally ND values, with some low-level values as appropriate depending on the data history of the particular COC/well combination. In two cases higher values were supplied; those higher values are conjectures based on similar patterns in nearby wells.

Finally, the data history for every COC/well combination was reviewed toward the end of identifying the appropriate subset of the historical data to use as the background data. This was a labor-intensive process due to the time gaps in the data and apparent changes in conditions, slugs

of contaminant passing through, and in some cases increasing values up through the end of the background data period. Substantive issues surface in the background data determination; these are discussed, with examples, in DT discussion later in this report.

Other Data Preparation Observations

The GeoTrans team provides the following additional observations regarding data preparation:

- The sampling data have to be in a CSV file to import into the software. For Model Builder the format is "Date, SiteID, EastCoordinate, NorthCoordinate, COC1, COC2, …", where SiteID is the well identification For data tracker the east and north coordinates are optional, and the format of the CSV data is "Date, SiteID, COC1, COC2, …".

- Concentration units have to be consistent over time for each individual COC. Different COCs can have different concentration units.

- There is no constraint on the exact chemical name (whatever user enters is OK, which is not the case with MAROS).

- No detection limits and flags are utilized or allowed. However, a "graphing value" for non-detects needs to be assigned to serve as the concentration value for non-detects in the data that are going to be imported. This is similar to a process in MAROS where user enters "detection limit" for non-detects. The user must decide *a priori* whether to include or delete "high non-detects" (i.e., non-detects with reporting limits higher than actual data values).

- For SO (but not DT) all samples must be assigned to a sampling group in the data to be imported, such that every sample in that sampling event has the same sampling date within the software. For spatio-temporal data, the time lag between two adjacent sampling events has to be at least quarterly frequency, and the software User's Guide provides guidance for sampling event "frequency alignment". This data consolidation is done outside of the software prior to import. This is less flexible than MAROS, in which such data can be treated as separate events or consolidated as part of one "sampling event" defined by the user within the software.

Sampling Optimizer (Including Model Builder)

Brief Overview of Functionality

Sampling Optimizer provides users with six possible combinations of interpolation technique and data transformation. They are

- Two interpolation technique options
    - Inverse Distance Weighting

    - Kriging

- Three data transformation options
    - None (i.e., No transformation)
    - Logarithmic
    - Quantile

This results in six possible combinations for these basic options. Generally a user will only utilize one combination, and Summit suggests using kriging with quantile transformation. For the ESTCP project, GeoTrans tried all of the combinations, and did verify that kriging with quantile transformation provided the most reasonable representation of the plume distribution (discussed in more detail later).

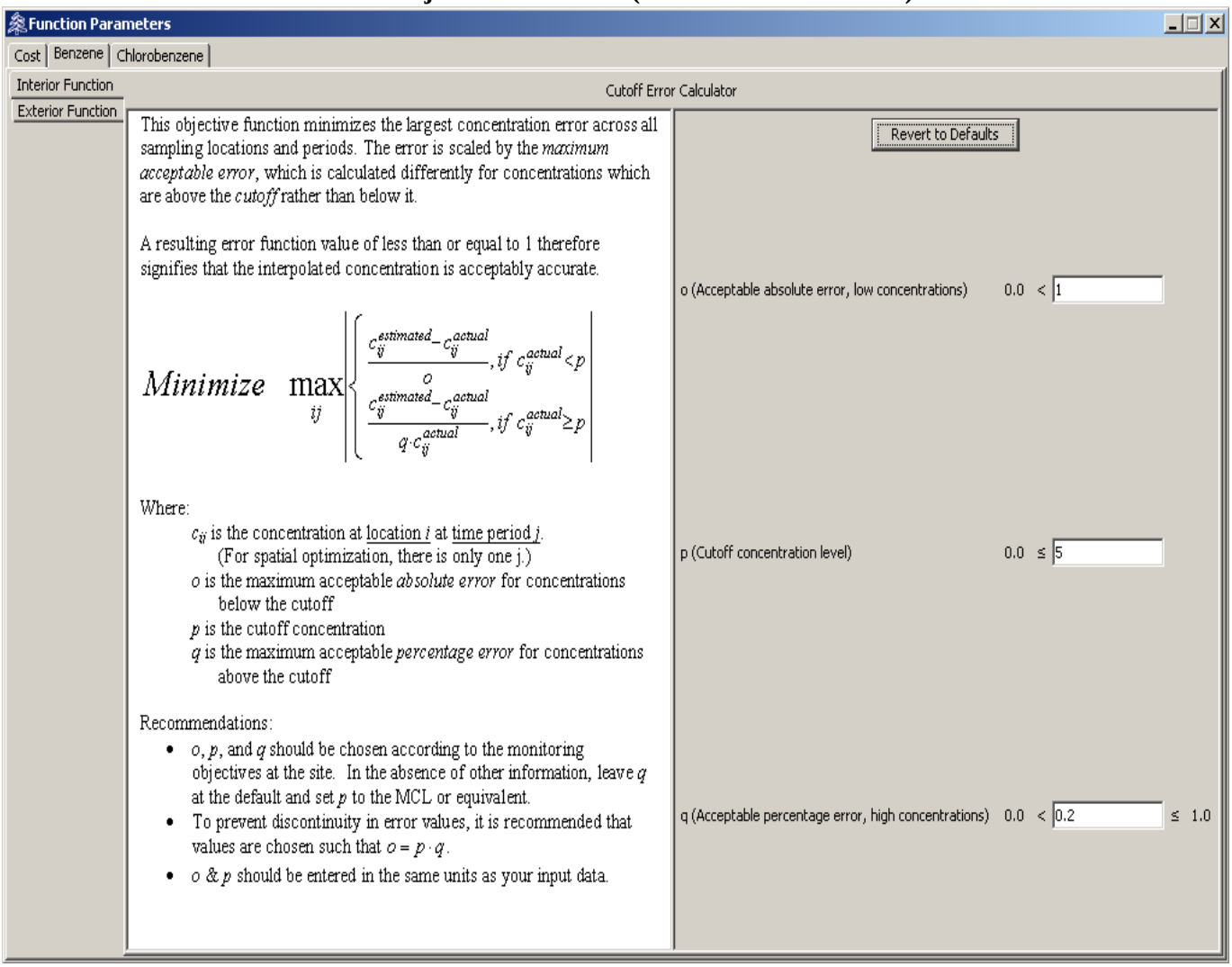
The Model Builder (MB) component of the software provides model fitting, visualization, and analysis functions, as well as maps of relative uncertainty. Within the MB component of the software the user defines options for the parameters of the interpolation technique selected by the user for the Optimizer. Both automated and manual model parameter fitting are supported for Kriging, while the user must manually specify the power to be used for inverse distance weighting. Within the MB component the user also specifies desired changes to the defaults on the "Model Builder Settings" screen, such as the number of vertical slices that defines the resolution of the image. If the data imported into the Sampling Optimizer has multiple events, Model Builder provides visualization for each event. The Optimizer module uses the model parameters specified within the Model Builder component.

Sampling Optimizer uses a genetic optimization algorithm to suggest favorable monitoring plan alternatives relative to the base sampling plan (i.e., where one or more of the samples are removed). In spatial optimization, the original model is based on one set of sampling data that do not vary in time, and the optimization results are with respect to sampling locations only. In spatiotemporal optimization, the original model consists of actual data that vary in space and time, and the optimization results are with respect to sampling location and sampling frequency. Temporal analysis is a subset of spatio-temporal analysis where wells cannot be removed. The user can utilize software defaults for the optimization algorithm (e.g., population size) or can specify values for these parameters in the "GA Settings" screen.

For this site the "errors" were calculated by using the "Cutoff Error Calculator" option for the objective function provided within Optimizer (the other option is the "Percentage Error Calculator"). The Cutoff Error Calculator incorporates a function for calculating error associated with samples that are removed that is displayed below.

{this gap is intentional}

Error Objective Calculator (From Software Manual)



The overall purpose of Error Calculator is to come up with an objective function value that represents the overall similarity of a new sampling plan to the baseline sampling plan. The Cutoff Error Calculator is designed so that "error" is calculated in a manner that makes deviations between interpolated and actual values more significant in areas of low concentration versus areas of high concentration. This is accomplished as follows:

- The user defines a cutoff concentration (p) for the actual data values that differentiates between low concentrations versus high concentrations, and also defines a value for Acceptable absolute error (o).

- When a low concentration data point is removed (i.e., below the cutoff), error is calculated as the absolute value of the actual value minus the interpolated value, divided by the acceptable absolute error. For example, if the actual value is 5 µg/l (i.e., below the cutoff concentration of 10 µg/l) and acceptable absolute error is 1.0, and the difference between the actual and interpolated value is 5 µg/l, then the error would be 5 / 1 = 5.

- When a high concentration data point is removed (i.e., above the cutoff), error is calculated as the absolute value of the actual value minus the interpolated value, divided by a percentage (q) of the actual value, where q is specified by the user. For example, if the actual value is 100 µg/l (i.e., above the cutoff concentration of 10 µg/l) and the percentage input by the user is 10%, and the difference between the actual and interpolated value is 5 µg/l, then the error would be 5 / (0.10 * 100) = 0.5.

In these examples, the difference between the actual value and the interpolated value was 5 µg/l in both cases, but in the first case the calculated error is 5.0 whereas in the second case it is only 0.5. This illustrates how the calculation increases the significance of deviation between actual and interpolated values in the lower concentration areas of the plume.

Observations Regarding Use of Sampling Optimizer Including Model Builder

The following observations were made by GeoTrans based on application of Sampling Optimizer (including Model Builder) in conjunction with spatial analysis:

- The software is very easy to use. However, the post-software analysis of results can take quite some time to analyze the resulting monitoring plans, especially when multiple COCs are present. Two COCs (i.e., TCE, RDX) are included for evaluation of this site.

- The software allows the user to easily save a project and re-open it later. However, some updates to the software made during the project prevented previously saved projects from opening.

- The software allows the user to enter "run titles" which are used as the part of the file names when exporting the results for both Model Builder and Optimizer.

- The software uses a "seed" value for the random number generator used in many calculations. A genetic algorithm does not guarantee an optimal solution, just one that has high probability of being very close to optimal, and a different solution may be obtained if a different seed value is utilized. The software uses default seed values that are fixed for both Model Builder and Optimizer, which ensures that the same results can be obtained by different users with the same parameter settings. The user can change the seed manually if desired (the optimization results will change with different seeds).

- Model Builder provides visualization of plume distribution and plume relative uncertainty. It gives users the option to post either well names or measured concentrations on the visualizations. This is a useful feature, more advanced than MAROS.

- The "visualization resolution" (the user controls this by defining the number of vertical and border slices for the image) has a big impact on whether the plume generated by Model Builder can be correctly displayed. The visualization resolution can be modified in "# of vertical slices for image" and "# of border slices" of "Visualization" settings with Model Builder. We used the default value for # of vertical slices, and we increased the default value of 10 for # of border slices to 50 to create a larger margin at the border.

- For the SO module, the "Well Constraint" feature allows the user to specify the maximum sampling frequency and the minimum sampling frequency for each well. For spatial analysis, this feature allows the user to specify which wells cannot be removed from the system (which may be specified in the optimization formulation). This can also be useful for abandoned wells which may be part of the historical data but cannot be sampled in the future, by specifying such wells as "always off".

- The software includes some new features regarding the objective function: (1) *Function Selection* – this enables the user to specify what calculator the Optimizer is using for calculation. There are currently two calculators available: Cutoff Error Calculator and Percentage Error Calculator; (2) *Location Group Assignments* – the user tells the software what wells are characterized as "Interior" and what are "Exterior". This new feature allows the user additional flexibility to develop the objective function; (3) *Function Parameter* – it allows the user to specify parameters for the Calculator selected in *Function Selection*. The user can also specify the cost per sample, if known; (4) *Combined COC Objectives* – for multiple COCs, the software provides the user options to use combined COC objectives; these include "Maximum Error Across COC's" and "Additive Error Across COC's" options. This is a very helpful feature if there are multiple COCs.

- A "population size" (utilized for the genetic algorithm) of 1,000 for SO was recommended by Summit to ensure that "good" solutions can be found. As a new feature, the software asks the user after each run if he/she wants to seed an additional run with the results from the last run with the population size doubled. We applied this feature to our spatial optimization (will discuss later).

- Plume visualization for both Model Builder and Optimizer also allows users to change the zoom scale and color scale. The color scale is a linear scale allowing users to define the minimum and maximum concentrations for each COC. Then the software can plot the plume maps in color based on the minimum and maximum concentrations defined. However, it does not provide an option for a logarithmic scale, thus, for sites with a very big range in concentration, it cannot plot both high-end concentrations and low-end concentrations with sufficient detail (though multiple plots with different ranges could be made independently).

- The software allows the user to export the following files:

    - The plume maps, uncertainty maps, and variogram charts (for kriging model type only) can be exported as image files (.png files).

    - A tradeoff curve for each COC can be exported as an image file.

    - A file containing, for each optimization sampling plan on the tradeoff curve, which wells are recommended to be "on" or "off", the maximum concentration error for each COC, and the sampling cost (i.e., number of wells which are on) can be exported as a CSV file.

    - For each optimization sampling plan on the tradeoff curve, an individual listing of which wells are on and which wells are off can be exported as a CSV file (for

spatio-temporal analysis the frequencies are also exported).

- For each potential optimization sampling plan on the tradeoff curve (and for the current sampling plan), a plume map can be exported as image file for each COC with symbols indicating which wells are on and which wells are off, with "+" indicating wells that are recommended to be removed from the monitoring network and "o" indicating remaining active wells.

Computation Time for Model Builder and Optimizer

The computation time depends on the size of the dataset (e.g., number of wells), model type selected (e.g., kriging versus inverse distance, plus the type of data transformation), and CPU speed. Computation time for Model Builder also increases significantly with the increase in resolution (i.e., number of vertical slices for image), and computation time increases with increased population size and number of generations for Optimizer. An estimate of the amount of time it takes to apply Model Builder and Optimizer to evaluate the plumes for Mead was as follows:

- EnviroStat spent several days evaluating the data and preparing data files for import into the software and preparing additional preliminary analyses and reports, such as data availability summaries, time series plots (TSPlots) of water elevation, wells with interesting concentration history, etc. The data supplied by Mead were in the standard "one record per value (regular or field duplicate) per well per COC per date" format.

- The data provided by EnviroStat were already in the right format required by the software. Any revisions to these data prior to import only took a few minutes.

- Importing the concentration data into the actual software took seconds.

- Next, it took minutes to enter the facility ID and choose the model type (inverse distance or kriging) and data transformation type (quantile, log, or none).

- For the Site, which had on the order of 80 sampling locations per layer, computation time for visualizing a plume in Model Builder and running Sampling Optimizer was as follows:

{this gap is intentional}

Summary of Model Quality and Computation Time for Each Combination of Interpolation technique and Data Transformation, Spatial Data

Interpolation	**Data Transformation**	**Computation Time[1]**		**Comment Regarding Model Quality**
		Model Builder	Sampling Optimizer	
Inverse Distance Weighting (IDW)	None	5 minutes[3]	10~20 minutes[2]	Bad
	Logarithm	5 minutes[3]	10~20 minutes[2]	Acceptable
	Quantile	5 minutes[3]	10~20 minutes[2]	Acceptable
Kriging	None	5 minutes[3]	15~50 minutes[2]	Bad
	Logarithm	5 minutes[3]	15~50 minutes[2]	Good
	Quantile	5 minutes[3]	15~50 minutes[2]	**Best**

Note: *1. Desktop with Pentium 4, 3.2 GHz PC and 2.0 GB RAM was used for computation*
2. Spatial optimization computation time depends on the dataset size, the number of generations, and population size
3. Time for Model Builder (i.e., plume visualization) depends heavily on the resolution of the image

- After performing Model Builder, it took several minutes to set up the well constraints and objective function for Optimizer.

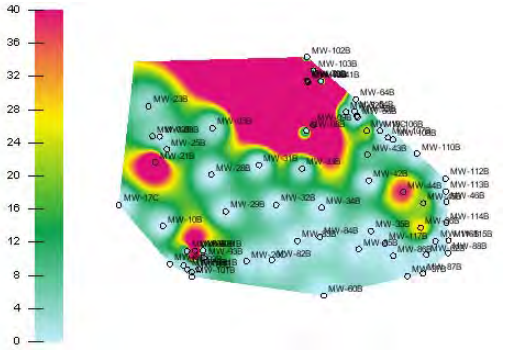
- Spatial optimization with Sampling Optimizer speed took approximately 10-40 minutes depending primarily on the population size and number of generations specified in the software for the genetic algorithm.

- Finally, reviewing plume maps for the potential plans and exporting them to image files took a few minutes.

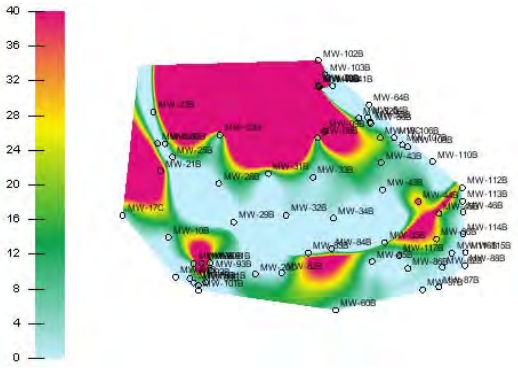
Spatial Analysis Results

To determine which of the six combinations for interpolation and transformation is the best for visualizing plumes at the Site, each was applied to the TCE plume in the shallow aquifer, which has 81 sampling locations. The plume was then visualized in Model Builder for further review and comparison. Plume maps from Model Builder for each of the six combinations are presented below.

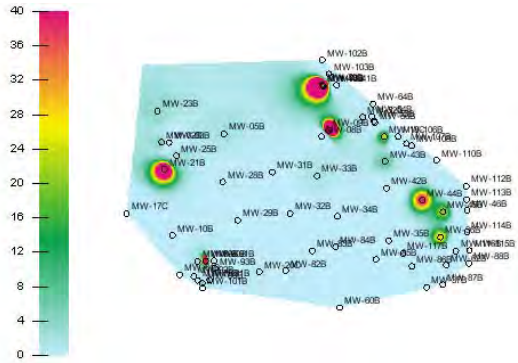
IDW-None



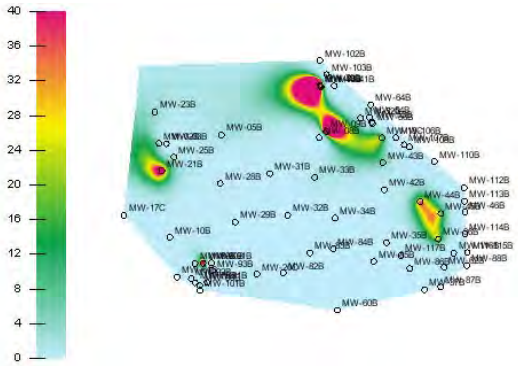
Kriging-None



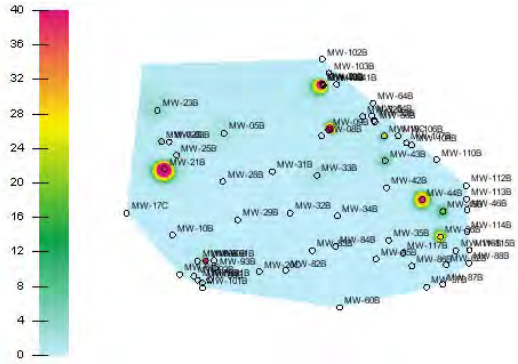
IDW-Logarithm



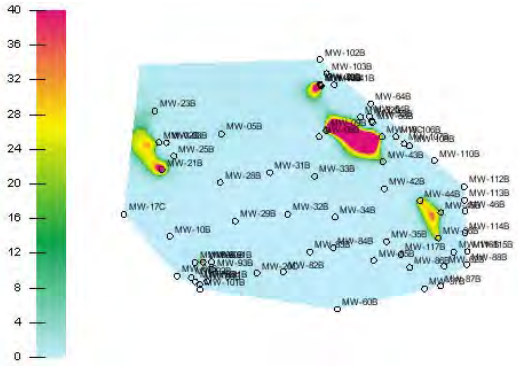
Kriging- Logarithm



IDW-Quantile



Kriging-Quantile



In general, the kriging method with quantile or logarithm transformation generates the best plume representation. Kriging with quantile transformation is recommended by Summit and was also determined to be the best option during analysis of the Camp Allen and George Sites.

Observations made regarding visualizing the generated plume maps are summarized below:

- Of the six combinations of interpolation and data transformation, four generally produced acceptable visual representations of the plume. These are:
    - Kriging with quantile transformation,
    - Kriging with logarithmic transformation,
    - Inverse distance weighting with logarithm transformation, and
    - Inverse distance weighting with quantile transformation

- Both inverse distance weighting with no data transformation, and kriging with no data transformation, resulted in model outputs that are biased to higher concentration, e.g., interpolated concentrations along the northern boundary of the map are much higher than believed to actually be the case. For example, in the figure above labeled IDW-None, concentrations on the northern edge near clean well MW-102B are modeled to be on the order of 40µg/l or more. Additionally, the area of high concentrations (in red) is much larger than with the other transformation options, and is not consistent with how most people would choose to interpret the actual data values.

Our analyst qualitatively ranked the combinations for interpolation technique and data transformation from best to worst, as follows

- Kriging with quantile transformation
- Kriging with logarithm transformation
- Inverse distance weighting with logarithm transformation
- Inverse distance weighting with quantile transformation
- Kriging with no data transformation
- Inverse distance weighting with no data transformation

For consistency with previous analysis for George and by making comparisons to plume maps in Mead site reports, Kriging-Quantile is considered to provide the best representation of plumes at the Site and was chosen for visualizing the plume maps and optimizing monitoring plans.

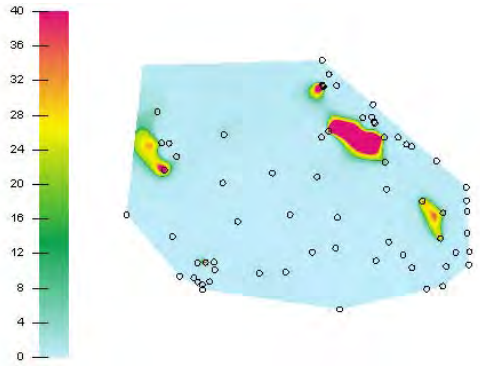
As discussed previously, only TCE and RDX were evaluated for Spatial Optimization (SO). Samples were collected as early as Spring 1993 and through Spring 2007. Data earlier than 2003 were omitted, because we believe that these data were irrelevant to the current plume interpretations. All the sampling dates were assigned as “March 15, 2007” for the spatial analysis, which requires a single sampling date. We also included LL1 in addition to the Main Plume (LL2, LL3, and LL4) in our SO analysis. Furthermore, the three aquifers (i.e., shallow, intermediate, deep) were analyzed separately when performing optimization.

A common question arises whether one should analyze COCs together or independently in cases where there are two or more COCs to be evaluated. To answer this question we attempted a variety of approaches, summarized below.

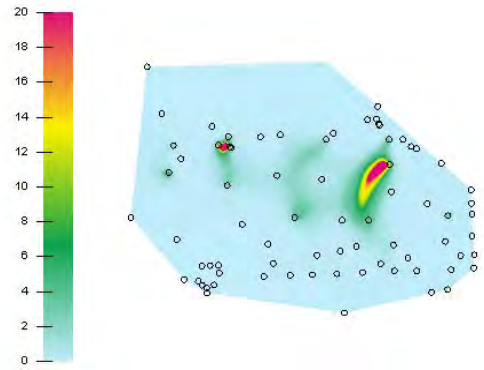
TCE		RDX	
A1	interpolated TCE plume with all sampling data (baseline)	B1	interpolated RDX plume with all sampling data (baseline)
A2	interpolated TCE plume with combined maximum error of TCE/RDX less than 1.0	B2	interpolated RDX plume with combined maximum error of TCE/RDX less than 1.0
A3	interpolated TCE plume with maximum TCE error less than 0.5	B3	interpolated RDX plume with maximum RDX error less than 0.5
A4	interpolated TCE plume with wells recommended to be removed for the plan of maximum RDX error less than 0.5	B4	interpolated RDX plume with wells recommended to be removed for the plan of maximum TCE error less than 0.5
A5	interpolated TCE plume with common wells recommended to be removed by evaluating each COC independently (i.e., separate optimization runs) with maximum error of 0.5 for each COC	B5	interpolated RDX plume with common wells recommended to be removed by evaluating each COC independently (i.e., separate optimization runs) with maximum error of 0.5 for each COC

The resulting plume maps for each of the approaches listed above are presented below (the selected plan in each case is the plan along the tradeoff curve with maximum acceptable error for that optimization run, as described in the table above).

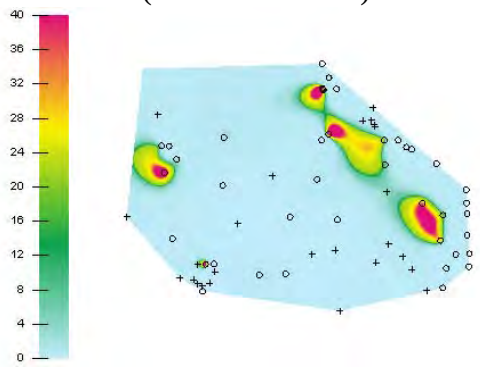
A1



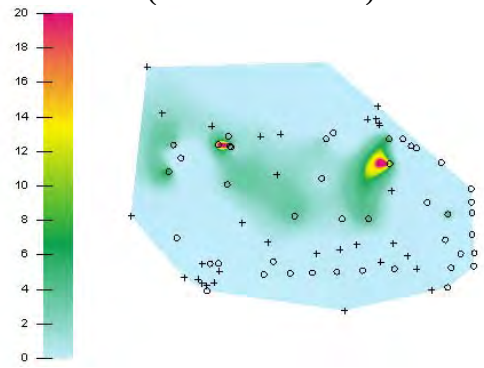
B1



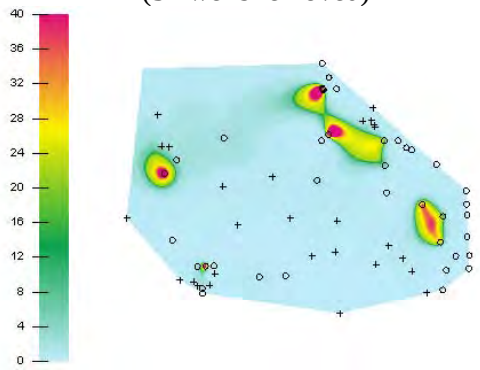
A2
(31 wells removed)



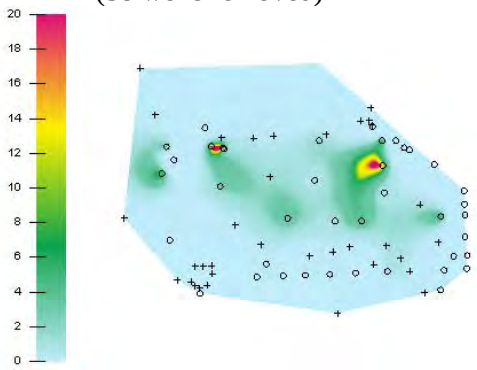
B2
(31 wells removed)



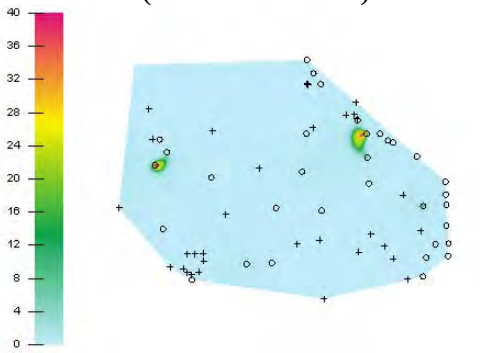
A3
(37 wells removed)



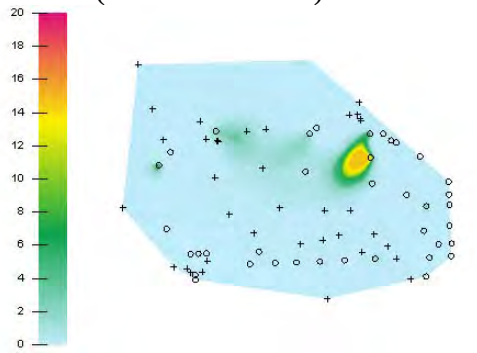
B3
(38 wells removed)



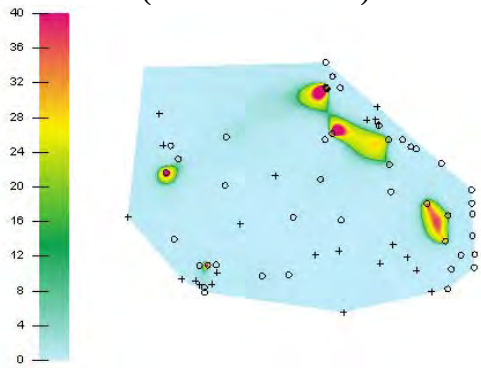
A4
(38 wells removed)



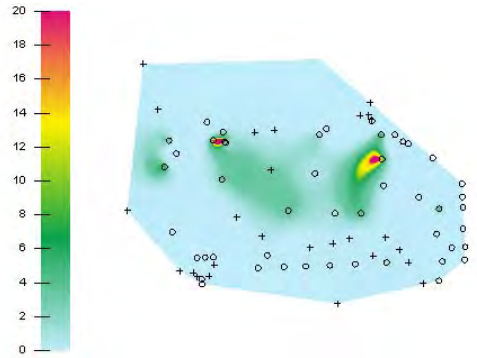
B4
(37 wells removed)



A5
(27 wells removed)



B5
(27 wells removed)



Note:

1. The symbol "+" indicates wells that are recommended to be removed by the Optimizer, while the symbol "O" denotes wells that are recommended to keep.

A comparison of these plume maps (which pertain to the approaches described in the table before the plume maps) is summarized in the table below:

Approach	Plume representation	# of wells removed	Comment
A2/B2	Good	31	Low cost, good plume interpolation
A3/B3	Good	37/38*	Low cost, good plume interpolation
A4/B4	Bad	38/37*	Bad plume interpolation
A5/B5	acceptable	27	Conservative, acceptable plume interpolation

Note: * 38 is the # of wells recommended to be removed by evaluating RDX only; 37 is the # of wells recommended to be removed by evaluating TCE only.

The most important result is that A4 and B4, where the wells recommended for removal based on one of the COCs are then removed for the other COC, result in poor plume interpolation. The reason is that wells removed for one constituent may be critical for accurate representation of different constituent. In approaches A2 and B2, where both constituents are analyzed together, the resulting plume representations remain accurate. This illustrates the benefit of performing the optimization with multiple COCs using the combined error (rather than optimizing them individually). We subsequently used approach A2/B2 (i.e., **combined maximum error of TCE/RDX**) for detailed spatial optimization analysis.

Tradeoff curves (i.e., sampling cost versus error) for each COC are generated by Optimizer. In contrast to the older version of the software, the newer version used for this site asks the analyst after each run if an additional run is needed, ensuring convergence of optimization result. First, the user starts with the default values for population size, number of generations, and random seed. Second, after each run is completed the software will ask whether the user wants to "seed this run with the results of the last run, and double the population size" – if yes, the mutation probability will automatically re-adjust for the new population size; the user makes a decision whether the tradeoff curve changes significantly during the run – if it does, then an additional run is necessary; otherwise the optimization result has converged. This new feature enables the user to perform convergence tests without guessing initial population size, number of generations, and random seed. We implemented this approach for our spatial analysis.

In an effort to compare spatial optimization results obtained by applying different combinations of interpolation and data transformation, spatial optimization was performed for the shallow aquifer using combined TCE/RDX error using each of the six combinations. Results are summarized below.

Summary of "Optimal Plans"– Spatial Optimization (Shallow Aquifer)

Aquifer	Interpola-tion	Data Transform-tion	# of Plans w/ Errors < 0.5	# of Plans w/ Errors < 1.0	# of Plans w/ Errors < 1.5	Min Cost w/ errors < 0.5 ($)	Min Cost w/ errors < 1.0 ($)	Min Cost w/ errors < 1.5 ($)
Shallow	Inverse Distance Weighting	None	7	11	13	75,000	71,000	69,000
		Logarithm	26	29	32	53,000	51,000	48,000
		Quantile	23	27	30	54,000	50,000	47,000
	Kriging	None	13	17	19	67,000	62,000	59,000
		Logarithm	27	30	33	54,000	51,000	48,000
		Quantile	25	30	33	55,000	50,000	47,000

Notes: 1. shallow aquifer (Baseline model has a total of 81wells)
2. cost per sample = $1,000
3. error is combined error of TCE and RDX

From the above table, it can be concluded that, given the same maximum error, interpolation with data transformations (i.e., logarithm and quantile) generates considerably more optimization plans than with no data transformation. Additionally, interpolation with no data transformation gives more expensive sampling plans than those with data transformation. Interpolations with the two data transformation methods (i.e., logarithm and quantile) yield similar sampling plans in that their number of plans and sampling costs are close. As discussed previously, our analyst prefers the plume representations provided by kriging with quantile transformation. Furthermore, a majority of the optimized plans were found within the combined maximum error of 0.5. Finally, it is noted that the analyst determined that interpolations with no data transformation ere found to be poor representation of the plume, so that optimization based on that data models with no data transformation are probably flawed and should not generally be performed.

{this gap is intentional}

1. Detailed SO Results, Shallow aquifer

The table below presents the results from a number of convergence runs for spatial analysis of the shallow aquifer.

Max Combined TCE/RDX Error versus Sampling Cost (Shallow Aquifer)

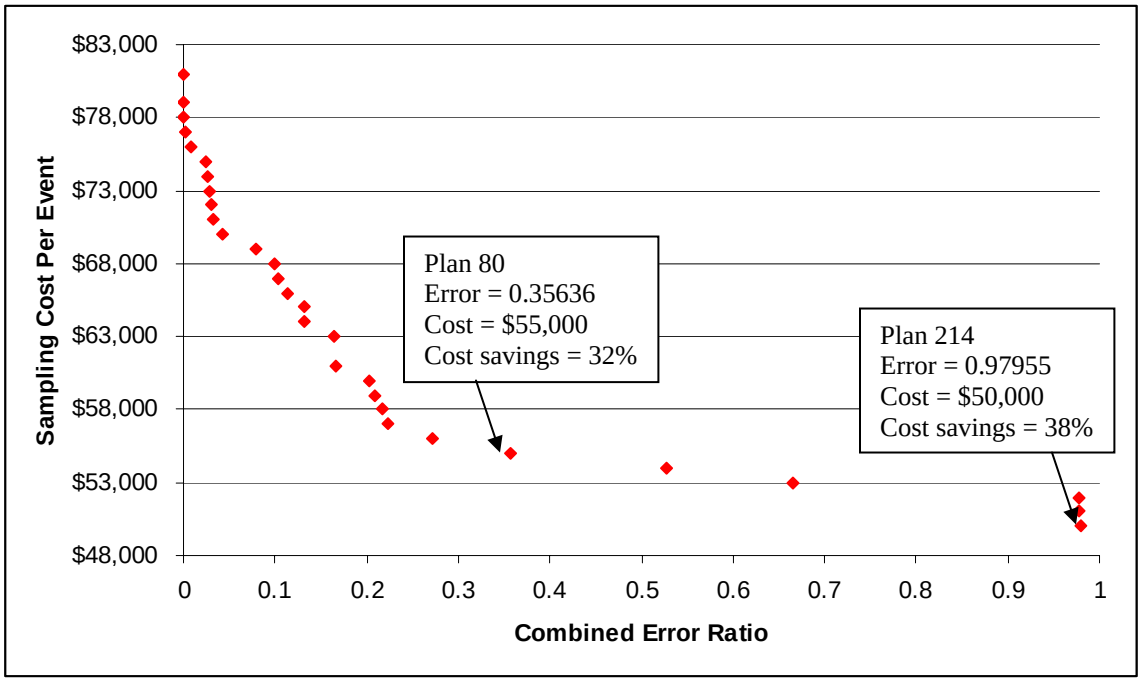
1st run (114, 162)*		2nd run (228, 162)		3rd run (456, 162)	
Combined Error	Sampling Cost Per Event	Combined Error	Sampling Cost Per Event	Combined Error	Sampling Cost Per Event
0	$81,000	0	$81,000	0	$81,000
0	$80,000	0	$79,000	0	$79,000
0.0078	$78,000	0.00034	$78,000	0.00034	$78,000
0.02899	$74,000	0.00266	$77,000	0.00266	$77,000
0.03175	$72,000	0.02367	$76,000	0.00789	$76,000
0.0884	$71,000	0.0273	$75,000	0.02367	$75,000
0.13193	$69,000	0.02827	$74,000	0.02729	$74,000
0.16511	$63,000	0.02911	$73,000	0.02848	$73,000
0.22017	$60,000	0.03093	$72,000	0.03093	$72,000
0.25512	$58,000	0.03178	$71,000	0.03178	$71,000
0.27275	$56,000	0.04168	$70,000	0.04168	$70,000
0.45099	$55,000	0.07841	$69,000	0.07841	$69,000
0.86251	$54,000	0.09845	$68,000	0.09845	$68,000
0.8934	$53,000	0.10291	$67,000	0.10291	$67,000
0.97981	$51,000	0.10682	$69,000	0.11282	$66,000
		0.13103	$66,000	0.13123	$65,000
		0.13134	$65,000	0.1318	$64,000
		0.16511	$63,000	0.16483	$63,000
		0.16881	$62,000	0.16602	$61,000
		0.20027	$61,000	0.20306	$60,000
		0.20306	$60,000	0.20906	$59,000
		0.20906	$59,000	0.21711	$58,000
		0.21952	$58,000	0.22321	$57,000
		0.22321	$57,000	0.27099	$56,000
		0.271	$56,000	0.35636	$55,000
		0.3585	$55,000	0.52835	$54,000
		0.52858	$54,000	0.66583	$53,000
		0.66583	$53,000	0.97837	$52,000
		0.97837	$52,000	0.97861	$51,000
		0.97861	$51,000	0.97955	$50,000
		0.97955	$50,000		

Notes:
1. optimized plans are selected within maximum error of 1.0
*2. *first number is population size, second is generation number; the random seed =347,182*
3. cost per sample assigned as $1,000

The above table shows that the 2nd and the 3rd runs yield more and better plans (i.e., plans of the same cost but with lower error) than the 1st. Moreover, the 2nd run yields practically identical results as the 3rd run, indicating convergence by the third run.

The figure below shows the resulting cost-error tradeoff curve for the shallow aquifer. Only optimal plans with errors less than 1.0 for the combined TCE/RDX error are included.

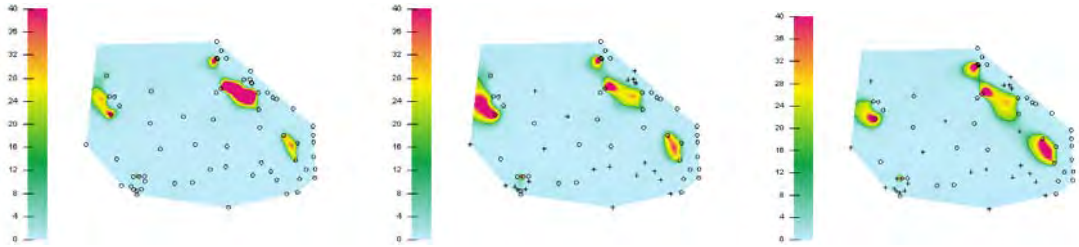
Tradeoff Curve for Shallow Aquifer
(Number of wells = 81; Number of non-removable wells = 25)



Plan 80 and Plan 214 were selected by our analyst for detailed evaluation (note one might also chose to evaluate plans "to the left" of Plan 80 with lower error but only slightly higher cost). Plan 80 reduces number of wells from 81 to 55 (32%) while Plan 214 reduces number of wells from 81 to 50 (38%) but with somewhat more error. It is important to note that 25 of the 81 wells cannot be removed based on the formulation, making the optimization results that much more impressive. For instance, Plan 214 has 31 wells removed from 56 that can potentially be removed (55%). Plume illustrations for these two plans, versus the base sampling plan with all locations, are presented below.

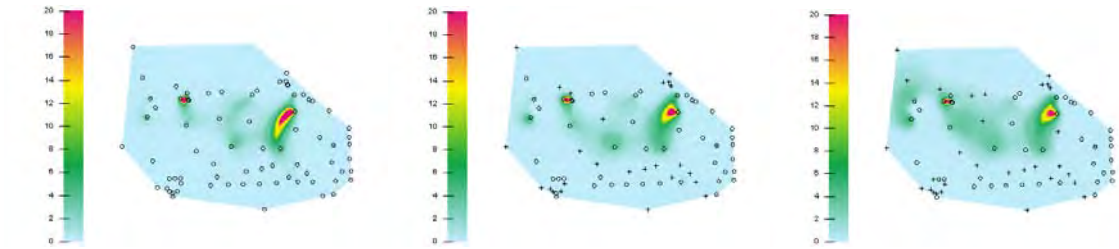
TCE Plume Map

All sampling locations Plan 80 Plan 214



RDX Plume Map

All sampling locations Plan 80 Plan 214



Comparing sampling costs (number of wells multiplied by $1000 per sample) and errors on the tradeoff curve, and visually inspecting the plume maps for selected plan(s), Plan 80 and Plan 214 were both considered by our analyst to be acceptable because they are both generally similar to the map with all sampling locations for both the higher concentration areas and the lower concentration areas. Of course, it would ultimately be up to site stakeholders to decide if either plan with reduced number of wells is acceptable. The table below lists the wells recommended to be removed for these two plans.

Wells recommended to be removed, Plan 80		**Wells recommended to be removed, Plan214**	
MW-05B	MW-56B	MW-07B	MW-56B
MW-117B	MW-60B	MW-117B	MW-60B
MW-118B	MW-61B	MW-118B	MW-61B
MW-16C	MW-64B	MW-16C	MW-64B
MW-17C	MW-79B	MW-17C	MW-79B
MW-27B	MW-81B	MW-23B	MW-80B
MW-29B	MW-83B	MW-27B	MW-81B
MW-31B	MW-84B	MW-29B	MW-83B
MW-35B	MW-85B	MW-30B	MW-84B
MW-37B	MW-86B	MW-31B	MW-85B
MW-52B	MW-92B	MW-35B	MW-86B
MW-54B	MW-93B	MW-37B	MW-89B
MW-55B	MW-94B	MW-42B	MW-92B
		MW-52B	MW-93B
		MW-54B	MW-94B
		MW-55B	

2. Detailed SO Results, Intermediate aquifer

The table below presents the results from a number of convergence runs for spatial analysis of the intermediate aquifer.

Max Combined TCE/RDX Error versus Sampling Cost (Intermediate Aquifer)

1st run (118, 168)*		2nd run (236, 168)		3rd run (472, 168)	
Combined Error	Sampling Cost Per Event	Combined Error	Sampling Cost Per Event	Combined Error	Sampling Cost Per Event
0	$84,000	0	$84,000	0	$84,000
0.00415	$83,000	0	$80,000	0	$80,000
0.01279	$82,000	0.00019	$79,000	0.00019	$79,000
0.01771	$80,000	0.00293	$78,000	0.00293	$78,000
0.02258	$77,000	0.00582	$77,000	0.00582	$77,000
0.02713	$74,000	0.0059	$76,000	0.0059	$76,000
0.0286	$73,000	0.00765	$75,000	0.00765	$75,000
0.02976	$72,000	0.00941	$74,000	0.00941	$74,000
0.03368	$70,000	0.01832	$73,000	0.01832	$73,000
0.05039	$67,000	0.02229	$72,000	0.02229	$72,000
0.09	$66,000	0.02745	$71,000	0.02283	$71,000
0.10563	$60,000	0.02748	$70,000	0.02361	$70,000
0.13158	$59,000	0.02847	$69,000	0.02656	$69,000
0.13187	$58,000	0.03033	$68,000	0.0299	$68,000
0.20353	$57,000	0.03224	$66,000	0.03122	$67,000
0.21814	$56,000	0.03378	$65,000	0.03224	$66,000
0.25681	$55,000	0.08096	$62,000	0.03378	$65,000
0.257	$54,000	0.10491	$60,000	0.04075	$64,000
0.32799	$53,000	0.11048	$59,000	0.08089	$62,000
0.46847	$51,000	0.12679	$58,000	0.08664	$61,000
0.50134	$49,000	0.16859	$57,000	0.10491	$60,000
0.59419	$48,000	0.20722	$56,000	0.11048	$59,000
0.60446	$47,000	0.21312	$55,000	0.12679	$58,000
0.6059	$46,000	0.22216	$54,000	0.16859	$57,000
0.69153	$45,000	0.25685	$53,000	0.20703	$56,000
0.78577	$44,000	0.27053	$52,000	0.212	$55,000
0.91822	$43,000	0.37985	$51,000	0.21635	$54,000
		0.39543	$50,000	0.22342	$53,000
		0.46884	$49,000	0.25704	$52,000
		0.47838	$48,000	0.37985	$51,000
		0.52739	$47,000	0.39543	$50,000
		0.60521	$46,000	0.46884	$49,000
		0.69153	$45,000	0.47483	$48,000
		0.78577	$44,000	0.52739	$47,000
		0.91822	$43,000	0.60521	$46,000
				0.69153	$45,000
				0.77479	$44,000
				0.91822	$43,000

Notes:

1. optimized plans are selected within maximum error of 1.0

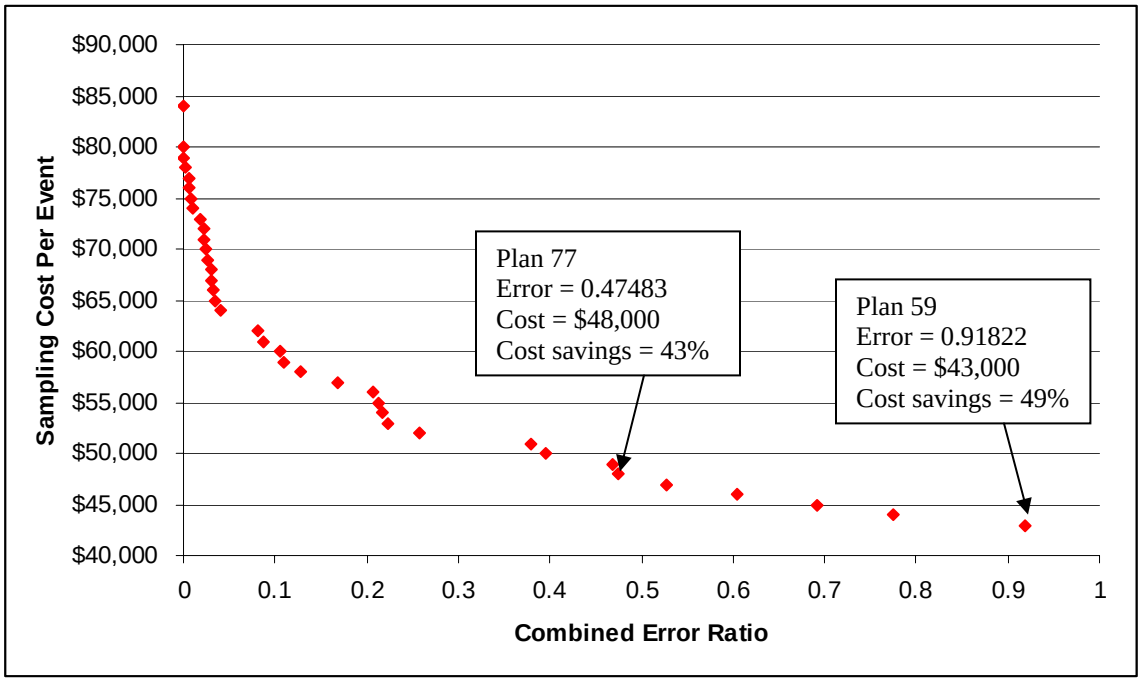
*2. *first number is population size, second is generation number; the random seed =347,182*

3. cost per sample assigned as $1,000

The above table shows that the 2nd and the 3rd runs yield more and better plans (i.e., plans of the same cost but with lower error) than the 1st. Moreover, the 2nd run yields practically identical results as the 3rd run, indicating convergence by the third run.

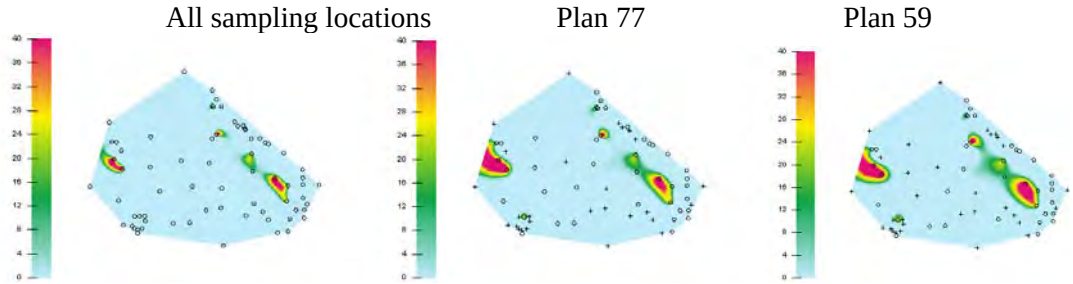
The figure below shows the resulting cost-error tradeoff curve for the intermediate aquifer. Only optimal plans with errors less than 1.0 for the combined TCE/RDX error are included.

Tradeoff Curve for Intermediate Aquifer
(Number of wells = 84; Number of non-removable wells = 25)



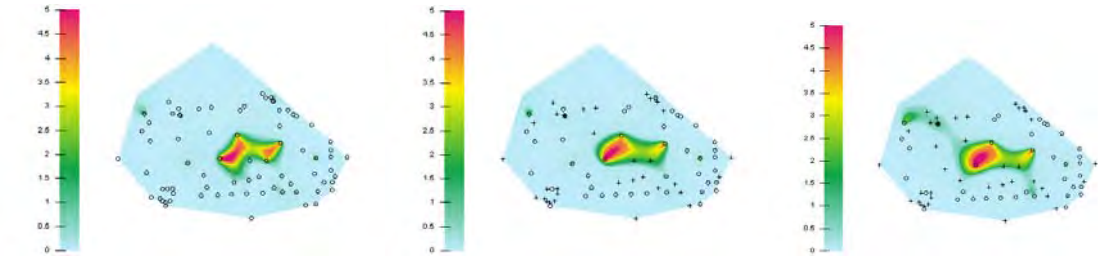
Plan 77 and Plan 59 were selected by our analyst for detailed evaluation (note one might also chose to evaluate plans "to the left" of Plan 77 with lower error but only slightly higher cost). Plan 77 reduces number of wells from 84 to 48 (43%) while Plan 59 reduces number of wells from 84 to 43 (49%) but with somewhat more error. It is important to note that 25 of the 84 wells cannot be removed based on the formulation, making the optimization results that much more impressive. For instance, Plan 59 has 41 wells removed from 59 that can potentially be removed (69%). Plume illustrations for these two plans, versus the base sampling plan with all locations, are presented below.

TCE Plume Map



RDX Plume Map

All sampling locations Plan 77 Plan 59



Comparing sampling costs (number of wells multiplied by $1000 per sample) and errors on the tradeoff curve, and visually inspecting the plume maps for selected plan(s), Plan 77 and Plan 59 were both considered by our analyst to be acceptable because they are both generally similar to the map with all sampling locations for both the higher concentration areas and the lower concentration areas. For RDX in Plan 59 there is somewhat higher concentrations estimated in the northwestern portion compared to the map based on all sampling locations, however our analyst felt those concentrations were still acceptably low. Of course, it would ultimately be up to site stakeholders to decide if either plan with reduced number of wells is acceptable. The table below lists the wells recommended to be removed for these two plans.

Wells recommended to be removed, Plan 77		**Wells recommended to be removed, Plan 59**	
MW-04A	MW-55A	MW-04A	MW-52A
MW-07A	MW-56A	MW-05A	MW-54A
MW-11	MW-60A	MW-07A	MW-55A
MW-117A	MW-61A	MW-11	MW-56A
MW-118A	MW-65A	MW-117A	MW-60A
MW-17B	MW-67A	MW-118A	MW-61A
MW-18B	MW-79A	MW-17B	MW-65A
MW-19B	MW-80A	MW-18B	MW-67A
MW-23A	MW-81A	MW-19B	MW-79A
MW-25A	MW-83A	MW-23A	MW-80A
MW-27A	MW-84A	MW-25A	MW-81A
MW-30A	MW-85A	MW-27A	MW-83A
MW-31A	MW-86A	MW-28A	MW-84A
MW-34A	MW-89A	MW-29A	MW-86A
MW-37A	MW-91A	MW-30A	MW-89A
MW-38A	MW-92A	MW-31A	MW-91A
MW-39A	MW-93A	MW-34A	MW-92A
MW-52A	MW-99A	MW-35A	MW-93A
		MW-37A	MW-94A
		MW-38A	MW-99A
		MW-39A	

3. Detailed SO Results, Deep aquifer

The table below presents the results from a number of convergence runs for spatial analysis of the deep aquifer.

Max Combined TCE/RDX Error versus Sampling Cost (Deep Aquifer)

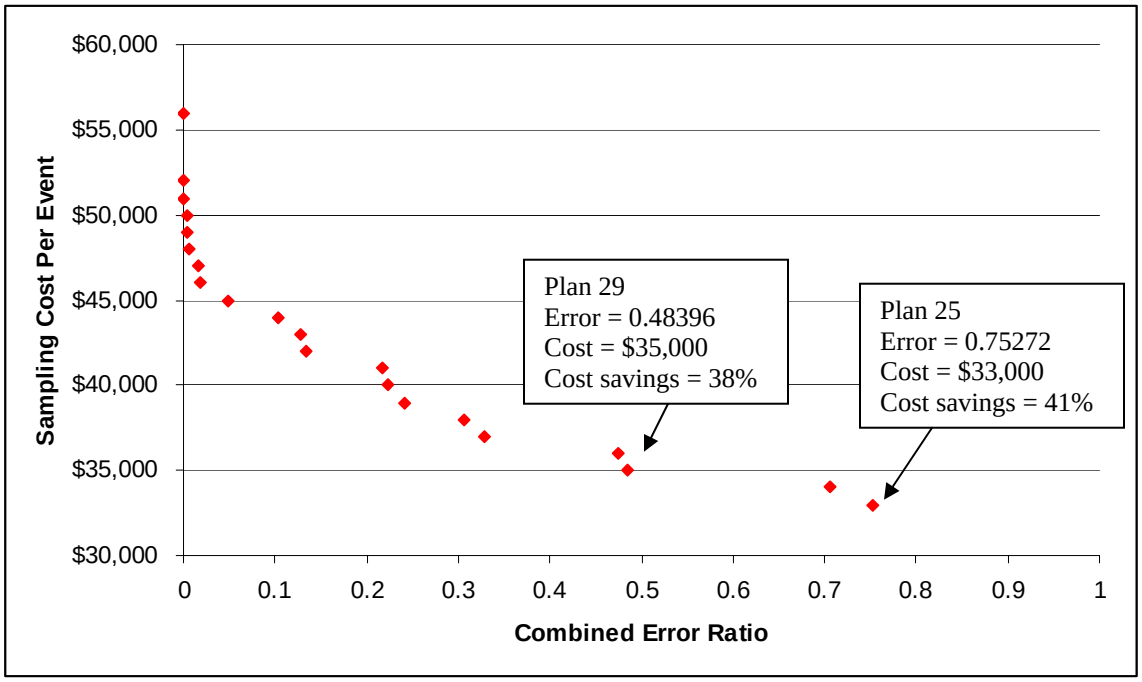
1st run (80, 112)*		**2nd run (160, 112)**		**3rd run (320, 112)**	
Combined Error	Sampling Cost Per Event	Combined Error	Sampling Cost Per Event	Combined Error	Sampling Cost Per Event
0	$52,000	0	$56,000	0	$56,000
0	$56,000	0.00021	$51,000	0	$52,000
0.00021	$51,000	0.00324	$50,000	0.00021	$51,000
0.00681	$48,000	0.00681	$48,000	0.00324	$50,000
0.10408	$45,000	0.01573	$47,000	0.00494	$49,000
0.12778	$44,000	0.01726	$46,000	0.00681	$48,000
0.12926	$43,000	0.04895	$45,000	0.01573	$47,000
0.22316	$40,000	0.12255	$44,000	0.01726	$46,000
0.24067	$39,000	0.12864	$43,000	0.04895	$45,000
0.30562	$38,000	0.13484	$42,000	0.10342	$44,000
0.32831	$37,000	0.21632	$41,000	0.12864	$43,000
0.4822	$36,000	0.22316	$40,000	0.13484	$42,000
0.48396	$35,000	0.24067	$39,000	0.21632	$41,000
0.70606	$34,000	0.30562	$38,000	0.22316	$40,000
0.75272	$33,000	0.32831	$37,000	0.24067	$39,000
		0.47593	$36,000	0.30562	$38,000
		0.48396	$35,000	0.32831	$37,000
		0.70602	$34,000	0.47523	$36,000
		0.75272	$33,000	0.48396	$35,000
				0.70602	$34,000
				0.75272	$33,000

Notes:
1. optimized plans are selected within maximum error of 1.0
*2. *first number is population size, second is generation number; the random seed =347,182*
3. cost per sample assigned as $1,000

The above table shows that the 2nd and the 3rd runs yield more and better plans (i.e., plans of the same cost but with lower error) than the 1st. Moreover, the 2nd run yields practically identical results as the 3rd run, indicating convergence by the third run.

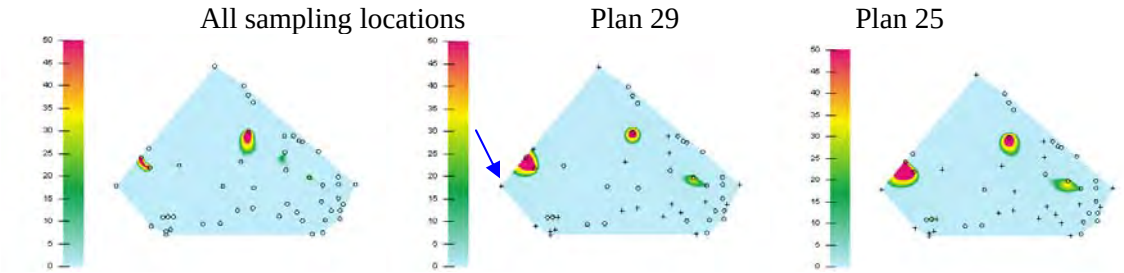
The figure below shows the resulting cost-error tradeoff curve for the deep aquifer. Only optimal plans with errors less than 1.0 for the combined TCE/RDX error are included.

Tradeoff Curve for Deep Aquifer
(Number of wells = 56; Number of non-removable wells = 22)



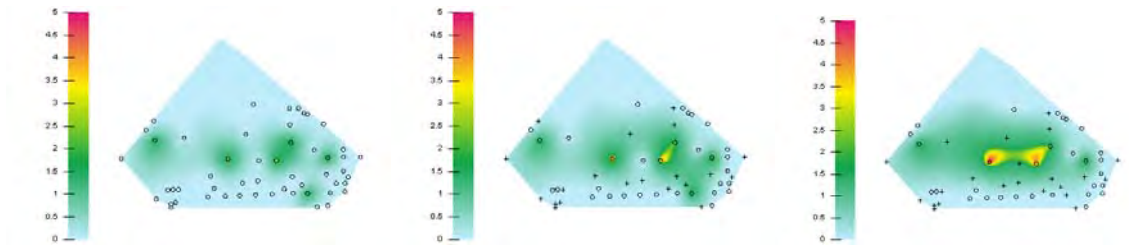
Plan 29 and Plan 25 were selected by our analyst for detailed evaluation (note one might also chose to evaluate plans "to the left" of Plan 29 with lower error but only slightly higher cost). Plan 29 reduces number of wells from 56 to 35 (38%) while Plan 25 reduces number of wells from 56 to 33 (41%) but with somewhat more error. It is important to note that 22 of the 56 wells cannot be removed based on the formulation, making the optimization results that much more impressive. For instance, Plan 25 has 23 wells removed from 34 that can potentially be removed (68%). Plume illustrations for these two plans, versus the base sampling plan with all locations, are presented below.

TCE Plume Map



RDX Plume Map

All sampling locations Plan 29 Plan 25



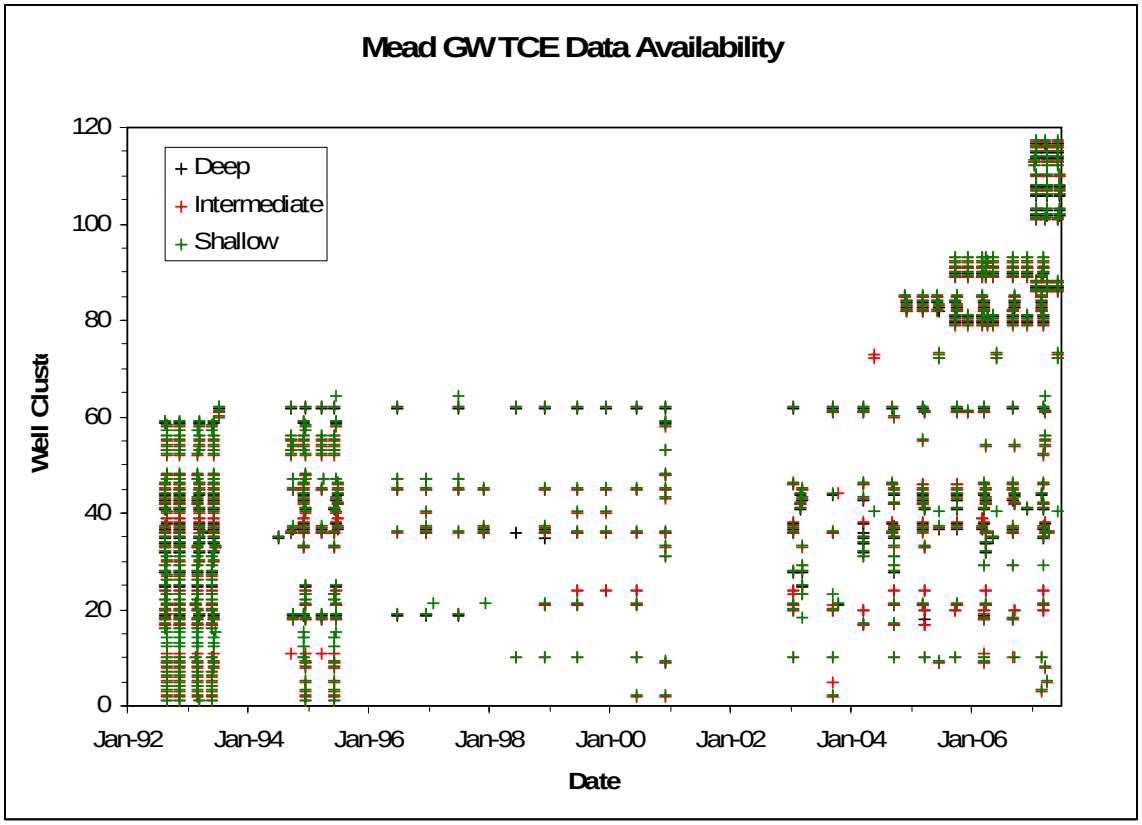
Comparing sampling costs (number of wells multiplied by $1000 per sample) and errors on the tradeoff curve, and visually inspecting the plume maps for selected plan(s), Plan 29 was considered by our analyst to be acceptable because it is generally similar to the map with all sampling locations for both the higher concentration areas and the lower concentration areas. For RDX in Plan 25 there is somewhat higher concentrations estimated in the central portion compared to the map based on all sampling locations, and our analyst felt those concentrations might be unacceptably different. Of course, it would ultimately be up to site stakeholders to decide if either plan with reduced number of wells is acceptable. The table below lists the wells recommended to be removed for these two plans.

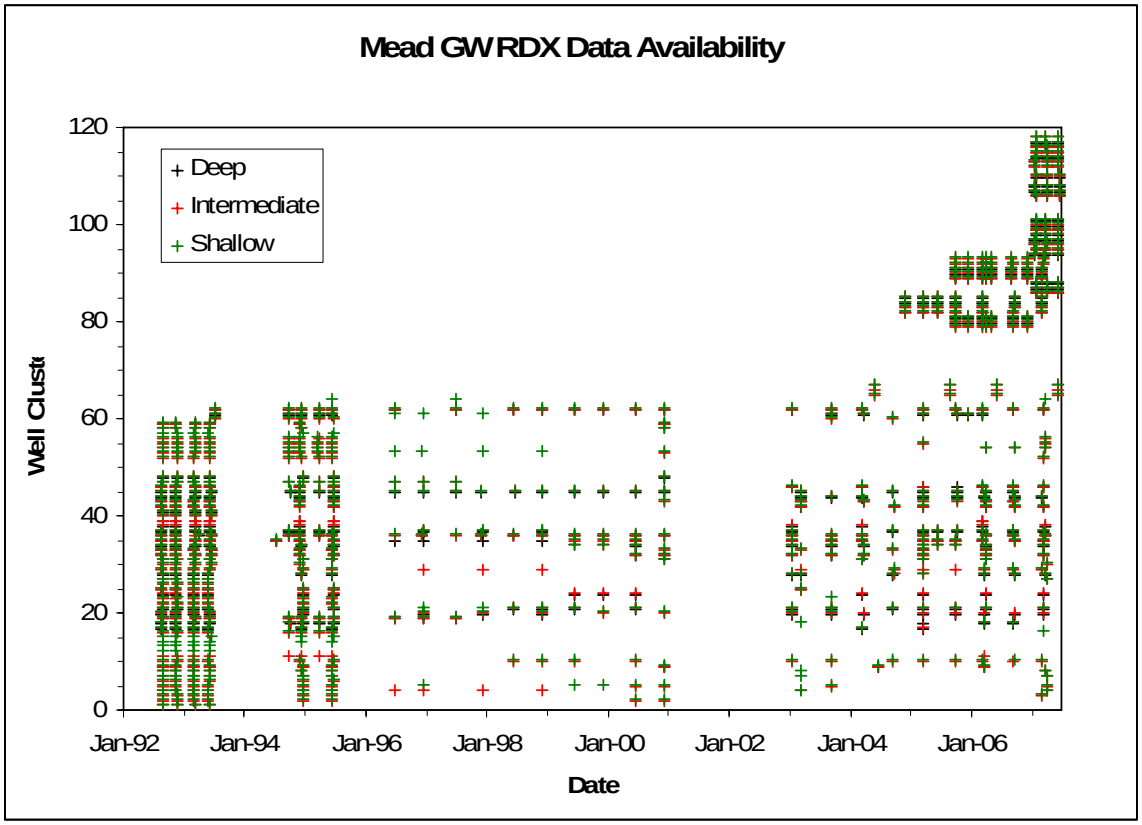
Wells recommended to be removed, Plan 29		**Wells recommended to be removed, Plan 25**	
MW-101D	MW-39D	MW-101D	MW-39D
MW-117D	MW-43D	MW-117D	MW-43D
MW-17A	MW-61D	MW-17A	MW-61D
MW-18A	MW-80D	MW-18A	MW-80D
MW-19A	MW-81D	MW-19A	MW-81D
MW-25D	MW-83D	MW-28D	MW-83D
MW-33D	MW-84D	MW-33D	MW-84D
MW-35D	MW-85D	MW-34D	MW-85D
MW-36D	MW-91D	MW-35D	MW-86D
MW-37D	MW-94D	MW-36D	MW-91D
MW-38D		MW-37D	MW-94D
		MW-38D	

Also, as can be seen from the above plume maps of the deep aquifer, removal of a clean well in the western corner (pointed toward by a blue arrow in the TCE map for Plan 29) results in a larger representation of the plume, indicating that although the interpolated concentration at that particular well location is within the specified error threshold, the concentrations beyond that point are possibly overestimated. Site stakeholders might choose not to eliminate that clean well even though the software indicates that eliminating it results in acceptable error.

Spatio-Temporal Analysis Results

The charts below illustrate the sporadic nature of sampling at specific wells over time at the site. The available data from most wells indicated NDs or very low values, and the site has not found it worthwhile to sample the majority of locations on a regular schedule.


Mead GW TCE Data Availability
Deep
Intermediate
Shallow
Well Clust
0
20
40
60
80
100
120
Jan-92
Jan-94
Jan-96
Jan-98
Jan-00
Jan-02
Jan-04
Jan-06
Date


Mead GW RDX Data Availability
Deep
Intermediate
Shallow
Well Clust
0
20
40
60
80
100
120
Jan-92
Jan-94
Jan-96
Jan-98
Jan-00
Jan-02
Jan-04
Jan-06
Date

This type of historical sampling pattern makes a spatial-temporal evaluation with this software product difficult. The software calculates error for spatio-temporal analysis as the maximum spatial error at any sample location removed in any event. Since most of the events for this site have many fewer wells than the total number of wells in the spatial analysis, spatial errors caused by removing samples in those events would tend to result in very large spatial errors within those events, and therefore larger overall errors. The only way to address this would be to fill in data values artificially for those that are missing, and given the sporadic nature of sampling over time at this site, that does not seem practical or reasonable (as discussed in the Data Preparation section presented earlier).

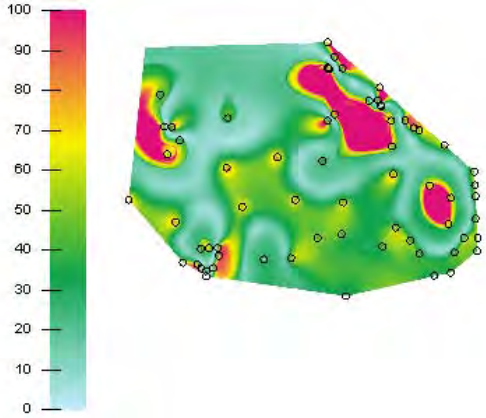
Relative Uncertainty Analysis

Model Builder can provide maps of uncertainty for plume visualization. In the new version of the software, this feature has changed to "Visualize Relative Uncertainty", replacing "Visualize Uncertainty" in the previous version. The idea behind the relative uncertainty is that root mean square error at each pixel comprising the plume image is calculated and then is divided by the arithmetic mean value at that pixel. It is designed to normalize the uncertainty metric so that the values over different sampling events and sites can be compared.

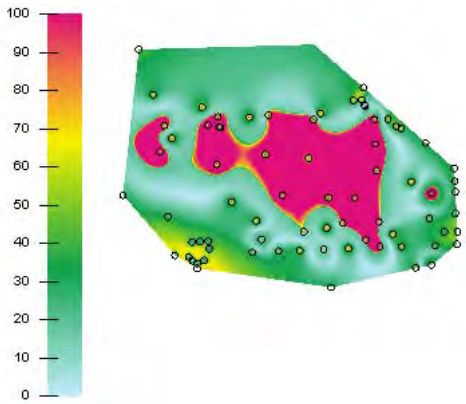
The figures below illustrate the plume uncertainty maps using four combinations of interpolation technique and transformation, using the spatial data for the shallow aquifer.

Relative Uncertainty Map (Kriging-Quantile)

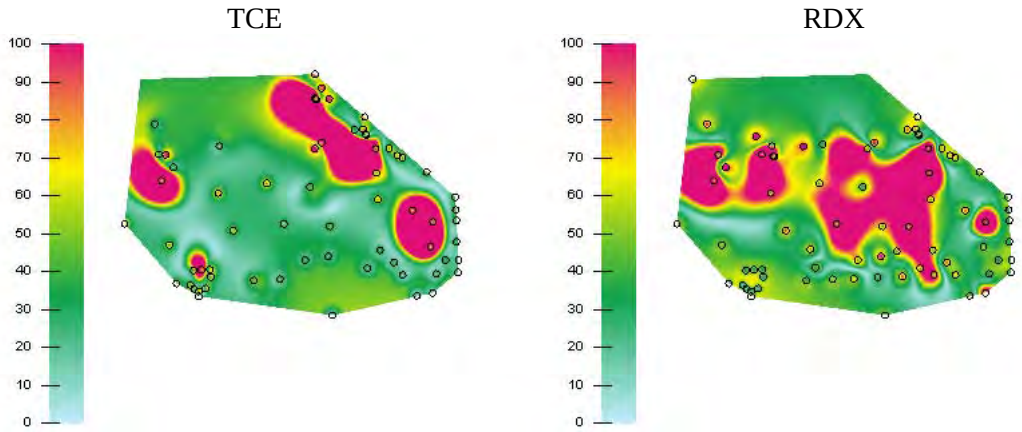
TCE



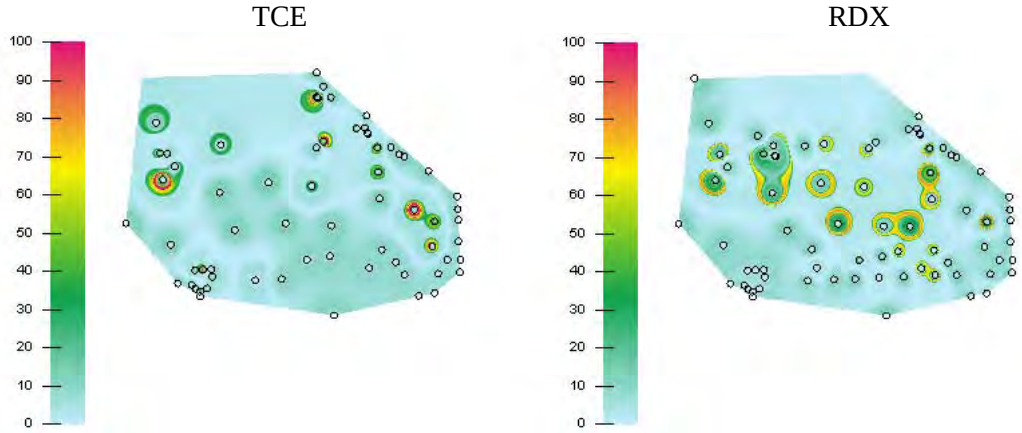
RDX



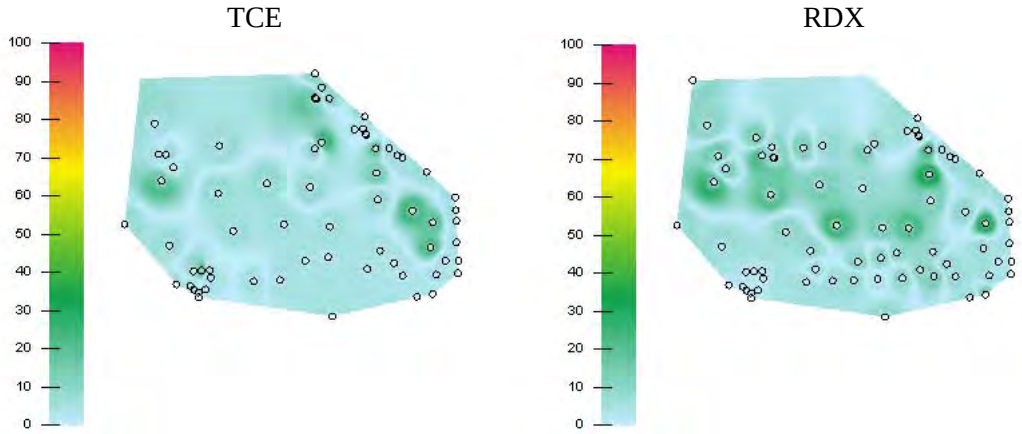
Relative Uncertainty Map (Kriging-Logarithm)



Relative Uncertainty Map (IDW-Quantile)



Relative Uncertainty Map (IDW-Logarithm)



Consistent with the findings of the previous analysis for the George site, the high uncertainty zones are again found to exist among wells with high measured concentrations (based on comparison with plume map figures). By comparing the maps for different transformations within each interpolation technique, the results seem relatively similar (i.e., IDW results for both quantile and logarithm transformations are similar) However, by comparing the relative uncertainty maps between the two interpolation techniques (i.e., kriging versus inverse distance weighting), it is observed that the high relative uncertainty area for kriging is much larger than that for IDW. It is unclear why there would be such a pronounced difference, or exactly what the higher values of uncertainty indicated for kriging would mean to the user.

It is also not clear to the GeoTrans analyst how these uncertainty maps can be used to specifically indicate what an acceptable amount of uncertainty is, whether that varies spatially, and how the software can be used to determine how many new wells might be need to reduce the uncertainty to an "acceptable" degree, and where to locate those wells.

Data Tracker Results (Reported by Charles Davis, EnviroStat)

Overview and Summary

Data Tracker (DT) is a feature of the Summit Monitoring Tools designed to aid in identifying anomalies of potential interest while sorting through monitoring data reports. DT has two modes: tracking data values for individual COCs at individual wells, and tracking other types of metrics computed using more than one COC and/or more than one well. The irregular historical data collection patterns at the Mead site make it very difficult to implement the latter functionality, however, in that a large number of data values would have to be interpolated for, in many cases, long periods of time. Accordingly, only the tracking of individual COCs at individual wells was performed at this site.

For tracking individual COCs at individual wells, DT compares new ("current") data values with selected background data from that well. When a current value is not consistent with expectations based on prior data, it should be flagged for evaluation. These expectations are formulated as prediction bounds (upper and lower) for the current observation. DT has two modes: static and time-dependent. If DT finds a statistically significant decreasing trend in the background data, it uses time-dependent bounds; otherwise it uses static bounds. The situation where the recent historical data show increasing values is of concern and will be discussed.

Using DT

DT works as follows:

- The user screens the historical data to remove atypical values that are not representative of current conditions, such as erratic values that are not repeated, nondetects (NDs) with unusually high reporting limits (RLs), or in some cases early data where concentrations have changed over the years (examples are provided below). The idea is that the "background" data used should be representative of values to be expected in future monitoring. Future values inconsistent with those background data will be flagged for inspection. This is done separately for each COC/well combination. The user may set the minimum number of background observations; DT requires an absolute minimum of

four background observations. Field duplicates are averaged with regular measurement values at this stage; DT allows at most one measurement per COC per well per date. Actual sample collection dates are used, so there is no need to align dates across wells in this mode.

- The user must supply a nominal value for each ND. For the Mead site, EnviroStat prepared the background dataset used by GeoTrans. In this dataset NDs were generally assigned a common arbitrary value: 0.5 µg/L for the VOCs (TCE, DCP, and MCl, for which 99% of the RLs ranged from 0.119 to 2.000) and 0.1 µg/L for the explosives (RDX, DNT, TNB, and TNT, for which 99% of the RLs ranged from 0.011 to 0.250). Samples with RLs above these ranges were generally omitted, unless values were needed to provide the minimum of four data values required by DT. In the latter case either the common ND value or half the RL was used, depending on the configuration of the remaining data value for that COC/well combination.

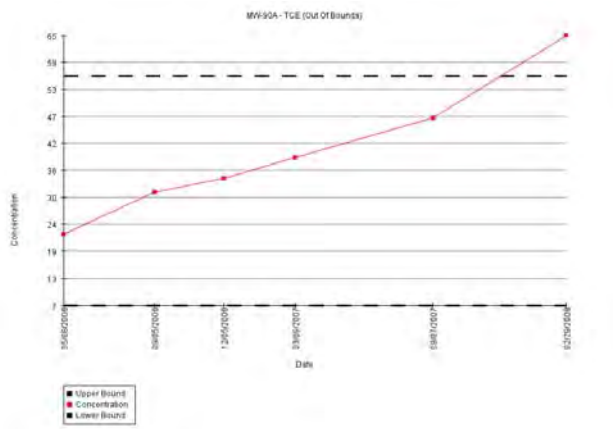
- The background data form one dataset, stored in CSV format. Historical values that are not included in the background data are not included in this dataset; the version of DT used at Mead does not yet have provision for displaying such data (ideally one could read in such data but not use them to calculate bounds, and display such data with a separate symbol). This issue was discussed in the DT report prepared for the former George AFB (GAFB).

- Prediction limits (PL bounds) are prepared for each well/COC combination with at least the minimum number of background values. These can be either static or time-dependent. If static, they are computed as nominal two-sided 95% PL bounds using the original data. If time-dependent, they are computed as nominal 95% PL bounds based on an exponentially decaying fit to the background data. The selection of type of bound is made automatically by DT. If a statistically significant <u>decreasing</u> trend is detected, the time-dependent option is used; otherwise static bounds are computed. DT will not provide time-dependent <u>increasing</u> bounds. The nominal prediction confidence is user-configurable; with the default 95% prediction confidence, so long as the assumptions are reasonably met, only around 5% of observations from truly steady-state or truly exponentially decaying processes should be out-of-bounds. In the static case the nominal confidence level is based on assumptions of (a) steady-state variation, (b) normal distributions of data, and (c) uncensored data; in the time-dependent case it is based on assumptions of (a') steady-state variation about a decreasing straight line on the log scale with (b') normally distributed errors and (c') uncensored data. In actuality, of course, real data often have trends that are not log-linear, outliers, NDs, and non-normal distributions; hence the nominal confidence level is somewhat approximate.

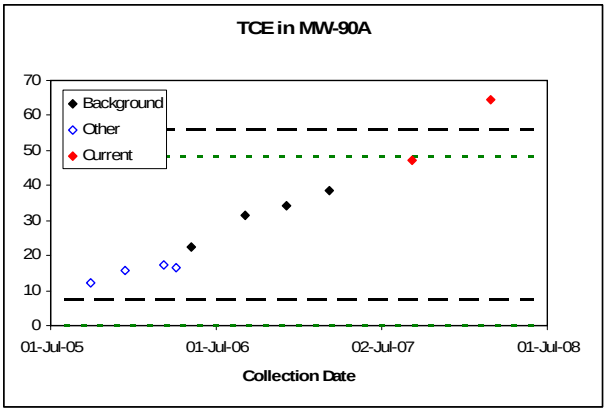
- When data for one or more new monitoring events arrive, another CSV file of these current data is prepared. DT compares each current value with its PL bounds. It prepares two tables, one of the COC/well combinations that are in-bounds for all dates in the current data file, and the other of COC/well combinations that are out-of-bounds for at least one measurement in the current file. The user can then click on a COC/well combination in either table to see a Time Series Plot (TSPlot) of the historical and current data along with the bounds. The data and the TSPlot can be saved for future reference or embedding in documents.

As discussed elsewhere, the data available at Mead are somewhat sporadic and do include significant changes in data patterns over the sixteen years (1992-2008). This situation demands that one screen the historical data carefully with an eye toward selecting the data that best represent the patterns that one anticipates in the future for each COC/well combination. This activity is much more labor-intensive than simply using all historical data as background data. Several examples are given to illustrate the effects of background data selection on the effectiveness of DT. Some of these examples, or artificial examples illustrating the situations involved, may be useful inclusions in a Users' Guide for the software.

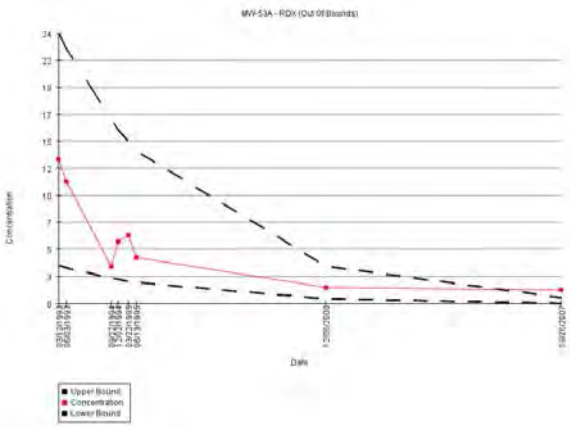
At the right is one example with Mead data (TCE in MW-90A). Static bounds are used, since the background data (four values from May 2006 through March 2007) do not have a statistically significant decreasing trend. The bounds are approximately (7, 56); these are fairly wide, which reflects the small number and spread of the background data values. The September 2007 value (47) is inbounds, but that of February 2008 (64.5) is out-of-bounds.



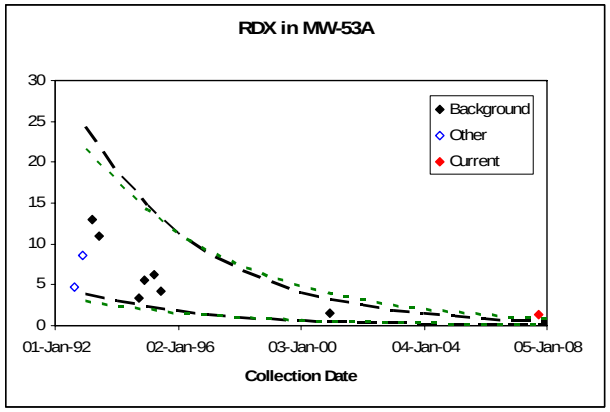
In this particular example, the four background data values used are only a portion of the historical data actually available from that well. The entire history is shown in the adjacent plot. There is in fact a statistically significant increasing trend whether one looks at only the last four historical values (p = 0.015) or all eight values (p = 0.000). This situation is problematic, in that a basic assumption involved in groundwater monitoring optimization is that conditions at the site are essentially in a steady state even though random. In this case, EnviroStat elected to use the minimum number of most recent values as the background data; the continued increase produced a value outside of the prediction bounds as early as the second value past the background period. In this figure the prediction bounds computed by DT are shown as dashed lines. Those that would have been computed using all historical data are dotted lines; the lower bound was slightly negative and was therefore set to 0.



The next DT plot (to the right) shows RDX in MW-53A. This well was sampled eight times between August 1992 and June 1995, and then again in December 2000, and then not until the current sample in September 2007.



Again, not all of the actual historical data were selected for use as background data; see the following plot. It appears that a minor slug of RDX came across this well in the winter of 1992-1993. A reasonable way of handling this situation in DT is to start the background data at the peak of the slug; this is the default treatment adopted by EnviroStat in preparing background datasets with apparent slugs passing through. Doing so, there is a statistically significant downward trend (p = 0.028), so exponentially decreasing bounds are used. These bounds necessarily decay toward zero; the September 2007 data value (1.3 µg/L) is above the upper bound, and hence is flagged. Note that this current data value is considerably above 99% of the reporting limits for RDX.



For comparison, the second plot shows all available data for that COC/well combination. The dashed prediction bounds are those computed by DT using the selected background data; the dotted lines are the bounds that would be computed using all historical data. In this case the bounds are similar. In other cases, particularly where the background data are more balanced in time before and after the peak of the slug, using the entire historical dataset might result in very wide but static bounds, which would not accurately reflect one's expectation that the slug should continue to dissipate with time. Of course, since the exponentially decreasing bounds do necessarily tend toward zero, eventually any data values will exceed the bounds. At that time it will be appropriate to re-evaluate the background data selection for that COC/well combination. A more sophisticated approach might be for DT to alter the time-dependent bounds once they fall below the typical RL; that would require that DT receive and process the RL information, which it does not do at present.

Simply flagging a data value as in- or out-of-bounds is not in itself a decision or action. The intent of the tables and plots produced by DT is to allow the user to concentrate on the out-of-bounds values, hence focusing attention on the COC/well combinations likely to require attention or action. Since the bounds are two-sided, in some cases one will find values that are lower than anticipated, and in others higher than anticipated. Either might indicate anomalies of interest; it is left to site personnel to make the actual decisions regarding each flagged value.

Testing DT using the Mead data

The operation of DT was tested using Mead data for seven COCs: three VOCs (TCE, DCP, and MCl) and four explosives (RDX, DNT, TNB, and TNT). Six versions of the reserved data (July 2007-June 2008) were prepared and used with background data from August 1992 through April 2007. One version contains the actual reserved data, unaltered except for averaging in field duplicate values and replacing NDs with the common value, as discussed previously; "J" flags are disregarded. The analyst at GeoTrans did not know which dataset was the actual site data. Artificial anomalies are introduced into the other five datasets, following plausible scenarios of interest at Mead. Descriptions of these scenarios (see below) were presented *a priori* and discussed with site personnel.

There were a large number of COC/well combinations with only three historical observations, rather than the minimum of four required by DT. In these situations, EnviroStat added an artificial data value and date in order to allow DT to function with that COC/well combination. In most cases these were wells located outside the plumes and reasonably presumed to be free of the COC; many were wells recently added. ND values were added in 451 such situations, along with low-level values consistent with the other values in eleven cases. There were two cases (TCE in MW-72B and MW-73B) where the added values were elevated (7250 and 155 µg/L, respectively) and hence clearly conjectural.

Details: Preparing Datasets with Artificial Anomalies

These datasets were generated from the July 2007 – June 2008 data received from Mead. We have TCE data for 135 wells: 82 wells for both events, 22 wells for Fall 2007 only, and 31 wells for Spring 2008 only. There were 88 wells with adequate amounts of background data but which were not sampled in Fall 2007 or Spring 2008; conversely, two wells whose samples had previously been analyzed only for explosives were analyzed for TCE for the first time. For RDX we have data for 151 wells: 73 wells for both events, 36 wells for Fall 2007 only, and 42 wells for Spring 2008 only. There were 90 wells with adequate background data that were not sampled in Fall 2007 or Spring 2008, and two wells whose samples had previously been analyzed only for VOCs were analyzed for RDX for the first time. In nearly every case DCP and MCl were measured along with TCE, and similarly for the other explosives along with RDX.

The six versions of the current (July 2007 – April 2008) data are labeled Ant, Bear, Cat, Deer, Elk, and Fox. Deer is the unaltered current dataset. The first step in creating the other five versions was to "jitter" all values (except NDs) slightly, to avoid having, say, all values except the Fox dataset value being identical for a given well, which would be an obvious clue that the Fox value was the artificial one. This was accomplished by multiplying each original value by a random number between 0.75 and 1.25, then rounding the result the same way as the original value was rounded. In a few cases NDs were replaced by "J" values and vice versa. Artificial anomalies were then added to the five versions other than Deer, following the scripts previously proposed.

The artificial anomaly scenarios

Narrative descriptions of the artificial anomaly scenarios follow. Various combinations of these were incorporated into the five datasets.

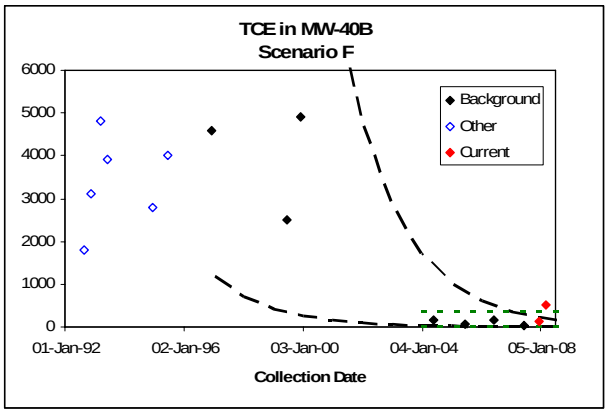
A. Due to a laboratory difficulty, all RDX measurements made for samples collected 20-Sep-07 are reported as NDs. Other explosives measurements are unaltered. The problem affects only samples collected this date.

B. MW-23B and MW-23A at the north end of the LL1 plume both show increasing levels of TCE (other COCs remaining as they are); nearby wells MW-22A and MW-22B start showing traces of TCE. These are sampled only in Fall 2007.

C. MW-36B and MW-36A at the south end of the LL4 plume show increases in TCE, and MW-62B, MW-62A, and MW-38A (still outside the plume as of October 2007) start having low levels of TCE. All increase from Fall 2007 to Spring 2008.

D. Labels on the sample bottles for MW-72A and MW-72B are switched for Fall 2007. Shallow well MW-72B has values in the ppm range (2500-4700 µg/L), whereas intermediate well MW-72A has values from 5 to 71 µg/L. These wells have been sampled only since 2004; they are located in the north-central portion of the LL4/Atlas plume.

E. MW-14 is sampled in Spring 2008; TCE is higher than its historical values (~10 µg/L, say), whereas DCP is lower (~3 µg/L, but still above the RL).



F. MW-40B TCE jumps up to around 10% of the 1995 levels, and DCP and MCl follow. MW-40A follows along. MW-40B TCE has a curious history, with ppm values through 2000, a gap in the sample record, and values less than 180 µg/L starting in 2004; see the adjacent plot. The selection of background data will determine the success of DT for the COC/well combination of Scenario F. For purposes of illustration, EnviroStat elected to use the last seven values as background data, which produce sharply decreasing time-dependent bounds (the black dashed lines). One could also reasonably use only the last four values (170, 61, 155, and 25), producing static bounds (the dotted green lines). Again, the selection of background data should reflect the patterns anticipated in future data; in this case, the static bounds based on only the more recent data would be the better choice.

G. MW-84A, MW-84B, MW-83A, and MW-83B (intermediate and shallow wells just outside of the south end of the LL2 plume) have increasing values of RDX in Fall 2007 and Spring 2008; MW-97B and MW-98B (shallow wells somewhat further south) have increases in Spring 2008.

H. MW-46 cluster wells (east of the south end of the LL4 plume) start to resemble those of MW-45 (centered longitudinally in the south end of the LL4 plume) in Fall 2007, increasing in Spring 2008; MW-112 and MW-113 (a bit north of MW-46) start to become impacted by Spring 2008 (one more, one less).

I. MW-100A and MW-100B have increases in Spring 2008 due to leaks from the Main Treatment Plant; MW-85A and MW-85B increase as well. These are located just outside the ½-mile buffer zone south of the southeast end of the LL3 plume.

J. TCE values for samples collected 26 September 2007 are five times too high, except for the NDs; other dates are not affected.

These ten scenarios were assigned to the five datasets as shown in the table below. Capital letters indicate the inclusion of the anomaly as stated; lower-case letters indicate situations where typical data values were supplied for completeness of the overall dataset. The wells involved in each of the datasets are also indicated in the table.

	Scenario									
Dataset	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Ant		**B**		**d**	**e**	**F**	**G**		**i**	**J**
Bear		**B**	**c**	**D**	**E**			**H**	**I**	**J**
Cat	**A**		**C**	**d**	**E**	**F**	**g**	**H**		
Deer										
Elk			**C**		**E**	**F**	**g**		**I**	
Fox	**A**	**B**	**c**	**D**			**G**		**I**	
	Wells affected									
	MW-100A	MW-22A	MW-36A	MW-72A	MW-14	MW-40A	MW-83A	MW-112A	MW-100A	MW-52A
	MW-100B	MW-22B	MW-36B	MW-72B		MW-40B	MW-83B	MW-112B	MW-100B	MW-52B
	MW-100D	MW-23A	MW-38A				MW-84A	MW-113A	MW-85A	MW-53A
	MW-83A	MW-23B	MW-62A				MW-84B	MW-113B	MW-85B	MW-53B
	MW-83B		MW-62B				MW-97B	MW-46A		MW-56B
	MW-83D						MW-98B	MW-46B		
	MW-85A							MW-46D		
	MW-85B									
	MW-85D									

DT-GeoTrans Results

In addition to reporting the COC/well combinations flagged by DT, GeoTrans suggested an interpretation for each flagged value. In its interpretations GeoTrans took into account the cleanup goals for the specific COCs, historical data and values from the other current event where there were two current events, and values at neighboring wells. In characterizing the out-of-bounds values, GeoTrans assigned one of six codes, as follows.

A: The current concentration is much higher than the historical data and above cleanup goal, more likely bad data than plume migration.

B: The current concentration is much lower than the historical data which were above cleanup goal, could be bad data.

C: The current concentration is higher than historical data and above cleanup goal but following an increasing trend.

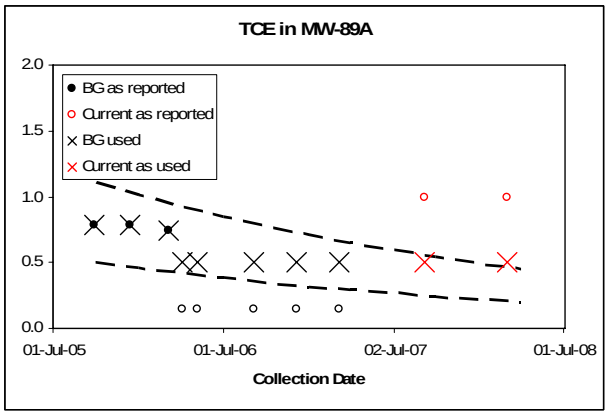
D: The current concentration is lower than historical data but following a decreasing trend.

E: The current concentration is out of bounds, but not a concern.

F: Current concentrations are higher than previous data, more likely plume migration than bad data.

GeoTrans provided a spreadsheet for each dataset listing the out-of-bounds COC/well combinations, the cleanup goal (for TCE and RDX), the assigned code, and a narrative comment in some cases. It also provided the plots given by DT in a PowerPoint file. In many cases the DT output was sufficient to allow the GeoTrans team to identify scenarios involving adjacent wells and/or two or more COCs at the same well.

Detecting actual anomalies in the Mead data



DT and GeoTrans flagged 41 anomalies in the actual Mead data (the Deer dataset). Of these 18 involve either background or current values below typical RLs (i.e., "J" values). Twelve of these involve static bounds; the other six involve time-dependent bounds. Perhaps the most curious of the latter is the situation with TCE in MW-89A, shown in the adjacent plot. In this plot the hollow symbols represent RLs for reported NDs and the crosses represent the values used. Because of the initial "J" values DT finds a statistically significant downward trend, and by February 2008 the upper prediction bound has dipped below the common value (0.5 µg/L) used for the NDs. In this case we have both a statistically significant decreasing trend and a high out-of-bounds value with data all below the usual RL! As mentioned previously, DT does not distinguish between actual values, "J" values, and common values used to represent NDs. The actual situation is clear from the DT plot, though, so long as the user is aware of the typical RLs for TCE.

Otherwise, DT and GeoTrans flagged 23 anomalies that represent real differences between expected and current values. In 17 of these at least one current value was higher than anticipated based on background data, very slightly in two cases and substantially in one (TCE in MW-42A). In six of these the values had previously been decreasing, but those decreases appear to have been reversed with the current data. One of these 17 cases disclosed an error in the data retrieval software; the errant value was actually for a different well, but was reported for two wells.

The remaining six cases involve current data values lower than expected. One should note that in four cases of the 23 actual anomalies, the background data were taken mostly or entirely prior to 2001, and therefore may not reliably represent current conditions. The narrative and categorical (A-F) descriptions provided by GeoTrans were accurate. In one case GeoTrans noted that, although the increase (in TNB in MW-21B) was slight, it was suspicious because of higher

concentrations in nearby MW-02B. An ultimate goal for a program such as DT would be to make it very easy for the user to notice such spatial connections.

Detecting the artificial anomalies

These will be discussed by anomaly rather than by dataset, since the DT/GeoTrans response was virtually the same for a given anomaly regardless of dataset, even though the data values were jittered somewhat between the datasets.

A. DT found many of these sudden dips in RDX to ND levels. Those giving it difficulty were cases where variable background data made the lower prediction bound (LPB) negative (effectively zero). Interpretation was a bit difficult, in that this scenario interacted with scenarios G and I in some of the datasets.

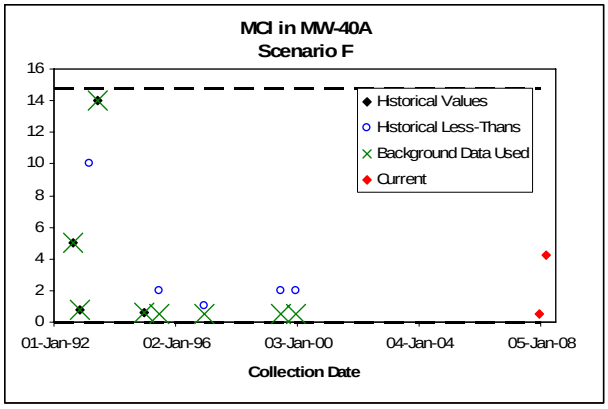
B. DT found these increases in TCE in MW-23B and MW-23A; interpretation was not a problem.

C. DT found the increases in TCE in MW-36B, MW-38A, and MW-62A/B. Those in MW-36A were more subtle relative to the background variability (data values around 35 µg/L).

D. DT found the anomalies in TCE in MW-72A and MW-72B; GeoTrans easily identified switched labels as the likely cause. Recall that MW-72B was the case with only three historical values in the ppm range, although MW-72A had four historical values in the <100 ppb range. The switch would probably have been identified without the artificially added data point, unless MW-72B had been deleted completely from the background data file due to insufficient data.

E. DT found the changes in TCE, not DCP; the background data were too variable for the latter.

F. DT found the increases in TCE in MW-40B and MW-40A and in MCl in MW-40B; see the discussion of the background data selection presented previously for TCE. It did not find the increase in MCl in MW-40A; here is another case where DT's ability to find an anomaly depends critically on the selection of background data and the treatment of "less than" data values. The increase would have been found if only the recent low-level NDs had been used as the background data. With a few early (1992-1993) higher values included, DT did not find the decrease to be statistically significant, and so used static prediction bounds. These were too wide to allow efficient detection of small increases in the current data. A better choice of background data would have been to include only the recent NDs.



G. DT found these increases in RDX; GeoTrans identified the possible relationship between these and a possible migration of the LL2 plume.

H. DT flagged most of these increases.

I. DT flagged these increases in RDX in the vicinity of the Main Treatment Plant. GeoTrans compared the values with the cleanup goals in making its determinations of the significance, but their evaluation did not take into account the fact that these wells are outside of the ½-mile buffer zone around the southern edge of the plume. As with other situations, a very useful enhancement to DT would be to not only flag an apparent anomaly but to also to identify its location on a site map, to aid the user in interpreting its potential significance.

J. This scenario had to do with a systematic lab error making all TCE values for samples collected 26 September 2007 five times higher than they would have been otherwise. Five COC/well combinations were impacted; DT found three of them. In the others the background data were too variable, so the change was not detected.

DT- Discussion

DT and GeoTrans identified numerous anomalies present in the actual current (July 2007 – April 2008) data (the Deer dataset) and nearly all of the artificially introduced anomalies in the other datasets. In some cases the anomalies were too small relative to the variation in background data to be found. In several scenarios GeoTrans was able to correlate anomalies spatially, outside of the software itself.

As noticed with the data from the former George Air Force Base (GAFB), an annoyance is that DT often flags very low values when the background data are mostly NDs and/or "J" values. EnviroStat had prepared the background data for Mead, replacing low-level NDs with a common value and omitting higher-level NDs in nearly all cases. Nonetheless, 18 of 41 flagged COC/well combinations in the actual data involved such low-level data of little or no real interest. A user who is aware of the reasonable range of RLs for the constituents in question would have little difficulty in sorting out these cases from the plots provided by DT.

The GAFB data did prompt a discussion of the need for care in selecting background data, and these Mead data reinforce that need; several examples are presented in the preceding material. A situation arising in the Mead data that had not come up previously is where recent concentrations for a particular COC/well combination are increasing, either from the beginning of its history or after a relatively static period of low-level and/or ND measurements. Such a data history may represent a slug of contaminant passing through the well. In one case in the actual data the latest current data point may be the beginning of the turn-around of such a slug.

At any rate, one possible useful enhancement to DT might be to flag any COC/well combination that shows a statistically significant increase in the background data, regardless of the value of the current data. DT does not use time-dependent increasing bounds; if the values in a COC/well combination are increasing, DT should bring that situation to the attention of the user. Even with static bounds, so long as the bounds are not automatically updated with each new run, a continually increasing trend will eventually be flagged. Flagging a COC/well combination with a statistically significant increasing trend in background data will accomplish that result more quickly, however.

The issue of when and how to update the background data was addressed to some extent in the DT reports for the first two sites. The follow-up data analyses to be conducted at GAFB will provide an opportunity to explore this issue further.

Finally, the following recommendation from the DT report for GAFB is repeated. Summit should consider including a discussion of the issues surround background data selection, and perhaps a tutorial using these or similar examples, in the Users' Guide and related materials. There has been some debate within the ESTCP project team about the level of knowledge and experience needed to successfully use DT. It may well be that the background data selection step requires more care and familiarity with the issues than does the routine screening of a current dataset. That initial background selection would take place when DT is used for the first time at a given facility. This recommendation presupposes that there is a mechanism for communicating the background data selection decisions (and possibly other decisions as well) between uses of DT to screen successive new datasets.

Mass Metric and Mass Flux Results

Using the software to evaluate changes in mass over time requires a spatio-temporal dataset that allows mass in each time period to be estimated based on spatially-distributed data values in that time period. Because of the sporadic sampling frequency over time discussed earlier that precludes spatio-temporal evaluation at this site, we are unable to perform mass calculations over time for this site.

Model Validation

The reserved dataset prepared by EnviroStat for validation is for Fall 2007 and Spring 2008. Some of the wells were sampled twice (i.e., Fall 2007 and Spring 2008) for TCE and RDX. Because only one sampling event is required for input data for model validation, and because concentrations between Fall 2007 and Spring 2008 are similar, the values for the two sampling events were averaged, forming a new validation dataset that has only one concentration value per location (assigned as 1/1/2008).

The following general procedures were applied for performing model validation:

- Create a full dataset which includes all reserved sampling data; make sure the locations (i.e., well IDs) are identical to those included in the original spatial dataset. For wells that were not sampled in either Fall 2007 or Spring 2008, the latest data value (e.g., Spring 2007 data) was assigned to it, so that the computational domain remains the same.

- For wells that are recommended to be shut off by the Optimizer based on the spatial optimization, set both their Max/Min sampling frequency to be "off"; for the remaining wells, set both their Max/Min sampling frequency to be "on". This way no sampling optimization will be performed by the software since the sampling frequency for each individual well has been predetermined.

- Run Optimizer with the above settings with only one generation, and display the interpolated concentrations at removed well locations.

- Compare the interpolated values against actual values, and evaluate the loss of information due to the removal of those wells.

The reason we need to use Optimizer in this process is that Model Builder does not extrapolate concentration values at the locations outside of the convex hull of the data points, nor does it interpolate concentration value at a removed location. Therefore, if points located at the plume edge are removed, the area surrounding the removed point is not interpolated and appears to be "missing" in Model Builder. Thus, it is hard to compare the concentration values at removed locations to the actual values purely based on plume maps using Model Builder.

Model Validation for Spatial Analysis

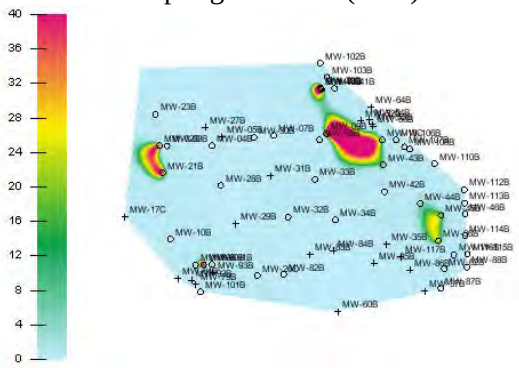
Shallow Aquifer

Plan 80 (26 of 81 wells removed) and Plan 214 (31 of 81 wells removed) from the Spatial Analysis results were used for the model validation analysis. The figures below illustrate the plume maps generated for all sampling locations versus interpolated plumes for Plan 80. The posted values are the interpolated values at the removed wells.

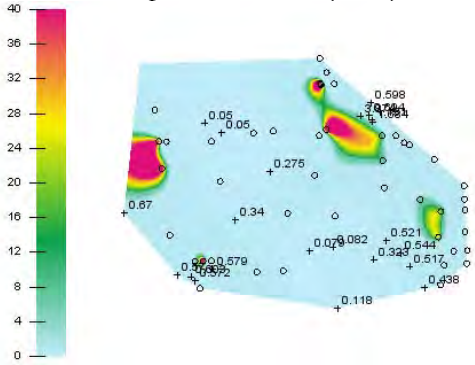
{this gap is intentional}

All Sampling Locations versus Plan 80

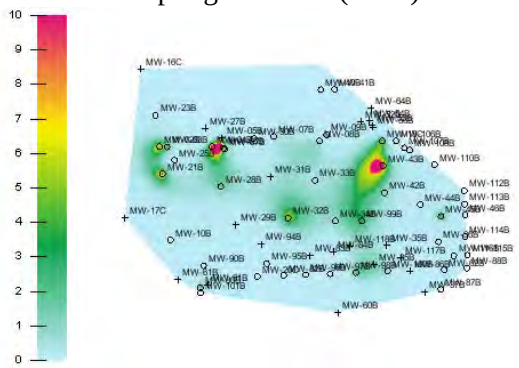
All sampling locations (TCE)



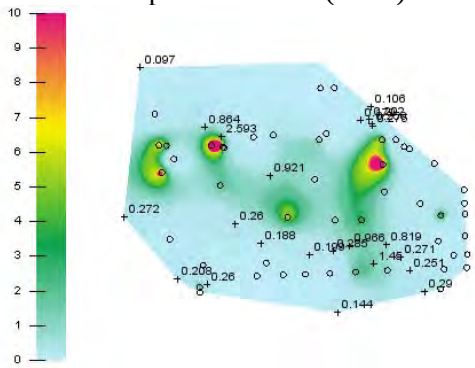
Interpolated values (TCE)



All sampling locations (RDX)



Interpolated values (RDX)



Visually, the RDX plume maps look similar, validating that there is little loss of information caused by removal of the wells. The TCE plume maps, however, differ somewhat in that the LL1 plume along the western boundary of the domain expanded in the interpolated plume maps. This plume misrepresentation is caused by the removal of MW-17C, a low concentration well, resulting in overestimated concentration in its vicinity. As a result, although Plan 80 suggests removing MW-17C, one should be cautious of doing so because of its sensitivity to plume interpolation accuracy. The table below presents a comparison of the interpolated concentrations at the removed locations versus the actual measured concentrations at removed wells in the validation data for Plan 80.

Interpolated Concentrations versus Actual Concentrations at Removed Wells
Optimal Plan 80

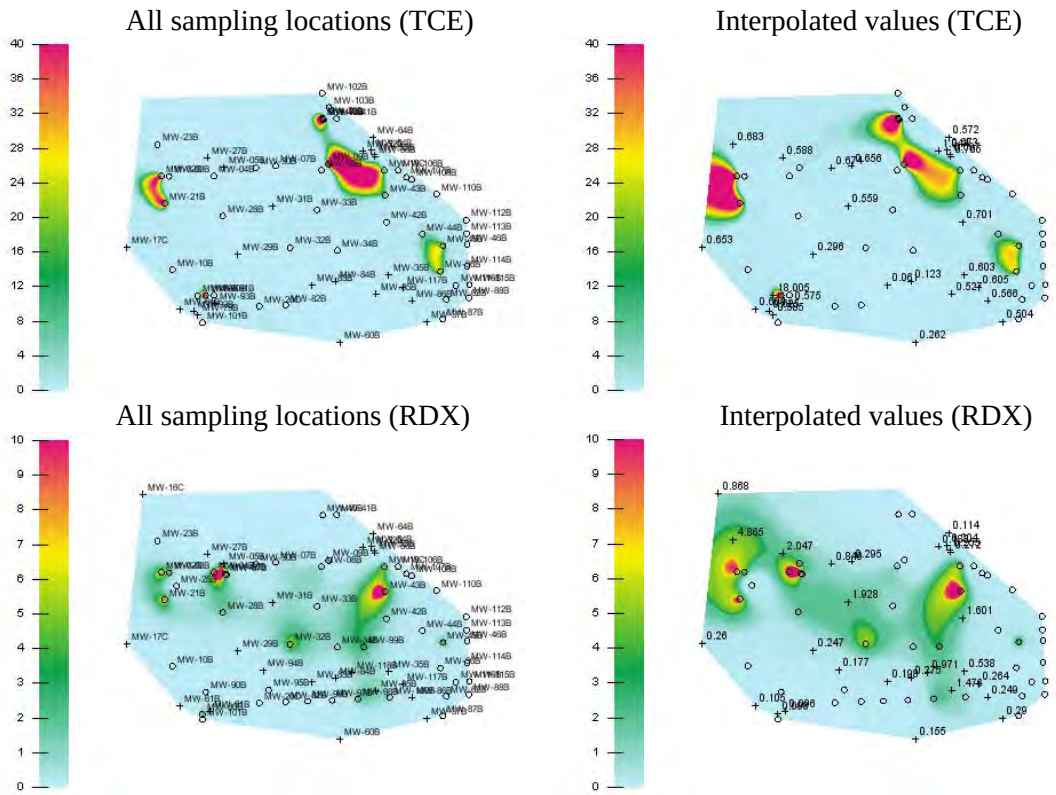
Removed Wells	Interpolated Concentrations (µg/l)		Actual Concentrations (µg/l)		Absolute Deviation (µg/l)	
	TCE	RDX	TCE	RDX	TCE	RDX
MW-05B	0.05	2.593	2.06	2.02	2.01	0.573
MW-117B	0.544	0.271	0.5	0.1	0.044	0.171
MW-118B		0.966		0.62		0.346
MW-16C		0.097		0.005		0.092
MW-17C	0.67	0.272	0.05	0.005	0.62	0.267
MW-27B	0.05	0.864	0.05	0.005	0	0.859
MW-29B	0.34	0.26	0.05	0.1	0.29	0.16
MW-31B	0.275	0.921	0.5	0.46	0.225	0.461
MW-35B	0.521	0.819	0.5	0.23	0.021	0.589
MW-37B	0.438	0.29	0.05	0.005	0.388	0.285
MW-52B	3.87	0.179	0.57	0.1	3.3	0.079
MW-54B	0.694	0.202	0.5	0.1	0.194	0.102
MW-55B	1.131	0.256	0.5	0.69	0.631	0.434
MW-56B	1.684	0.275	0.875	0.1	0.809	0.175
MW-60B	0.118	0.144	0.05	0.005	0.068	0.139
MW-61B	0.57	0.208	0.05	0.005	0.52	0.203
MW-64B	0.598	0.106	0.5	0.1	0.098	0.006
MW-79B	0.572		0.4		0.172	
MW-81B		0.26		0.5		0.24
MW-83B	0.079	0.199	0.05	0.265	0.029	0.066
MW-84B	0.082	0.289	0.05	0.1	0.032	0.189
MW-85B	0.323	1.45	0.05	1.4	0.273	0.05
MW-86B	0.517	0.251	0.5	0.1	0.017	0.151
MW-92B	0.6		0.5		0.1	
MW-93B	0.579		0.5		0.079	
MW-94B		0.188		0.1		0.088

Note that some of the values are blank because that well was not historically sampled for that parameter and is therefore not included in the original spatial dataset for that parameter.

Our analyst concluded from these results that there were no major concerns with the interpolation with wells removed (as per the spatial optimization recommendations) versus the actual data because of the low concentrations at the wells, but that would ultimately be up to site personnel.

The figures below illustrate the plume maps generated for all sampling locations versus the locations recommended by Plan 214. This plan has slightly more error than Plan 80 because more wells are removed. The posted values are the interpolated values at the removed wells.

All Sampling Locations versus Plan 214



One observation is that, for TCE, the high concentration area on the western boundary is enlarged when wells are removed as per the spatial optimization recommendations. The most important factor appears to be the removal of MW-17C. In the discussion of spatial optimization results it was suggested by our analyst that it might be a good idea to keep sampling at this well for this very reason.

Another observation is that, for RDX, removing wells causes significantly higher estimated concentrations in the western portion of the site. For instance, as illustrated on the figure above, the interpolated value of 4.865 µg/L at MW-23B is much higher than the actual sampled value at that well. The table below presents a comparison of the interpolated concentrations at the removed locations versus the actual measured concentrations at those locations in the validation data, for Plan 214.

.

Interpolated Concentrations versus Actual Concentrations at Removed Wells
Optimal Plan 214

Removed Wells	Interpolated Concentrations (µg/l)		Actual Concentrations (µg/l)		Absolute Deviation (µg/l)	
	TCE	RDX	TCE	RDX	TCE	RDX
MW-07B	0.656	0.295	0.05	0.525	0.606	0.23
MW-117B	0.605	0.264	0.5	0.1	0.105	0.164
MW-118B		0.971		0.62		0.351
MW-16C		0.005		0.005		0
MW-17C	0.653	0.26	0.05	0.005	0.603	0.255
MW-23B	0.683	4.865	0.21	0.1	0.473	4.765
MW-27B	0.588	2.047	0.05	0.005	0.538	2.042
MW-29B	0.296	0.247	0.05	0.1	0.246	0.147
MW-30B	0.647	0.846	0.05	0.005	0.597	0.841
MW-31B	0.559	1.928	0.5	0.46	0.059	1.468
MW-35B	0.603	0.538	0.5	0.23	0.103	0.308
MW-37B	0.05	0.29	0.05	0.005	0	0.285
MW-42B	0.701	1.601	0.5	0.98	0.201	0.621
MW-52B	1.044	0.183	0.57	0.1	0.474	0.083
MW-54B	0.673	0.204	0.5	0.1	0.173	0.104
MW-55B	0.607	0.255	0.5	0.69	0.107	0.435
MW-56B	0.705	0.272	0.875	0.1	0.17	0.172
MW-60B	0.05	0.155	0.05	0.005	0	0.15
MW-61B	0.604	0.105	0.05	0.005	0.554	0.1
MW-64B	0.572	0.114	0.5	0.1	0.072	0.014
MW-79B	0.585		0.4		0.185	
MW-80B		0.898		2.55		1.652
MW-81B		0.096		0.5		0.404
MW-83B	0.05	0.198	0.05	0.265	0	0.067
MW-84B	0.05	0.275	0.05	0.1	0	0.175
MW-85B	0.527	1.476	0.05	1.4	0.477	0.076
MW-86B	0.566	0.249	0.5	0.1	0.066	0.149
MW-89B	18.005		1.7		16.305	
MW-92B	0.625		0.5		0.125	
MW-93B	0.575		0.5		0.075	
MW-94B		0.177		0.1		0.077

Our analyst concluded from these results that there were no major concerns with the interpolation with wells removed (as per the spatial optimization recommendations) versus the actual data for TCE, but for RDX there is some concern in the western portion of the site.

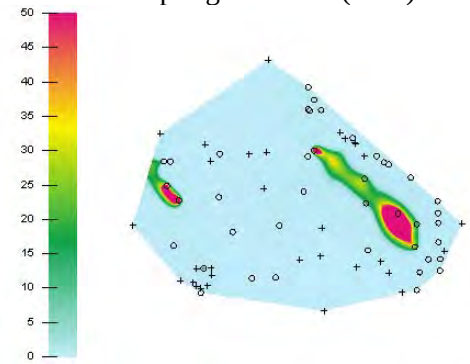
As expected, there is more overall deviation with Plan 214 (more wells removed) than with Plan 80 (fewer wells removed).

Intermediate Aquifer

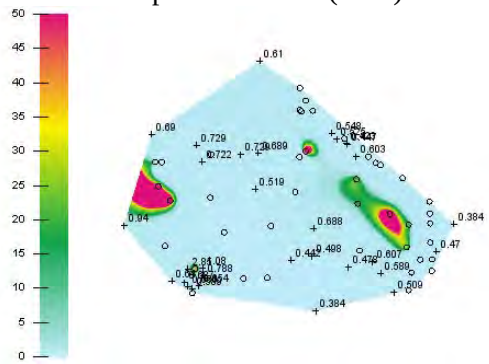
Plan 77 (36 of 84 wells removed) and Plan 59 (41 of 84 wells removed) from the Spatial Analysis results were used for the model validation analysis. The figures below illustrate the plume maps generated for all sampling locations versus interpolated plumes for Plan 77. The posted values are the interpolated values at the removed wells.

All Sampling Locations versus Plan 77

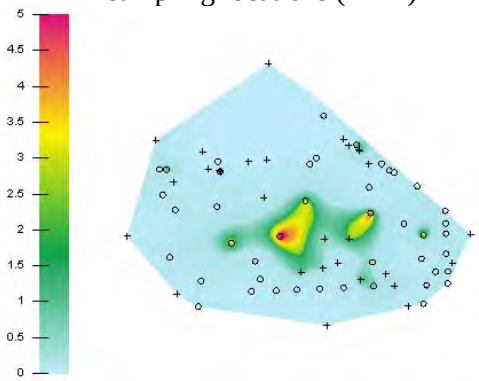
All sampling locations (TCE)



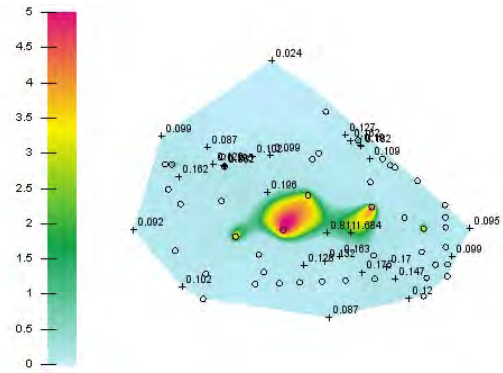
Interpolated values (TCE)



All sampling locations (RDX)



Interpolated values (RDX)



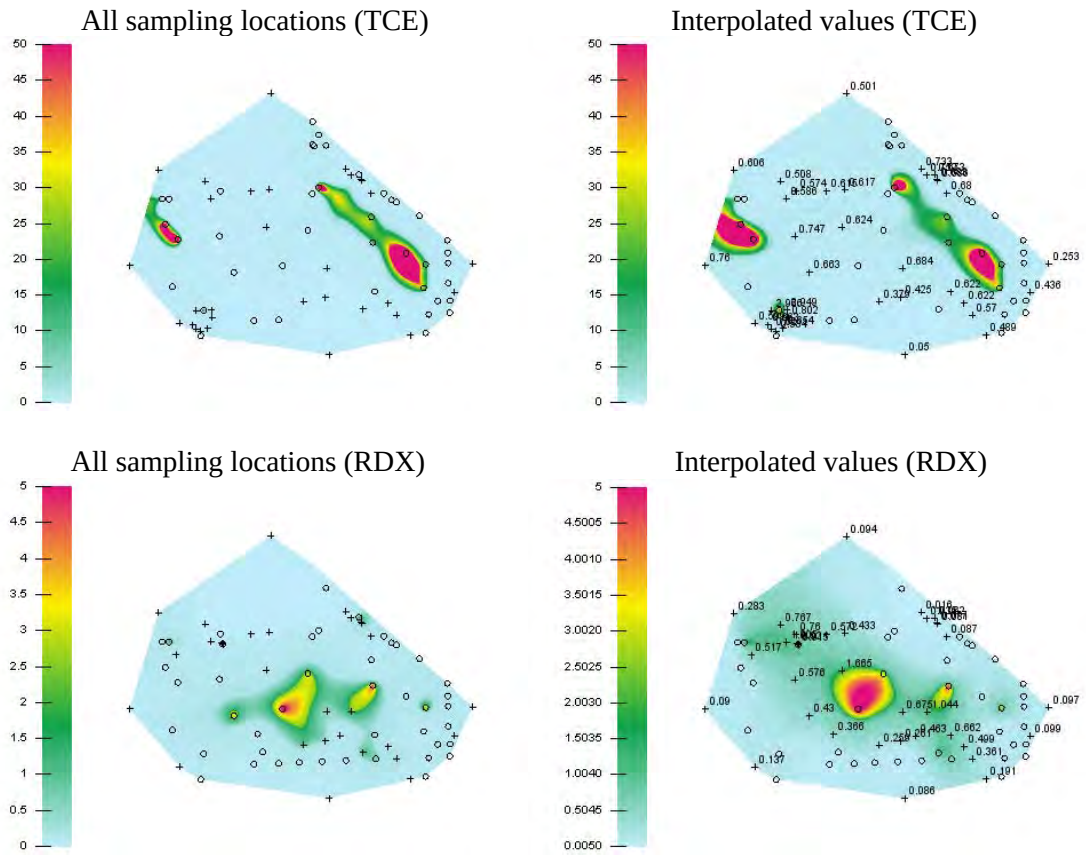
Visually, there are some differences for both TCE and RDX between the full dataset and the interpolated values. For TCE, the area of higher concentrations on the western part of the site is larger, and for RDX the zone of higher concentrations in the middle part of the site is higher. Our analyst believes that site personnel could potentially be concerned by these differences. The table below presents a comparison of the interpolated concentrations at the removed locations versus the actual measured concentrations at those locations in the validation data for Plan 77.

Interpolated Concentrations versus Actual Concentrations at Removed Wells
Optimal Plan 77

Removed Wells	Interpolated Concentrations (µg/l)		Actual Concentrations (µg/l)		Absolute Deviation (µg/l)	
	TCE	RDX	TCE	RDX	TCE	RDX
MW-04A	0.722	0.122	0.05	0.005	0.672	0.117
MW-07A	0.689	0.099	0.05	0.005	0.639	0.094
MW-11	0.548	0.127	0.05	0.005	0.498	0.122
MW-117A	0.607	0.17	0.5	0.1	0.107	0.07
MW-118A		0.163		0.1		0.063
MW-17B	0.94	0.092	0.05	0.005	0.89	0.087
MW-18B	0.603	0.109	0.5	0.1	0.103	0.009
MW-19B	0.61	0.024	0.05	0.005	0.56	0.019
MW-23A	0.69	0.099	0.5	0.1	0.19	0.001
MW-25A		0.162		0.1		0.062
MW-27A	0.729	0.087	0.05	0.005	0.679	0.082
MW-30A	0.729	0.102	0.05	0.005	0.679	0.097
MW-31A	0.519	0.196	0.5	0.1	0.019	0.096
MW-34A	0.688	0.811	0.05	0.43	0.638	0.381
MW-37A	0.509	0.12	0.05	0.005	0.459	0.115
MW-38A	0.47	0.099	0.5	0.1	0.03	0.001
MW-39A	0.384	0.095	0.05	0.005	0.334	0.09
MW-52A	0.525	0.162	0.695	0.1	0.17	0.062
MW-55A	0.423	0.19	0.5	0.565	0.077	0.375
MW-56A	0.447	0.182	0.5	1.1	0.053	0.918
MW-60A	0.384	0.087	0.05	0.005	0.334	0.082
MW-61A	0.64	0.102	0.05	0.005	0.59	0.097
MW-65A		0.195		0.1		0.095
MW-67A		0.862		0.1		0.762
MW-79A	0.636		0.5		0.136	
MW-80A	0.589		0.35		0.239	
MW-81A	0.654		0.5		0.154	
MW-83A	0.442	0.128	0.05	1	0.392	0.872
MW-84A	0.498	0.132	0.05	0.5	0.448	0.368
MW-85A	0.478	0.175	0.05	1.1	0.428	0.925
MW-86A	0.589	0.147	0.5	0.73	0.089	0.583
MW-89A	2.85		0.5		2.35	
MW-91A	1.08		0.5		0.58	
MW-92A	0.687		0.5		0.187	
MW-93A	0.788		0.5		0.288	
MW-99A		1.684		2.6		0.916

The figures below illustrate the plume maps generated for all sampling locations versus the locations recommended by Plan 59. This plan has slightly more error than Plan 77 because more wells are removed. The posted values are the interpolated values at the removed wells.

All Sampling Locations versus Plan 59



One observation is that, for TCE, the high concentration area on the western boundary is enlarged when wells are removed as per the spatial optimization recommendations. The most important factor appears to be the removal of MW-17C. In the discussion of spatial optimization results it was suggested by our analyst that it might be a good idea to keep sampling at this well for this very reason. The table below presents a comparison of the interpolated concentrations at the removed locations versus the actual measured concentrations at those locations in the validation data, for Plan 59.

Interpolated Concentrations versus Actual Concentrations at Removed Wells
Optimal Plan 59

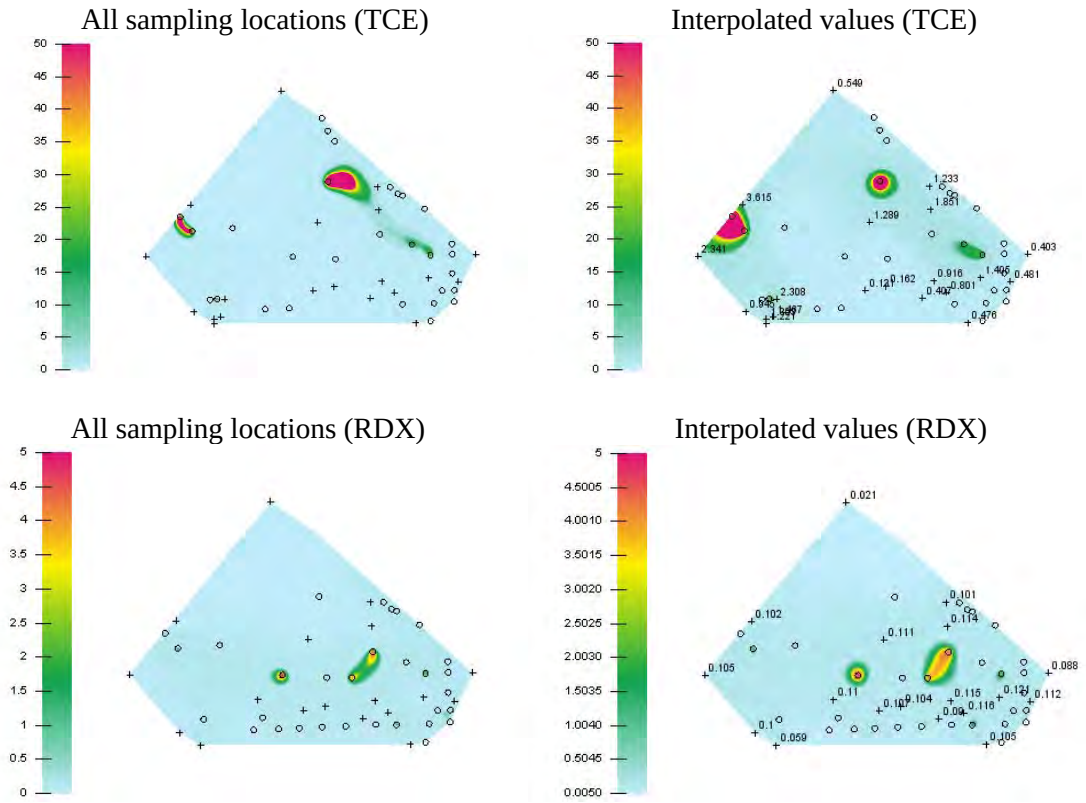
Removed Wells	Interpolated Concentrations (µg/l)		Actual Concentrations (µg/l)		Absolute Deviation (µg/l)	
	TCE	RDX	TCE	RDX	TCE	RDX
MW-04A	0.586	1.005	0.05	0.005	0.536	1
MW-05A	0.574	0.76	2.02	0.005	1.446	0.755
MW-07A	0.617	0.433	0.05	0.005	0.567	0.428
MW-11	0.733	0.016	0.05	0.005	0.683	0.011
MW-117A	0.622	0.499	0.5	0.1	0.122	0.399
MW-118A		0.463		0.1		0.363
MW-17B	0.76	0.09	0.05	0.005	0.71	0.085
MW-18B	0.68	0.087	0.5	0.1	0.18	0.013
MW-19B	0.501	0.094	0.05	0.005	0.451	0.089
MW-23A	0.606	0.283	0.5	0.1	0.106	0.183
MW-25A		0.517		0.1		0.417
MW-27A	0.508	0.767	0.05	0.005	0.458	0.762
MW-28A	0.747	0.576	0.05	0.1	0.697	0.476
MW-29A	0.663	0.43	0.05	3.7	0.613	3.27
MW-30A	0.615	0.572	0.05	0.005	0.565	0.567
MW-31A	0.624	1.665	0.5	0.1	0.124	1.565
MW-34A	0.684	0.675	0.05	0.43	0.634	0.245
MW-35A	0.622	0.662	0.5	0.2	0.122	0.462
MW-37A	0.489	0.191	0.05	0.005	0.439	0.186
MW-38A	0.436	0.099	0.5	0.1	0.064	0.001
MW-39A	0.253	0.097	0.05	0.005	0.203	0.092
MW-52A	0.752	0.035	0.695	0.1	0.057	0.065
MW-54A	0.673	0.062	0.35	1.4	0.323	1.338
MW-55A	0.683	0.077	0.5	0.565	0.183	0.488
MW-56A	0.688	0.081	0.5	1.1	0.188	1.019
MW-60A	0.05	0.086	0.05	0.005	0	0.081
MW-61A	0.599	0.137	0.05	0.005	0.549	0.132
MW-65A		0.92		0.1		0.82
MW-67A		0.915		0.1		0.815
MW-79A	0.62		0.5		0.12	
MW-80A	0.584		0.35		0.234	
MW-81A	0.654		0.5		0.154	
MW-83A	0.379	0.295	0.05	1	0.329	0.705
MW-84A	0.425	0.261	0.05	0.5	0.375	0.239
MW-86A	0.57	0.361	0.5	0.73	0.07	0.369
MW-89A	2.966		0.5		2.466	
MW-91A	2.949		0.5		2.449	
MW-92A	0.663		0.5		0.163	
MW-93A	0.802		0.5		0.302	
MW-94A		0.366		0.1		0.266
MW-99A		1.044		2.6		1.556

Thus, for the intermediate aquifer, the validation exercise does suggest that there might be some minor issues switching to one of the plans with reduced sampling.

Deep Aquifer

Plan 29 (21 of 56 wells removed) and Plan 25 (23 of 56 wells removed) from the Spatial Analysis results were used for the model validation analysis. The figures below illustrate the plume maps generated for all sampling locations versus interpolated plumes for Plan 29. The posted values are the interpolated values at the removed wells.

All Sampling Locations versus Plan 29



Visually, there are some differences for TCE in the western portion of the site (similar to the shallow aquifer results). For RDX the results indicate little difference between the full dataset and the interpolated values. The table below presents a comparison of the interpolated concentrations at the removed locations versus the actual measured concentrations at those locations in the validation data, for Plan 29.

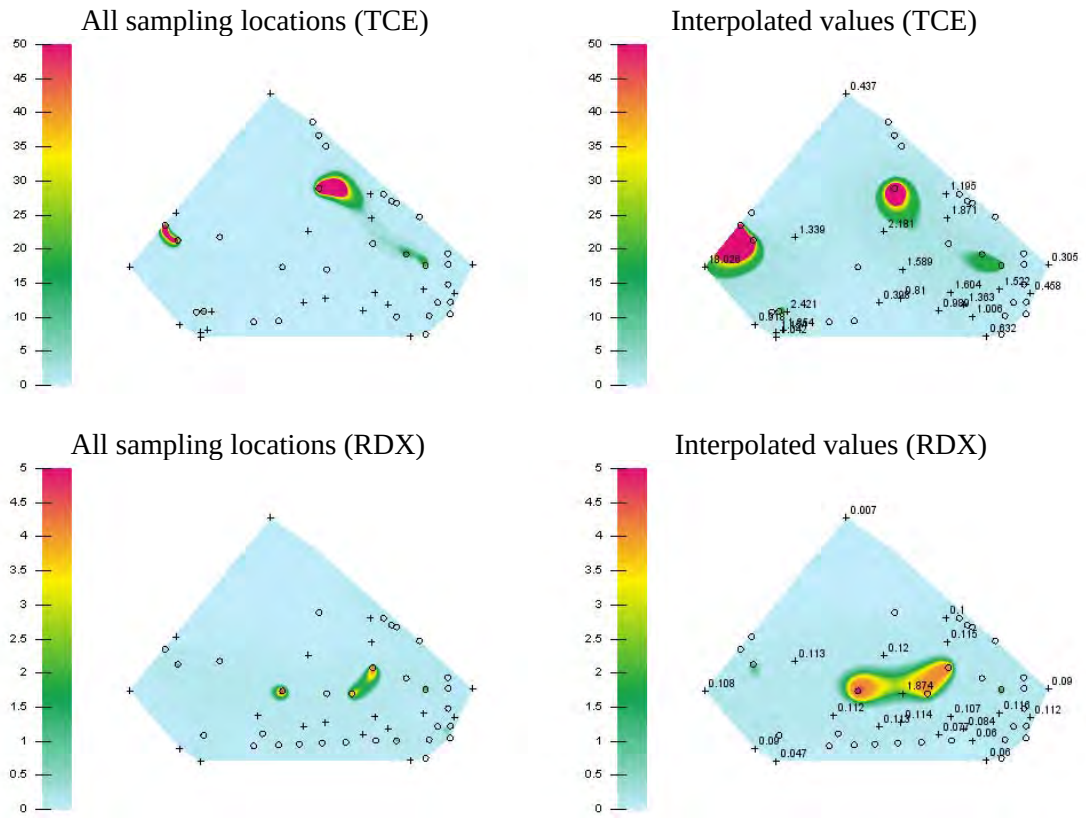
Interpolated Concentrations versus Actual Concentrations at Removed Wells
Optimal Plan 29

Removed Wells	Interpolated Concentrations (µg/l)		Actual Concentrations (µg/l)		Absolute Deviation (µg/l)	
	TCE	RDX	TCE	RDX	TCE	RDX
MW-101D	1.221	0.059	0.5	0.1	0.721	0.041
MW-117D	0.801	0.116	0.5	0.1	0.301	0.016
MW-17A	2.341	0.105	0.05	0.005	2.291	0.1
MW-18A	1.233	0.101	0.55	0.1	0.683	0.001
MW-19A	0.549	0.021	0.05	0.005	0.499	0.016
MW-25D	3.615	0.102	0.05	0.005	3.565	0.097
MW-33D	1.289	0.111	0.13	0.005	1.159	0.106
MW-35D	0.916	0.115	0.5	0.1	0.416	0.015
MW-36D	1.405	0.121	0.05	0.005	1.355	0.116
MW-37D	0.476	0.105	0.05	0.005	0.426	0.1
MW-38D	0.481	0.112	0.5	0.1	0.019	0.012
MW-39D	0.403	0.088	0.05	0.005	0.353	0.083
MW-43D	1.851	0.114	2.88	0.005	1.029	0.109
MW-61D	0.945	0.1	0.05	0.005	0.895	0.095
MW-80D	1.303		0.5		0.803	
MW-81D	1.467		0.5		0.967	
MW-83D	0.131	0.107	0.05	0.1	0.081	0.007
MW-84D	0.162	0.104	0.05	0.1	0.112	0.004
MW-85D	0.407	0.09	0.05	0.1	0.357	0.01
MW-91D	2.308	0.11	0.5	0.1	1.808	0.01
MW-94D						

Our analyst concluded from these results that there were no major concerns with the interpolation with wells removed (as per the spatial optimization recommendations) versus the actual data because of the low concentrations at the wells, but that would ultimately be up to site personnel.

The figures below illustrate the plume maps generated for all sampling locations versus the locations recommended by Plan 25. This plan has slightly more error than Plan 77 because more wells are removed. The posted values are the interpolated values at the removed wells.

All Sampling Locations versus Plan 25



Visually, there are again some differences for TCE in the western portion of the site (similar to Plan 29). For RDX the results indicate more substantial differences in the middle portion of the site than Plan 29, because more wells are removed in Plan 25. The table below presents a comparison of the interpolated concentrations at the removed locations versus the actual measured concentrations at those locations in the validation data, for Plan 29.

Interpolated Concentrations versus Actual Concentrations at Removed Wells
Optimal Plan 25

Removed Wells	Interpolated Concentrations (µg/l)		Actual Concentrations (µg/l)		Absolute Deviation (µg/l)	
	TCE	RDX	TCE	RDX	TCE	RDX
MW-101D	1.042	0.047	0.5	0.1	0.542	0.053
MW-117D	1.363	0.084	0.5	0.1	0.863	0.016
MW-17A	18.026	0.108	0.05	0.005	17.976	0.103
MW-18A	1.195	0.1	0.55	0.1	0.645	0
MW-19A	0.437	0.007	0.05	0.005	0.387	0.002
MW-28D	1.339	0.113	0.05	0.1	1.289	0.013
MW-33D	2.181	0.12	0.13	0.005	2.051	0.115
MW-34D	1.589	1.874	0.05	0.1	1.539	1.774
MW-35D	1.604	0.107	0.5	0.1	1.104	0.007
MW-36D	1.522	0.116	0.05	0.005	1.472	0.111
MW-37D	0.632	0.06	0.05	0.005	0.582	0.055
MW-38D	0.458	0.112	0.5	0.1	0.042	0.012
MW-39D	0.305	0.09	0.05	0.005	0.255	0.085
MW-43D	1.871	0.115	2.88	0.005	1.009	0.11
MW-61D	0.918	0.09	0.05	0.005	0.868	0.085
MW-80D	1.154		0.5		0.654	
MW-81D	1.354		0.5		0.854	
MW-83D	0.398	0.113	0.05	0.1	0.348	0.013
MW-84D	0.81	0.114	0.05	0.1	0.76	0.014
MW-85D	0.989	0.077	0.05	0.1	0.939	0.023
MW-86D	1.006	0.06	0.5	1.5	0.506	1.44
MW-91D	2.421		0.5		1.921	
MW-94D		0.112		0.1		0.012

Our analyst concluded that site personnel might be more concerned with RDX representation in Plan 25 (more wells removed) than with Plan 29 (fewer wells removed).

Discussion of Validation Results

It is noted that, when comparing the interpolated concentrations (at the removed locations) to the actual values, the interpolated values are more often than not higher than the actual values. This might result if the optimization preferentially removes wells with low concentration values. However, it is not clear that this is the cause, or if this is a general result or a site-specific result.

Overall, the validation results do suggest that the sampling plans from the optimization results (i.e., with reduced number of sampling locations) generally represent the contamination patterns, but there are some deviations in some locations that might be of concern to site personnel. As expected, as more wells are removed, the deviations become greater. It would ultimately be up to site personnel to determine if the validation results for any of the plans recommended by the spatial optimization (i.e., for each aquifer) would be acceptable. If not, plans with less error (i.e., with even fewer wells removed) could be selected from the tradeoff curves and subsequently validated.

APPENDIX D:

SITE-SPECIFIC ANALYSIS – CAMP ALLEN SITE

ESTCP Project: LTMO Optimization Software by Summit
Write-Up for Camp Allen Site

Sections

Background Site Information **2**

LTMO Software Demonstrated **6**

Optimization Formulation **7**

Data Preparation **8**

Sampling Optimizer (Including Model Builder) **11**
Brief Overview of Functionality 11
Observations Regarding Use of Sampling OptimizerIncluding Model Builder 14
Computation Time for Model Builder and Optimizer 16
Results for Model Builder and Optimizer 18
Spatial Analysis Results 18
Spatio-Temporal Analysis Results 30
Uncertainty Analysis Results 35

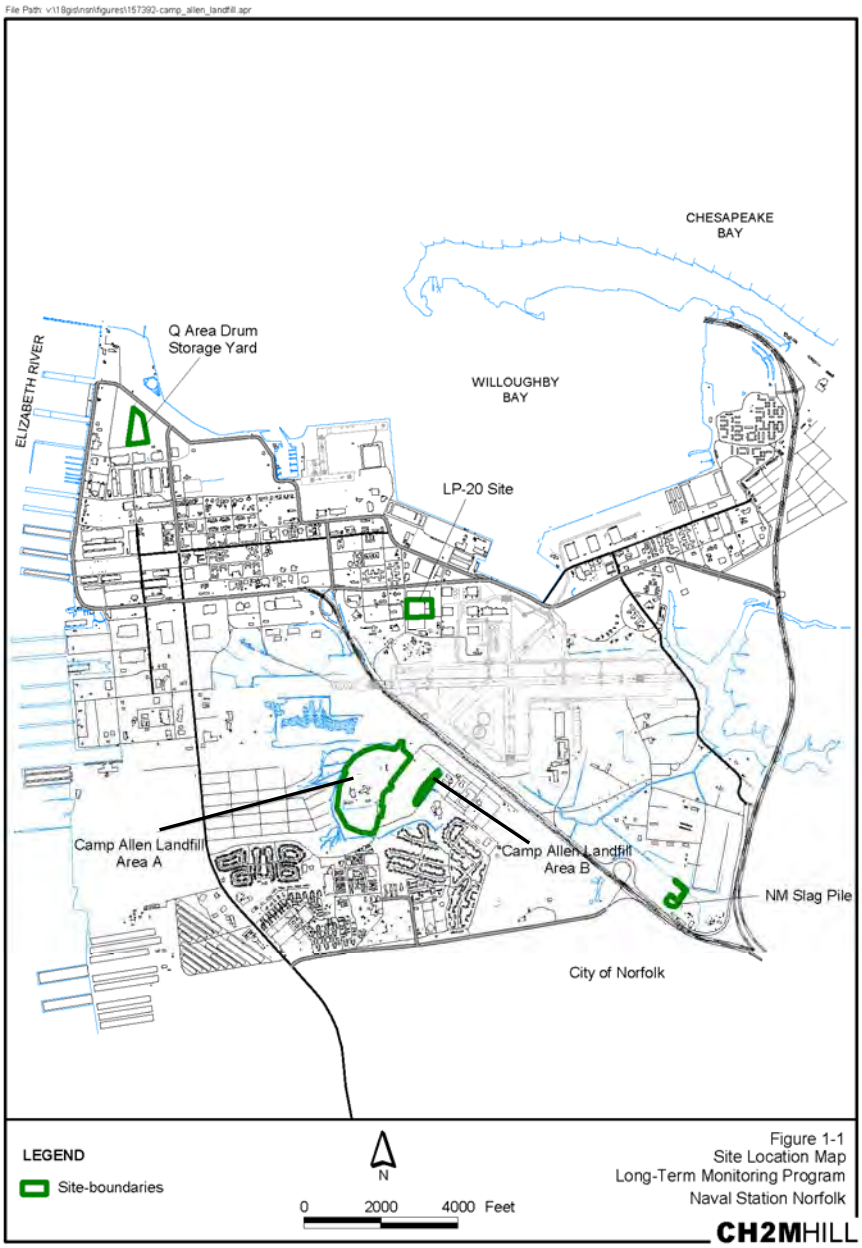
Data Tracker Results (Reported by Charles Davis, EnviroStat) **36**
Overview and Summary 36
Preparing Datasets with Artificial Anomalies for DT 38
Other Data Preparation Issues 40
Results/Conclusions from Applying DT 40
Discussion of DT Results 47

Model Validation **49**
Model Validation for Spatial Analysis 50
Model Validation for Spatio-Temporal Analysis 58

ESTCP Project: LTMO Optimization Software by Summit
Preliminary Draft Write-Up for Camp Allen Site

Background Site Information

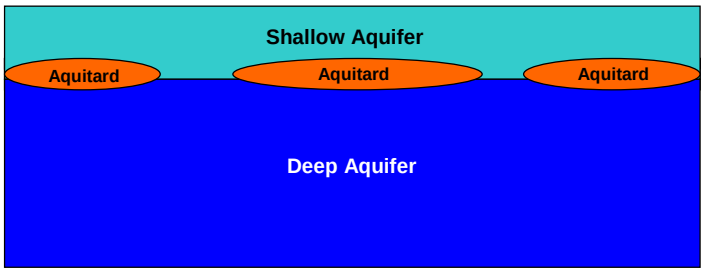
The Camp Allen Landfill is located at the Naval Base Norfolk, Norfolk, Virginia (see Figure below).



The demonstration site is comprised of Landfill Area A (approximately 45 acres), Landfill Area B (approximately 3 acres), and a Salvage Yard located between these two landfill areas. The Camp Allen Landfill site is located in a mixed-use, urban land setting. Military facilities are located atop and adjacent to the landfill areas. There are nine Constituents of Concern (COCs):

Primary COCs	Other COCs
cis-1,2-dichloroethene (c12DCE) **Tricloroethene (TCE)** **Vinyl chloride (VC)**	**1,2-dichloroethane (12DCA)** **1,1,1-Trichloroethane** **Benzene** **Tetrachloroethane (PCE)** **Toluene** **Xylenes**

Two aquifer systems are impacted by the Camp Allen Landfill: the shallow aquifer and the underlying deep aquifer, as illustrated in the schematic figure below.



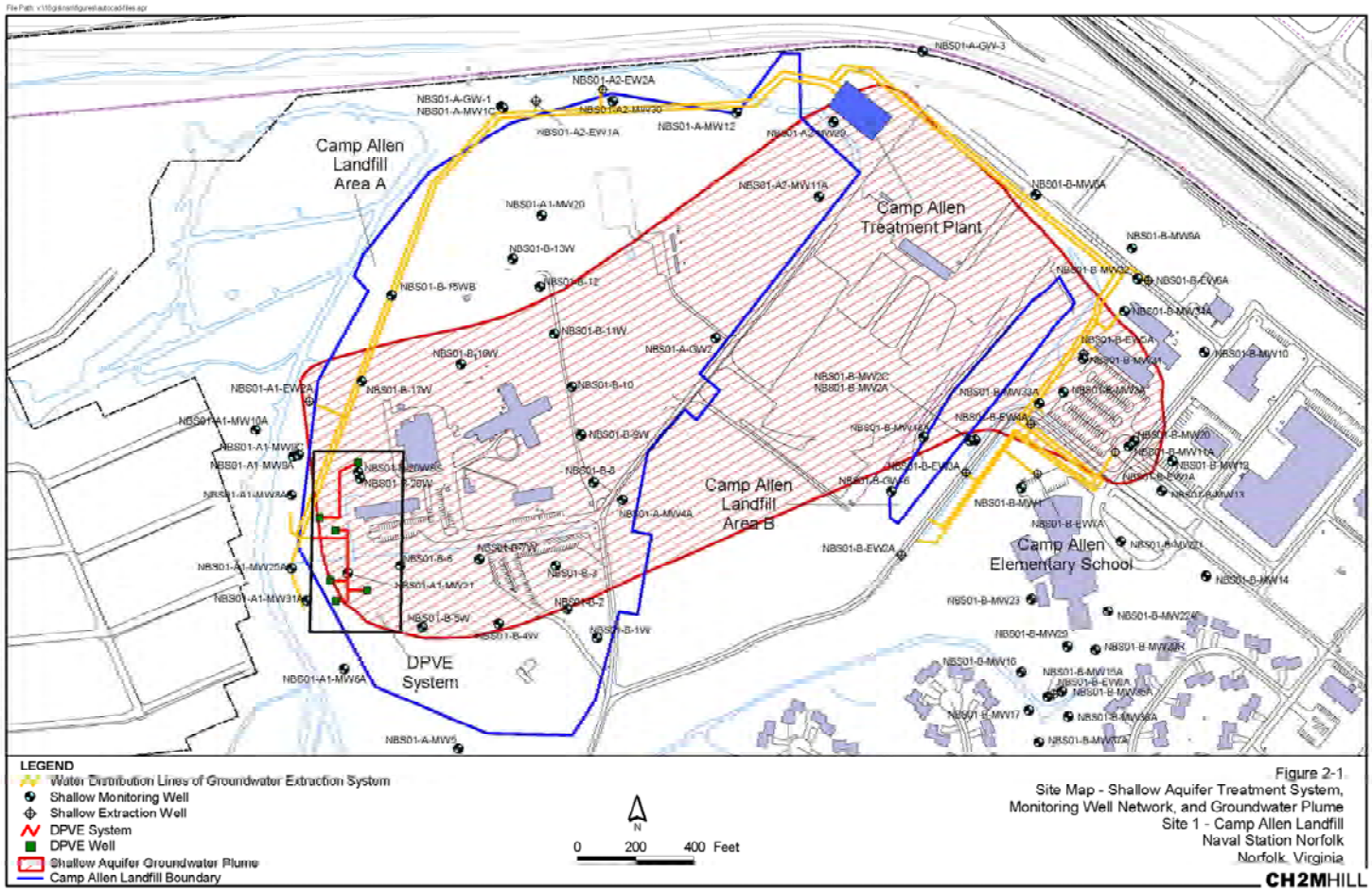
Overall plume extents for the shallow and deep aquifers are shown on figures on the following twp pages. Groundwater flow patterns at this site are complex. The site team indicated that the plume boundaries have been stable for past 3-4 years. The remedial measures at the Camp Allen Landfill consist of groundwater extraction from both aquifers (up to 15 extraction wells total, in both Area A and Area B) and treatment via air stripping. The principle objectives of the remediation systems are:

- To prevent further migration of contaminated groundwater; and
- To restore contaminated aquifers.

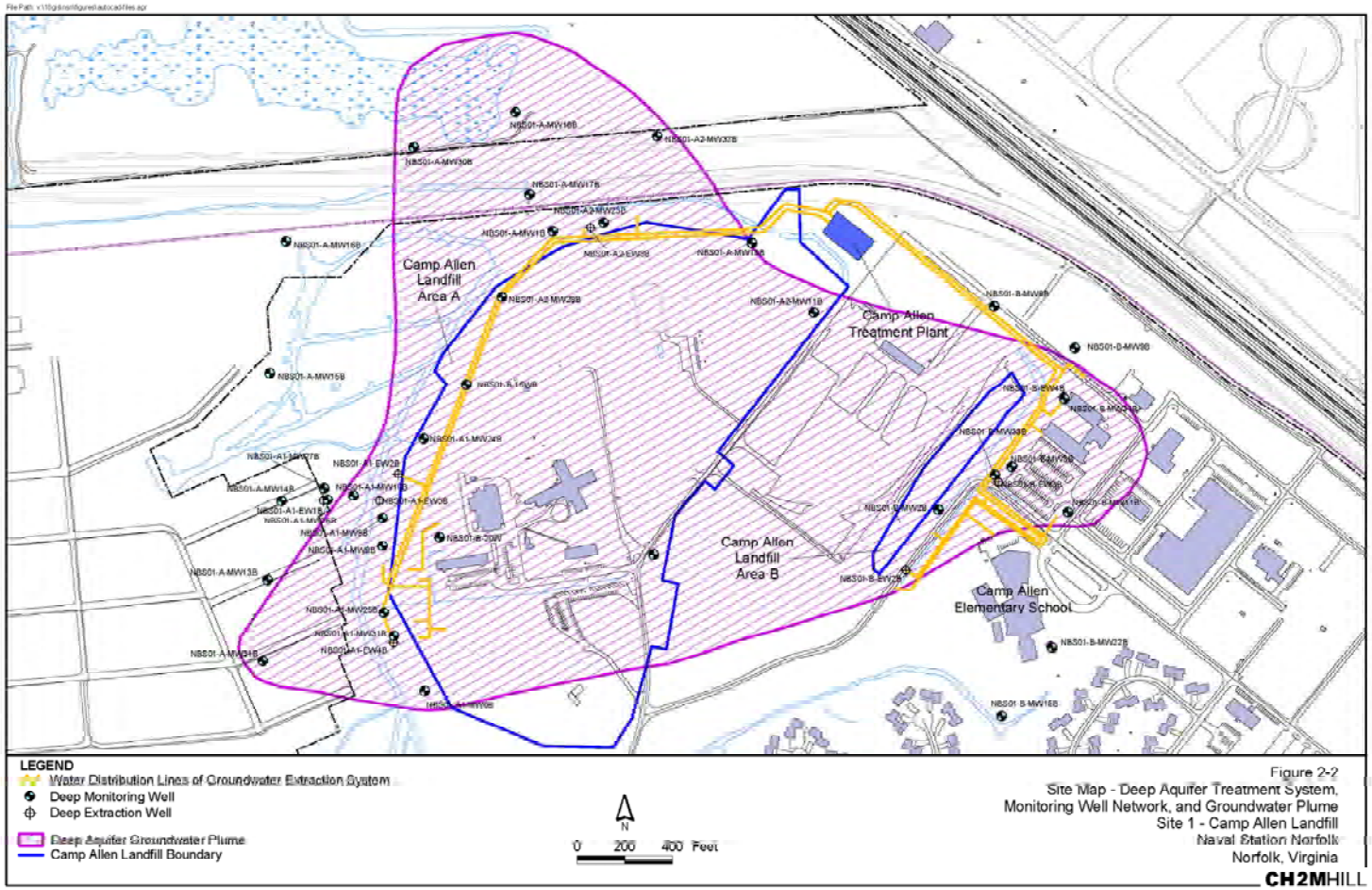
The cleanup goals are different for the shallow and deep aquifers. For the deep aquifer, the cleanup goals are to achieve the federal maximum contaminant levels (MCLs) to protect the aquifer as a potential future drinking water source. For the shallow aquifer, the cleanup goals are based on the assumptions that the shallow aquifer is not a source of potable water and that the exposure will be by incidental ingestions and dermal absorption of contaminants during such outdoor activities as washing cars. Therefore, the cleanup goals for the shallow aquifer are much higher than those for the deep aquifer (e.g., 15,000 µg/l for c12DCE and 1,600 µg/l for TCE). However, during a site visit on January 17, 2007, the remedial project manager (RPM) stated that the cleanup goals for the shallow aquifer may be changed to be consistent with, or closer to, the MCLs.

There are approximately 110 monitoring wells in total. However, some wells are used only for water level measurements. Approximately 50 monitoring wells are currently sampled for groundwater quality. Of these, 45 wells are sampled annually and approximately 5 additional wells are recommended each period (based on a decision process) for monitoring potential plume boundary migration and data gaps.

Plume Boundary in Shallow Aquifer, Camp Allen Landfill



Plume Boundary in Deep Aquifer, Camp Allen Landfill



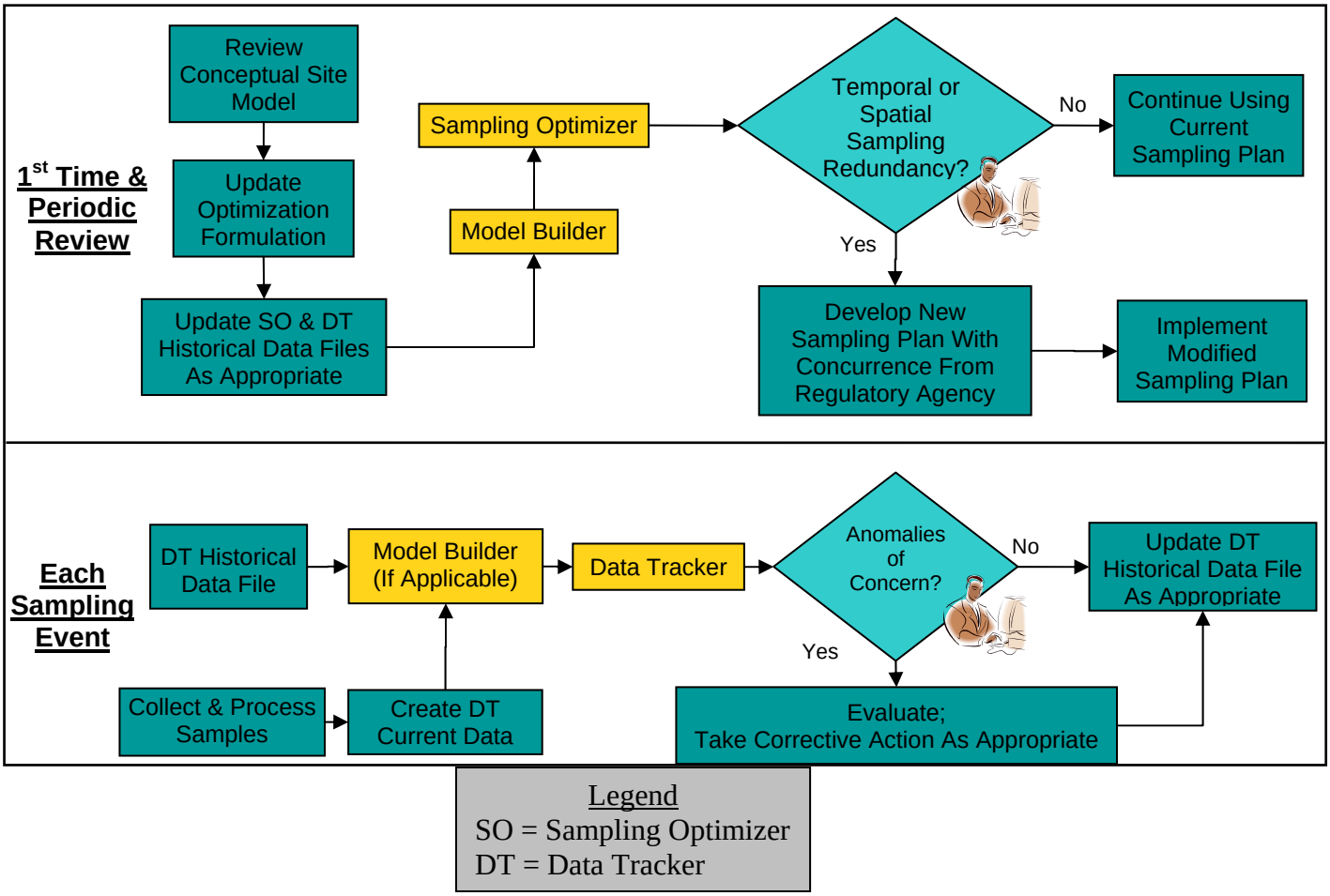
LTMO Software Demonstrated

The goal of this project is to demonstrate and validate the use of the Sampling Optimizer tools (Summit Envirosolutions, Inc.) to reduce cost and improve effectiveness of long term monitoring at DoD sites. Two major modules comprise the Summit software: Sampling Optimizer and Data Tracker.

- Sampling Optimizer identifies redundant sampling locations and/or frequencies in historical data. This module identifies redundant sampling locations and/or frequencies using a multi-objective genetic algorithm to obtain monitoring designs that represent optimal tradeoffs among two or more monitoring objectives, such as minimizing the number of samples and minimizing the concentration error (i.e., error typically increases as number of wells decreases, resulting in a tradeoff)

- Data Tracker allows current monitoring data to be reviewed against historical data to identify cases where current data deviate from expectations that are based on the historical values and patterns.

Model Builder is an additional component within the software that is utilized by the Sampling Optimizer, and in some cases, by Data Tracker. Model Builder has two sections: one for model fitting, visualization, and analysis (with kriging or inverse distance weighting), and another for visualizing relative uncertainty. A general flowchart of the software modules is presented below.



In addition, the MAROS software was applied at Camp Allen to allow for comparison (to the extent possible) of software features and results.

Optimization Formulation

A site visit and meeting with site personnel was conducted on January 17, 2007. Based on this meeting, considerations for optimization objectives and constraints included the following.

- LTM optimization analysis would include both the shallow and deep aquifers in both Area A and Area B.

- The site personnel were interested in using the three primary COCs (c12DCE, TCE, and VC) for Optimizer, and all nine COCs for Data Tracker.

- For Optimizer, accuracy (i.e., low interpolation error) is more important in the plume boundary area than in the plume interior (especially near the source areas).

- For Data Tracker, the site personnel were most interested in data trends over time for monitoring wells (particularly near the plume boundary), as well as related changes in the leading edge of the plume over time.

- Site personnel provided the ESTCP project team with a list of wells that should not be eliminated from the sampling program.

Area A, Area B, and the salvage yard are treated as one combined plume in each aquifer for this project. For each aquifer, concentration data for each primary COC are loaded into the software and considered individually, but the optimization regarding selection of sampling locations and/or frequency (i.e., Sampling Plan Optimizer) is performed simultaneously for all three primary COCs (i.e., there is not one set of wells selected for one COC and another set of wells selected for a different COC).

The Optimizer provides “tradeoff curves” associated with two competing objectives:

- Minimize the number of sampling points; and

- Minimize the maximum concentration errors (emphasizing the importance of errors near the plume boundary) that result from removing specific sampling locations. Since three COCs are evaluated simultaneously, the error for each COC is ultimately combined into a single “merit function” that represents the combined error for all three COCs.

The error function built into the objective function in the software accounts for the different significance of errors near the plume boundary versus the plume interior, placing greater emphasis on the significance of errors near the plume boundary. The user defines a cutoff concentration between low values (i.e., plume boundary) and high values (i.e., plume interior). For locations where the actual value is below the cutoff, the error is calculated as the difference between actual and estimated concentration, divided by the “acceptable error for low concentrations” for the specific parameter. The “acceptable error for low concentrations” is frequently assigned as the MCL. For locations where the actual value is above the cutoff, the error is calculated as the difference between actual and estimated concentration, divided by a specified percentage of the actual value. Details are provided in the software documentation. This function scales the difference between observed and estimated

values by a different amount for different contaminants in the plume boundary area (e.g., by the MCL for each contaminant), and also diminishes the importance of errors in high concentration regions. The following values were specified for the objective function:

	c12DCE	TCE	VC
Acceptable error level for low concentrations (cleanup goals)	70	5	2
Cutoff between low and high concentrations	500	250	100
Acceptable percentage error for high concentrations (%)	50	50	50

Wells were defined as "Plume Interior" if concentrations historically exceeded the following values for any of the three primary COCs since 2000:

- TCE: 250 ppb
- c12DCE: 500 ppb
- VC: 100 ppb

For the Shallow Aquifer, the following wells were in the "Plume Interior":

B-MW3A B-MW11A B-MW15A
B-MW33A B-MW35A B-20W

For the Deep Aquifer, the following wells were in the "Plume Interior":

B-MW3B B-MW11B B-MW19B (abandoned before 2002)
B-MW33B A-MW1B A1-MW31B
A2-MW28B

All other wells were classified as "Plume Boundary". A list of wells that could not be eliminated from future sampling plans was provided by the installation, as follows:

A-MW13B A-MW14B A-MW15B A-MW16B A-MW17B
A-MW31B B-MW15A B-MW11B B-MW2B

Data Preparation

Data were received from Camp Allen personnel on a CD-ROM containing the following information:

- Historical Volatile Organic Compound (VOC) data for 1997-2005;
- Historical water level data for 1997-2005;
- VOC data for 2006;
- Water level data for 2006;

- Well location and construction information and a spreadsheet listing wells not to be removed from the monitoring program; and

- Various other documents and maps, including sets of plume maps for c12DCE, TCE, and VC for the unconfined (shallow) and confined (deep) aquifers for 1999, 2003, 2004, 2005, and 2006, along with a variety of ancillary Arcview files and maps.

Note that water levels were not used within the Summit software, but rather were reviewed along with concentration data as part of the preliminary evaluation of data for reasonableness.

Camp Allen also provided several other useful reports, ranging from the Final Decision Document of 17 July 1995 and the Revised Long-Term Monitoring Plan of 16 September 1998 to the 2006 Analytical Summary Report for 2006. Most useful were the 2004 and 2005 Long-Term Monitoring Annual Reports, particularly the maps and tables therein. Various assessment activities had taken place from 1982 through 1995 at Camp Allen. The 1997 sampling is characterized as "baseline" or "startup" sampling in the 1998 Revised Long-Term Monitoring Plan.

EnviroStat used these data files and reports to prepare CSV files to be used as input for the Summit software as well as other information displays for use by the project team. There were several steps.

- First, a master EXCEL file containing several worksheets was created. The VOC data worksheets are in approximately the format required by the Summit Monitoring Tools. Separate worksheets were created for the historical (1997-2005) and current (2006) data. Worksheets were created for the well location and depth information and for the water level histories. This step required reformatting the data from the format supplied by Camp Allen.

- As part of a preliminary evaluation of data for reasonableness, Well IDs and sampling dates were compared among the VOC data files, the water level files, and the various documents. Some omissions were found, along with discrepancies in data values. These were corrected to the extent possible, with some consultation with Camp Allen personnel. This resulted in adding four wells (B-15WA, B-20W, A-MW19B, and A-MW1B) and over a dozen sampling events to the master file based on information found in maps and tables in the reports. Also, inconsistencies in well IDs were resolved.

- Similarly, well location information was not included in the CD-ROM files for 18 of the 110 wells and piezometers. Northing and Easting were obtained by reading coordinates from maps in reports, in arbitrary units, and interpolating northing and easting from wells included in the files. Well screen depth information is not available for these wells.

- Summaries of data availability were prepared and included in the master file, as were Time Series Plots (TSPlots) of VOC data (1997-2005) for wells with "interesting" histories. TSPlots of water level were included as well. Note that this step is performed to determine if there are potential data quality issues prior to using the software. Likely errors in the database were identified and resolutions suggested.

- Data were reviewed for inclusion or exclusion with regard to use for by Optimizer (SO) and Data Tracker (DT). The master data file includes all data in one set of columns for each VOC, along with an additional set of columns containing the EnviroStat recommendations for data to actually be used.

- A version of the master file excluding the current (2006) VOC and water level data was supplied to GeoTrans and the rest of the project team.

- The master file was then used to prepare the CSV files of "historical data" (five versions - one for spatial SO, two for spatiotemporal SO, and two for DT). The two versions of historical data for spatiotemporal optimization and DT have different treatments of a pair of shallow and deep wells whose data may have been swapped for 2003 and 2004. Normally there would be one historical file for spatial SO, one historical file for spatiotemporal SO, and one historical file for DT.

- Just for our project, CSV files of the "current data" for the DT analysis were supplied to GeoTrans as six different files (one with the actual data, and five containing artificial anomalies). Normally, there would be just one CSV file with the actual current data for DT.

- The previous two steps required some data massaging, such as handling field duplicates, typical non-detects (NDs), and NDs with elevated reporting limits (RLs).

    - Data prior to 2000 were excluded because EnviroStat felt they were not comparable to more recent data for a variety of reasons.

    - For spatial optimization only, which requires one sample value per location, the latest data value for each well was used, so long as that latest value was from 2001 or later, and sample date was artificially set equal to a common value. This is not the only rule that could be used, others might choose to use average values at each location for instance.

    - Values from "Duplicate" QA/QC samples were averaged. For Model Builder, NDs were replaced by a "graphing value" (0.05) that is less than any RL or J value. For DT, values equal to the detection limit were provided to GeoTrans, who subsequently tried several different things for NDs, such as the detection limit, half the detection limit, etc.

    - For spatiotemporal optimization input data, since sampling event dates must be synchronized, the data were grouped into yearly increments, and assigned the artificial date 3/15/yyyy. In the few cases where there were two samples taken during the same year, these were averaged to produce the single value included in the CSV file.

    - For DT, historical data files prepared by EnviroStat contain all historical data from 2001 through 2005, and the current data were from 2006. Actual sampling dates are used by DT.

    - For DT, EnviroStat supplied a column recommending inclusion or non-inclusion of each data values when computing prediction limit bounds within DT. DT uses the prediction limit bounds calculated from the historical data to indicate if the current data are "in bounds" or "out of bounds". In theory, the historical data included for computing the prediction limit bounds would be called "background" data, and would be a subset of the historical data. Historical values that are included in the CSV file but not indicated as to be included in the bounds computation (such as data values where the result was ND but the detection limit is high) would nonetheless be included in data plots that DT creates, but not used to calculate the prediction limit

bounds. That feature is not yet available in the current implementation of DT.

This data preparation effort required several person-days due to the incompleteness of the data files, discrepancies between values in the files and those in the maps and tables in the reports, discrepancies in well IDs, sparse data histories extending over nine years (1997-2005) in some cases, and so on. Of course, such data clean-up is required before data are usable in any optimization analysis, and is not specific to the Summit Monitoring Tools. The only data preparation item specific to the Summit Tools is an artificial alignment of sample dates to designate events (discussed in more detail below); other software products may have their own conventions in this regard.

Some additional observations regarding the data preparation step provided by GeoTrans from using the software include the following:

- The sampling data have to be a CSV file to import into the software. For Model Builder the format is "Date, SiteID, EastCoordinate, NorthCoordinate, COC1, COC2, …". For data tracker the east and north coordinates are optional, and the format of the CSV data is "Date, SiteID, COC1, COC2, …".

- Concentration units have to be consistent over time for each individual COC.

- Different COCs can have different concentration units.

- There is no constraint on the exact chemical name (whatever user enters is OK, which is not the case with MAROS).

- No detection limits and flags are utilized or allowed. However, a "graphing value" for non-detects needs to be assigned to serve as the concentration value for non-detects in the data that are going to be imported. This is similar to a process in MAROS where user enters "detection limit" for non-detects. The user must decide a priori whether to include or delete "high non-detects" (i.e., non-detects with reporting limits higher than actual data values).

- For SO (but not DT) all samples must be assigned to a sampling group in the data to be imported, such that every sample in that sampling event has the same sampling date within the software. For spatiotemporal SO, the time lag between two adjacent sampling events has to be at least quarterly frequency. This data consolidation is done outside of the software prior to import. This is less flexible than in MAROS, because in MAROS such data can be treated as separate events or consolidated as part of one "sampling event" defined by the user within the software.

Sampling Optimizer (Including Model Builder)

Brief Overview of Functionality

Data are loaded, and the user has several options to select for building the model. One set of options pertains to interpolation technique and data transformation:

- Two options for interpolation techniques
    - Inverse distance
    - Kriging

- Three options for data transformation
    - Quantile
    - Log
    - None

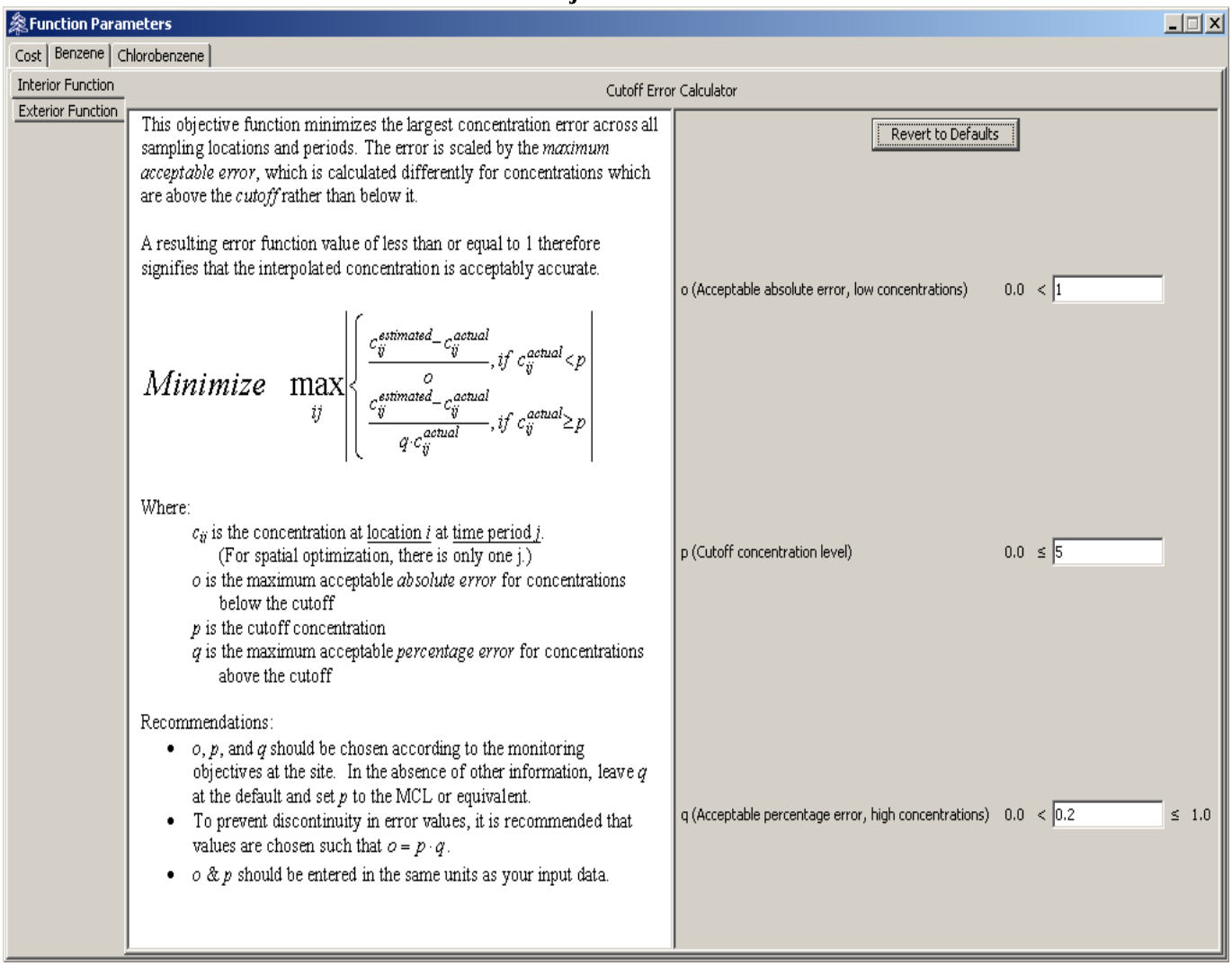
This results in six possible combinations for these basic options. Generally a user will only utilize one combination, and Summit suggests using kriging with quantile transformation. For the ESTCP project, GeoTrans tried all of the combinations, and did in fact determine during their use of the software that kriging with quantile transformation provided the most reasonable representation of the plume distribution (discussed in more detail later).

The Model Builder component of the software provides model fitting, visualization, and analysis functions, as well as maps of relative uncertainty. Within the Model Builder component of the software the user defines options for the parameters of the interpolation technique selected by the user for the Optimizer. Both automated and manual model parameter fitting are supported for Kriging, while the user must manually specify the power to be used for inverse distance weighting. Within the Model Builder component the user also specifies desired changes to the defaults on the "Model Builder Settings" screen, such as the number of vertical slices that defines the resolution of the image. If the data imported into the Sampling Optimizer has multiple events, Model Builder provides visualization for each event. The Optimizer module uses the model parameters specified within the Model Builder component.

Sampling Optimizer uses a genetic optimization algorithm to suggest favorable monitoring plan alternatives relative to the base sampling plan (i.e., where one or more of the samples are removed). In spatial optimization, the original model is based on one set of sampling data that do not vary in time, and the optimization results are with respect to sampling locations only. In spatiotemporal optimization, the original model consists of actual data that vary in space and time, and the optimization results are with respect to sampling location and sampling frequency. Temporal analysis is a subset of spatio-temporal analysis where wells cannot be removed. The user can utilize software defaults for the optimization algorithm (e.g., population size) or can specify values for these parameters in the "GA Settings" screen.

For this site the "errors" were calculated by using the "Cutoff Error Calculator" option for the objective function provided within Optimizer (the other option is the "Percentage Error Calculator"). The Cutoff Error Calculator incorporates a function for calculating error associated with samples that are removed that is displayed below.

Error Objective Calculator



The overall purpose of Error Calculator is to come up with an objective function value that represents the overall similarity of a new sampling plan to the baseline sampling plan. The Cutoff Error Calculator is designed so that "error" is calculated in a manner that makes deviations between interpolated and actual values more significant in areas of low concentration versus areas of high concentration. This is accomplished as follows:

- The user defines a cutoff concentration (p) for the actual data values that differentiates between low concentrations versus high concentrations, and also defines a value for Acceptable absolute error (o).

- When a low concentration data point is removed (i.e., below the cutoff), error is calculated as the absolute value of the actual value minus the interpolated value, divided by the acceptable absolute error. For example, if the actual value is 8 µg/l (i.e., below the cutoff concentration of 10 µg/l) and acceptable absolute error is 1.0, and the difference between the actual and interpolated value is 5 µg/l , then the error would be 5 / 1 = 5.

- When a high concentration data point is removed (i.e, above the cutoff), error is calculated as the absolute value of the actual value minus the interpolated value, divided by a percentage (q) of the actual value, where q is specified by the user. For example, if the actual value is 100 µg/l (i.e., above the cutoff concentration of 10 µg/l) and the percentage input by the user is 10%, and the difference between the actual and interpolated value is 5 µg/l, then the error would be 5 / (0.10 * 100) = 0.5.

In these examples, the difference between the actual value and the interpolated value was 5 µg/l in both cases, but in the first case the calculated error is 5.0 whereas in the second case it is only 0.5. This illustrates how the calculation increases the significance of deviation between actual and interpolated values in the lower concentration areas of the plume.

Observations Regarding Use of Sampling OptimizerIncluding Model Builder

The following observations were made by GeoTrans based on application of Sampling Optimizer (including Model Builder) for Camp Allen in conjunction with both spatial and spatiotemporal analysis:

- The software is very easy to use. However, the post-software analysis of results can take quite some time to analyze the resulting monitoring plans, especially when multiple COCs are present such as at Camp Allen. This is because the software provides a tradeoff curve for each COC, each of which consists of numerous potential monitoring plans. To make sense of this, the user ideally wants to look at one tradeoff curve that represents the results for all three COCs, and this can only be done outside the software using MS Excel in conjunction with software output, by defining a "merit function (e.g., sum of errors for all three COCs plotted versus number of wells). This approach allows the user to merge the results for multiple COCs into one combined tradeoff curve. This is discussed more later, and Summit is considering adding this type of functionality within future versions of the software.

- The software allows the user to easily save a project and re-open it later. However,, it was noted that some updates to the software made during the project prevented previously saved projects from opening.

- The software allows the user to enter run titles which are used as the part of the file names when exporting the results for both Model Builder and Optimizer.

- The software uses a "seed" value for certain calculations. A genetic algorithm does not guarantee an optimal solution, just one that has high probability of being very close to optimal, and a different solution may be obtained if a different seed value is utilized. The software uses default seed values that are fixed for Sampling Optimizer, which ensures that the same results can be obtained by different users with the same parameter settings. The user can change the seed manually if desired.

- Model Builder provides visualization of plume distribution and plume uncertainty. It gives users the option to post either well names or measured concentrations. This is a useful feature, more advanced than MAROS.

- The "visualization resolution" (the user controls this by defining the number of vertical slices for the image) has a big impact on whether the plume generated by Model Builder can be

correctly displayed. The visualization resolution can be modified in "# of vertical slices for image" of "Model Builder Setting". To better match the actual data when viewing the plume maps, it was found that a higher resolution than the default value of 50 was appropriate. For Camp Allen, resolution of 200 was used.

- For the Sampling Optimizer (SO) module, the "Well Constraint" feature allows the user to specify the maximum sampling frequency and the minimum sampling frequency for each well. For spatial analysis, this feature allows the user to specify which wells cannot be removed from the system (which may be specified in the optimization formulation). This can also be useful for abandoned wells which may be part of the historical data but cannot be sampled in the future, by specifying such wells as "always off".

- The software currently has one general type of objective function available, allowing the user to enter an acceptable error level for lower concentration points, the cut-off concentration between high and low concentration points, and the acceptable error percentage for high concentration points. This was consistent with the formulation for Camp Allen.

- For the Camp Allen site, a population size (utilized for the genetic algorithm) of 1,000 for SO was recommended by Summit to ensure that "good" solutions (i.e., calculated maximum errors for each COC at removed points, which are scaled within the software based on the objective function, are less than 1.0) can be found, but using a population size that large significantly increased the computational time and caused the software to come to a near-halt when saving the project file. Thus, a population size of 800 was used for *spatial analysis* for Camp Allen, because it's close to recommended value of 1000 but the software did not come to a near-halt and generally finished in minutes to several hours, depending on the interpolation technique. However, with population size of 800 for *spatial-temporal* analysis, the software came to a near-halt after 7 – 8 hours running. Using population size of 500 prevented the software from the near-halt in some optimizations, but not in others (that is for identical runs, suggesting a potential bug). Computation times are discussed in more detail later.

- Plume visualization for Optimizer provides users the option to post either the well names, the measured concentrations at all locations, or the "interpreted" concentrations at the removed locations, which are the basis for the error calculations. This is a really good feature.

- Plume visualization for both Model Builder and Optimizer also allows users to change the zoom scale and color scale. The color scale is a linear scale allowing users to define the minimum and maximum concentrations for each COC. Then the software can plot the plume maps in color based on the minimum and maximum concentrations defined. However, it does not provide an option for a logarithmic scale, thus, for sites with a very big range in concentrations such as Camp Allen, it cannot plot both high-end concentrations and low-end concentrations with sufficient detail.

- For multiple COCs, the software generates a tradeoff curve for each COC. The tradeoff is the sampling cost (i.e., number of wells) versus the maximum errors for each COC at removed points. The error is a value that is scaled within the software based on the parameter values input as part of the objective function. These results for plans that define the tradeoff curve within the software can be exported to a CSV file and imported into MS Excel. For Camp Allen, this CSV file generally contained several hundred potential plans. Summit indicated that "good" solutions have error less than 1.0 for each COC. Based on that rule, potential plans were narrowed down from several hundred plans that define the tradeoff curves to

around 50 plans with acceptable errors. A single "merit function" was then be developed outside of the software (i.e., in MS Excel) to combine the errors for each COC for these narrowed-down plans into a composite error for all COCs (e.g., add the individual errors for each to get one composite value for error).

- The software allows the user to export the following files:
    - The plume maps, uncertainty maps, and variogram charts (for kriging model type only) can be exported as image files (.png files).
    - A tradeoff curve for each COC can be exported as an image file.
    - A file containing, for each optimization sampling plan on the tradeoff curve, which wells are recommended to be "on" or "off", the maximum concentration error for each COC, and the sampling cost (i.e., number of wells which are on) can be exported as a CSV file.
    - For each optimization sampling plan on the tradeoff curve, an individual listing of which wells are on and which wells are off can be exported as a CSV file.
    - For each potential optimization sampling plan on the tradeoff curve (and for the current sampling plan), a plume map can be exported as image file for each COC with symbols indicating which wells are on and which wells are off, with "+" indicating wells that are recommended to be removed from the monitoring network and "o" indicating remaining active wells.

Computation Time for Model Builder and Optimizer

The computation time depends on the size of the dataset (e.g., number of wells) and model type selected (e.g., kriging versus inverse distance, plus the type of data transformation). Computation time for Model Builder also increases significantly with the increase in resolution (i.e., number of vertical slices for image), and computation time increases with increased population size for Optimizer. An estimate of the amount of time it takes to apply Model Builder and Optimizer to evaluate one plume (i.e., one plume with multiple constituents in one aquifer) at Camp Allen, after the initial "learning curve" was overcome, is as follows:

- The data provided by EnviroStat were already in the right format required by the software. To import these data into Model Builder, all we needed to do was to separate shallow aquifer wells and deep aquifer wells (changes to sampling dates for grouping into events was already done as part of data preparation). This took a few minutes.
- Importing the concentration data into the actual software took seconds.
- Next, it took minutes to enter the facility ID and choose the model type (inverse distance or kriging) and data transformation type (quantile, log, or none).
- For the Camp Allen site with 30 – 40 sampling locations in each aquifer, longer computational time is expected for Model Builder with increasing resolution (i.e., number of vertical slices for image).

- With resolution of 50 (default value), it took about 5 minutes for inverse distance weighting and 30 to 40 minutes for kriging to build the model 1GB RAM PC (Pentium 4, 3.2 GHz);

- With resolution of 200, it took about 10 – 20 minutes for inverse distance weighting and 2 – 4 hours for kriging to build the model 1GB RAM PC (Pentium 4, 3.2 GHz).

- To ensure adequate results for the interpreted models, resolution of 200 was used for all the simulations, otherwise solutions with acceptable errors were not always found.

- Visualizing the plume maps and uncertainty maps inside the software, and exporting them if desired, took minutes.

- After performing Model Builder, it took several minutes to set up the well constraints and objective function for Optimizer.

- Optimizer took several minutes up to several hours to complete the optimization process (three COCs) depending on model type and population size. Computation time is also heavily influenced by PC RAM.

    - Increasing population size from 300 (default value) to 800 with inverse distance weighting increased computation time for spatial analysis from about 5 minutes to about 20 minutes on 1GB RAM PC (Pentium 4, 3.2 GHz).

    - With population size of 800, Optimizer for spatial analysis took about 20 minutes for inverse distance weighting and 30 – 60 minutes for kriging on 1GB RAM PC (Pentium 4, 3.2 GHz).

    - When running on 512MB RAM PC (Pentium 4, 3.0 GHZ), it took several hours to complete the spatial analysis optimization process for inverse distance weighting with population size of 800, while it only took 20 minutes on 1GB RAM PC (Pentium 4, 3.2 GHz).

    - With spatio-temporal analysis and population size of 500, it took 5 – 6 hours for both inverse distance weighting and kriging on 1GB RAM PC (Pentium 4, 3.2 GHz) with more frequent near-halt occurring than the spatial analysis.

    - For final results, population size of 800 was used for spatial analysis and population size of 500 was used for spatio-temporal analysis.

    - Large population size can cause the software to occasionally come to a near-halt.

        - With population size of 1000 (originally recommended by Summit to ensure that good solutions can be found) for spatial analysis, the software came to a near-halt when saving the project after completing the Optimizer. Thus, a population of 800 was used for spatial analysis.

        - With several attempts for population size of 800 for spatio-temporal analysis, the software always came to a near-halt when running Optimizer. Thus, a population of 500 was used for spatio-temporal analysis.

- With population size of 500 for spatio-temporal analysis using inverse distance weighting, the software only completed Optimizer after several failed attempts. Failure appears to be somewhat random since the same optimization ultimately can be run to completion with identical input parameters.

- Finally, reviewing plume maps for the potential plans (which we limited to plans with errors for each COC less than 1.5) and exporting them to image files took several minutes to up to several hours, and took several iterations to choose a good zoom scale and color scale.

Results for Model Builder and Optimizer

A total of 14 optimizations were performed with different combinations of interpolation techniques and data transformation (see table below).

Analysis	Aquifer	Interpolation	Data Transformation	Optimization Name
Spatial	Shallow Aquifer	Inverse Distance Weighting	Quantile	S-SIQ
			Logarithm	S-SIL
			None	S-SIN
		Kriging	Quantile	S-SKQ
			Logarithm	S-SKL
			None	S-SKN
	Deep Aquifer	Inverse Distance Weighting	Quantile	S-DIQ
			Logarithm	S-DIL
			None	S-DIN
		Kriging	Quantile	S-DKQ
			Logarithm	S-DKL
			None	S-DKN
Spatio-temporal	Deep Aquifer	Inverse Distance Weighting	Quantile	ST-DIQ
		Kriging	Quantile	ST-DKQ

Of these, 12 of 14 optimizations were performed for spatial analysis, with 6 of those 12 optimizations for the shallow aquifer and the other 6 optimizations for the deep aquifer. Two of 14 optimizations were performed for spatio-temporal analysis, for the deep aquifer only, since there are not enough data to conduct a spatio-temporal analysis for the shallow aquifer. Note that we did this many variations to test the software, and generally the user would select one combination for interpolation and transformation (e.g., kriging with quantile transformation) rather than trying all six combinations.

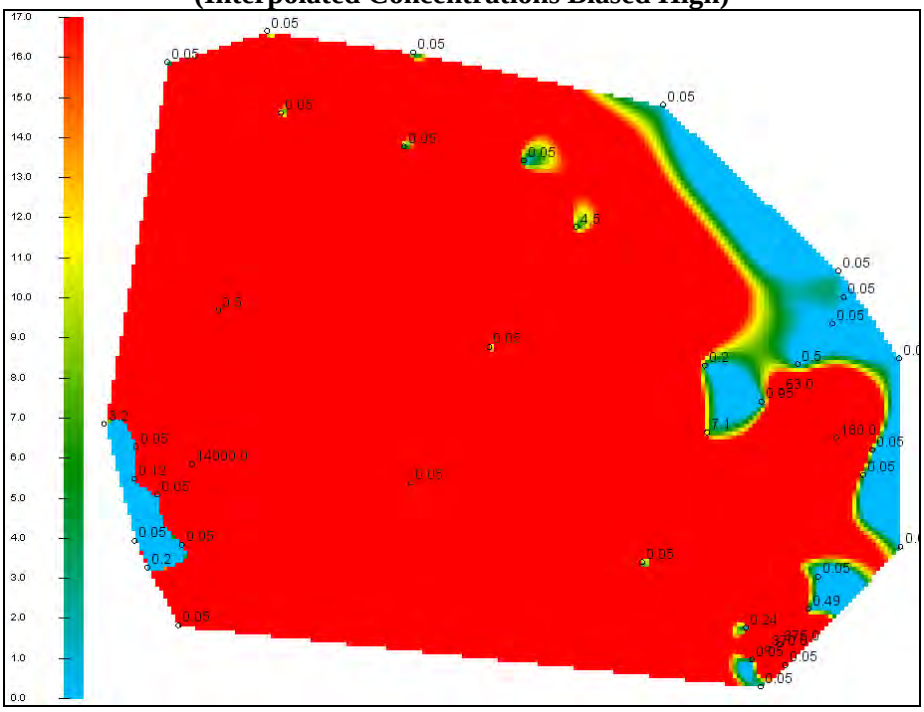
Spatial Analysis Results

A total of 6 spatial analysis optimizations were performed for the shallow aquifer and 6 spatial analysis optimizations were performed for the deep aquifer. Each of the six optimizations had a different combination of interpolation technique and data transformation. Observations regarding Model Builder are summarized below for the different combinations.

- Of the six unique combinations of different interpolation and data transformation for each aquifer, the following three generally produced more reasonable visual representations of the plume: 1) kriging with quantile transformation; 2) inverse distance weighting with quantile transformation; and 3) inverse distance weighting with logarithm transformation.

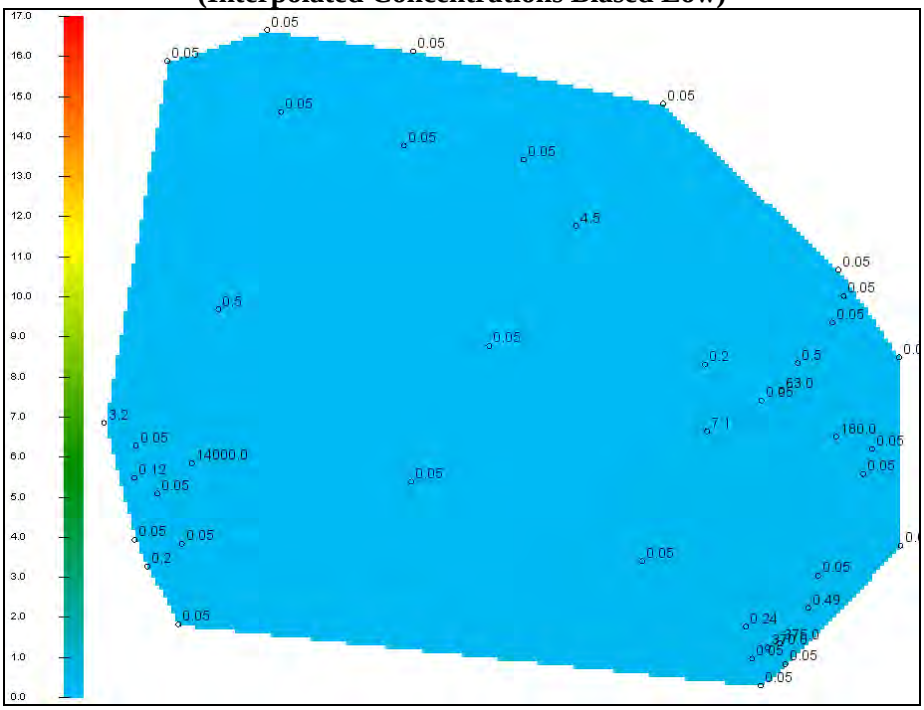
- Both inverse distance weighting with no data transformation and kriging with no data transformation resulted in models that are biased to the high concentration, i.e., concentrations at the locations without actual measurements are interpreted much higher that they should be (see figure below for example, where high values seem to dominate areas that are in fact dominated by NDs).

Example: Kriging with no Data Transformation, Shallow Aquifer, VC (Interpolated Concentrations Biased High)



- Kriging with logarithm transformation may also not provide reasonable representations. For example, for VC the concentrations are biased too low (see figure below, where no high concentration areas are plotted around the high measured concentrations). Note the concentrations posted at the measured locations are the actual measured concentrations.

Example: Kriging with no Log Transformation, Shallow Aquifer, VC (Interpolated Concentrations Biased Low)



The table below summarizes, for each combination of interpolation type and data transformation, the conclusion regarding model quality along with an indication of associated computation time for Optimizer.

Summary of Model Quality and Computation Time* for Each Combination Of Interpolation technique and Data Transformation, Camp Allen Data (Spatial Optimization)

Interpolation	Data Transformation	Comments
Inverse Distance Weighting	Quantile	Reasonable models obtained, in 10 – 20 minutes, with Model Builder 10 – 15 minutes for Optimizer
	Logarithm	Reasonable models obtained, in 10 – 20 minutes, with Model Builder Approximately 20 minutes for Optimizer
	None	Bad models obtained, in 10 – 20 minutes, with Model Builder 10 – 20 minutes for Optimizer
Kriging	Quantile	Best models obtained, but required 2 – 4 hours, with Model Builder Approximately 30 minutes for Optimizer
	Logarithm	Questionable model for VC, and required 2 – 4 hours, with Model Builder 30 – 60 minutes for Optimizer
	None	Bad models obtained, and required 2 – 4 hours, with Model Builder 30 – 60 minutes for Optimizer

**1GB RAM PC (Pentium 4, 3.2 GHz)*

For each Optimizer run, the software resulted in tradeoff curves for each COC containing several hundred plans. Post-software analysis is required to select a subset of the plans and to generate one single tradeoff curve that represents all COCs (if desired). The following post-software analysis was conducted:

- Select the plans with errors less than 1.5 for each COC. Note we used 1.5 instead of the 1.0 error limit recommended by Summit to allow additional optimal plans with larger errors to be considered.

- Create a "combined" tradeoff curve in Excel for the selected plans by adding the errors for each COC to get one composite error.

- Select final plan(s) based on the combined tradeoff curve and a subjective review of the plume maps generated within the software for each COC for that specific plan (i.e., if the selected plan does not provide a "reasonable" representation of the plume based on visual inspection, a different plan along the composite tradeoff curve is selected).

The following table lists number of optimal plans and minimum number of wells within error scale of 0.5, 1.0, and 1.5 for each COC. The smaller the minimum number of wells, the greater the number of wells that can be removed from the monitoring network.

{this gap is intentional}

Summary of "Optimal Plans" Identified by Optimizer – Spatial Optimization

Simula -tion	Interpola-tion	Data Transfor-mation	# Plans w/ Errors < 0.5 for Each COC	# Plans w/ Errors < 1.0 for Each COC	# Plans w/ Errors < 1.5 for Each COC	Min. # of Wells in Plan w/ errors < 0.5 for each COC	Min. # of Wells in Plan w/ errors < 1.0 for each COC	Min. # of Wells in Plan w/ errors < 1.5 for each COC
Shallow Aquifer (Baseline Model has 42 Wells)								
S-SIQ	Inverse Distance Weighting	Quantile*	32	40	52	19	18	16
S-SIL		Logarithm*	19	29	37	27	22	20
S-SIN		None	3	6	6	38	35	35
S-SKQ	Kriging	Quantile*	44	50	62	17	16	14
S-SKL		Logarithm	44	69	92	22	17	14
S-SKN		None	1	1	1	42	42	42
Deep Aquifer (Baseline Model has 31 Wells) **								
S-DIQ	Inverse Distance Weighting	Quantile*	3	9	10	28	26	26
S-DIL		Logarithm*	2	12	17	30	23	22
S-DIN		None	1	3	3	31	30	30
S-DKQ	Kriging	Quantile*	5	18	29	28	21	21
S-DKL		Logarithm	6	27	49	25	20	18
S-DKN		None	2	2	2	30	30	30

** combinations where Model Builder was found to provide "reasonable" representations of plume based on measured values*
*** two abandoned wells are included when performing Optimizer, otherwise no plan could be found with errors less than 1.5 for each COC, this is discussed in more detail later*

Based on the evaluations on applications of Model Builder and Optimizer, different combinations of interpolation and data transformation are subjectively ranked below from the best to the worst:

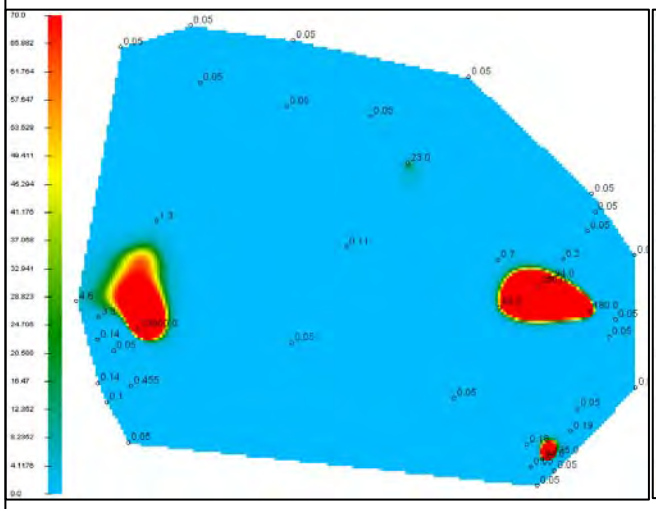
- Kriging with quantile transformation
- Inverse distance weighting with quantile transformation
- Inverse distance weighting with logarithm transformation
- Kriging with logarithm transformation
- Inverse distance weighting with no data transformation
- Kriging with no data transformation

Although kriging with quantile transformation (the "top choice") took longer than the next two choices for both Model Builder and Optimizer, we preferred it because we liked the way it

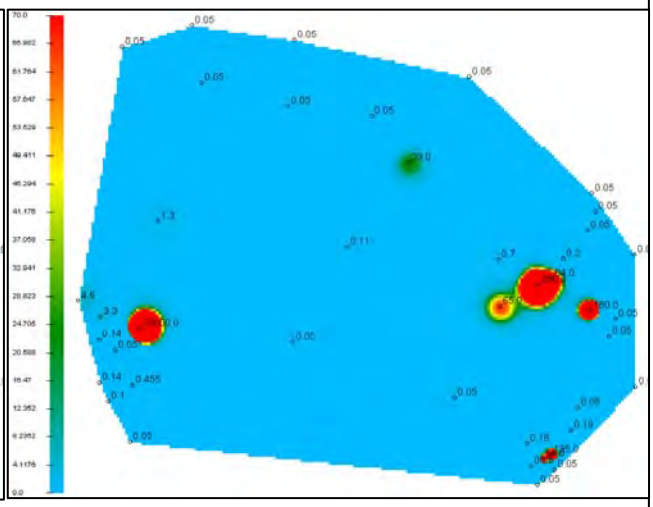
transitioned from highest concentrations to lower concentrations and it provided many potential plans with errors less than 1.0 for each COC. We chose Optimizer results using kriging with quantile transformation for our post-software analysis.

Example Model Builder Results, c12DCE, Shallow Aquifer
(for three preferred combinations of interpolation technique and data transformation)

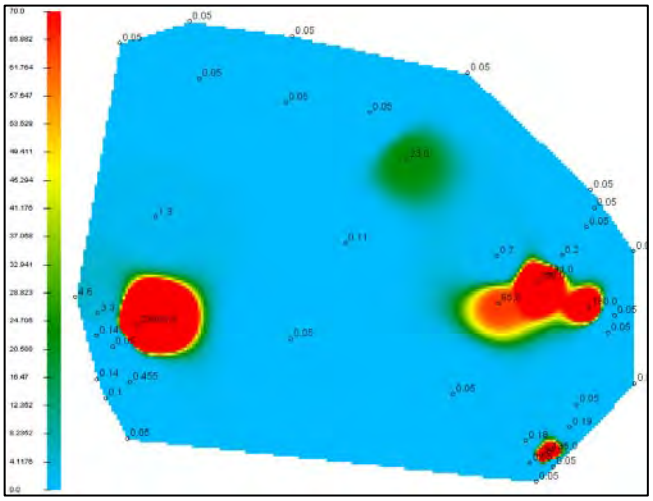
Kriging with quantile transformation



Inverse distance weighting with quantile transformation



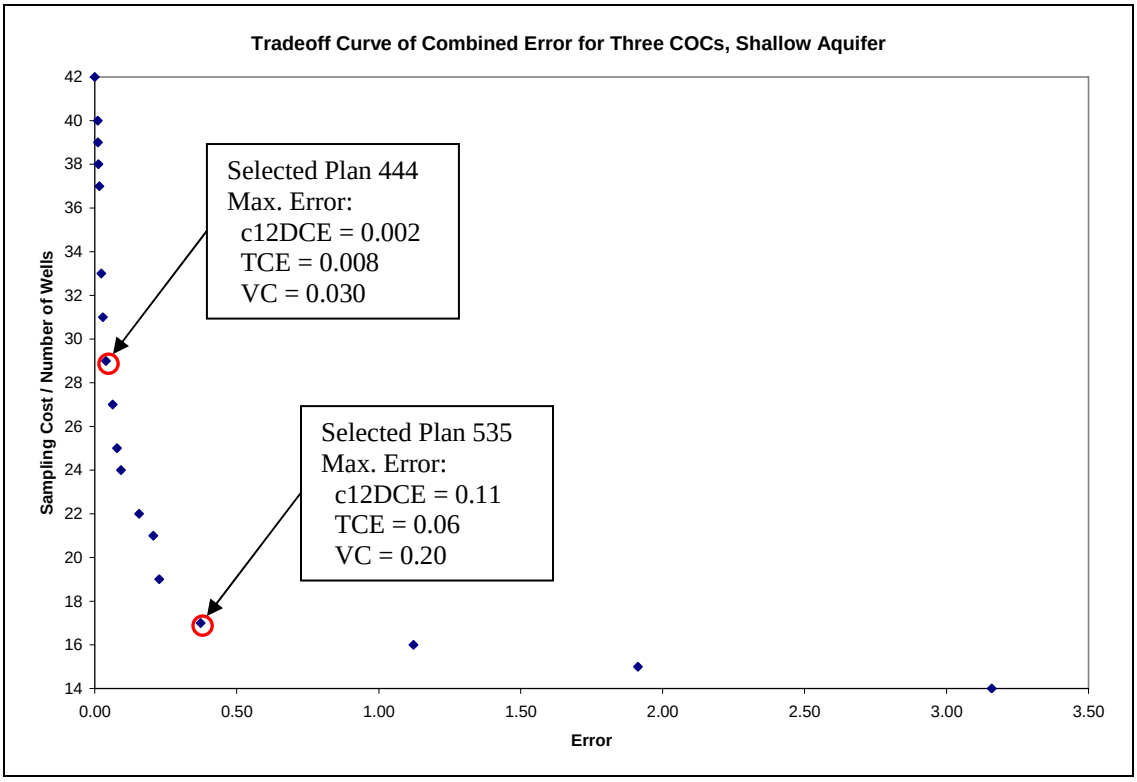
Inverse distance weighting with log transformation



Shallow Aquifer Tradeoff Curves

Below is the combined tradeoff curve for the shallow aquifer (computed outside the Summit software, using MS Excel to post-process results exported from the Summit software), showing all the plans with errors less than 1.5 for each COC (i.e., the sum could be as high as 4.5). By comparing sampling cost (number of wells) and errors on the combined tradeoff curve, and by visually inspecting the plume maps for selected plan(s) and comparing to the plume maps with all wells, Plan 535 was selected as a promising plan. This plan reduces the number of wells (relative to the base model for the shallow aquifer) from 42 to 17, with little increase in error relative to plans with a greater number

of wells. Other identified plans with 17 wells or less have significantly greater error. Plan 444 represents a plan with more wells (29 versus 17) but less error. For comparison, Camp Allen actually sampled 25 wells in the Shallow Aquifer during 2006.

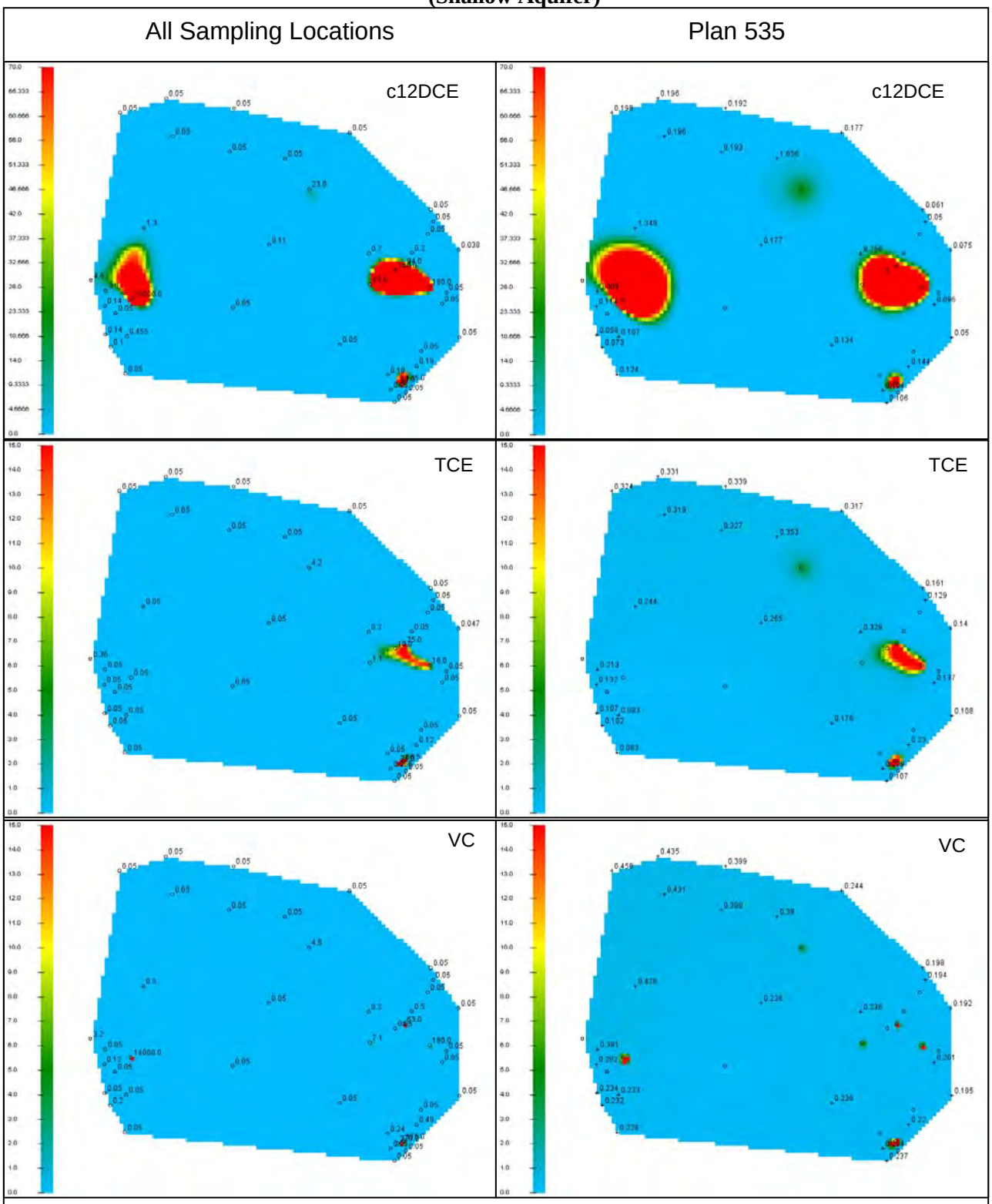


The table below lists wells that are recommended to be kept and wells that are recommended to be removed for Plan 535, as well as the maximum error for each COC. The subsequent figures illustrate the interpolated plumes for Plan 535 (17 wells) versus the baseline model (42 wells), and Plan 444 (29 wells, lower error than Plan 535) versus the baseline model (42 wells). The user or installation can compare the interpolated concentrations for the two plans to see if the interpolations for the plan with only 17 wells are acceptable, and if not, perhaps consider the plan with 29 wells.

Shallow Aquifer, Plan 535

Wells recommended to be kept		Wells recommended to be removed		Maximum Error
A1-MW10A	B-MW22A	A-GW-3	B-15WA	c12DCE = 0.108
A1-MW20	B-MW2A	A-MW12	GW-2	
A1-MW4A	B-MW31	A-MW17A	B-GW-6	TCE = 0.061
A2-MW11A	B-MW33A	A-MW18A	B-MW10	
B-20W	B-MW34A	A-MW30A	B-MW13	VC = 0.204
B-MW11A	B-MW35A	A1-MW21	B-MW14	
B-MW12	B-MW36A	A1-MW25A	B-MW17	
B-MW15A	B-MW3A	A1-MW31A	B-MW19A	
B-MW16		A1-MW6A	B-MW29R	
		A1-MW8A	B-MW32	
		A1-MW9A	B-MW37A	
		A2-MW30	B-MW9A	
		A2-MW34A		

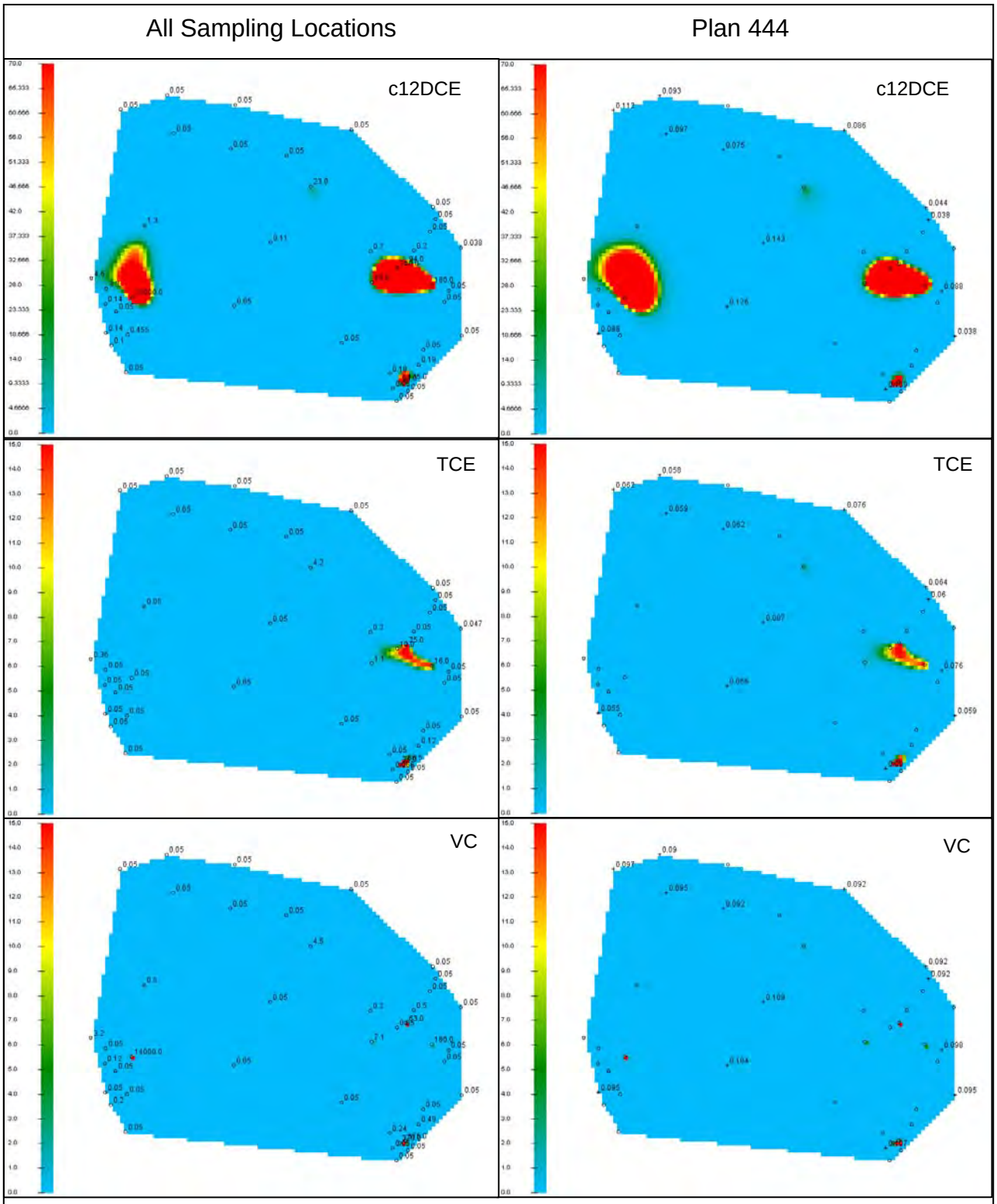
Interpolated Concentrations of Plan 535 (17 wells) versus Baseline Model (42 wells) (Shallow Aquifer)



Note:

1. Symbol "+" indicates wells that are recommended to be removed. Symbol "o" indicates wells that are recommended to keep;
2. Only interpreted concentrations at the removed locations are posted for the selected plan.

Interpolated Concentrations of Plan 444 (29 wells) versus Baseline Model (42 wells)
(Shallow Aquifer)



Note:

3. Symbol "+" indicates wells that are recommended to be removed. Symbol "o" indicates wells that are recommended to keep;
4. Only interpreted concentrations at the removed locations are posted for the selected plan.

Shallow Aquifer, Plan 444

Wells recommended to be kept		Wells recommended to be removed		Maximum Error
A-MW12	B-MW13	A-GW-3		c12DCE = 0.002
A1-MW10A	B-MW15A	A-MW17A		
A1-MW20	B-MW16	A-MW18A		TCE = 0.0008
A1-MW21	B-MW19A	A-MW30A		
A1-MW31A	B-MW22A	A1-MW25A		VC = 0.030
A1-MW6A	B-MW29R	A1-MW4A		
A1-MW8A	B-MW2A	A2-MW30		
A1-MW9A	B-MW31	B-MW12		
A2-MW11A	B-MW33A	B-MW14		
A2-MW34A	B-MW34A	B-MW17		
B-15WA	B-MW35A	B-MW32		
B-20W	B-MW36A	B-MW9A		
B-GW-6	B-MW37A	GW-2		
B-MW10	B-MW3A			
B-MW11A				

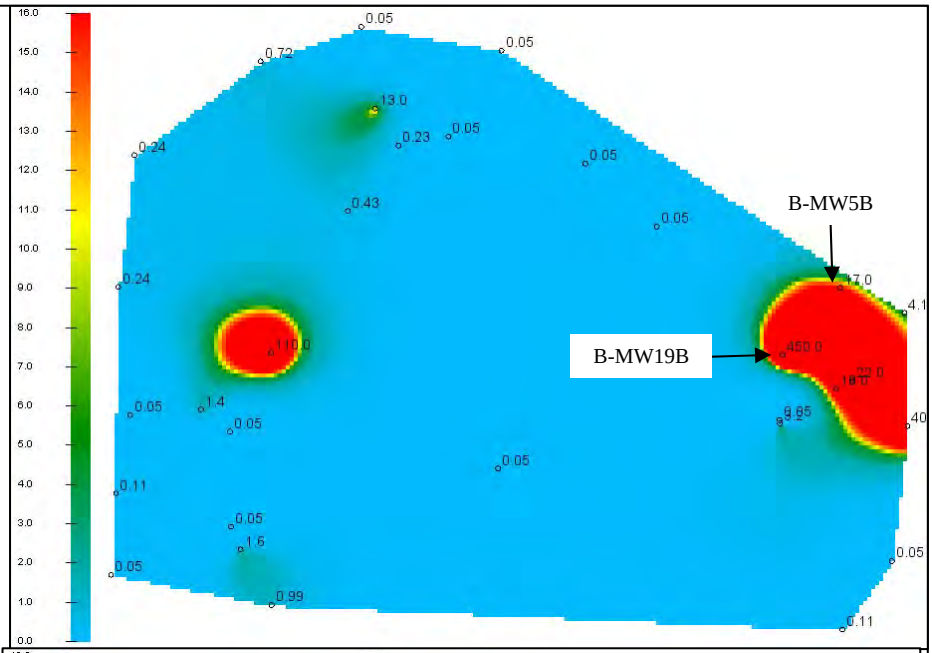
Deep Aquifer Tradeoff Curves

For the deep aquifer, two wells (B-MW5B and B-MW19B) have been abandoned since 2002 but they are still included in the input data for spatial analysis. The figure below illustrates the TCE plume map generated using kriging with quantile transformation with and without these two abandoned wells.

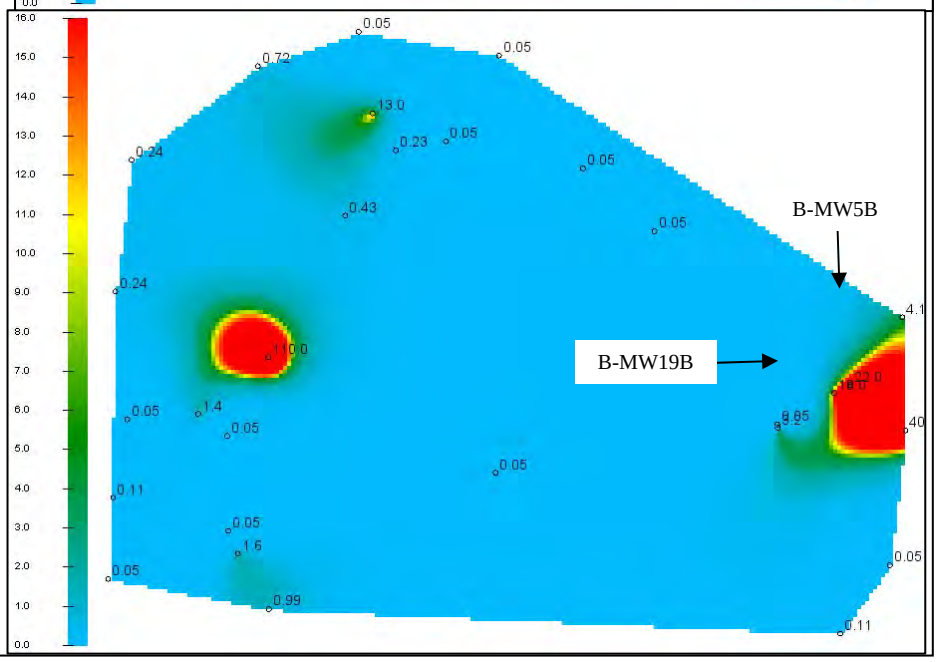
{this gap is intentional}

Model Builder Results With and Without the Two Abandoned Wells
Deep Aquifer

TCE Plume Map, kriging with quantile transformation, with two abandoned wells



TCE Plume Map, kriging with quantile transformation, without two abandoned wells

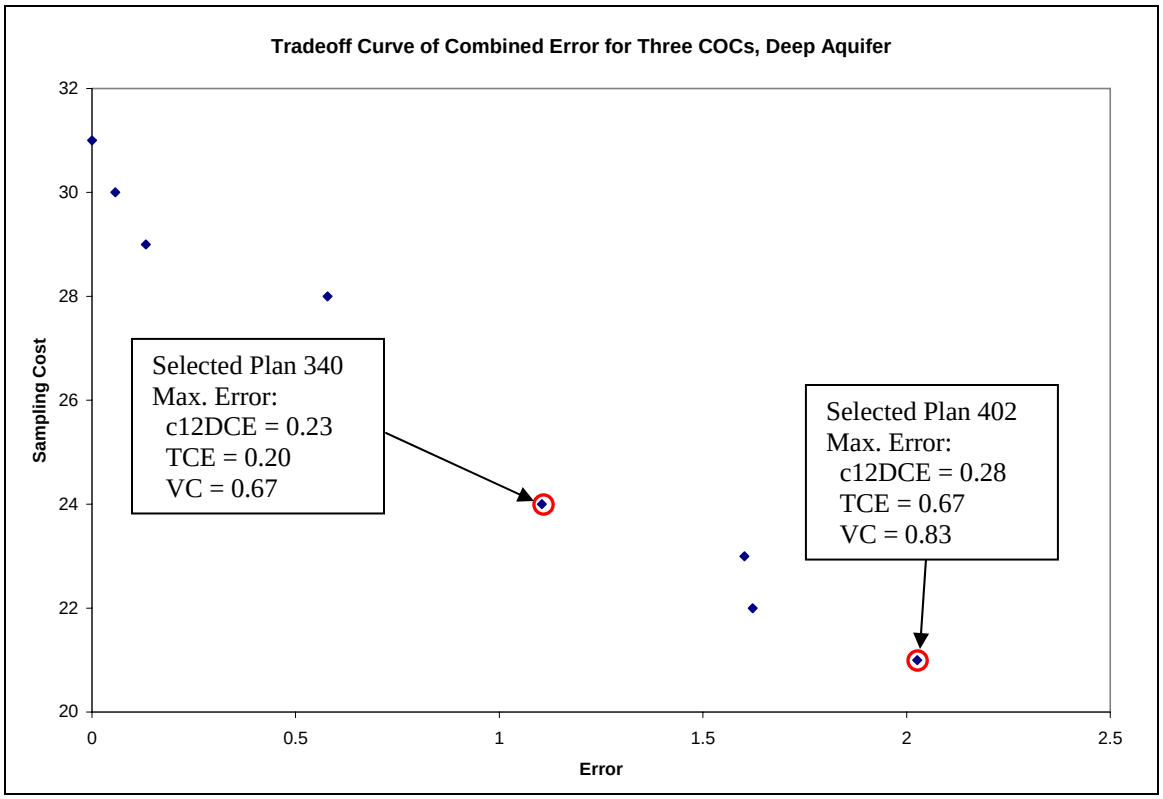


As illustrated, the highest TCE concentration of 450 µg/l in the baseline model occurred at abandoned well B-MW19B, and concentration of 17 µg/l at abandoned well B-MW5B is also above the TCE cleanup criteria. Without measurements at these two wells, the interpreted concentrations in that area are biased toward ND.

Optimizer was first performed with these two wells set as always off to represent these wells as abandoned. However, no plan with errors less than 1.5 for each COC was found, for all three preferred choices (i.e., kriging with quantile transformation, inverse distance weighting with

quantile transformation, and inverse distance weighting with log transformation). This is because the interpreted concentrations at these two abandoned well locations are much lower than the measured concentrations (which are included in the baseline "model" generated by Model Builder), thus the errors calculated at these two locations are much higher than 1.5.

When allowing these two wells to be active (conceptually, adding replacement wells), Optimizer identified 10 – 30 plans (depending on modeling approach) with errors less than 1.5 for each COC. The figure below illustrates the combined tradeoff curve with errors less than 1.5 for all three COCs for kriging with quantile transformation. If further evaluation was desired, we might select two potential plans (Plan 340 and Plan 402). However, that would assume new wells near the locations of the abandoned wells.



The conclusion is there is no well in the deep aquifer that can be removed from the monitoring network given that B-MW19B and B-MW5B are inactive. However, if these two wells are replaced, other wells can potentially be removed. For example, the number of wells is reduced from 31 to 24 for Plan 340 and from 31 to 21 Plan 402. That is, 7 to 10 wells might be recommended to be removed, relative to the baseline model, if these two abandoned wells are replaced and included in the monitoring network.

Deep Aquifer, Plan 402

Wells recommended to be kept		Wells recommended to be removed		Maximum Error
A-MW13B A-MW14B A-MW15B A-MW16B A-MW17B A-MW1B A-MW31B A1-MW10B A1-MW24B A1-MW25B A1-MW31B	A1-MW9B A2-MW11B A2-MW23B A2-MW28B B-MW11B B-MW19B B-MW22B B-MW2B B-MW34B B-MW5B	A-MW18B A-MW19B A-MW30B A1-MW4B A1-MW6B A2-MW32B B-MW15B B-MW2C B-MW33B B-MW3B		c12DCE = 0.23 TCE = 0.20 VC = 0.67

Deep Aquifer, Plan 340

Wells recommended to be kept		Wells recommended to be removed		Maximum Error
A-MW13B A-MW14B A-MW15B A-MW16B A-MW17B A-MW1B A-MW31B A1-MW10B A1-MW24B A1-MW25B A1-MW31B A1-MW9B	A2-MW11B A2-MW23B A2-MW28B B-MW11B B-MW15B B-MW19B B-MW2B B-MW2C B-MW33B B-MW34B B-MW3B B-MW5B	A-MW18B A-MW19B A-MW30B A1-MW4B A1-MW6B A2-MW32B B-MW22B		c12DCE = 0.28 TCE = 0.67 VC = 0.83

Spatio-Temporal Analysis Results

For a reliable analysis, spatio-temporal analysis with the Sampling Optimizer has the following requirement for the input data:

- For a specific COC at a specific well to be included in the dataset to be analyzed, there must be at least 4 samples at that well for that COC. The software will warn the user if there are only 4 to 7 samples for a COC at a specific well by indicating that the data may be insufficient for reliable analysis (i.e., 8 or more samples per well is preferable).

- For a specific COC to be analyzed as a part of a specific sampling event, there must be 15 samples of a COC for that sampling event. The software will warn if there are only 15 to 19 samples for that event by indicating that the data may be insufficient for reliable analysis (i.e., 20 or more samples per event is preferable).

Samples not used for either of the above reasons will be grayed out on the data input screen within the software.

There was some uncertainty identified by EnviroStat regarding data from 2003 and 2004 at wells B-MW33A and B-MW33B. The concentrations at B-MW33A were lower in these events than in

previous years, and the concentrations at B-MW33B were higher in these events than in previous years. These wells are close together, but the "33B" well is in the deep aquifer while the "33A" well is in the shallow aquifer. There is some possibility that these data were "switched" due to sampling or reporting error, but that could not be confirmed by the site team when questioned by EnviroStat.

Concentrations (µg/l) at B-MW33B and B- MW33A
(shading indicates values that may be "switched")

Date	SiteID	c12DCE	TCE	VC
3/15/2002	B-MW33B	4.1	9.4	0.05
3/15/2003	B-MW33B	64	190	4.5
3/15/2004	B-MW33B	130	290	5.4
3/15/2005	B-MW33B	8.9	18	0.05
3/15/2001	B-MW33A	460	59	25
3/15/2002	B-MW33A	240	26	15.5
3/15/2003	B-MW33A	6.4	1.2	0.76
3/15/2004	B-MW33A	6.7	6.1	0.78
3/15/2005	B-MW33A	250	12	0.05

Because of the uncertainty, we excluded these data points from 2003 and 2004 from the spatiotemporal analysis. As a result, each well had fewer than four sampling events that were available for the spatio-temporal analysis, and thus these wells were not used within that analysis. As explained below, there actually was not enough data to perform spatio-temporal analysis for the shallow aquifer even if well B-MW33A had been included, so there was no impact to the analysis. For the deep aquifer, excluding well B-MW33B from the spatio-temporal analysis implies that this well would be recommended for inclusion in future sampling events (i.e., it would be added back into any optimal plan recommended as a result of the spatio-temporal evaluation), and this seems appropriate given the uncertainty of past sampling results at that well.

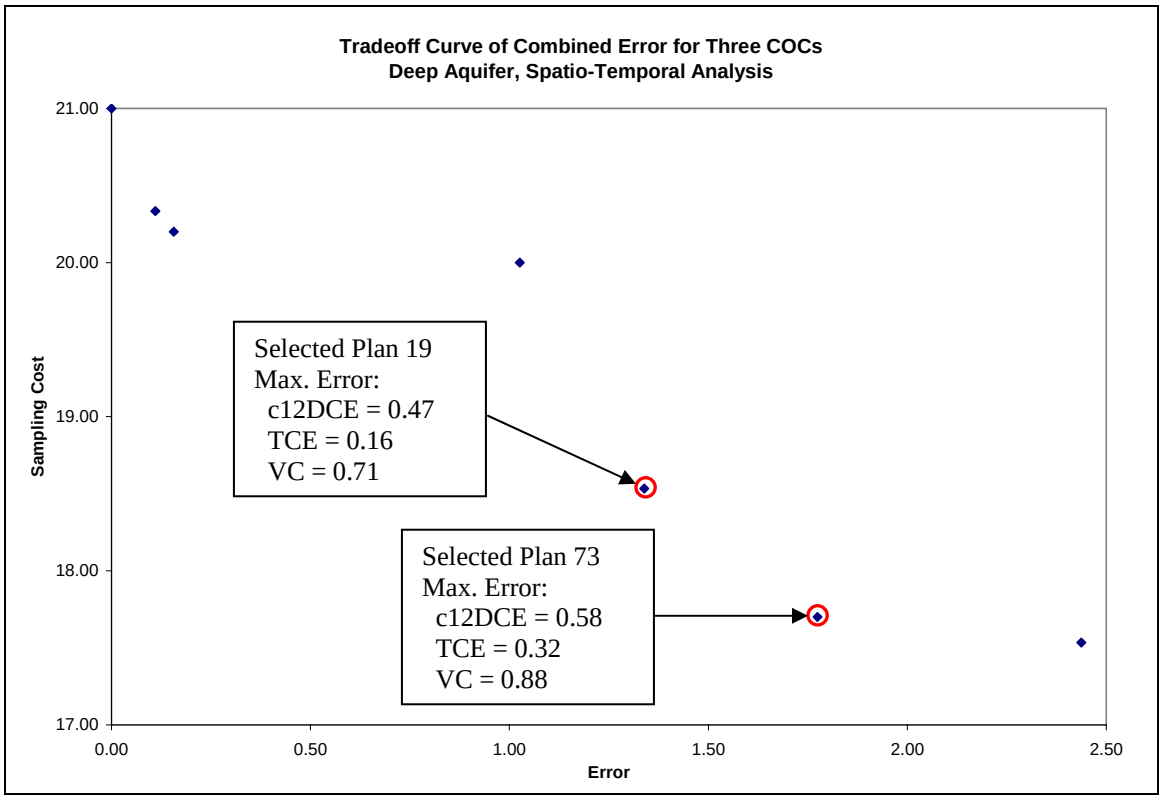
Based on the criteria above, there are not enough data for the shallow aquifer at Camp Allen site to conduct a spatio-temporal analysis. For the deep aquifer, there are only 21 of 29 wells that are used for spatio-temporal analysis based on the criteria above. Conceptually, existing wells in the baseline model that are not included in the spatio-temporal analysis, such as new wells that do not have a long data history, will be added back into any plan suggested by the spatio-temporal optimization and continue to be monitored. Two spatio-temporal optimizations were performed for the deep aquifer. The two top choices of combinations of interpolation technique and data transformation determined during the spatial analysis are used for the spatio-temporal analysis. They are: 1) inverse distance weighting with quantile transformation; and 2) kriging with quintile transformation.

Summary of "Optimal Plans" Identified by Optimizer – Spatial Optimization (Deep Aquifer)

Simula-tion	Interpola-tion	Data Transfor-mation	# Plans w/ errors < 0.5 for each COC	# Plans w/ errors < 1.0 for each COC	# Plans w/ errors < 1.5 for each COC	Min. Wells w/ errors < 0.5 for each COC	Min. Wells w/ errors < 1.0 for each COC	Min. Wells w/ errors < 1.5 for each COC
ST-DIQ	Inverse Distance Weighting	Quantile	2	3	5	20.5	20.2	19.53
ST-DKQ	Kriging	Quantile	5	11	17	20.2	17.7	17.53

** minimum number of wells is reported as decimal because there are a different number of wells for different sampling events in spatio-temporal optimization.*

As with the spatial analysis, we chose to illustrate Optimizer results using kriging with quantile transformation for our post-software analysis. The combined tradeoff curve is illustrated below. By comparing sampling cost (number of wells) and errors on the combined tradeoff curve, and by visually inspecting the plume maps for selected plan(s) and comparing to the plume maps with all wells (baseline model), two potential plans: Plan 19 and Plan 73 are identified. The sampling cost (average number of wells) is reduced from 21 to 18.53 for Plan 19 and to 17.70 for Plan 73.



The table below lists recommended sampling frequency and the wells that are recommended to be removed (i.e., always “off”), as well as the maximum error for each COC.

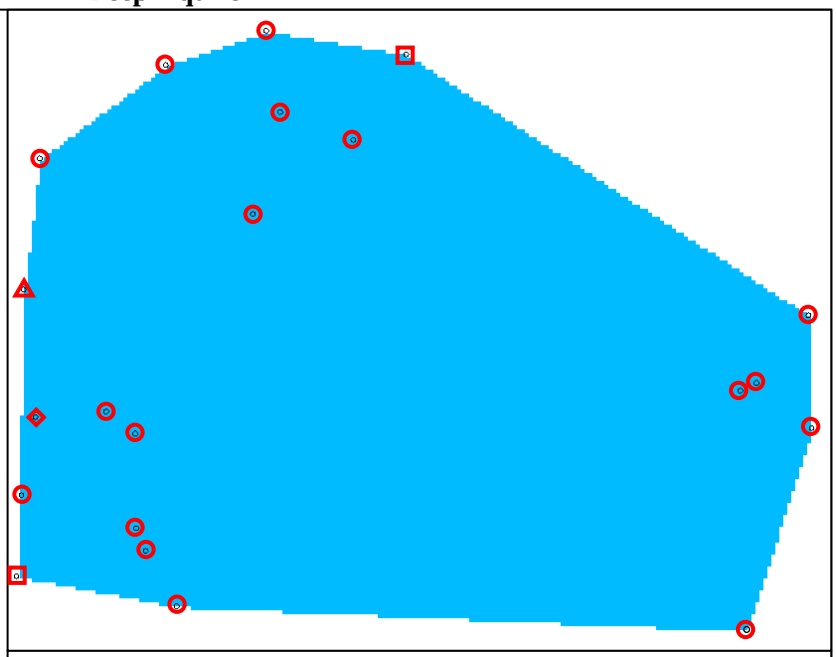
Spatio-Temporal Analysis Recommendations
Deep Aquifer

Plan #	Recommended Sampling Frequency						Max. Error
	Annually		Every 2 Year	Every 3 Year	Every 5 Year	Off	
19	A-MW13B A-MW16B A-MW17B A-MW18B A-MW30B A1-MW10B A1-MW25B A1-MW31B A1-MW6B	A1-MW9B A2-MW23B A2-MW28B B-MW11B B-MW15B B-MW33B B-MW34B B-MW3B	A-MW31B A2-MW32B	A-MW15B	A-MW14B		c12DCE = 0.471 TCE = 0.161 VC = 0.707
73	A-MW13B A-MW15B A-MW16B A-MW17B A-MW18B A-MW30B A1-MW10B A1-MW25B A1-MW31B	A1-MW6B A1-MW9B A2-MW23B A2-MW28B B-MW11B B-MW33B B-MW34B B-MW3B	B-MW15B		A-MW14B	A-MW31B A2-MW32B	c12DCE = 0.580 TCE = 0.318 VC = 0.876

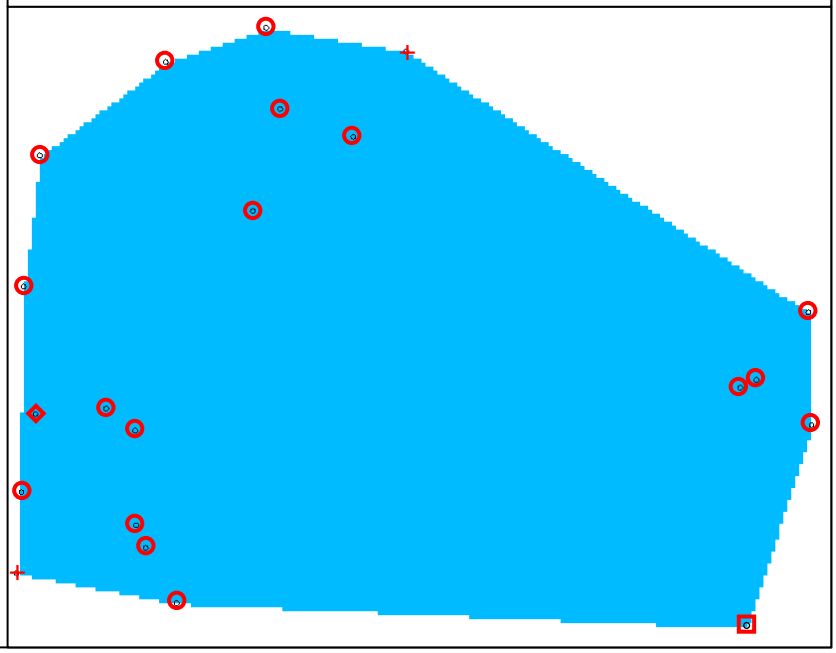
The figure below illustrates sampling frequency recommended by the Optimizer for the two selected Plan 19 and Plan 73.

Spatio-Temporal Analysis – Recommended Sampling Frequencies
Deep Aquifer

Recommended Sampling Frequency, Plan 19 (More Wells, Lower Error)



Recommended Sampling Frequency, Plan 73 (Less Wells, Higher Error)



Sampling Frequency Legend

- ○ Annually
- □ Every 2 Year
- △ Every 3 Year
- ◇ Every 5 Year
- + Off

GeoTrans created PowerPoint files (not shown here) to illustrate the side-by-side comparison of the plume maps illustrating the results of the spatio-temporal analysis, for each year from 2001 to 2005, for Plan 19 and Plan 73. They also included the corresponding plume maps with all available data. In each case, the models were kriging with quantile transformation (the approach used for spatio-temporal optimizations). The following observations were made:

- For c12DCE in 2003, interpreted concentration along the northeast boundary for Plan 73 is close to the cleanup goal of 70 µg/l mainly due to the removal of A2-MW32B, whereas

for “all points available” and Plan 19 the concentration for c12DCE in that same area is interpreted as 20 – 30 µg/l along that boundary

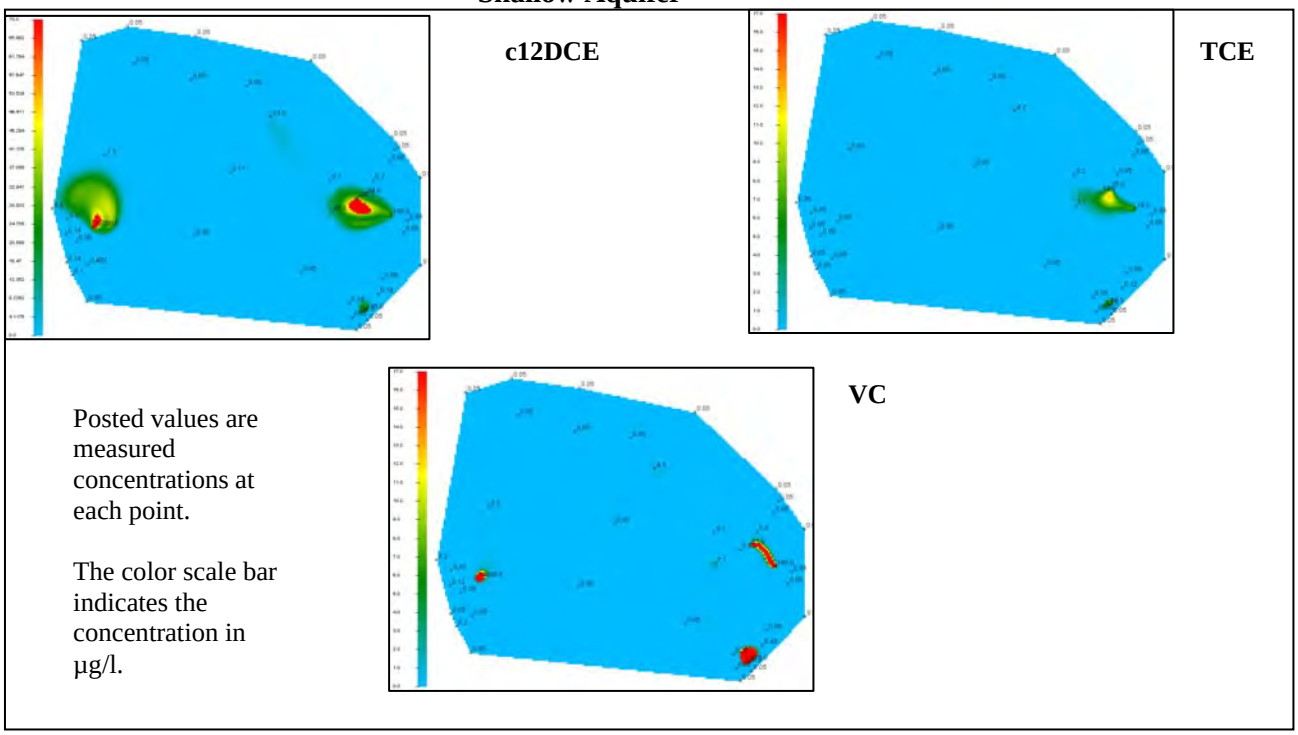
- For VC in 2004, a large area along the northwest boundary for Plan 73 is interpreted in the range of 4 – 5 µg/l (above the cleanup goal of 2 µg/l) whereas for “all points available” and Plan 19 the size of the area interpreted above the cleanup goal is much smaller.

These observations are consistent with the results of the tradeoff curve (i.e., Plan 73 has increased error relative to Plan 19). The user (e.g., installation) would need to determine if these types of observed differences are acceptable, such that Plan 73 would be preferred due to slightly lower cost. An additional useful analysis would be to perform a similar comparison after adding back in to Plan 73 and Plan 19 the existing monitoring locations that were not included in the spatio-temporal analysis due to inadequate data history. We did not have time to perform that analysis.

Uncertainty Analysis Results

Model Builder also provides visualization of uncertainty. The figure below illustrates the uncertainty maps obtained using kriging interpolation with quantile data transformation for the shallow aquifer. GeoTrans reported that is was not clear to them how to make practical use of these results.

Example Uncertainty Results, Kriging with Quantile Transformation
Shallow Aquifer



Data Tracker Results (Reported by Charles Davis, EnviroStat)

Overview and Summary

Data Tracker (DT) is a feature of the Summit Monitoring Tools designed help site personnel identify anomalies of potential interest in new monitoring data. DT functions as follows:

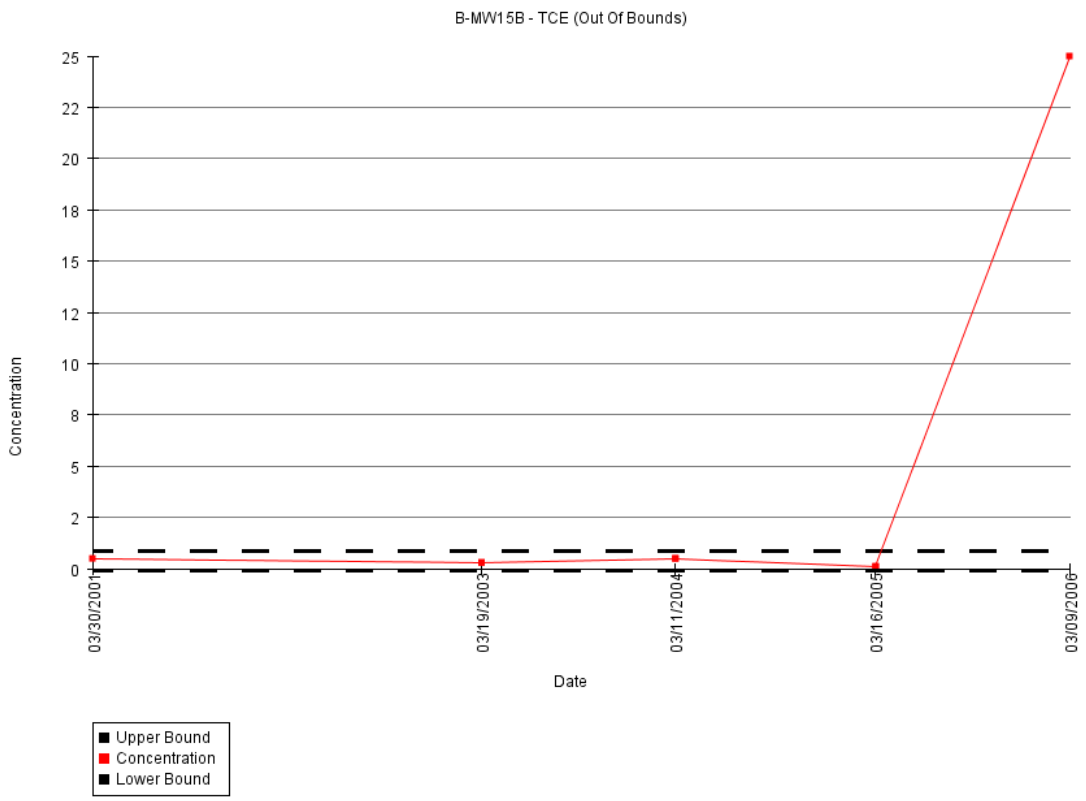
- First, historical data are manually screened to remove values deemed to be "atypical", such as erratic values that are not repeated and non-detects with unusually high reporting limits. The idea is that the remaining data values, termed "background" data, should be representative of values to be expected in future monitoring. Future values inconsistent with the background data will be flagged for inspection. This is done separately for each COC and for each well. The user may set the minimum number of background observations required for a particular well/COC combination; there is an absolute minimum of four background observations required by DT.

- The background data form one dataset, stored in a CSV file. In future versions of DT, it is anticipated that historical values that are not included in the background data may be included in this dataset; if present, they will be included in data plots, but not used in computing bounds in the next step (this was not available in this version of DT).

- Prediction limits (PL bounds) are prepared for each well/COC combination with at least the specified minimum number of background values. At present these are static (not time-dependent). At Camp Allen nominal 90% prediction limits were used; this means that, so long as the assumptions are reasonably met, only 10% of observations from truly steady-state processes should be "out-of-bounds". The prediction confidence level is user-configurable; with more background observations than available at Camp Allen, one would tend to use a higher confidence level, say 95%. Although the version of DT used at Camp Allen does not allow for time-dependent bounds, these are anticipated in a future version. The stated nominal confidence level is based on assumptions of (a) steady-state variation, (b) normal distributions of data, and (c) uncensored data, whereas the actual data often have trends, outliers, and non-detects (NDs) reported as "<RL" where RL is a reporting limit.

- When the data for a new monitoring event are received (i.e., "new data"), another CSV file of the new data is prepared and imported into DT. DT compares each value with its PL bounds and prepares two tables, one with well/COC combinations that are "in-bounds" and the other with combinations that are "out-of-bounds". The user can then click on a well/COC combination in either table to see a Time Series Plot (TSPlot) of the historical and current data along with the bounds. The data and the TSPlot can be saved for future reference or embedding in documents.

An example of a DT plot for an "out-of-bounds" sample is provided below.



Concentrating on the out-of-bounds cases allows site personnel to focus on the well/COC combinations likely to require attention or action. Since the bounds are two-sided, in some cases one will find out-of-bounds values that are lower than anticipated, and in other cases higher than anticipated. Either case might indicate anomalies of interest; it is left to site personnel to make the actual decisions regarding each flagged value.

For the purposes of this ESTCP project, in order to test the operation of DT, six versions of the current (2006) data were prepared and used with the historical dataset based on 2001-2005 data. One of these is the actual 2006 data. Artificial anomalies were introduced into the other five datasets by EnviroStat, following plausible scenarios that may be of interest to Camp Allen. The anomalies were not known by GeoTrans, who applied the DT software. The narratives (see below) were presented *a priori* to and discussed with site personnel and members of the project team not directly involved in using the software. Note that in a typical application of the software only one dataset for new data would be imported into DT (i.e., the actual data).

DT successfully flagged virtually all anomalies of interest. It turned out that the actual 2006 data had a few curiosities and one major error, which DT identified. DT identified nearly all of the artificial anomalies added to the other datasets as well. The software cannot by itself "imagine" what the cause of a particular anomaly or set of anomalies might be, but by sorting the lists of flagged well/COC combinations by well and then referring as needed to site maps, one may conceptualize what the cause of the anomalies might be.

DT also flagged a number of atypical values that are ultimately not of particular interest. These occurred with well/COC combinations where all historical concentrations were extremely low, with little variability, and the current value therefore appeared high (or even low, in some cases,

with J values which are values below a lab's usual reporting limit, reported as uncertain numerical values rather than simply as non-detects). Refining the DT algorithms to avoid diverting attention to such cases has been suggested.

Several logistical issues with DT were identified for future refinement. The present version of DT requires that the same wells be present in both background and current data files, whereas not all wells have been or (particularly in an optimized monitoring program) will be sampled during each event. Also, at present the current dataset must be manually screened for field duplicates, and any field duplicates removed or averaged or treated in some fashion so that there is at most one value per well/COC combination. (An individual well/COC combination may be missing, however.) DT does not check to ensure that data are provided for all required wells, or for all required COCs at the required wells. Most significantly, at present DT does not have built-in facility for handling NDs; this must be done by the user in preparing both historical and current datasets. Refinement in these regards is anticipated.

Preparing Datasets with Artificial Anomalies for DT

These datasets were generated from the 2006 data received from Camp Allen. Not all of the wells with historical data were sampled during 2006. Conversely, and more importantly perhaps, three wells without historical data were sampled during 2006: shallow well A-GW-1, shallow well B-20WSS, and extra-deep well B-15WB. Shallow well B-20WSS is particularly curious, in that it has very low concentrations of all COCs, but is only 21 feet from, and screened at the same depth as, shallow well B-20W which has extremely high concentrations of six of the nine COCs.

As a preliminary step, the few field duplicates were compared with their partner sample data values. The differences being quite minor at most, the field duplicate values were discarded. Also, J flags are ignored. The six versions of the 2006 data are labeled Red, Orange, Yellow, Green, Blue, and Egg. Orange is the original Camp Allen 2006 data.

The first step in creating the other five versions was to "jitter" all values (except NDs) slightly, to avoid having, say, all values except the Red dataset value being identical for a given well, which would be an obvious give-away that the Red value was the artificial one. This was accomplished by multiplying each original value by a random number between 0.85 and 1.15, then rounding the result the same way as the original value was rounded. Artificial anomalies were then added to the five versions, following the scripts previously proposed.

The narrative descriptions of the artificial anomalies added follow.

Red

1. A previously unknown migration pathway is beginning to allow contaminants found in deep wells A1-MW25B and A1-MW31B to travel to extra-deep well A1-MW9C, which has previously not been contaminated.
2. The data value for Vinyl Chloride (VC), one of the three primary COCs, is missing for required well A-MW13B.
3. Shallow well B-20W, which has had recent sharp increases in all COCs, continues its upward trends.

Orange

1. This is the original 2006 dataset. (Well B-20W reversed those recent upward trends; see plots below.)

Yellow

1. The data value for VC is missing for required well A-MW13B.
2. Shallow well B-20W, which has had recent sharp increases in all nine COCs, continues its upward trends.
3. A previously unknown source in the northeast Area B begins to impact shallow wells (B-MW9A and B-MW34A), involving Benzene, Toluene, and total Xylenes (BTX) but none of the other COCs.

Green

1. The sampling crew inadvertently used the bottle labeled B-20WSS for the sample from well B-20W and vice versa. B-20WSS had not been sampled between 1997 and 2005. It is 21 feet from B-20W, and its 10-ft screened interval overlaps the bottom half of the 20-ft screened interval of B-20W. But B-20W has very high concentrations, whereas B-20WSS values are basically NDs or very low concentrations. (Since B-20WSS was not present in the historical dataset, its value had to be manually removed from the current dataset.)
2. A previously unknown source in the northeast Area B begins to impact shallow wells (B-MW9A and B-MW34A), involving Benzene, Toluene, and total Xylenes (BTX) but none of the other COCs.

Blue

1. A previously unknown migration pathway is beginning to allow contaminants found in deep wells A1-MW25B and A1-MW31B to travel to extra-deep well A1-MW9C, which has previously not been contaminated.
2. The sample from well A1-MW21 is contaminated by GC/MS column carryover from the previous sample from B-MW35A.

Egg

1. Shallow well B-20W, which has had recent sharp increases in all nine COCs, continues its upward trends.
2. Due to a clerical/software error all 1,2-Dichloroethane (12DCA) values become 20 times lower than actual values.
3. The sample from well A1-MW21 is contaminated by the previous sample from B-MW35A in the lab.
4. Area A deep sentinel well A-MW14B, previous nearly always ND, starts to see detects in the three primary COCs.

There is some overlap in the anomalies included in the datasets. This is intentional; otherwise one might be able to deduce readily by majority vote which dataset contained the true 2006 data.

Other Data Preparation Issues

It had been noted that the historical data from one shallow-deep pair of wells (B-MW33A and B-MW33B) appeared to have been switched for 2003 and 2004. In preparing the historical datasets two approaches were used. One was to indicate "do not use" for the apparently incorrect data; unfortunately, that would leave only two or three background observations for those wells, too few for DT. The other approach was to swap the values for those wells for those years. EnviroStat prepared historical datasets using both approaches. GeoTrans opted to use the first approach for the most part, but did use both for the Red dataset.

Also, since the current version of DT does not have provision for handling NDs, GeoTrans in most cases simply replaced NDs in all data with 0.5 µg/l, the lowest RL found in the historical dataset. A common alternative approach is to replace each ND with half its RL; GeoTrans used this alternate approach in addition with the Red dataset. (In all, GeoTrans did three different versions of the analysis with the Red dataset.) In preparing the historical/background datasets, EnviroStat had recommended including historical NDs with RLs no higher than 13 µg/l, but had indicated "do not use" for NDs with higher RLs.

Incidentally, although the lowest RLs in the data are 0.5 µg/l, there are numerous J values in both historical and current datasets lower than that. There are reported historical values as low as 0.038 µg/l, over ten times lower than the lowest RL. There are current data values as low as 0.10 µg/l.

Results/Conclusions from Applying DT

As stated, DT and GeoTrans identified several anomalies in the actual 2006 data as well as nearly all of the artificial anomalies added in the various versions of the data, and in addition identified a few low-level artifacts of the ND treatment. GeoTrans made subjective judgments about each flagged value as to whether or not it was truly anomalous, and provided a recommended interpretation or action. In its interpretations GeoTrans took into account the cleanup goals for the specific COCs.

In characterizing the natures of the out-of-bounds values, GeoTrans assigned one of six descriptions, as follows.

A = Current concentration is much higher than the historical data and above cleanup goal, could be bad data.

B = Current concentration is much lower than the historical data which were above cleanup goal, could be bad data.

C = Current concentration is higher than historical data and above cleanup goal but following an increasing trend.

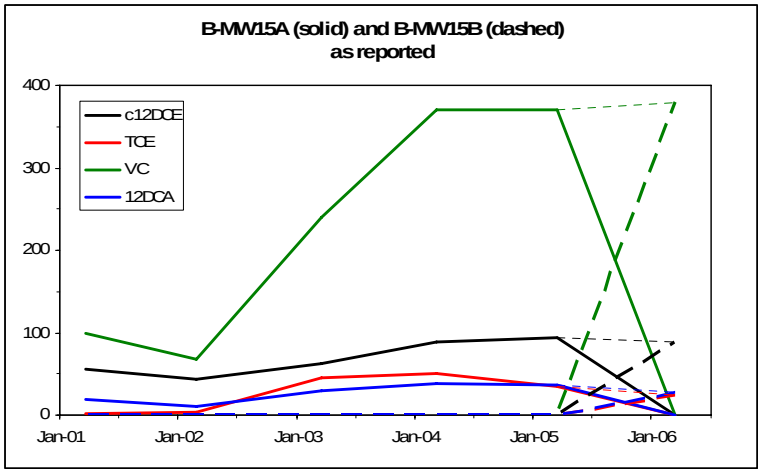
D = Current concentration is lower than historical data but following a decreasing trend.

E = Current concentration is out of bounds, but not a concern. No action is required.

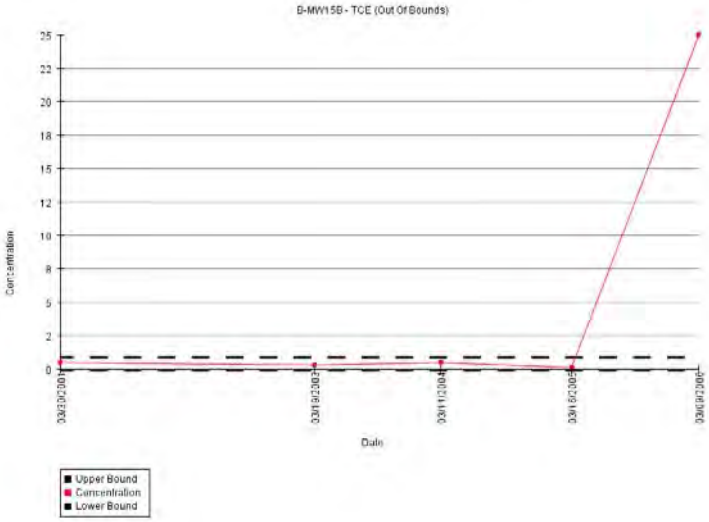
F = No anomaly apparent. No further attention needed.

GeoTrans provided a spreadsheet for each dataset listing the flagged values, the lower and upper PL bounds, the cleanup goal, and their interpretation. It also provided the plots given by DT. Their results for any given situation are essentially the same regardless of which dataset is involved, so the following discussion organizes the results by situation rather than dataset.

Anomalies in the Actual 2006 Data

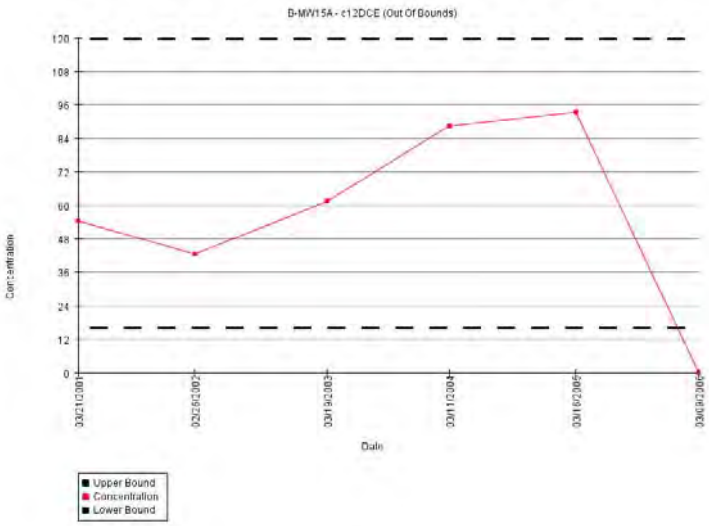


The data from shallow well B-MW15A and deep well B-MW15B were apparently switched; recall the suspected switching in 2003 and 2004 of a similar pair of shallow-deep wells. The northing and easting coordinates of these wells are the same. The shallow well has a history of elevated concentrations, whereas the deep well has seen only NDs or J values less than 0.5 µg/l historically. But for 2006 the reported B-MW15A values are <0.5 for all COCs, whereas the B-MW15B values match the B-MW15A historical data very nicely, as indicated by the dotted lines on the adjacent plot. EnviroStat suggested that the first two values for TCE for B-MW15A be excluded from the background data due to their low, atypical values; that left only three background values, not enough for DT so use. All values were used in the background data for the remaining COCs.



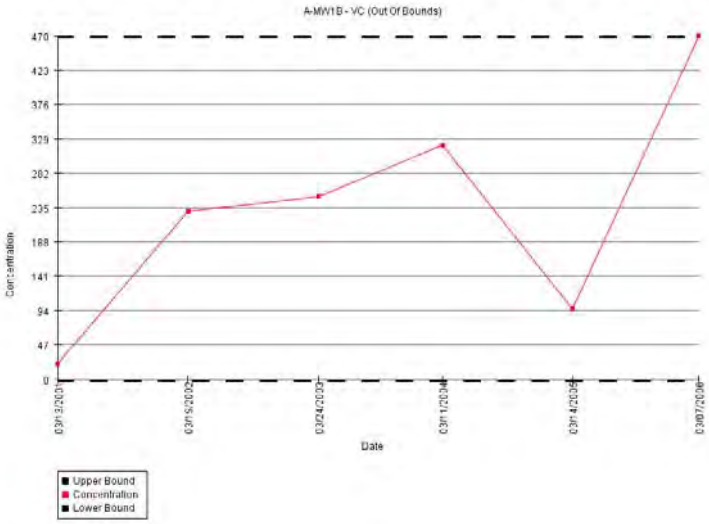
DT and GeoTrans found these large increases in B-MW15B in every case. As an example, see the DT plot at right for TCE. The only one missed was 12DCA in the Green dataset, since in that dataset all 12DCA values were artificially deflated to mimic a lab transcription or software error. For four COCs the GeoTrans interpretation is "A", with the comment that these could be bad data. DT and GeoTrans also found the increase in Benzene (from ND to 3.5), but provided an "E" rating, noting although the value was clearly above background, it is below the cleanup goal (5 µg/l). If one were viewing only the Benzene data histories this would be a reasonable conclusion, but when combined with the other COCs, and especially the B-MW15A data, the conclusion "bad data" is quite clear.

Paradoxically, at B-MW15A (which had lower than normal concentrations due to the switch), DT only found a decrease for c12DCE (see the adjacent plot). With such a small amount of background data, only that parameter has historical values that are consistent enough (and away from zero) that the lower PL bound is higher than a value of ND. For other parameters, ND values are still within the expected range. If the user chose to exclude the low historical values for other parameters in the historical data, the abnormally low values in new data would be more readily detected by DT.



DT and GeoTrans also found increases for five COCs in B-MW2A, with interpretations "C", "A", "E", "A", and "E" for c12DCE, TCE, 111TCA, 12DCA, and Benzene respectively. The interpretations vary depending on whether the data appear to be continuing an increasing trend and whether the current value is above the cleanup goal or not. These differing interpretations seem reasonable; c12DCE was already elevated above prior levels in 2005, whereas 111TCA had previously been all ND, and so on.

The background data are quite variable for VC for well A-MW1B, so that the PL bounds are quite wide (0 to 469.78). The actual background data range from 21 to 320. Nonetheless, the 2006 value at 470 is just above the upper bound. In this case it is difficult to interpret the situation satisfactorily. GeoTrans concluded "C", that there is an increasing trend. A reasonable action would be to continue to track this well/COC combination with perhaps more interest than usual. At this same well c12DCE has historically had the highest values of all the COCs, and its 2006 value of 160 seems to be part of a <u>decreasing</u> trend from a value of 1000 in 2002. This could be related to reductive dechlorination of c12DCE to VC.



DT and GeoTrans identified seven additional out-of-bounds values in the actual 2006 data, and appropriately rated them all "D" or "F", as all involve very low values or NDs. When background data include numerous NDs, and all NDs are simply replaced by the same value, PL

bounds can be excessively narrow, leading to out-of-bounds values. Four of these seven involve J values being found lower than a history of NDs or, conversely, an ND being found to be higher than a history of J values. In the remaining three cases the values are outside of, but close to, the PL bounds; recall that these are nominally 90% bounds, so one expects 10% of observations to be out-of-bounds with steady-state randomness. (Not all of these were found out-of-bounds in the jittered versions of the data used in the other datasets.)

Artificial Anomaly – A1-MW9C (Red and Blue)

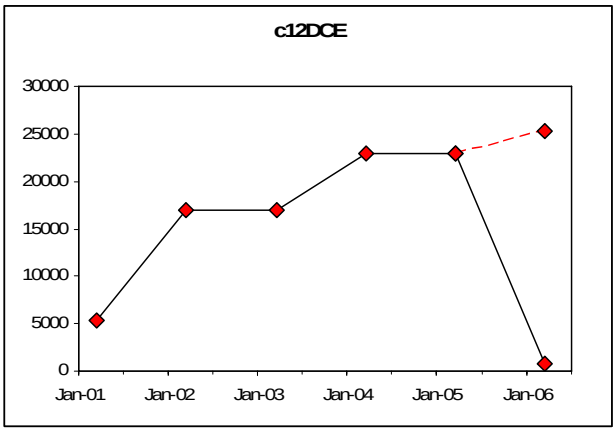
The first of the artificial anomalies involves extra-deep well A1-MW9C, which has had only NDs or J values less than 0.5 in the past. This well's screen is centered at 99 feet below MSL; typical mid-screen elevations for Area A deep wells are 63 to 37 feet below MSL. Contaminated deep wells A1-MW25B and A1-MW31B are located 362 and 496 feet to the south, respectively. This scenario involves contaminants from the latter two deep wells finding a migration pathway to A1-MW9C. The primary COCs in those two wells are c12DCE (14 to 50 in A1-MW25B, 2.8 to 40 in A1-MW31B) and VC (ND to 55 in A1-MW25B, 2.9 to 24 in A1-MW31B), along with traces of most of the other COCs.

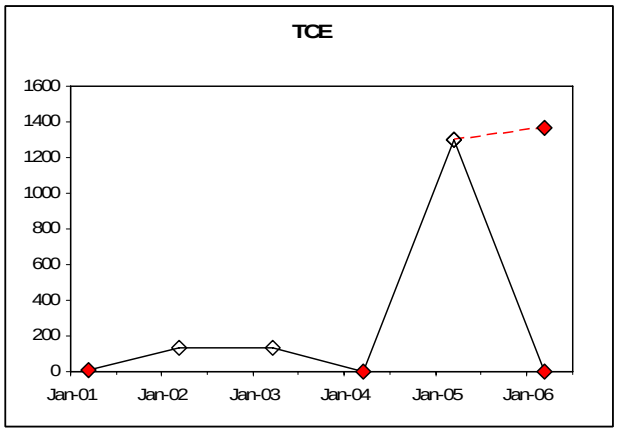
A1-MW9C was not actually sampled during 2006. The artificial anomalies added in datasets Red and Blue involve adding records with (c12DCE, VC) = (15, 25) for Red and (13, 29) for Blue (i.e., higher than the NDs typically found at well A1-MW9C), with trace levels of other constituents. DT and GeoTrans found increases in three of the COCs in both the Red and Blue datasets (where the anomalies were added), regardless of the ND treatment used for the Red dataset. This particular artificial anomaly is quite large relative to the background data, and was easily found. GeoTrans interpretations are "A" for VC and "E" or "F" otherwise, even for c12DCE (because the anomaly value is still well below the cleanup goal of 70). Again, these interpretations were apparently made for each well/COC combination separately; looking at all COCs for this well together, and especially considering its depth, one would surely decide that the data are clearly inconsistent with the past and hence a cause for alarm, or at least a good deal of additional scrutiny.

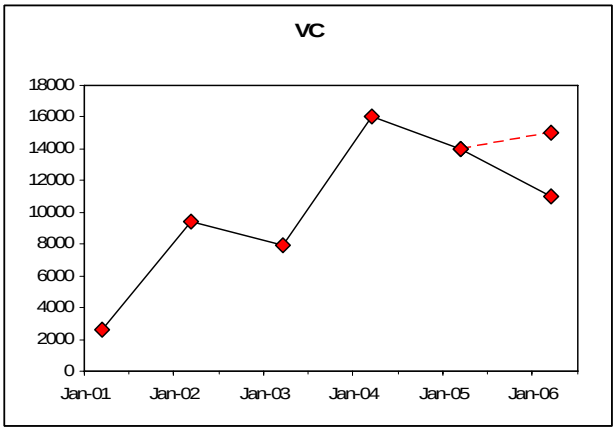
Artificial Anomaly – B-20W (Red, Yellow, and Egg)

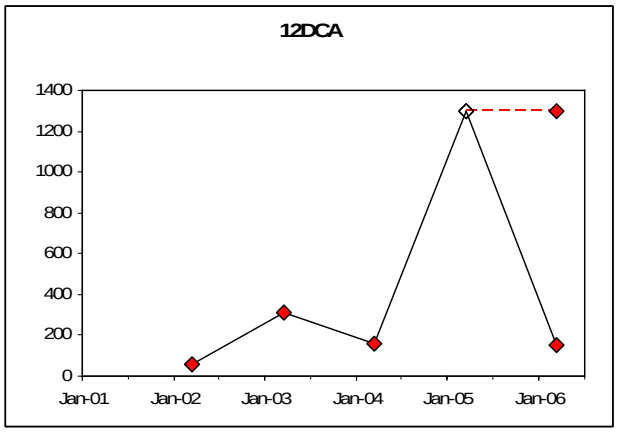
Values have been high to extremely high and increasing for c12DCE and VC. Data for other COCs have been a blend of moderately high values and NDs with high RLs (ranging up to 1300). GeoTrans ignored the suggestion to discard the "<130" and "<1300" NDs for TCE and 12DCA, otherwise there would not have been enough data to run DT for these COCs. Rather, they replaced all NDs with 0.5.

The actual 2006 data reverse the increasing trends somewhat. The plots below show both the actual data (solid lines) which illustrate the lower 2006 values, and the artificial data (dotted lines) which have higher values. Filled-in symbols represent actual values and outline-only symbols represent NDs, plotted at the RL.

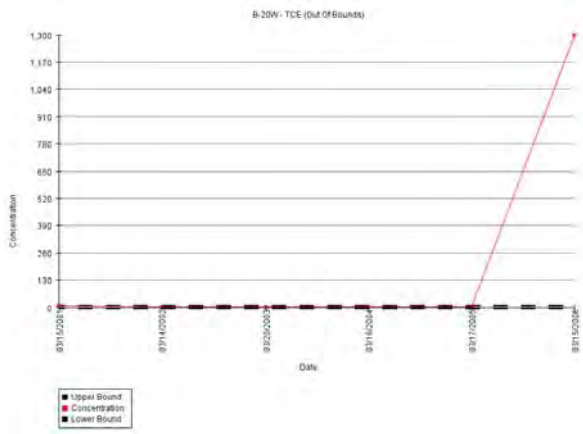


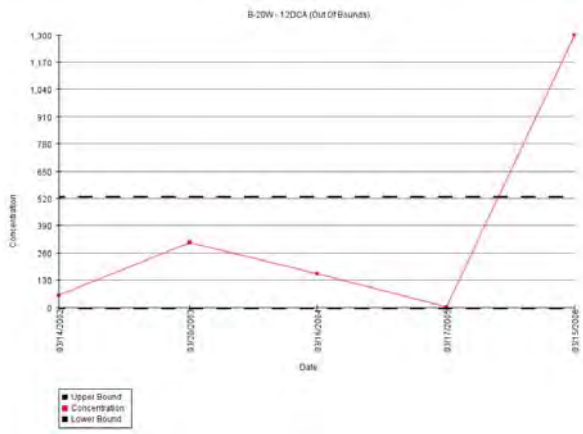






DT indicated out-of-bounds values for TCE and 12DCA; it found those because the high NDs in the background data were treated as 0.5 in computing the PL bounds. GeoTrans characterized these out-of-bounds values as “A”, with comments such as “previous high only 8 ppb” and “sudden increase after three years of decreasing trend”. As the plots show, these interpretations are appropriate for the data given the ND replacement value of 0.5. However, it is also possible the high RLs in the 2005 data could be associated with high actual values in 2005, and if the ND was treated as the RL rather than a low value such as 0.5, the 2006 data may not have been flagged as out-of-bounds. The plots from DT, using the 0.5 value for the NDs, are provided below for TCE and 12DCA.





There exist numerous treatments for ND data that are more sophisticated than simple replacement by an arbitrary value or by RL/2, and eventually it would be desirable to program those directly into DT.

<u>Artificial Anomaly – B-MW9A and B-MW34A</u> (Yellow and Green)

This involves a new source of releases starting up in the furthest northeast Area B shallow wells. The COCs involved are Benzene, Toluene, and Xylenes (BTX). Actual increases in BTX would be of considerable concern, since these are typically associated with petroleum operations rather than cleaning operations, as are the chlorinated hydrocarbons. Real increases in these COCs would therefore signal releases from a previously unknown source; historical values for BTX in these wells have been all less than 1.0, mostly NDs and J values.

The artificial anomalies introduced are shown in the table at the right. These are considerably lower in magnitude than those introduced in A1-MW9C.

		Benzene	Toluene	Xylenes
Yellow	B-MW9A	4.6	2.3	4.2
	B-MW34A	5.1	0.27J	3.3
Green	B-MW9A	3.3	2.4	2.7
	B-MW34A	2.3	<0.5	3.2

Toluene has not found its way to B-MW34A, in this scenario. There are only three background values for Xylenes for B-MW9A, so it was not evaluated by DT.

DT and GeoTrans found all of the anomalies for which there is enough background data. The GeoTrans interpretations are “A” for Benzene at B-MW34A (since the anomaly value is above the cleanup goal) and “E” for Benzene at B-MW9A, and is “F” for Toluene and Xylenes. “E” and “F” include “no further attention required”. Possibly that conclusion might be modified if it had been noticed that these wells are at the fringe of the site and only 217 feet apart. A potential future upgrade to DT might be to make it easier to make such geographical associations.

Toluene in well MW-9A is affected by the excessively tight PL bounds in the real 2006 data (Orange dataset). Hence, virtually any value above detection could have been found to be out-of-bounds, whether of actual interest or not. It would be desirable for DT to be able to flag this case, with actual increases, but not flag J values and NDs.

Artificial Anomaly – B-20W and B-20WSS (Green)

In the actual 2006 data well B-20WSS was sampled for the first time. Water levels have been taken in this well routinely, but there are no historical COC concentration values. B-20W was also sampled during 2006, as has been discussed.

All historical values from B-20WSS in 2006 are very low; the highest being 1.3 for VC and Toluene. In spatial context this is quite curious, since the well is only 21 feet from B-20W (recall the extremely high values there) and is screened at a similar depth.

At any rate, this anomaly involves again swapping labels, so that the bottle labeled B-20W actually contained the sample from B-20WSS. DT found the drop for c12DCE at B-20W, and GeoTrans appropriately gave it a "B" interpretation, with the comment that the data might be bad. DT did not find the drop in VC at B20W, and presumably the PL bounds for VC include NDs and J values.

Artificial Anomaly – A1-MW21 (Blue and Egg)

The scenario modeled here is that the laboratory instrument had some cross-contamination from the sample from well B-MW35A affecting the sample from A1-MW21. A1-MW21 has historically had low values of c12DCE and VC, and otherwise nearly all NDs and J values. B-MW35A has had values above 100 for c12DCE, TCE, and VC, the latter as high as 375.

The anomalous values are around 19, 5.5, and 98 for c12DCE, TCE, and VC respectively. All were readily detected by DT. The GeoTrans interpretation is "E" for c12DCE, since the anomalous values are below the cleanup goal of 70, and "A" otherwise. Since there is no spatial context for such an occurrence, the only reasonable action would be resampling.

Artificial Anomaly – 12DCA (Egg)

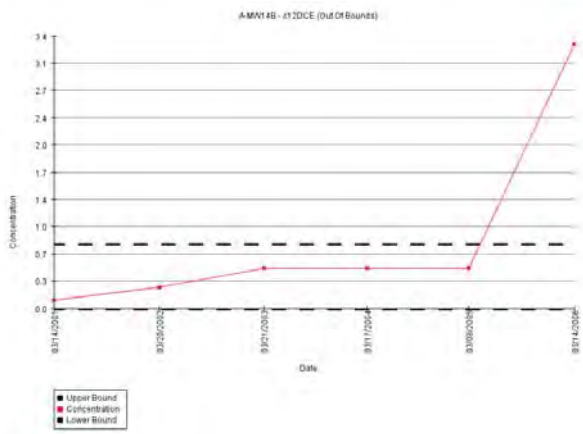
In this scenario, due to clerical or programming error all 12DCA values are reported as 20 times lower than they should be. This can occur, for example, if a needed volume adjustment is performed twice or not at all.

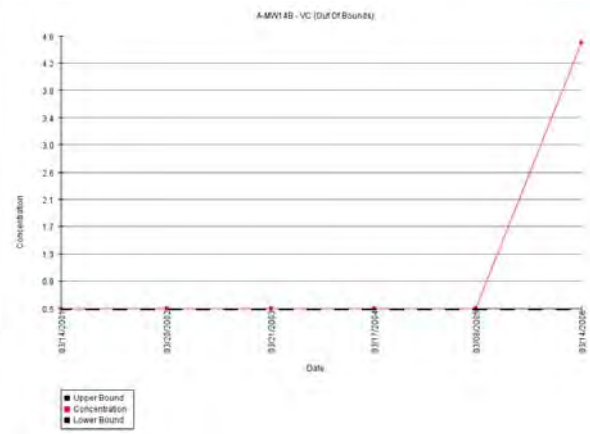
This affected seven wells with somewhat elevated historical values of 12DCA; B-MW33B was an eighth, but since GeoTrans elected not to use the re-swapped version of the historical data for the Egg dataset, there are only three background values. None were detected by DT. On examination of the data, this is appropriate. In three cases NDs remained NDs after dividing by 20. In three other cases the background data are widely variable (in two cases downward-trending time-dependent bounds would be appropriate), so the PL bounds would include even NDs and J values. And for B-20W, dividing the already anomalously high value by 20 left it unremarkable; given the wide data variation in the background data, it also would not be noticed in any case.

As a side note, one often hears values outside the expected range of variation being labeled statistical "outliers". Outliers may be valid data, but are suspect. The converse concept is that of "inliers", which are bad data that are not noticed because they are unremarkable. The 12DCA values in the Egg dataset are inliers.

Artificial Anomaly – A-MW14B (Egg)

In this final scenario A-MW14B, a deep sentinel well at the western fringe of the site, starts seeing low levels of the primary COCs. The historical data are all NDs and J values. Values are quite low: 3.3, 1.2, and 4.5 for 12DCE, TCE, and VC respectively. DT detected all. GeoTrans awarded the VC value an "A" since it is above the cleanup goal of 2; the others received the "F" interpretation due to their low levels.





Possibly the interest or urgency should be a bit higher because this is a sentinel well, and the cleanup goal is not so relevant for sentinel wells. DT does not presently provide information that would alert the user to this; rather, the user would have to know from other sources that this might be a situation of interest because of the geographical location of the well.

Discussion of DT Results

As stated, DT and GeoTrans identified all anomalies that could have been identified given the data available, in both the actual 2006 data (Orange) and the datasets with artificially introduced anomalies.

In addition, it flagged several cases where the ND treatment created very narrow PL bounds, to the extent that J values were found to be out-of-bounds low by comparison with a background of NDs, or a ND was found to be out-of-bounds high by comparison with a background of J values. It would be useful to improve the algorithms used to attempt to avoid these instances.

The GeoTrans interpretations are for the most part reasonable. In some cases it would be better to (a) look at all COCs for a given well simultaneously in arriving at an interpretation or recommended action, (b) look at the COC involved not only at the well in question but also at nearby wells, and/or (c) pay attention to the geographical location of the well involved as well as the cleanup goal in recommending an action. Suggestion (c) simply observes that the cleanup goal is of less importance at sentinel wells and previously uncontaminated deep and extra-deep wells than at wells in more central locations. At sentinel wells any increases are significant.

A number of logistical issues are identified. It is suggested that a future version of DT might take steps to become more flexible with regard to these issues, and possibly alert the user when it discovers these issues in the data. For example, if there are no data for the current monitoring event for a required COC at a required well, an alert is warranted. All of these suggestions are

consistent with the end goal of DT, which is to make it as easy as possible for the user to accurately and reliably screen large amounts of data and focus on potentially important issues.

A future implementation of DT is expected to include time-dependent PL bounds. None of the anomalies in these datasets would have been detected more readily with time-dependent bounds. It is possible, though, that a couple of the anomalies involving increasing trends might not have appeared remarkable with time-dependent increasing bounds. On the other hand, it is not clear that DT should deliberately ignore well/COC combinations with increasing trends.

One DT issue that is not part of the ESTCP dem/val is that of updating the historical/background datasets following the evaluation of one event and prior to screening the next event. EnviroStat recommends automatically adding all current data to the historical dataset following each event. Data that are unremarkable would be automatically added to the background data to be used in creating the PL bounds for the next event ONLY until there are eight background observations, however, in order to avoid masking slow trends in the data. Otherwise, and in any case where data in the current data have been flagged as potentially anomalous, the user should have the option of including the values in the background data or not.

Similarly, the user will on occasion want to update the background data by excluding the earliest observations. At Camp Allen, with only annual data and a fairly recent history, that should be done judiciously. In other cases, though, one possible decision following evaluation of an out-of-bounds value is that (a) the situation has indeed changed, and (b) the change is not of concern, but (c) it would be desirable to update the background data to avoid flagging this particular situation in the future.

Finally, a *caveat* regarding the confidence level used in the PL bounds is in order. For Camp Allen these bounds are nominally 90% prediction limits. As stated previously, with more background data one would like to increase the confidence level to, say, 95% for application in situations such as that at Camp Allen. Such situations exist where there is known contamination and out-of-bounds findings are used only by the facility as a source of information on the progress of a remediation program, or for similar internal information purposes. Such an application is rather different from the use of PL bounds in a formal monitoring program, such as groundwater detection monitoring conducted under RCRA (40 CFR Parts 264, 265, or 258) or similar state regulations. In those settings one uses background data to set upper prediction limits. An out-of-bounds value, by regulation and/or by permit, will trigger sometimes rather costly regulatory activity. In such settings the setting of the confidence level requires more sophistication. An extended discussion of this issue is beyond the scope of this report, but see "Ground-Water Monitoring Statistics Update: Part I: Progress Since 1988" (C.B. Davis and R.J. McNichols, *Ground Water Monitoring and Remediation* 1994, pp. 148-159) and articles referenced therein, for example.

Model Validation

NOTE – THIS ONLY APPLIES TO THE ESTCP PROJECTY ANALYSIS, AND WOULD NOT BE DONE FOR TYPICAL EVALUATIONS WITH OPTIMIZER

The baseline Camp Allen analysis with Optimizer was conducted on historical data from 2005 and earlier. The 2006 data that were reserved and not used in that evaluation were then used for "validation". The plumes for each COC were evaluated with and without the redundant data using the reserved dataset, and the predicted values at the eliminated points were compared to the actual values in order to quantify the loss of information resulting from eliminating the redundant data.

The 2006 reserved data (provided by EnviroStat) includes:

- 21 shallow sampling locations, and
- 25 deep sampling locations

The general procedure for validation is as follows:

- Create a full dataset containing 2006 reserved data, and run Model Builder on the full dataset to obtain the plume maps

- For Optimizer setup, set frequency for each data location as "fixed"

    - Set frequency as "always off" for the data points recommended to be removed (spatial and spatiotemporal) or to have less-than-annual frequency (spatiotemporal) from the full dataset based on the selected "optimal" plans

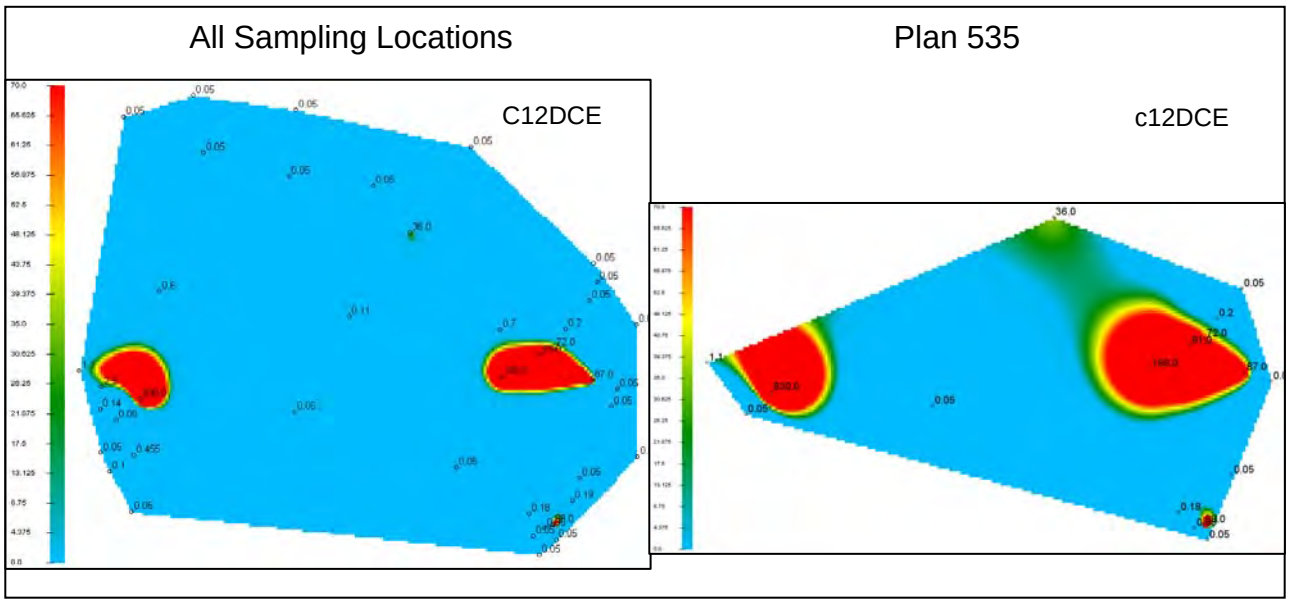
    - Set frequency as "always on" for all other data points from the full dataset

- Run Optimizer on "optimal" plans with only 1 generation to obtain the plume maps with interpreted concentrations at removed locations

- Compare results for the full dataset and the "optimal" plans to evaluate the loss of information from removing the data points

    - Maps
    - Estimated versus actual values at removed points

The reason we need to use Optimizer in this process is that Model Builder does not extrapolate concentration values at the locations outside of the convex hull of the data points, nor does it interpret concentration value at a specific (i.e., removed) location. Therefore, if points located at the plume edge are removed, the area surrounding the removed point is not interpreted and appears to be "missing" in Model Builder. Thus, it is hard to compare the concentration values at removed locations to the actual values purely based on plume maps using Model Builder. The following figure illustrates an example of a side-by-side comparison of plume maps of the full dataset versus one "optimal" plan for the shallow aquifer based on results from Model Builder only.



This clearly makes for an unsatisfying comparison! However, using the approach described above with Optimizer, which extrapolates concentration values and also provides estimated concentration at a specific (i.e., removed) location, we can make the comparison we wish to make. By doing tricks for Optimizer like setting frequency for all points as fixed and assigning only 1 generation, there is actually no optimization performed but estimated concentration values at removed locations can be obtained and the hull remains the same.

Model Validation for Spatial Analysis

The process of validation for the spatial analysis was as follows:

- Replace concentrations with 2006 data for the wells that were sampled in 2006.
    - For the shallow aquifer, 42 wells were used for original spatial analysis, but only 21 of them were sampled in 2006. Thus, 21 wells were replaced with 2006 data, and the original historical data were used for the other 21 wells, which creates the "full" dataset for validation.
    - For the deep aquifer, 31 wells were used for original spatial analysis, but only 25 of them were sampled in 2006. Thus, 25 wells were replaced with 2006 data, and the original historical data were used for the other 6 wells, which creates the "full" dataset for validation.
- Run Model Builder for the full dataset using kriging with quantile transformation to obtain the plume maps for the full dataset.
- Based on the model generated through Model Builder, set both the minimum and maximum frequency as "off" at wells that are recommended to be removed by the Optimizer from the full dataset, and as "on" for all other wells

- For the shallow aquifer, Plan 444 and Plan 535 were selected with 13 wells and 25 wells recommended to be removed, respectively. Thus, set minimum and maximum frequency as "off" for the 13 wells and the 25 wells from the full dataset, and set other wells with minimum and maximum frequency as "on".

- For the deep aquifer, Plan 340 and Plan 402 were selected with 7 wells and 10 wells recommended to be removed, respectively. Thus, set minimum and maximum frequency as "off" for the 7 wells and the 10 wells from the full dataset, and set other wells with minimum and maximum frequency as "on".

- Run Optimizer for the "optimal" plans with only 1 generation to obtain plume maps for the "optimal" plans with interpreted concentrations at the removed locations.

- Compare the plume maps and estimated concentration values versus the actual values between "full" dataset and the "optimal" datasets.

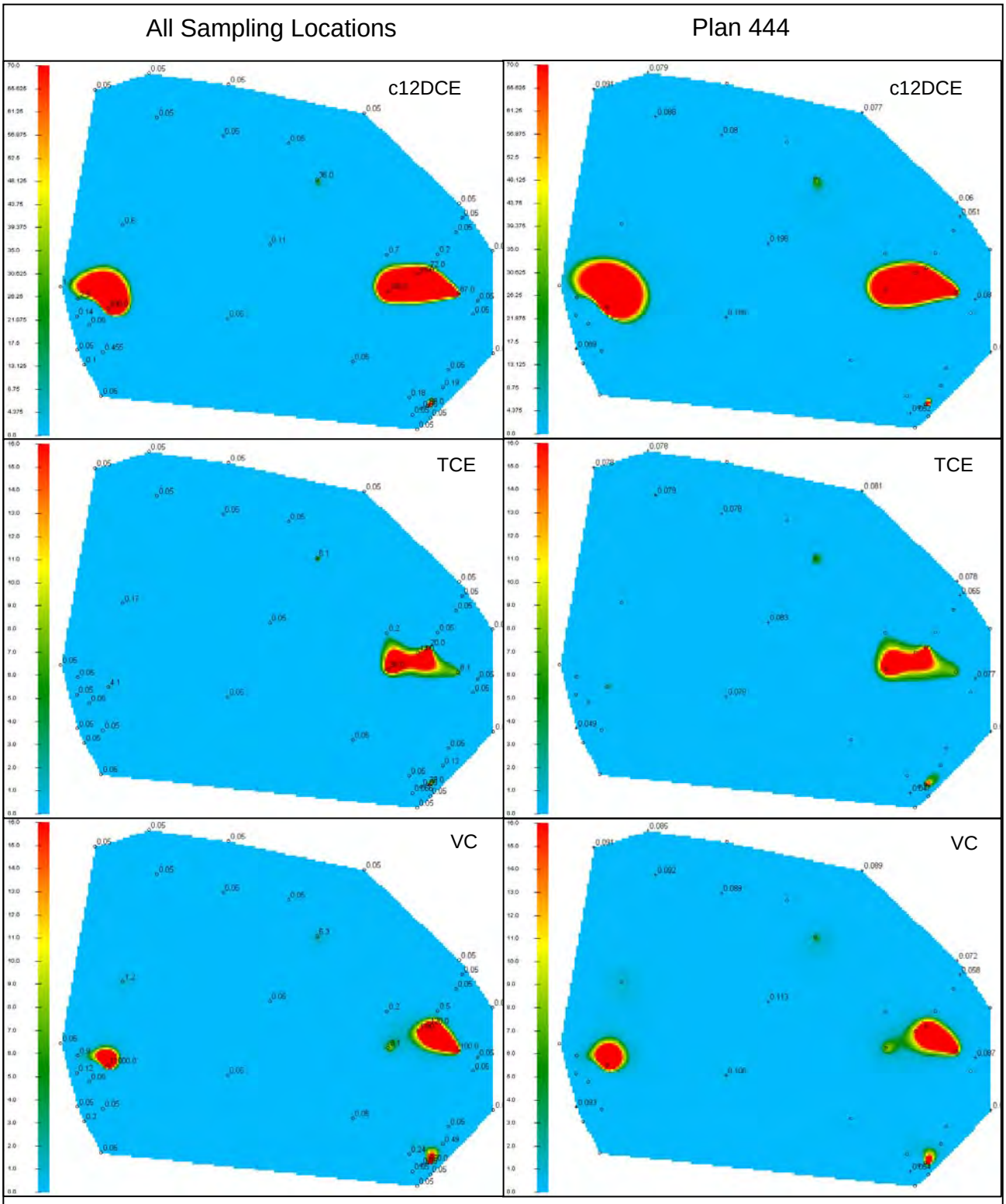
The following two figures illustrate the side-by-side comparison of plume maps for the full dataset and for the optimal plans 444 and 535 for the shallow aquifer.

- For Plan 444, the plume maps without redundant points reasonably represent the plume distribution with redundant points. The loss of information due to removing 13 points has little impact on interpretation of plume distribution.

- For Plan 535, removing 25 redundant points causes greater difference in interpreted concentration versus the full dataset. Concentrations up to 30 µg/l are interpreted in the northwest of c12DCE plume which is ND in full dataset, but are still interpreted below the cleanup goal of 70 µg/l. Also the size of the hot-spot areas are interpreted larger than the full dataset. The loss of information due to removing 25 points has a noticeable impact on interpretation of plume distribution, which may or may not be of concern to the installation or the regulators.

Note that there is greater deviation between Plan 535 and the full dataset in the validation sampling (2006 data) than for the dataset used in the baseline evaluation. The actual concentration values versus the actual values at the removed locations are listed in the tables following the figures.

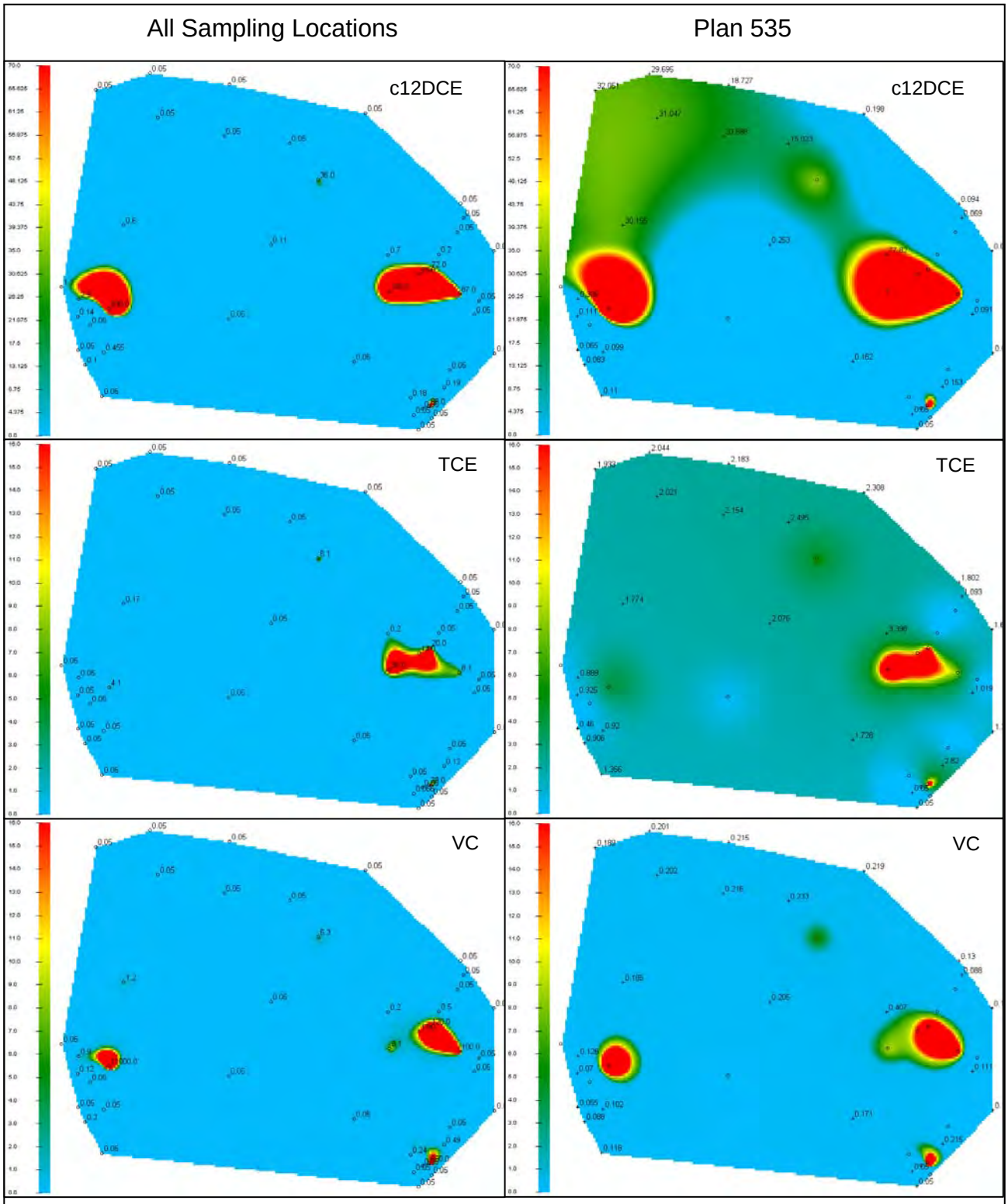
Plume Map Comparison (Spatial) with 2006 Reserved Data, Shallow Aquifer



Note:

- Symbol "+" indicates wells that are recommended to be removed. Symbol "o" indicates wells that are recommended to keep;
- Only interpreted concentrations at the removed locations are posted for the selected plan.

Plume Map Comparison (Spatial) with 2006 Reserved Data, Shallow Aquifer



Note:

- Symbol "+" indicates wells that are recommended to be removed. Symbol "o" indicates wells that are recommended to keep;
- Only interpreted concentrations at the removed locations are posted for the selected plan.

Interpolated Concentrations versus Actual Concentrations at Removed Wells
Optimal Plan 444

Removed Well	Actual Concentration (µg/l)			Estimated Concentration (µg/l)		
	c12DCE	TCE	VC	c12DCE	TCE	VC
A-GW-3	0.05	0.05	0.05	0.077	0.081	0.089
A-MW17A	0.05	0.05	0.05	0.088	0.079	0.092
A-MW18A	0.05	0.05	0.05	0.079	0.078	0.085
A-MW30A	0.05	0.05	0.05	0.091	0.078	0.091
A1-MW25A	0.05	0.05	0.05	0.089	0.049	0.093
A1-MW4A	0.05	0.05	0.05	0.186	0.079	0.106
A2-MW30	0.05	0.05	0.05	0.08	0.078	0.089
B-MW12	0.05	0.05	0.05	0.08	0.077	0.087
B-MW14	0.05	0.05	0.05	0.038	0.061	0.065
B-MW17	0.05	0.066	0.05	0.052	0.047	0.054
B-MW32	0.05	0.05	0.05	0.051	0.065	0.058
B-MW9A	0.05	0.05	0.05	0.06	0.078	0.072
GW-2	0.11	0.05	0.05	0.198	0.083	0.113

Interpolated Concentrations versus Actual Concentrations at Removed Wells
Optimal Plan 535

Removed Well	Actual Concentration (µg/l)			Estimated Concentration (µg/l)		
	c12DCE	TCE	VC	c12DCE	TCE	VC
A-GW-3	0.05	0.05	0.05	0.198	2.308	0.219
A-MW12	0.05	0.05	0.05	15.023	2.495	0.233
A-MW17A	0.05	0.05	0.05	31.047	2.021	0.202
A-MW18A	0.05	0.05	0.05	29.695	2.044	0.201
A-MW30A	0.05	0.05	0.05	32.051	1.933	0.189
A1-MW21	0.455	0.05	0.05	0.099	0.92	0.102
A1-MW25A	0.05	0.05	0.05	0.065	0.46	0.055
A1-MW31A	0.1	0.05	0.2	0.083	0.906	0.088
A1-MW6A	0.05	0.05	0.05	0.11	1.356	0.118
A1-MW8A	0.14	0.05	0.12	0.111	0.325	0.07
A1-MW9A	2.9	0.05	0.9	0.306	0.889	0.126
A2-MW30	0.05	0.05	0.05	20.888	2.154	0.216
A2-MW34A	0.05	0.05	0.05	18.727	2.183	0.215
B-15WA	0.6	0.17	1.2	30.155	1.744	0.185
B-GW-6	0.05	0.05	0.05	0.162	1.728	0.171
B-MW10	0.038	0.047	0.05	0.081	1.673	0.133
B-MW13	0.05	0.05	0.05	0.091	1.019	0.111
B-MW14	0.05	0.05	0.05	0.05	1.304	0.108
B-MW17	0.05	0.066	0.05	0.05	0.05	0.05
B-MW19A	0.7	0.2	0.2	77.82	3.398	0.407
B-MW29R	0.19	0.12	0.49	0.153	2.82	0.215
B-MW32	0.05	0.05	0.05	0.069	1.093	0.088
B-MW37A	0.05	0.05	0.05	0.05	0.05	0.05
B-MW9A	0.05	0.05	0.05	0.094	1.802	0.13
GW-2	0.11	0.05	0.05	0.253	2.076	0.205

The errors for Plan 535 are significantly higher than for Plan 444, as expected because fewer points are retained. Note that the cleanup goals are 70 µg/l for c12DCE, 5 µg/l for Benzene, and 2 µg/l for VC. Thus, the higher errors for C12DCE should be evaluated relative to the higher cleanup goal for that parameter. The larger errors for Plan 535 may or may not be of concern to the installation or the regulators.

For the deep aquifer, the following two figures illustrate the side-by-side comparison of plume maps with redundant data and without redundant data for two selected plans: Plan 340 (7 locations removed) and Plan 402 (10 locations removed). Plan 340 better represents the plume distribution with redundant data than Plan 402, particularly in the southeast corner. For Plan 402, a data point (B-MW15B) located in the southeast corner is recommended to be removed from the monitoring network. This location had concentrations of either ND or low values for all three COCs for the past years, but had much higher concentrations than cleanup goals for all three COCs in 2006. This was identified by Data Tracker as "anomalous" data, and this anomaly significantly impacts the validation exercise for Plan 402 in the southeast corner.

The estimated concentration values and the actual values at the removed locations for optimal Plan 340 and Plan 402 are listed in the following table.

Interpolated Concentrations versus Actual Concentrations at Removed Wells
Optimal Plan 340

Removed Well	Actual Concentration (µg/l)			Estimated Concentration (µg/l)		
	c12DCE	TCE	VC	c12DCE	TCE	VC
A-MW18B	0.92	0.05	0.05	12.468	1.124	0.3
A-MW19B	1.11	0.05	0.28	3.448	0.238	0.09
A-MW30B	21	0.84	2.8	9.944	0.311	1.831
A1-MW4B	0.05	0.05	0.05	15.214	0.291	1.468
A1-MW6B	7.1	0.7	0.21	9.51	1.891	0.753
A2-MW32B	0.05	0.05	0.05	9.895	0.751	0.373
B-MW22B	0.05	0.05	0.05	21.33	12.50	3.718

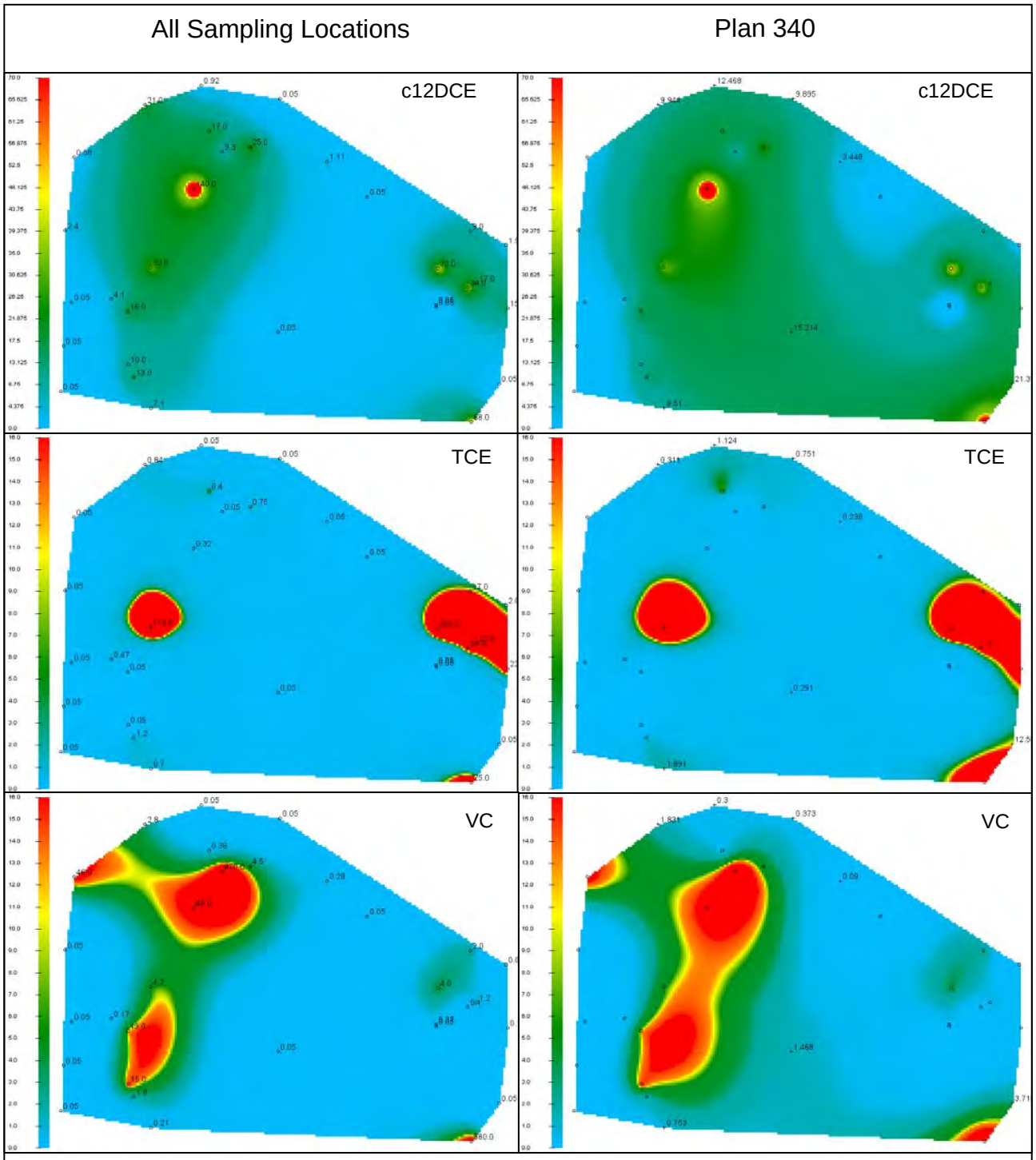
Interpolated Concentrations versus Actual Concentrations at Removed Wells
Optimal Plan 402

Removed Well	Actual Concentration (µg/l)			Estimated Concentration (µg/l)		
	c12DCE	TCE	VC	c12DCE	TCE	VC
A-MW18B	0.92	0.05	0.05	13.208	0.841	0.193
A-MW19B	1.11	0.05	0.28	3.248	0.219	0.111
A-MW30B	21	0.84	2.8	9.985	0.295	2.589
A1-MW4B	0.05	0.05	0.05	9.164	0.228	0.145
A1-MW6B	7.1	0.7	0.21	6.14	1.003	0.155
A2-MW32B	0.05	0.05	0.05	10.309	0.468	0.245
B-MW15B	88	25	380	0.246	0.132	0.054
B-MW2C	0.05	0.05	0.37	0.13	0.074	0.058
B-MW33B *	34	21	0.1	6.814	1.657	0.138
B-MW3B *	17	22	1.2	8.65	6.257	0.139

** plume interior wells*

Note the higher errors at B-MW15B in Plan 402, which as discussed earlier is due to the anomalous data and not caused by the software.

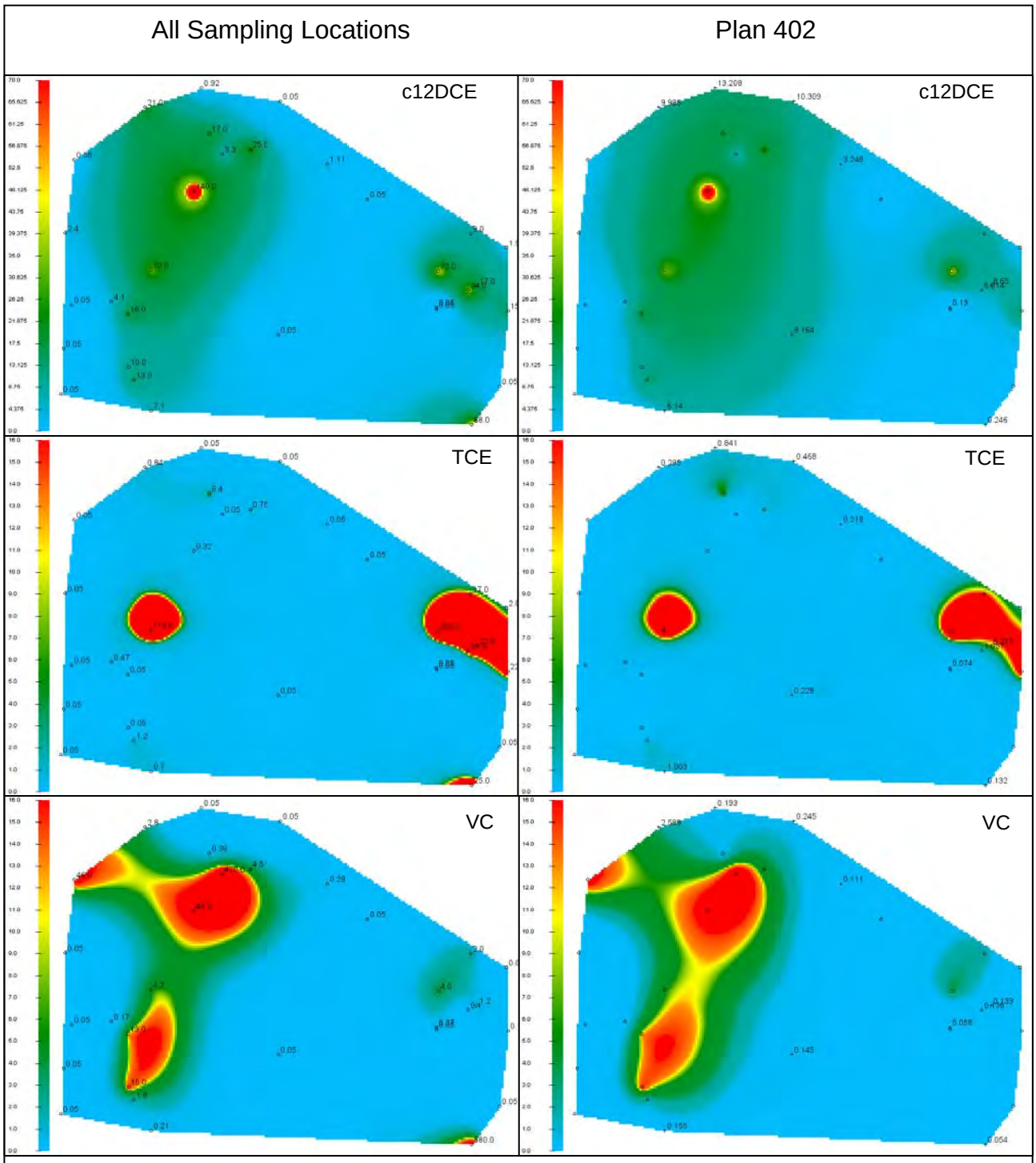
Plume Map Comparison (Spatial) with 2006 Reserved Data, Deep Aquifer



Note:

- Symbol "+" indicates wells that are recommended to be removed. Symbol "o" indicates wells that are recommended to keep;
- Only interpreted concentrations at the removed locations are posted for the selected plan.

Plume Map Comparison (Spatial) with 2006 Reserved Data, Deep Aquifer



Note:

- Symbol "+" indicates wells that are recommended to be removed. Symbol "o" indicates wells that are recommended to keep;
- Only interpreted concentrations at the removed locations are posted for the selected plan.

Model Validation for Spatio-Temporal Analysis

We believe the most appropriate way to do the validation for spatio-temporal is to compare the actual 2006 sampling data versus the 2006 sampling program(s) recommended by the spatio-temporal optimization. Recall that the spatio-temporal analysis conducted previously excluded some wells from the input data due to fewer than 4 sampling results for these wells (software requirement). Thus, outside of the software we would add those excluded wells back into the recommended plans. These wells are:

A1-MW4B A2-MW11B B-MW22B B-MW2B B-MW2C

We first checked to confirm these wells would continue to be sampled, which was the reason for adding them back into the recommended plans. We found that four of the five were in fact sampled in 2006. B-MW2C was not sampled in 2006, but B-MW2C and B-MW2B are clustered wells that have been treated as in the same aquifer, thus B-MW2B can provide enough information and missing B-MW2C in 2006 data does not likely cause information loss.

The subsequent process for validation for the spatio-temporal analysis was as follows:

- Use 2006 data as the "full" dataset for validation. Since no spatio-temporal analysis was conducted for the shallow aquifer, the validation is only performed for the deep aquifer.

- Run Model Builder for the full 2006 dataset (regardless of historical sampling frequency at each well) using kriging with quantile transformation to obtain the 2006 plume maps for the full dataset.

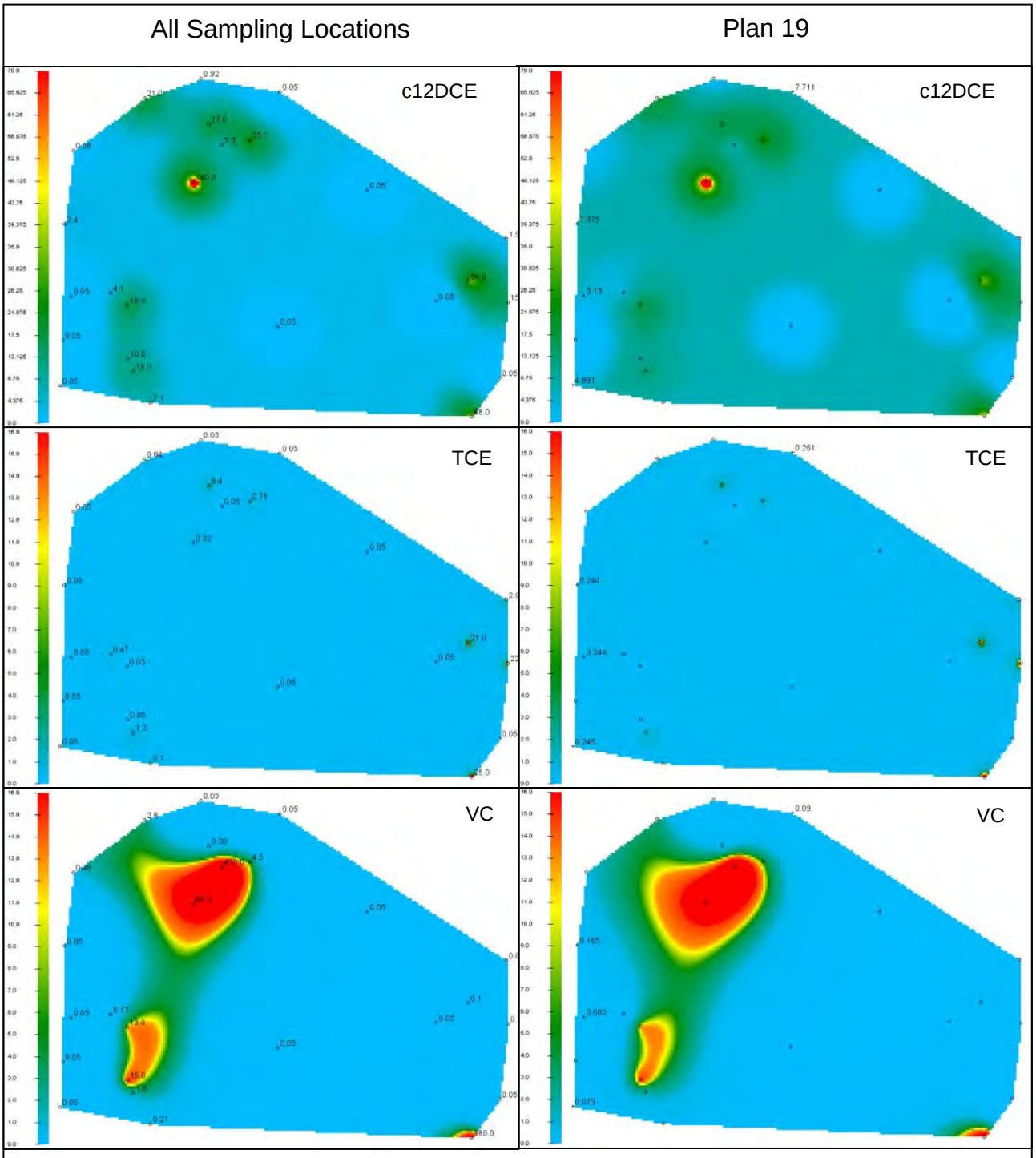
- Set both the minimum and maximum frequency as "off" at wells that are recommended to have less-than-annual sampling frequency or recommended to be removed, and as "on" for all other wells, for the two selected "optimal" Plan 19 and Plan 73.

- Run Optimizer for the "optimal" plans with only 1 generation to obtain plume maps for the "optimal" plans with interpreted concentrations at the removed locations.

- Compare the plume maps and estimated concentration values versus the actual values between "full" dataset and the "optimal" datasets.

The following two figures illustrate the side-by-side comparison of plume maps with redundant data and without redundant data for two selected plans: Plan 19 (no wells removed, but 4 wells recommended to be less than annual) and Plan 73 (2 wells removed, plus 2 wells recommended to be less than annual). In each case, all wells recommended to have less than annual sampling were removed in 2006. Plan 19 reasonably represents the plume distribution with all the 2006 data. Loss of information due to removing wells with recommended sampling frequency less than annual has little impact on the interpretation of plume distribution. For Plan 73, a data point (B-MW15B) located in the southeast corner is recommended to have biennial sampling. This location had concentrations of either ND or low values for all three COCs for the past years, but had much higher concentrations than cleanup goals for all three COCs in 2006. This was identified by Data Tracker as "anomalous" data, and this anomaly significantly impacts the validation exercise for Plan 73 in the southeast corner.

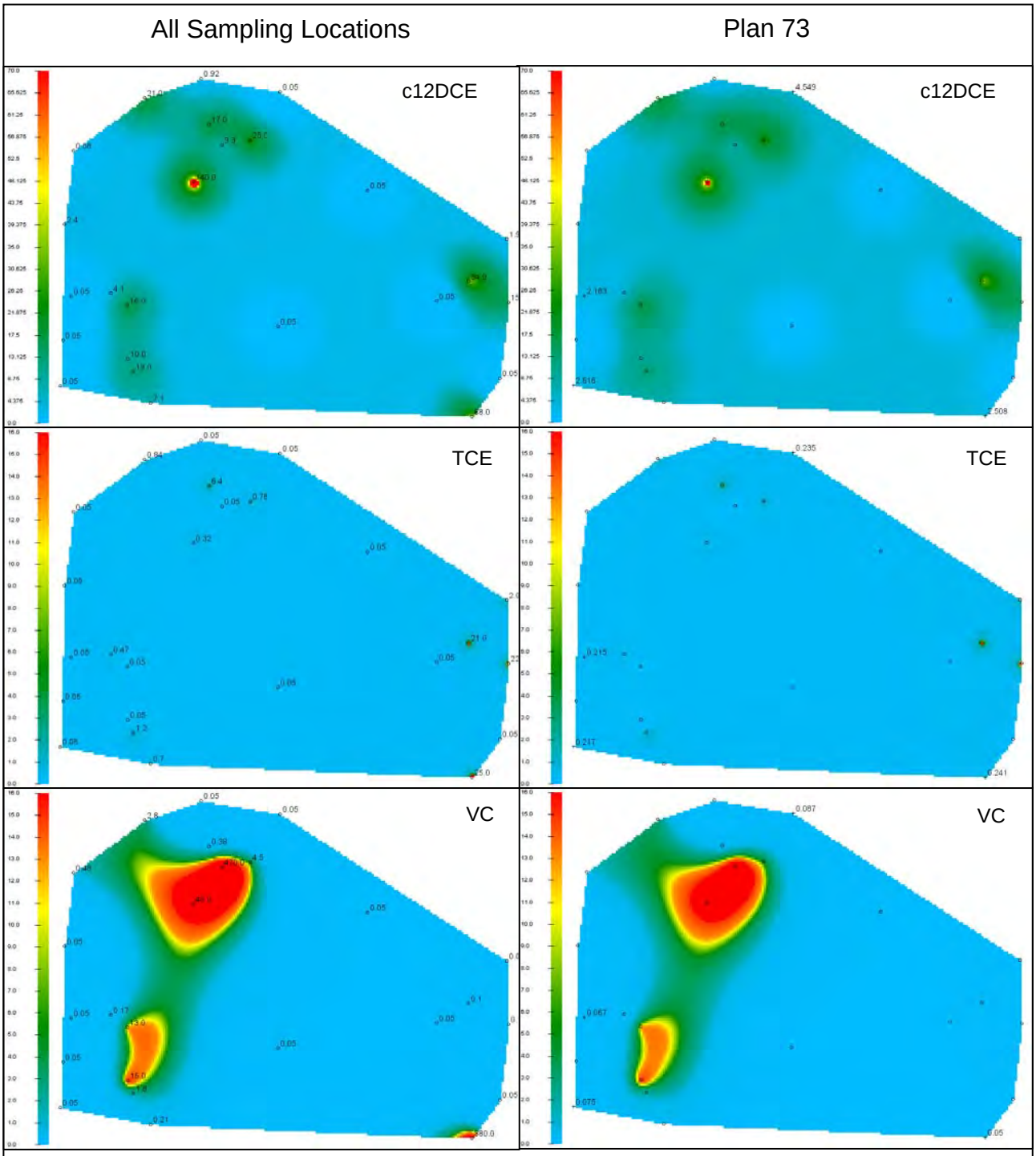
Plume Map Comparison (Spatio-Temporal) with 2006 Reserved Data, Deep Aquifer



Note:

- Symbol "+" indicates wells that are recommended to be removed. Symbol "o" indicates wells that are recommended to keep;
- Only interpreted concentrations at the removed locations are posted for the selected plan.

Plume Map Comparison (Spatio-Temporal) with 2006 Reserved Data, Deep Aquifer



Note:

- Symbol "+" indicates wells that are recommended to be removed. Symbol "o" indicates wells that are recommended to keep;
- Only interpreted concentrations at the removed locations are posted for the selected plan.

The estimated concentration values and the actual values at the "removed" locations for optimal Plan 19 and Plan 73 are listed in the following table.

Interpolated Concentrations versus Actual Concentrations at Removed Wells
Optimal Plan 19

	Actual Concentration (µg/l)			Estimated Concentration (µg/l)		
	c12DCE	**TCE**	**VC**	**c12DCE**	**TCE**	**VC**
A-MW14B	0.05	0.05	0.05	3.13	0.244	0.082
A-MW15B	2.4	0.05	0.05	7.375	0.244	0.165
A-MW31B *	0.05	0.05	0.05	4.881	0.246	0.079
A2-MW32B	0.05	0.05	0.05	7.711	0.261	0.09

* *plume interior wells*

Interpolated Concentrations versus Actual Concentrations at Removed Wells
Optimal Plan 73

	Actual Concentration (µg/l)			Estimated Concentration (µg/l)		
	c12DCE	**TCE**	**VC**	**c12DCE**	**TCE**	**VC**
A-MW14B	0.05	0.05	0.05	2.183	0.215	0.067
A-MW31B *	0.05	0.05	0.05	2.616	0.217	0.075
A2-MW32B	0.05	0.05	0.05	4.549	0.235	0.087
B-MW15B	88	25	380	2.508	0.241	0.05

* *plume interior wells*

Note the higher errors at B-MW15B in Plan 73, which as discussed earlier is due to the anomalous data and not caused by the software.

APPENDIX E:

EPA REGION V EVALUATION

ESTCP LTM-Optimization Project
Questionnaire for Software Application by Third-Party Participants

{please answer as many questions as possible}

1. Who applied the software (name, organization, level of expertise)?

   Rachel Shannon, hydrogeologist, SS Papadopulos & Assoc, Inc.
   First time using sampling frequency optimization software. Limited experience (<1 yr) with optimization programs in general (petroleum systems risk analysis)

   Matthew Tonkin, hydrogeologist, SS Papadopulos & Assoc, Inc.
   Previous experience (8 years) with sampling frequency optimization software, including MAROS, PAM, various statistical techniques, and custom in-house program development.

2. Please indicate site name and a *brief* description of the plume (contaminants, monitoring network, etc.).

   Site Name: Charlevoix, MI.
   Plume: Two plumes – one of TCE, one of PCE. Multiple sources of PCE; likely only one source of TCE. Source remediation (SVE, soil removal) previously completed. Monitoring network developed sporadically through time. Site evaluation part of third statutory USEPA 5 Year Review.

3. Please indicate which modules of the software you applied.

   Model Builder?
   Yes

   Sampling Optimizer – Spatial?
   No

   Sampling Optimizer – Spatiotemporal?
   Yes.

   Data Tracker?
   No.

4. For Sampling Optimizer, which constituents of concern did you evaluate?

   TCE
   PCE

5. For Data Tracker (if applied) which constituents of concern did you evaluate?

   N/A

6. Are the *Summit* monitoring tools easy to use and user-friendly? Did you find the user's Manual well structured, clear and complete? How long did it take for you to get comfortable with the software (hours, days, or weeks…or never)?

   The user interface was very easy to use. Export of results could be improved- suggestions are given in the response to question 16 below. User's manual was an excellent reference for set-up and execution, and it contained clear directions for navigating dialog boxes, setting parameters, formatting input files, etc.

   However, there was little information about theory and method. It would be useful to add to the manual or provide separately the following items, to avoid the software being used as a "black-box":

   1. Searchable index and terms glossary- (Genetic algorithm, non-denominated plans, objective functions, etc)
   2. Searchable list of error messages and suggestions for resolving the problem.
   3. Description of theory and method – e.g.., exactly how the input data and model parameters are used and how changes in parameters may affect results; why default parameter values are chosen; how the genetic algorithm is used, etc. References for additional reading may be helpful.
   4. Description of output plans, especially (a) how to tell which plan is which on the tradeoff curve, and (b) why there are varying numbers of plans for different scenarios.
   5. Tutorial

   It took only few hours to get comfortable using the software (import/export, model set up, running the program). The user's manual was very helpful in this aspect. It took a few days to fully understand the method, the effects of changes in parameter values, and the results.

7. Have you found any significant errors or bugs when applying the *Summit* monitoring tools? If any, what are they?

   1. Use of the software depends on screen resolution. We ran the program on several computers and in one case, the maximum screen resolution was too low and the program would not run. To work around this problem, we ran the program using remote desktop from another computer.

2. Computational demand is very large, which limits the computers that the code could be applied on.

8. The Model Builder offers six possible combinations of interpolation technique (kriging, inverse distance) and data transformation method (e.g., quantile, logarithmic, none). How many of these six combinations did you try? If you tried more than one, please rank them in terms of your preference, and briefly describe why?

   Kriging with logarithmic transform was used in all cases.

9. In Model Builder, other than specifying the interpolation technique and data transformation method, did you change any other settings or did you use all of the defaults? What did you change?

   Most settings were left as defaults, except the number of vertical slices, which was changed to 250. This seemed to be the best trade-off between resolution and drawing speed.

10. When importing data, what did you do regarding non-detect values?

    Two methods were used, depending on the scenario. Either, (a) the value was set to 1/10 of the assumed detection limit of 0.1 ug/L; or (b) a value was estimated based on linear interpolation in time.

11. Were the tradeoff curve(s) generated by the Sampling Optimizer understandable and/or useful?

    The trade- off curve was understandable, but it would be better if it was possible to set XY axis extents flexibly; i.e. at the end of the optimization, curve was zoomed in, and it was not possible to see the rest of the curve (see bugs list). We are uncertain if this has been remedied.

    Did the results from the Sampling Optimizer suggest sampling plans that would result in significant reduction in monitoring locations and/or frequency?

    The final plan did not show a substantial decrease in sampling frequency in most wells. Two wells were selected for abandonment. This may be because (a) there was a relatively small number of monitoring wells, given the size of the site; (b) the current sampling plan was arbitrary and led to very infrequent sampling at most wells; and, (c) several of the monitoring wells exhibited contamination above cleanup levels.

    How did you select potential plans to consider implementing based on the many plans that defined the tradeoff curve(s)? *and* How did the number of sampling

locations and sampling frequency in one or more "promising" sampling plans that were identified compare to the "base model"?

In order to evaluate the robustness of the results obtained using the optimizer, the optimization was executed several times with different scenarios – including (a) different data sets, and (b) different variogram parameters. The final sampling plan was based on (a) interpretation of the results of all analyses completed, and (b) interpretation of temporal trend statistics obtained using the USEPA PAM software (Subterranean Research, Inc).

In order to test the sensitivity of the results to several parameters, several scenarios were executed with different variograms; different time interpolation of data to prepare input data sets; and, selection of representative data for a specific location since the site comprises many vertically nested wells and the sample optimizer currently only considers two-dimensional (2D) data. For the latter, data sets were constructed using (a) the average concentration over all depths, and (b) the maximum concentration for any depth,

Each of these sensitivity-analysis scenarios resulted in several plans. In order to consider all plans from all scenarios, an analysis was completed by counting the number times a sampling frequency was selected for a well across all plans for all scenarios, and divided that number by the total number of plans. This yields a percentage or fraction of times a specific frequency was selected for a well from all analyses. The higher the percentage, the more robust and reliable the result was considered.

12. For Sampling Optimizer, did you change any settings (such as population size or number of generations for the genetic algorithm) or did you use all of the default settings? What did you change?

    These were left as default values.

For spatial analysis with Optimizer, which requires one value per location for the "base model", what rule did you use to assign that value?

We used two methods: one selected the maximum value at an XY location; the other selected the average value. Note that it was also necessary to perform some time interpolation since the current sampling strategy at the site is very arbitrary, and most monitoring events were not complete. We used time interpolation to construct reasonable estimates of concentration values at concurrent timeframes as an approximation to a single monitoring event.

13. If you performed spatio-temporal analysis, were some locations not included in the "base model" because there were not enough historical data at those locations? What, if anything, did you do to address this issue?

Yes, some locations were excluded. This was addressed by adjusting input data. Two methods were used:

1. All data was grouped into sampling rounds. All data is assigned the first date in the sampling round. If grouping resulted in duplicates, the higher concentration was used. Missing data were linearly interpolated through time based on data from surrounding wells.
2. All data was grouped into quarters. If grouping resulted in duplicates, the maximum value was used. If a well was missing a value for a quarter, it was assigned based on linear interpolation in time (i.e., estimated from measurements at the well before and after the missing quarter). Artificial "sampling rounds" were set- in this case, second and fourth quarter for each year. The final input dataset consisted of measured or interpolated data for each sampling round.

14. If you applied Data Tracker, what was the date range of "historical data" and approximately how many sampling events were in the "historical data"? Did Data Tracker identify any data anomalies in the "current data"? Were those anomalies determined to be of site-specific concern?

    N/A

15. Based on your experience applying the *Summit* monitoring tools, what do you consider to be the most effective, important, or useful features, and why?

    1. The ability to evaluate/produce an ensemble of potential sampling plans: this is not common in other software developed for LTM purposes.
    2. If the graphing of the pareto set was more flexible (i.e., scalable, labels, clearly visible, etc) this would be a very strong feature.

16. What areas of improvement, if any, would you suggest the programmers work on to further enhance the performance of the *Summit* monitoring tools?

    <u>Variography, Objective Functions</u>

    1. An independent analysis was made of the variogram for each data set, through time. Consistent with the general declining pattern of concentrations, the sill of the variogram declined in a monotonic pattern with time. Use of a single variogram calculated on the basis of all data through time misses this pattern. Would the ability to modify the variogram through time (principally, the sill, as the range changed little in this instance) change the results of the Optimizer?
    2. Objective functions: while the objective function capabilities are fairly flexible, it would be useful if additional objective functions could be used, such as:
        - Mass-related.
        - Foot-print (area) related.

- Area of plume beyond a defined site boundary or compliance point/line. This objective function would be of considerable interest since often low concentrations define the plume boundaries, so that they may not impact the current objective function markedly, but may impact the footprint of the contamination significantly.

File Management

1. Include comment lines in the input .csv files for the Optimizer.
2. Save paths for projects and default to save output to the same directory. When running more than one simulation at a time, it can be confusing which output belongs to which simulation.
3. Do not allow all settings reset to defaults when importing a new dataset. We would like to be able to set up a project (variogram, resolution, etc), copy the project, and run it again with different input datasets.

Results Post-Processing

1. Save model visualization and plume maps in an exportable output format (e.g., Surfer or ArcMAP ASCII grids, shapefiles, etc) so that it can be viewed in other programs and/or used for reporting purposes.
2. Add a numeric output display. Include average calculated frequency for each plan and standard deviation. This is essentially a post-processing capability that we implemented manually for our purposes to evaluate the robustness of the optimizer results given different data sets and assumptions.

Software Stability/Execution

1. Save periodically during the simulation so that Optimizer does not have to start over if there is a crash, loss of power, etc.
2. Speed up spatio-temporal optimization. In our case, each simulation took between four and six hours, depending on the computer and the number of sampling rounds in the input data. For sensitivity analysis, we ran a total of thirty scenarios, which used about six days of computing time. This rate may not be fast enough for one person/computer to do a thorough sensitivity analysis. In our case we separated TCE and PCE data and ran them independently. Overall time may have been improved if we had run them together; however, since these COCs originated from different sources and exhibited distinct footprints that only overlapped at their margins, we were interested in analyzing them independently.

17. If you have also applied MAROS, please provide any observations regarding the similarities and/or differences, and whether or not you prefer one versus the other (and why).

It has been a considerable time since either person has used the MAROS software, so efforts at a side-by-side comparison given that time lag would likely not be reasonable.

Assessment of the Summit Envirosolutions Sampling Optimizer for Groundwater Sampling Analysis by Application at the Charlevoix, Michigan, Site

S.S. PAPADOPULOS & ASSOCIATES, INC.
Environmental & Water-Resource Consultants

August 27, 2008

7944 Wisconsin Avenue, Bethesda, Maryland 20814-3620 • (301) 718-8900

Assessment of the Summit Envirosolutions Sampling Optimizer for Groundwater Sampling Analysis by Application at the Charlevoix, Michigan, Site

Prepared For:

U.S. EPA Region 5

Prepared By:

S.S. PAPADOPULOS & ASSOCIATES, INC.
Environmental & Water-Resource Consultants

August 27, 2008

7944 Wisconsin Avenue, Bethesda, Maryland 20814-3620 • (301) 718-8900

Table of Contents

Page

List of Figures ii

List of Tables ii

List of Appendices ii

Section 1 Introduction 1

Purpose 1

Site Background 2

Section 2 Optimizer Analysis 3

Input Variables Evaluated 3

Scenarios 3

Sample Data Preparation 4

Variogram Parameter Definition 5

Section 3 Optimizer Results 6

Plan Details/Trade-off Curves: 6

Summarizing Results 6

Section 4 Conclusions and Recommendations 7

Section 5 References 9

Figures

Tables

Appendices

List of Figures

Figure 1 Principal Site Features With Estimated Plume Extents
Figure 2 PCE Sample Variance Versus Time
Figure 3 Example Trade-off Curves
Figure 4 Excerpt From an Example Plan Details Table
Figure 5 Average Sampling Interval Using PCE Data
Figure 6 Average Sampling Interval Using TCE Data
Figure 7 Sampling Frequency Fractions Calculated Using PCE Scenarios
Figure 8 Sampling Frequency Fractions Calculated Using TCE Scenarios

List of Tables

Table 1 Sampling Optimizer Scenarios
Table 2 Calculated Overall Frequency Fractions

List of Appendices

Appendix A Selected Results of PAM Statistical Analyses

REPORT

Section 1
Introduction

Purpose

Under contract to the U.S.Enviornmental Protection Agency (USEPA), S.S. Papadopulos & Associates, Inc. (SSP&A) recently completed an evaluation of the groundwater monitoring network at the Charlevoix Site (the Site), Michigan, as part of the third statutory Five Year Review. This evaluation was undertaken to provide recommendations on suitable sampling locations and frequencies to monitor the migration and remediation of chlorinated solvents in groundwater, and hence to identify which, if any, wells in the current groundwater monitoring network are candidates for abandonment. Detailed results of the analyses were presented in the site report (SSP&A, 2008).

To support this evaluation, the Summit Envirosolutions (Summit) monitoring tools were employed (Summit, 2007). The Summit monitoring tools comprise (a) Model Builder, which creates geostatistical / statistical models of spatial and temporal data; (b) Sampling Optimizer, which identifies redundant sampling locations and frequencies in historical data; and (c) Data Tracker, which enables users to create time-dependent site-wide remediation targets or well-specific targets and evaluate new data relative to those targets.

The analyses described in this report focus on the use of the Sampling Optimizer (the Optimizer), which combines a multi-objective genetic algorithm with kriging interpolation to optimize the locations and frequencies of sampling within monitoring networks (NFESC and GeoTrans, Inc., 2007). The Optimizer produces an ensemble Pareto set of potential sampling plans, that expresses a trade-off between (i) a measure of cost and (ii) a measure of error. The method evaluates the relative redundancy of well samples at given locations and times to develop a “tradeoff curve” (Pareto plot) illustrating sample plans with the lowest “error” for a given expenditure (NFESC and GeoTrans, Inc., 2007).

Typically, optimization methods such as that incorporated in the Optimizer are executed once, or perhaps a small number of times, for a given set of assumptions (such as variogram parameters) and the ensemble of sample plans is reviewed to identify the suite of cost-effective plans that meet monitoring objectives. In the analysis undertaken for the Charelvoix Site the Optimizer was executed numerous times under a range of assumptions, in order to evaluate the robustness of the results to variations in subjective inputs within reasonable ranges.

This report includes:

- Background on the Charlevoix Site
- A description of sampling data at the Site and the steps undertaken to prepare these data for input into the Optimizer program

- A summary of the input data assumptions and model parameters (scenarios) used in the Optimizer simulations
- A description of the method developed to summarize the results of multiple Optimizer simulations and to evaluate the robustness of sampling plans calculated by the Optimizer
- Observations and discussion of Optimizer results as they were applied to the Site evaluation
- Conclusions and recommendations for the application of the Optimizer and similar sampling optimizer methodologies

Site Background

The Charlevoix Site (the Site), is located on an isthmus between Lake Michigan and Round Lake, along the shore of Lake Charlevoix (Figure 1). A municipal well site located on the shores of Lake Michigan, comprised of a shallow well connected to a horizontal flume buried beneath the beach, was abandoned in 1981 when trichloroethylene (TCE) and tetrachloroethylene (PCE) were found in the water supply. Mapping of recent Site sampling data indicates the existence of two relatively distinct plumes, one of TCE that appears to originate from a single source location, and one of PCE originating from multiple source locations (USEPA, 2007; SSP&A, 2008); however, it is noted that some wells located within the PCE contamination footprint historically exhibited TCE contamination that may represent a TCE source or degradation of the PCE. A groundwater monitoring network has developed sporadically over time. The last full sampling round at the site was completed in 2006.

The final Record of Decision (ROD) for the Site concluded that the contaminant plumes may be allowed to dilute through natural flow conditions to Lake Michigan, but required that long-term monitoring of the plumes continue during the natural purging period (USEPA 1985). The most recent USEPA Five-Year Review report determined that the Site was protective of human health and the environment (USEPA, 2001). However, data collected since that report suggest that PCE or TCE concentrations are above cleanup levels in some monitoring wells and/or some wells show increasing trends in concentration. Several monitoring wells have recently been abandoned at the Site, including one well for which the most recent sampling data showed PCE concentration above cleanup levels. The abandonment of certain wells may be inconsistent with one of the stated goals in the final ROD, to maintain long-term monitoring of the plume.

Section 2
Optimizer Analysis

Input Variables Evaluated

Spatio-temporal analysis in the Optimizer requires the following general inputs:

1. Sample data, in the form of a table of well coordinates, sampling dates and corresponding sampled values for the contaminants of concern
2. The parameters of a variogram to be used for interpolation
3. Definition of an objective function that is used to evaluate the error term
4. Definition of an objective function that minimizes the number of samples, a surrogate for minimizing cost (NFESC and GeoTrans, Inc, 2007)

Note: the parameters of the variogram can be estimated on the basis of the Site data using the Model Builder. However, this process was time consuming and during the present analyses it was determined that the variogram parameters calculated by Model Builder were not appropriate for the Site. The calculated nuggets were large, and plume estimates using the recommended variograms seemed somewhat unrealistic; e.g., the estimated plumes showed several small areas of high concentration only in the immediate vicinity of individual wells, rather than a continuous plume across the Site with smooth transitions between high and low concentrations. Therefore, the variogram parameters calculated by Model Builder were not used in this analysis. Instead, the variograms were defined using an empirical approach which is briefly described later in this report.

Since the present study was concerned with the robustness of results with regard to subjective inputs to the Optimizer, the analyses considered the effect of (a) different summaries of the sample data and (b) different variogram parameters on the sampling plans proposed by the Optimizer. As of the time the analysis was performed, the cost objective function was not configurable by the user. All scenarios used the same error objective function settings, listed below:

- σ (Acceptable error for low concentrations): 1.0
- p (Cutoff between low and high concentrations): 5
- q (Acceptable percentage error for high concentrations): 0.2

Scenarios

At the Charlevoix Site, the TCE and PCE contaminants originate from different sources, each plume has a relatively distinct footprint, and these footprints only potentially overlap at their margins (Figure 1). Therefore, the two constituents were analyzed separately. For each constituent, multiple spatio-temporal analyses ("scenarios") were conducted using (a) different

summaries of the sample data and (b) different variogram parameters in order to test the sensitivity of the results to these parameters and to evaluate the robustness of the calculated plans. In total, thirty scenarios were run: 18 for PCE and 12 for TCE (Table 1).

At the same time that these analyses were completed using the Optimizer, independent spatial analyses of the data were completed using various kriging techniques (SSP&A, 2008); and independent temporal analyses were completed using the USEPA PAM software developed by Subterranean Research, Inc. Results of these analyses were used to verify Optimizer results by identifying locations which may be candidates for frequent sampling due to high concentrations, upward trends, proximity to identified source areas, etc. Appendix A provides selected outputs from the PAM software including:

- Standard Test: compares the 95% upper confidence limit (UCL) calculated using the 4 most recent data points, or using data collected after a specified date, to Chemical of Concern (COC)-specific standards. The Standard Test reports the result (Compliance, Exceedance), the 95% UCL, and the COC-specific standard in consistent units.
- Trend Test: identifies upward or downward trends through time. The trend method used is the Sen's Test, a non-parametric trend analysis similar to the Mann-Kendall test. The trend statistics reported are the slope result (Upward, Downward, No Trend) and the slope estimate (in concentration units per year). Upward and Downward tests are each calculated at the 95% confidence level. Because the trend is calculated on the natural logarithm of the concentration, the slope estimate is reported in terms of the log of the concentration units per year.
- Baseline Test: compares the most recent datum to the upper prediction limit (UPL) calculated from a baseline subset of the data, the first 8 available samples collected at each point. The Baseline Test reports the result (Better, Worse, No Change); and the 95% prediction limit UPL.

Sample Data Preparation

PCE and TCE concentration data were available from sampling conducted from June 1996 to November 2006. A total of 39 locations were measured during that time, but the sampling plan was arbitrary and sampling was infrequent for many locations. Therefore, the following general approach was used to prepare a complete dataset for input into the Optimizer program for each of the PCE and TCE constituents:

1. Data were grouped into (a) annual or (b) semi-annual "sampling rounds." Concentration values were assigned to the sampling round closest to the actual measurement date. If grouping resulted in duplicates, the maximum value was used.
2. Since the spatio-temporal analysis is applicable only in two dimensions, depth-discrete sampling information was not used. For vertically nested wells possessing depth-discrete data, either (a) the maximum value or (b) the average value was used.

3. The spatio-temporal analysis requires at least four data points per location, and at least 15 locations per sampling round. For some sampling rounds and/or sampling locations, not enough measured data were available. In those cases, concentration values were estimated based on linear interpolation in time; i.e., if a sampling location had no measurements at the time of a sampling round, a value was assigned to that location and sampling round based on measurements taken at that location before and after the sampling round.

Unfortunately, the method reporting limits used for PCE and TCE were not available at the time that the analysis was undertaken. Therefore, reported non-detect values were handled one of two ways:

1. The detection limit was assumed to be 0.1 ug/L, and values for non-detects were set to 1/10 that value (0.01 ug/L); or,
2. A value was interpolated from measurements taken before and after the sampling round.

Variogram Parameter Definition

The Optimizer requires a variogram defined in terms of range, sill, nugget, and anisotropy. Based on the construction and plotting of empirical variograms for all data rounds, the following variogram parameters were defined:

- The sill was set to 3.0 (three)
- No anisotropy (X,Y) was specified (i.e., anisotropy = 1.0)
- The variogram range and nugget were varied within ranges estimated from the empirical variogram analyses.

The variogram parameter values for each scenario are shown in table 1.

When selecting the variogram parameters, an evaluation was made of the change in sample variance through time. Consistent with a maturing sample monitoring network, and with declining concentrations over time in response to remedial actions and/or MNA, it might be expected that the sample variance would decline over time. Figure 2 suggests that this is the case for the Charlevoix Site. While this is not a rigorous evaluation, it does suggest that if the error objective function incorporates the estimation error and that estimation error is a function of the kriging variance, then it may be appropriate for the error objective function to include a possibly time-varying sample variance. This was not done during the analyses presented here, and it was unclear if this is currently possible.

Section 3
Optimizer Results

Plan Details/Trade-off Curves:

For each scenario, multiple sampling frequency plans were calculated. Results of a spatio-temporal analysis are displayed two ways in the Optimizer: all plans are posted on a trade-off curve (Pareto plot), which shows sampling cost versus concentration error (Figure 3), and the plans are summarized in the Plan Details table (Figure 4). The table lists the sampling frequency for each location, and the estimated concentration error and cost associated with each plan.

Summarizing Results

Because several scenarios were analyzed and each scenario resulted in multiple plans, a method was developed to summarize the results in terms of proposed sampling frequencies, and to provide an indication of the sensitivity of the results to model inputs.

For each scenario, multiple sampling plans are generated as illustrated in Figure 4. The proposed sampling frequencies for each well were converted to numeric values representing the sampling interval in years between samples: semi-annual sampling = 0.5, annual sampling = 1, bi-annual sampling = 2. For the Charlevoix Site, quarterly sampling was not included, and all sampling intervals greater than every two years (including "off") were set to 5. To provide a qualitative indication of the appropriate relative sampling frequency for each location, the average value of these intervals was calculated. To evaluate the variability in selected sampling frequencies, frequency fractions were calculated for each location. This fraction represents the number of occasions on which a specified sampling frequency was determined in any Optimizer plan, divided by the total number of plans; i.e., the fraction (percent) of times the frequency was selected.

In order to combine and compare results using the scenarios listed in Table 1, average interval and frequency fractions were plotted on bar graphs which show results of all scenarios (figures 5-8). Overall frequency fractions were also calculated for each location (Table 2). This fraction represents the number of occasions on which a specified sampling frequency was determined in any plan in any scenario. Note that semi-annual sampling frequency results only from scenarios which used semi-annual time interpolated input data; for all other scenarios, the minimum sampling frequency is annual. Thus Figures 7a and 8a represent only scenarios using semi-annual time interpolated input data, and table 2 shows frequency fractions calculated separately for scenarios which used semi-annual and annual input data. (Note: due to the methods used to combine the summaries, the figures and tables as presented may not correspond in all instances. However, they illustrate the general approach that was undertaken to summarizing a relatively large number of independent Optimizer analyses.)

Section 4
Conclusions and Recommendations

The conclusions and recommendations described here focus on the semi-quantitative results of the Optimizer analyses, and supplement conclusions and recommendations that were described by SSP&A in responses to the "*ESTCP LTM-Optimization Project Questionnaire for Software Application by Third-Party Participants*" which are more focused on practical aspects of the software implementation.

Final sampling frequency was selected based on the highest overall frequency fraction for a specified location, and/or the average intervals. The sensitivity of the results to parameter variations represented by the scenarios was qualitatively evaluated using the bar charts which compare calculated average and frequency fractions for all scenarios (Figures 5-8). Similar average intervals and frequency fractions for all scenarios were considered to indicate a more robust overall result. Conversely, large variations in these values among scenarios were considered to indicate high sensitivity to subjective input parameters, and therefore a less robust result.

The following general conclusions are drawn on the basis of the results presented in Figures 5 through 8 and Table 2:

1. There are several wells for which the results obtained are robust:
    a. Wells 3, 406, 409, and MW101 should be sampled frequently for PCE.
    b. Wells 320, 502D and T2 should be sampled infrequently for PCE.
    c. Wells 212, 320, 406, 501D, 605S, 611 and T2 should be sampled frequently for TCE.
    d. Well 602 should be sampled infrequently for TCE.
2. The results at the following wells seem more sensitive to subjective input parameters:
    a. PCE sampling frequency results appear to depend on variogram range in well 501D. Longer range ("C" scenarios) result in less frequent sampling, and the shorter range ("A" and "B" scenarios) result in more frequent sampling. The shorter-range scenarios most commonly select semi-annual or annual sampling frequency, and the longer-range scenarios most commonly select a sampling interval of more than two years. This is consistent with the interpretation of a longer correlation length, as expressed by a longer-range variogram, indicating that less spatial (and by association temporal) density of sampling is required.
    b. TCE sampling frequency in well 3 appears to be dependant on time interpolation interval used in the input data. For scenarios using semi-annual time interpolation (A2, B2, and C2 scenarios), semi-annual sampling is chosen in most of the plans.

However, for input data using annual time interpolation, the majority of the plans select a sampling interval of more than two years.

c. TCE sampling frequency in wells 603, 606S, 607, MW101, MW104S, and MW604, and PCE sampling frequency in well 206 appear to depend on the choice of selected value for vertically nested wells. In most cases, selection of the average value results in less frequent sampling than selection of maximum value. However, for TCE scenarios in well 607, selection of the maximum value results in less frequent sampling.

3. Some of the apparently robust results described above are not, upon further analysis, consistent with conclusions that would be drawn from independent analyses of temporal concentration trends in wells, and the spatial location of the wells. Examples include the apparently robust results for wells 406 and 611 for TCE. Independent analyses of the sample data, of temporal trends at these wells using other software (PAM,MAROS, SSP&A Trend Analysis Tools), and of the spatial location of these wells would suggest that the TCE in those wells may be a degradation product from the PCE and that the current non-detect levels in these wells is indicative of the declining concentrations of parent material (PCE). As such, frequent monitoring of TCE in these wells would appear not to be justified.

In general, it appears that the Optimizer provides spatio-temporal sampling recommendations that are in concert with conclusions that might be drawn from independent analyses of spatial relationships and temporal trends at wells. However, on some instances the results of the Optimizer analyses (a) appear counterintuitive when compared to independent analyses of the data, and/or (b) are not robust when undertaken using different values for some inherently subjective inputs.

It is recommended that under any occasion that the Optimizer or a similar approach be employed, results not be based upon a single analysis but rather on a group of analyses that enable some evaluation of robustness to be made. Furthermore, use of the Optimizer or a similar approach does not eliminate the necessity for the practitioner to visually review the spatial and temporal patterns of data; to make independent analyses of the data, in particular temporal trends; and to consider the results in the context of Site constraints, potential pathways to receptors, etc. The analyses presented in this report are not comprehensive but provide an example of one possible procedure for assessing the robustness of sample plans calculated using the Optimizer.

Section 5
References

Naval Facilities Engineering Service Center, and GeoTrans Inc. 2007. Environmental Security Technology Certification Program Demonstration Plan for Adaptive Long-Term Monitoring at Environmental Restoration Sites.

S.S. Papadopulos & Associates Inc. 2008. Contaminant Distribution and Groundwater Sampling Analysis for the Charlevoix, Michigan, Site.

Summit Envirosolutions Inc. 2007. Monitoring Tools Sampling Optimizer & Data Tracker Reference Sheet.

U.S. Environmental Protection Agency. 1985. Record of Decision (ROD) (Signed) Remedial Action Selection, Charlevoix Municipal Well, Charlevoix, Charlevoix County, Michigan. U.S. Environmental Protection Agency. Washington, D.C.

U.S. Environmental Protection Agency. 2001. Five-Year Review Report, Charlevoix Municipal Well Superfund Site, Charlevoix, Michigan. U.S. Environmental Protection Agency. Washington, D.C.

U.S. Environmental Protection Agency. 2007. Fact Sheet for Charlevoix Municipal Well, USEPA Region 5, Charlevoix County, Charlevoix, Michigan. U.S. Environmental Protection Agency. EPA ID# MID980794390. Washington, D.C.

FIGURES

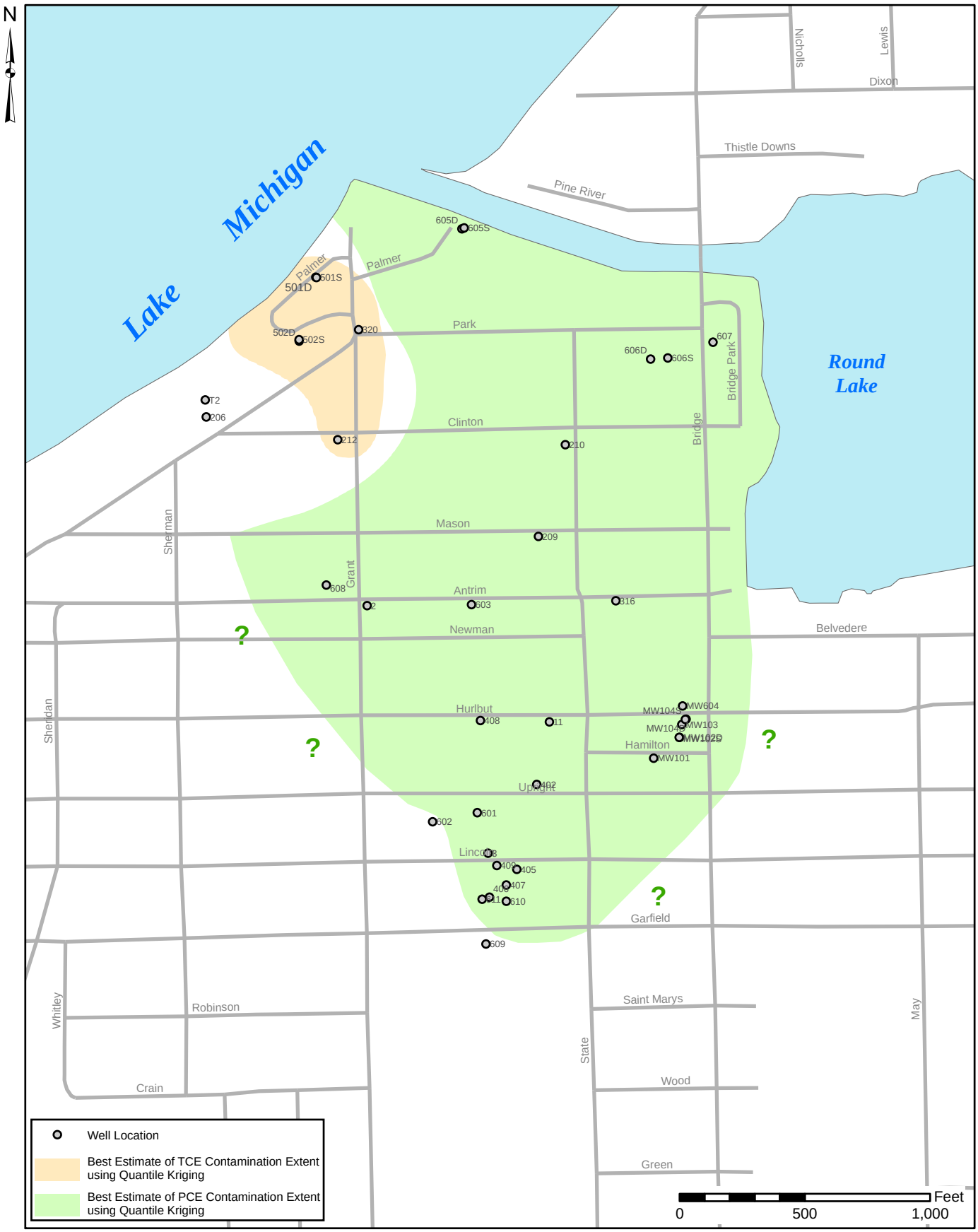


Figure 1 Principal site features with estimated plume extents
(modified from S.S. Papadopulos & Assoc. 2008)

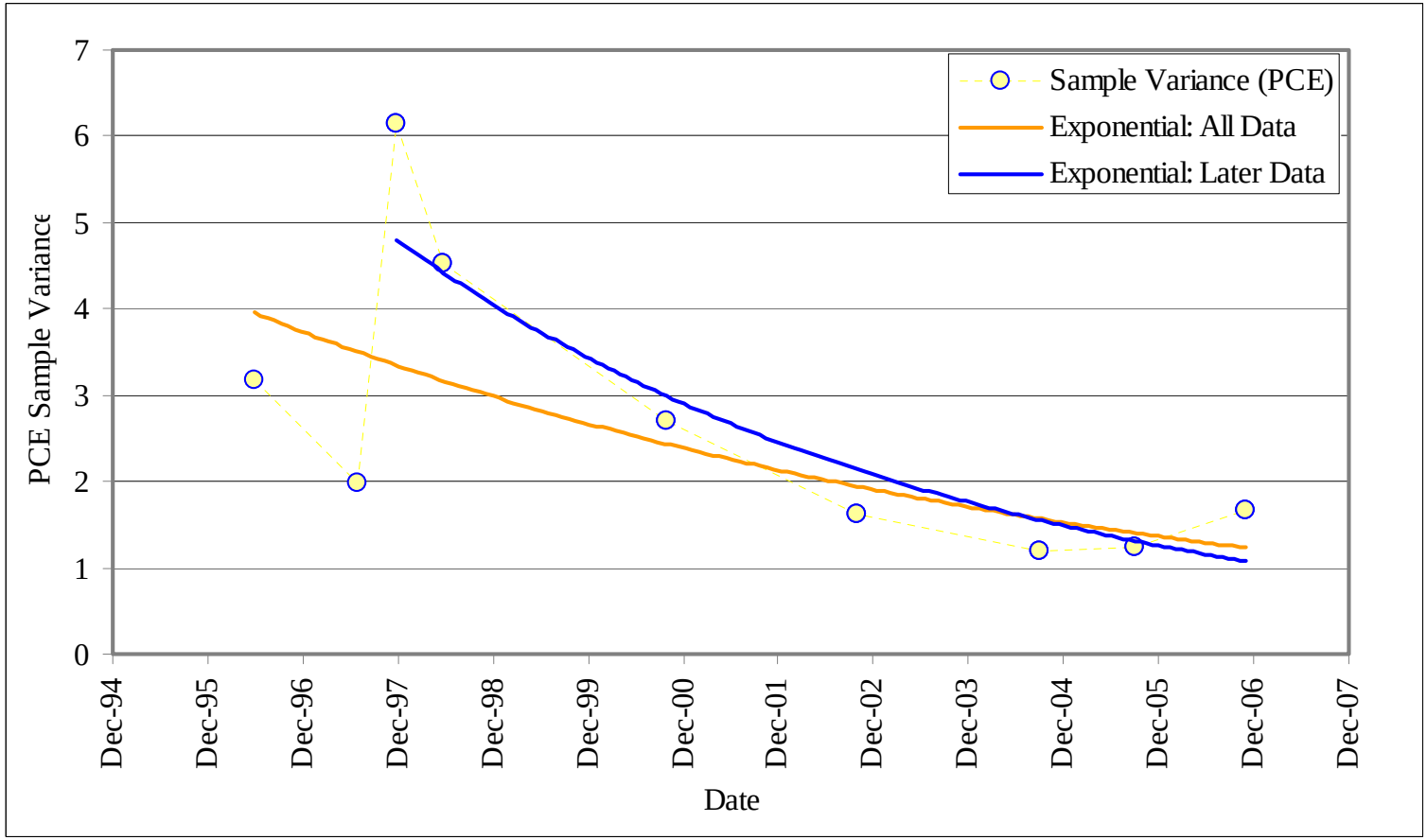
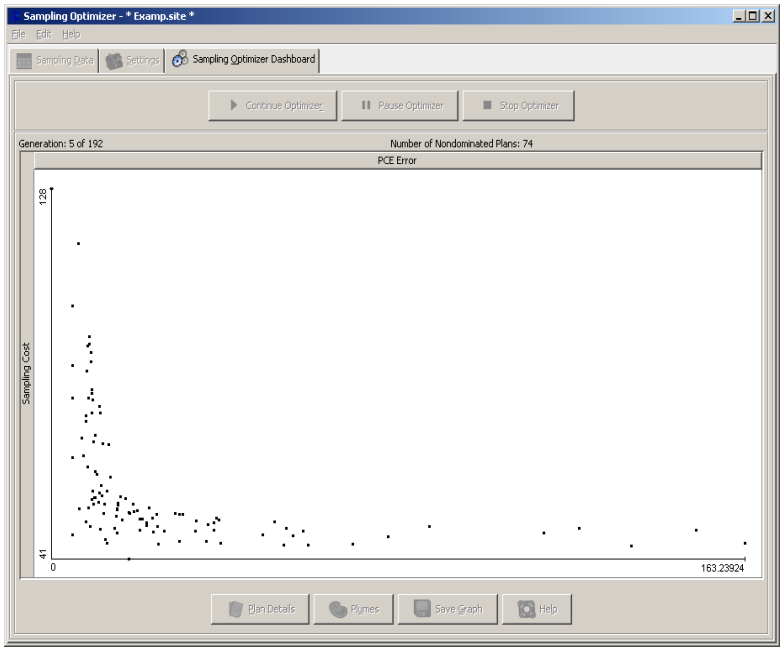
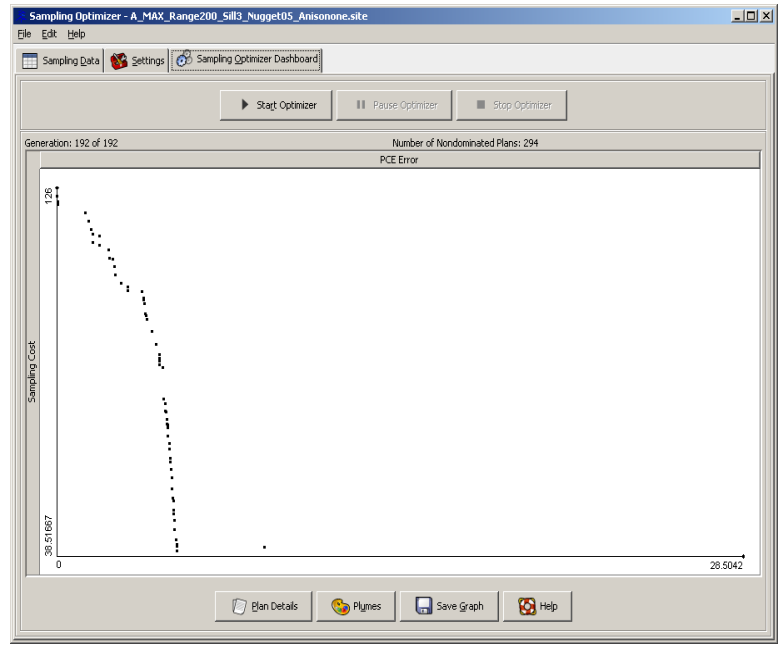


Figure 2: PCE sample variance versus time.



a.



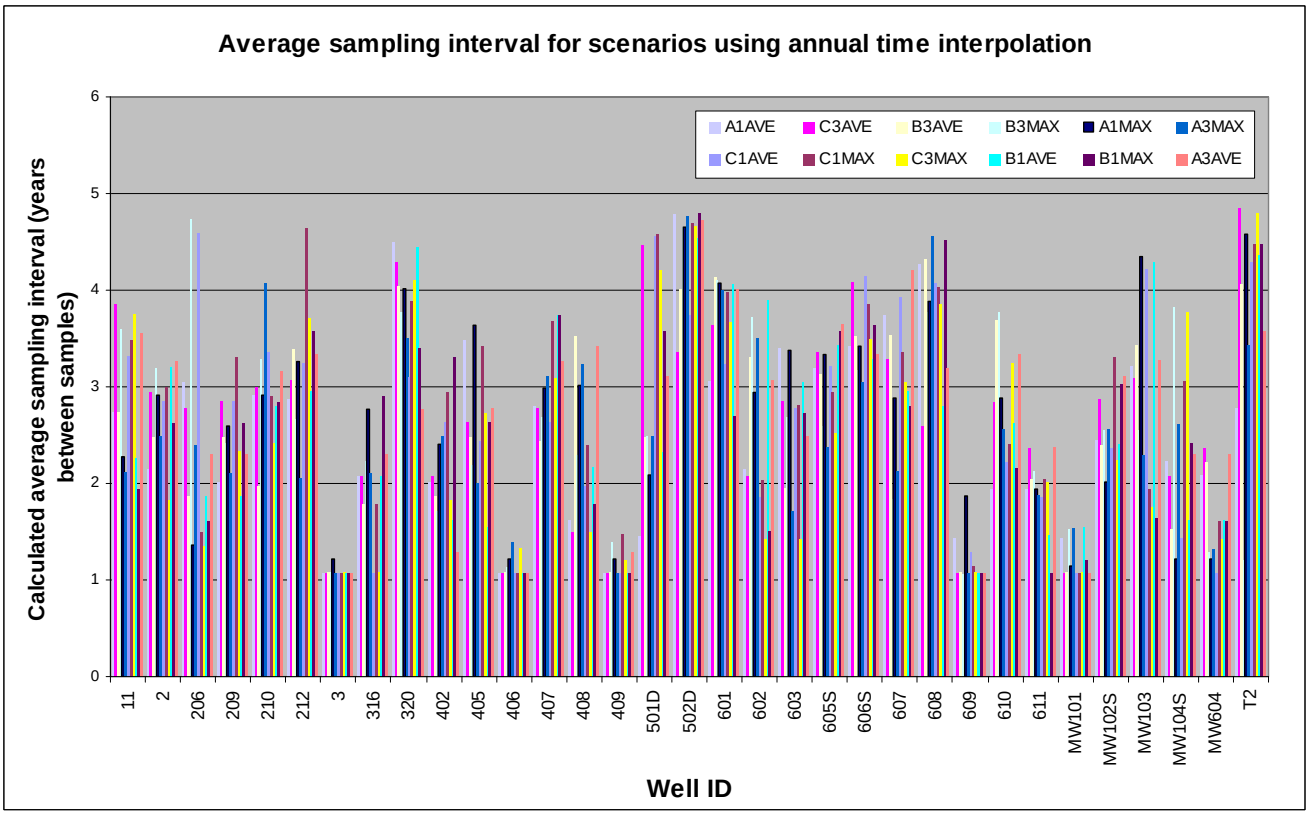
b.

Figure 3: Example trade-off curve showing PCE error versus cost for several sampling plans calculated using spatio-temporal analysis, a) after five generations, and b) at the end of the simulation, 192 generations.

Well/Plan	Well 11	Well 2	Well 206	Well 209	Well 210		PCE Error	Sampling Cost
244	Off	Every 4 years	Semi-Annually	Every 4 years	Every 3 years		4.974211	42.35
251	Every 3 years	Semi-Annually	Semi-Annually	Semi-Annually	Off		4.748738	59.1
25	Semi-Annually	Quarterly	Quarterly	Quarterly	Annually		4.123129	88.78333333
28	Semi-Annually	Quarterly	Quarterly	Quarterly	Quarterly		3.96838	91.78333333
61	Semi-Annually	Quarterly	Quarterly	Quarterly	Annually	...	3.750199	94.58333333
187	Off	Semi-Annually	Semi-Annually	Semi-Annually	Every 3 years		4.872366	48.58333333
37	Quarterly	Semi-Annually	Semi-Annually	Semi-Annually	Every 4 years		4.70782	61.6
259	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly		2.32268	109
2	Off	Every 4 years	Every 2 years	Every 4 years	Every 3 years		4.911441	44.71666667
15	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly		2.435965	105.2
66	Quarterly	Quarterly	Semi-Annually	Quarterly	Quarterly		0.03107	124
39	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly		2.694732	103.2

⋮

Figure 4: Excerpt from an example Plan Details table showing results of spatio-temporal analysis. Each row in the table lists the selected sampling frequency for each location, and the estimated error and cost associated with each plan.



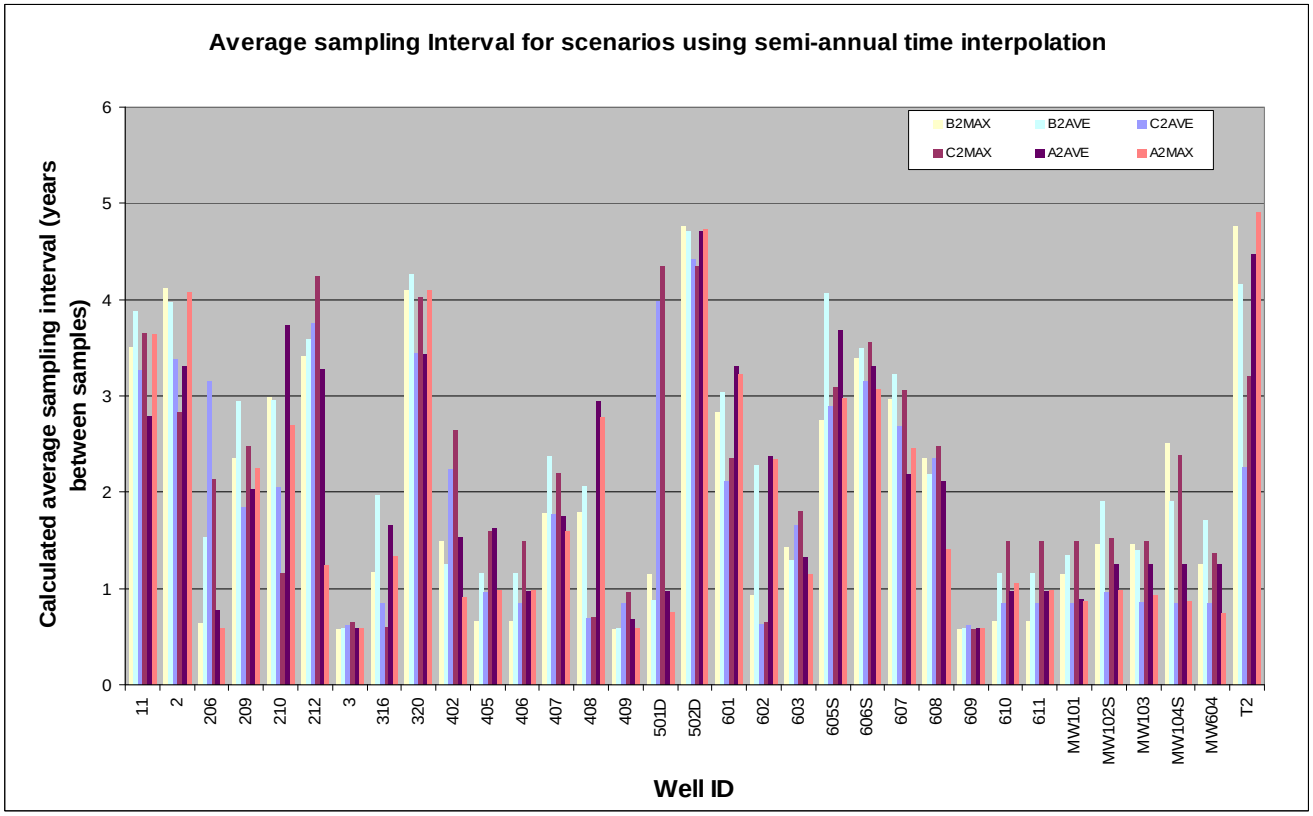
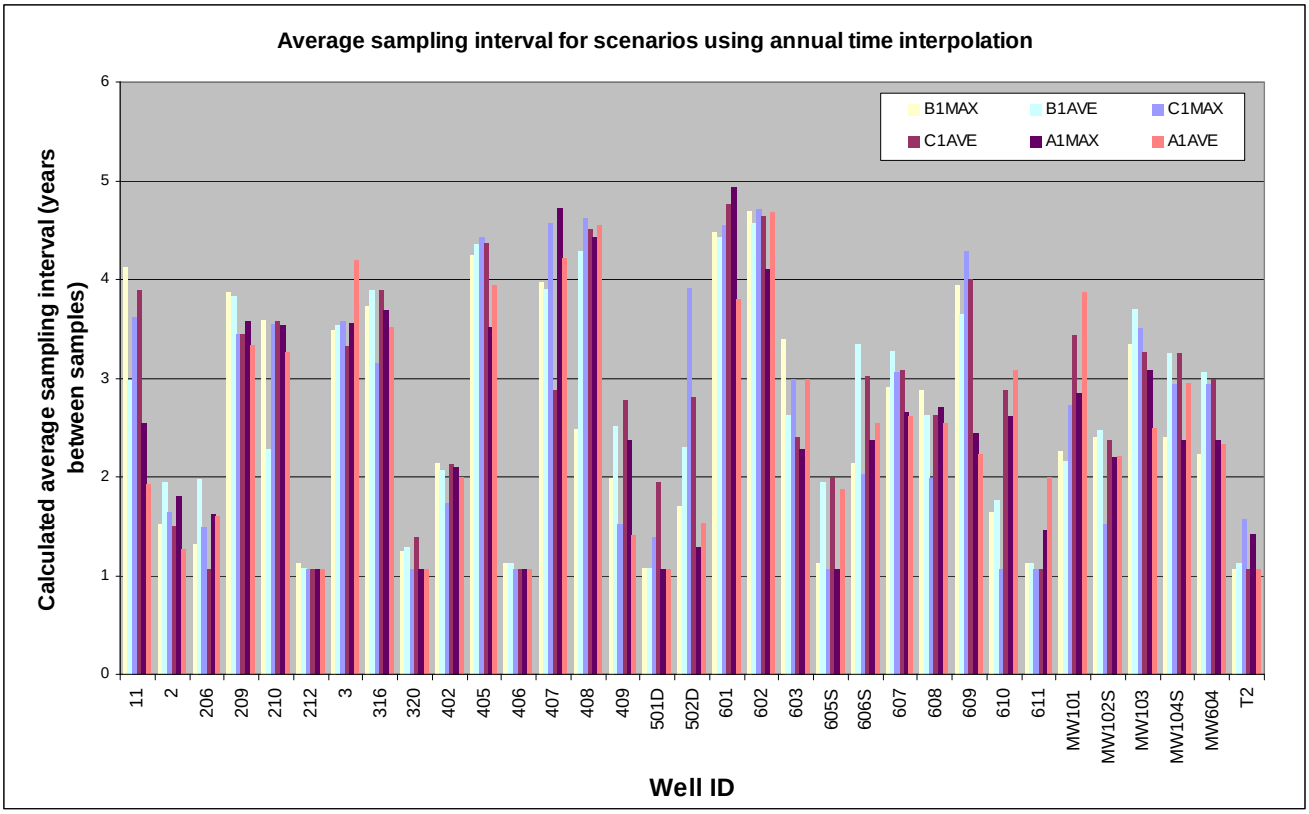


Figure 5: Average sampling interval calculated using PCE data, for scenarios using annual time interpolation (top) and using semi-annual time interpolation (bottom). Scenarios are described in table 1.



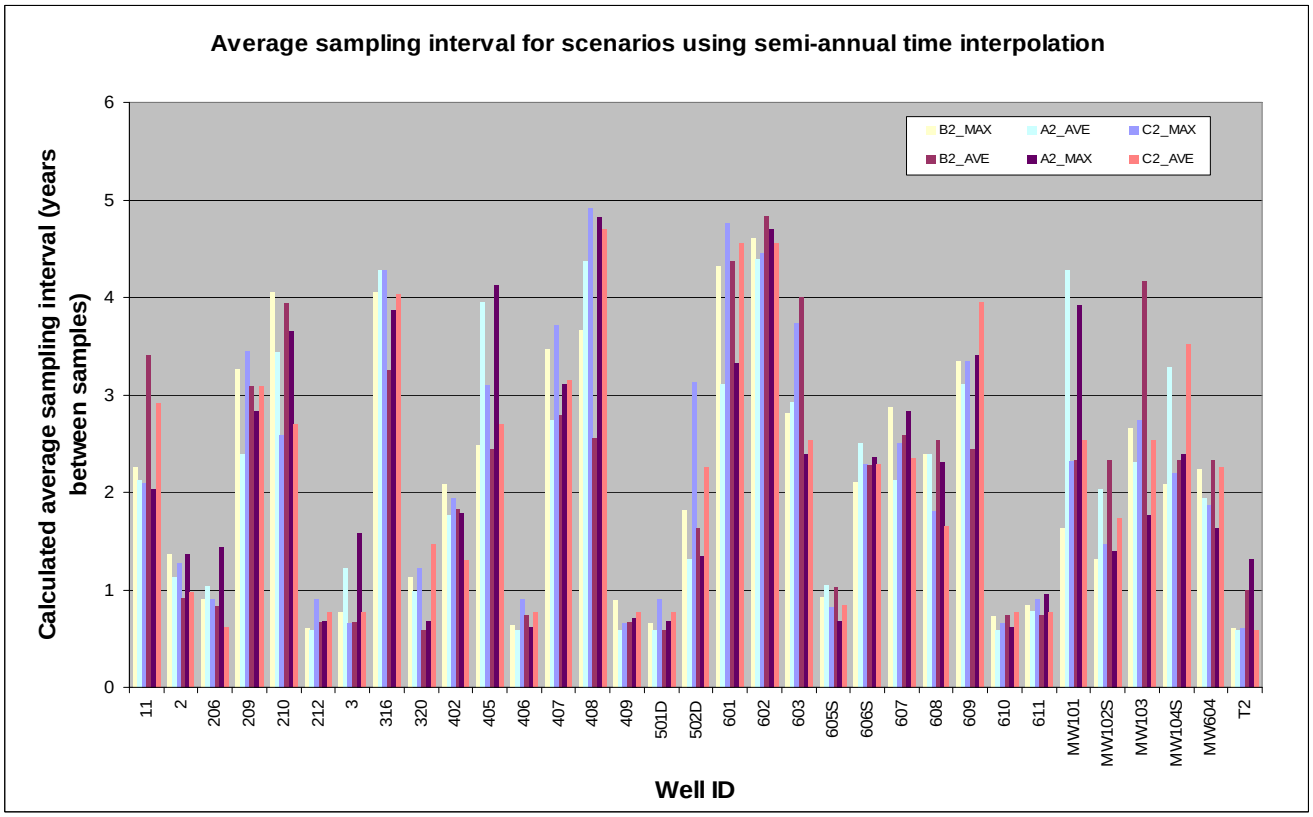


Figure 6: Average sampling interval calculated using TCE data, for scenarios using annual time interpolation (top) and using semi-annual time interpolation (bottom). Scenarios are described in table 1.

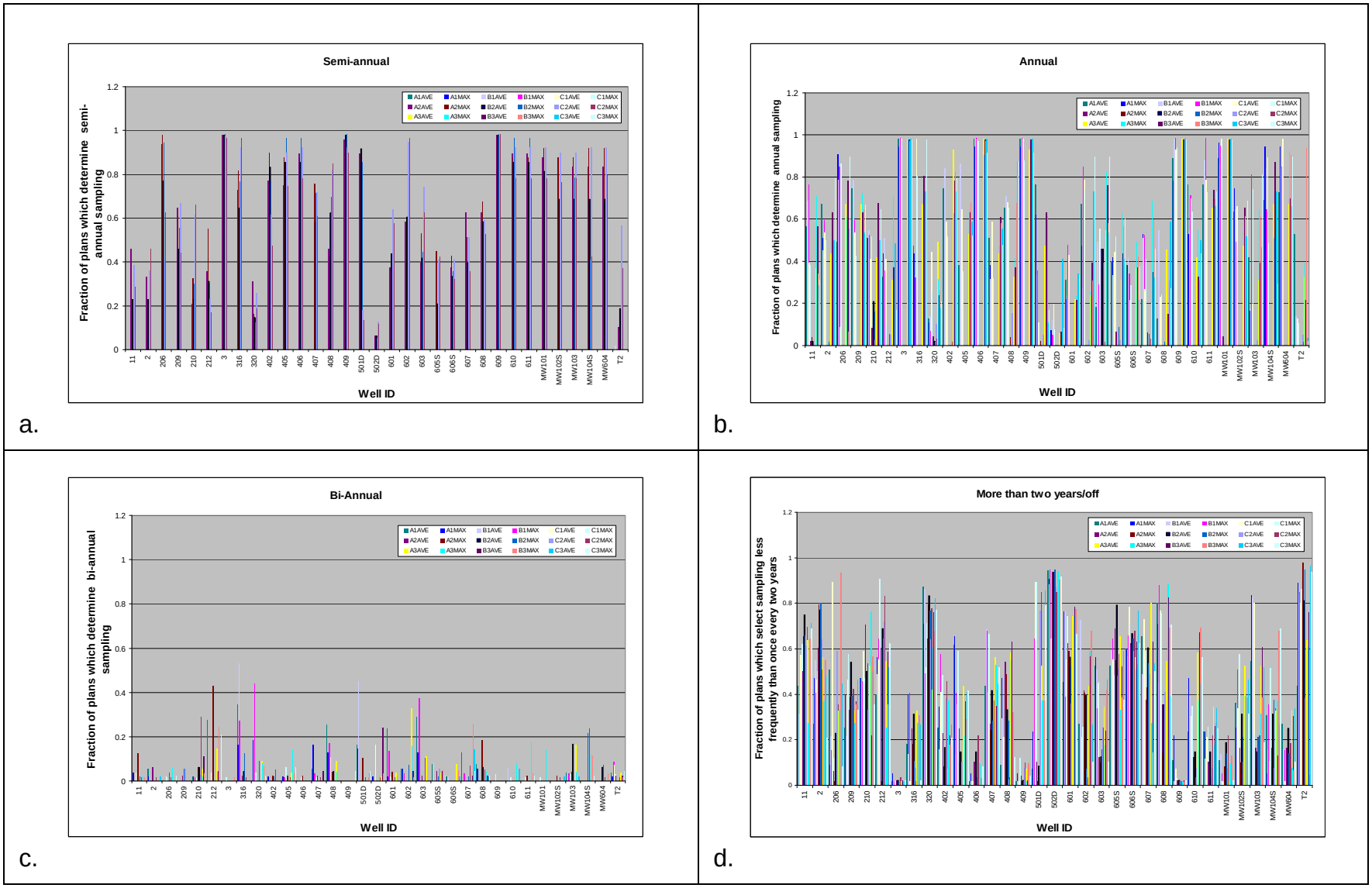


Figure 7: Fractions of calculated sampling frequencies determined from spatio-temporal analyses, using PCE scenarios listed in table 1: a) Semi-Annual, b) Annual, c) Bi-Annual, and d) Period between samples is two years or more.

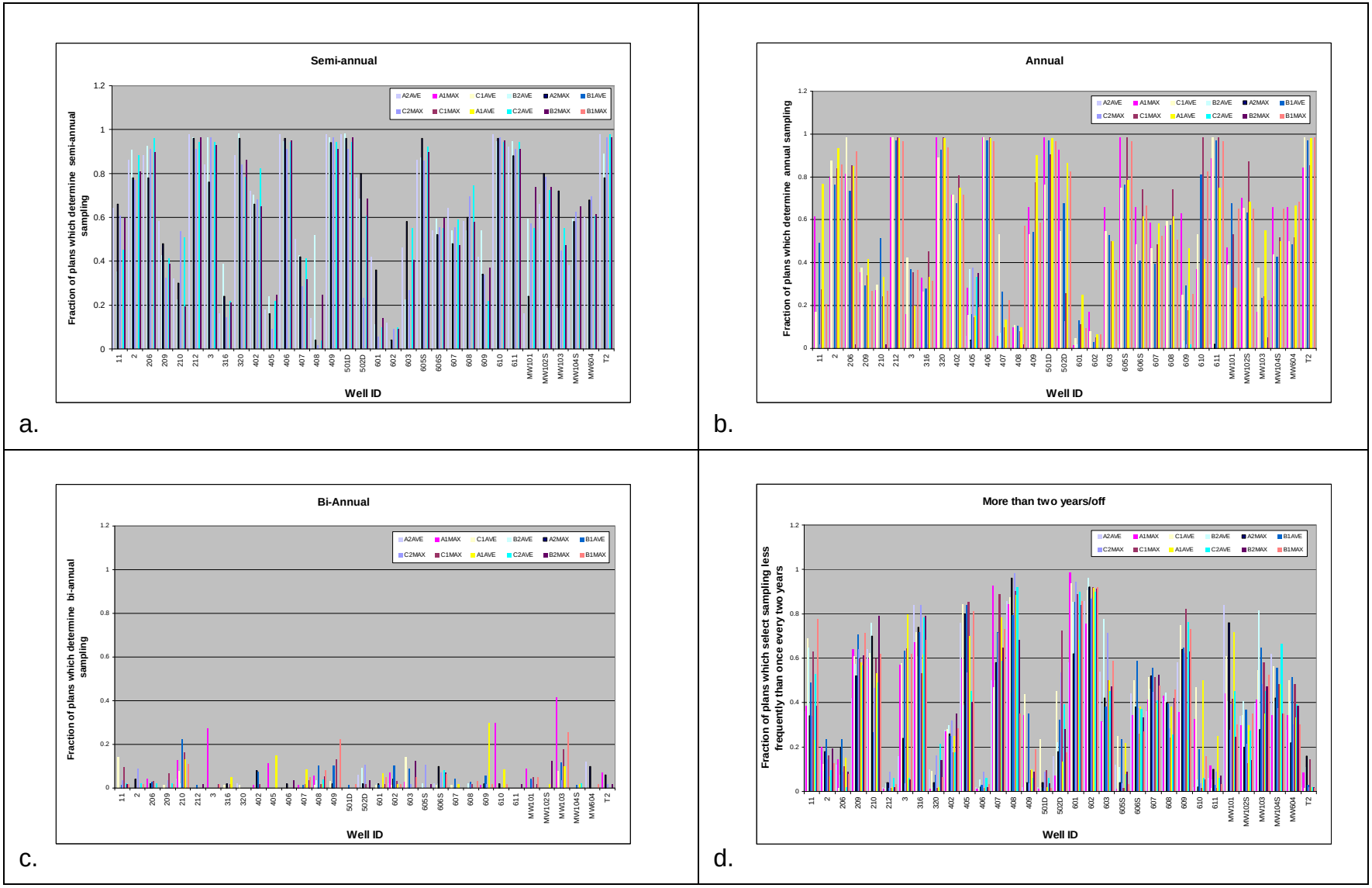


Figure 8: Fractions of calculated sampling frequencies determined from spatio-temporal analyses, using TCE scenarios listed in table 1: a) Semi-Annual, b) Annual, c) Bi-Annual, and d) Period between samples is two years or more.

TABLES

Table 1. Sampling Optimizer scenarios

		Input Data			Variograms			
Scenerio	**Constituent**	**Value at Nested Wells**	**Value Used for Non-Detects**	**Time Interpolation**	**Range**	**Sill**	**Nugget**	**Anisotropy**
A1 Ave PCE	PCE	Average	1/10 assumed detection limit	Annual	200	3	0.5	none
A2 Ave PCE	PCE	Average	1/10 assumed detection limit	Semi-Annual	200	3	0.5	none
A3 Ave PCE	PCE	Average	interpolated	Annual	200	3	0.5	none
B1 Ave PCE	PCE	Average	1/10 assumed detection limit	Annual	200	3	none	none
B2 Ave PCE	PCE	Average	1/10 assumed detection limit	Semi-Annual	200	3	none	none
B3 Ave PCE	PCE	Average	interpolated	Annual	200	3	none	none
C1 Ave PCE	PCE	Average	1/10 assumed detection limit	Annual	400	3	none	none
C2 Ave PCE	PCE	Average	1/10 assumed detection limit	Semi-Annual	400	3	none	none
C3 Ave PCE	PCE	Average	interpolated	Annual	400	3	none	none
A1 Max PCE	PCE	Maximum	1/10 assumed detection limit	Annual	200	3	0.5	none
A2 Max PCE	PCE	Maximum	1/10 assumed detection limit	Semi-Annual	200	3	0.5	none
A3 Max PCE	PCE	Maximum	interpolated	Annual	200	3	0.5	none
B1 Max PCE	PCE	Maximum	1/10 assumed detection limit	Annual	200	3	none	none
B2 Max PCE	PCE	Maximum	1/10 assumed detection limit	Semi-Annual	200	3	none	none
B3 Max PCE	PCE	Maximum	interpolated	Annual	200	3	none	none
C1 Max PCE	PCE	Maximum	1/10 assumed detection limit	Annual	400	3	none	none
C2 Max PCE	PCE	Maximum	1/10 assumed detection limit	Semi-Annual	400	3	none	none
C3 Max PCE	PCE	Maximum	interpolated	Annual	400	3	none	none
A1 Ave TCE	TCE	Average	1/10 assumed detection limit	Annual	200	3	0.5	none
A2 Ave TCE	TCE	Average	1/10 assumed detection limit	Semi-Annual	200	3	0.5	none
B1 Ave TCE	TCE	Average	1/10 assumed detection limit	Annual	200	3	none	none
B2 Ave TCE	TCE	Average	1/10 assumed detection limit	Semi-Annual	200	3	none	none
C1 Ave TCE	TCE	Average	1/10 assumed detection limit	Annual	400	3	none	none
C2 Ave TCE	TCE	Average	1/10 assumed detection limit	Semi-Annual	400	3	none	none
A1 Max TCE	TCE	Maximum	1/10 assumed detection limit	Annual	200	3	0.5	none
A2 Max TCE	TCE	Maximum	1/10 assumed detection limit	Semi-Annual	200	3	0.5	none
B1 Max TCE	TCE	Maximum	1/10 assumed detection limit	Annual	200	3	none	none
B2 Max TCE	TCE	Maximum	1/10 assumed detection limit	Semi-Annual	200	3	none	none
C1 Max TCE	TCE	Maximum	1/10 assumed detection limit	Annual	400	3	none	none
C2 Max TCE	TCE	Maximum	1/10 assumed detection limit	Semi-Annual	400	3	none	none

Table 2: Calculated overall frequency fractions.

Well Names	PCE Scenerios							TCE Scenerios						
	Time Interpolation: Annual			Time Interpolation: Semi-Annual				Time Interpolation: Annual			Time Interpolation: Semi-Annual			
	Annual	Every 2 years	> Every 2 years	Semi-Annual	Annual	Every 2 years	> Every 2 years	Annual	Every 2 years	> Every 2 years	Semi-Annual	Annual	Every 2 years	> Every 2 years
2	0.57	0.01	0.42	0.29	0.01	0.01	0.69	0.84	0.00	0.15	0.83	0.00	0.02	0.14
3	0.98	0.00	0.02	0.98	0.00	0.00	0.02	0.31	0.05	0.64	0.90	0.00	0.00	0.10
11	0.48	0.01	0.51	0.31	0.02	0.03	0.65	0.42	0.04	0.53	0.55	0.00	0.01	0.43
206	0.62	0.01	0.37	0.78	0.00	0.01	0.21	0.86	0.02	0.12	0.89	0.00	0.01	0.10
209	0.62	0.00	0.38	0.56	0.02	0.02	0.39	0.34	0.02	0.64	0.43	0.00	0.01	0.56
210	0.48	0.03	0.49	0.39	0.11	0.07	0.43	0.32	0.14	0.54	0.35	0.00	0.01	0.64
212	0.37	0.09	0.54	0.32	0.02	0.08	0.58	0.98	0.00	0.02	0.95	0.00	0.00	0.04
316	0.69	0.07	0.24	0.81	0.00	0.03	0.16	0.33	0.01	0.66	0.23	0.00	0.00	0.77
320	0.25	0.09	0.67	0.20	0.04	0.01	0.74	0.95	0.00	0.05	0.88	0.00	0.00	0.11
402	0.69	0.00	0.31	0.73	0.00	0.01	0.26	0.73	0.02	0.25	0.71	0.00	0.01	0.28
405	0.53	0.03	0.43	0.85	0.00	0.01	0.15	0.18	0.04	0.77	0.19	0.25	0.00	0.56
406	0.96	0.00	0.04	0.88	0.00	0.00	0.11	0.98	0.00	0.02	0.95	0.00	0.01	0.04
407	0.47	0.03	0.50	0.68	0.00	0.01	0.31	0.22	0.03	0.75	0.39	0.01	0.01	0.59
408	0.60	0.07	0.33	0.65	0.06	0.01	0.29	0.18	0.05	0.77	0.17	0.00	0.02	0.81
409	0.95	0.00	0.05	0.95	0.00	0.00	0.05	0.67	0.08	0.25	0.95	0.00	0.00	0.05
601	0.29	0.04	0.66	0.48	0.00	0.01	0.51	0.11	0.02	0.87	0.20	0.00	0.01	0.79
602	0.58	0.06	0.36	0.76	0.00	0.01	0.23	0.08	0.04	0.88	0.08	0.01	0.01	0.91
603	0.55	0.08	0.37	0.53	0.28	0.02	0.17	0.52	0.05	0.43	0.41	0.00	0.02	0.56
607	0.41	0.06	0.53	0.47	0.02	0.01	0.50	0.51	0.01	0.48	0.55	0.00	0.00	0.45
608	0.26	0.02	0.72	0.59	0.00	0.05	0.35	0.60	0.01	0.39	0.62	0.00	0.01	0.37
609	0.95	0.00	0.05	0.98	0.00	0.00	0.02	0.35	0.06	0.59	0.37	0.01	0.01	0.61
610	0.53	0.01	0.45	0.88	0.00	0.00	0.12	0.66	0.07	0.28	0.96	0.00	0.00	0.04
611	0.74	0.02	0.24	0.88	0.00	0.00	0.11	0.92	0.00	0.08	0.92	0.01	0.00	0.07
501D	0.42	0.04	0.55	0.64	0.01	0.02	0.33	0.93	0.01	0.07	0.96	0.00	0.00	0.04
501S														
502D	0.11	0.04	0.85	0.08	0.00	0.00	0.91	0.68	0.00	0.31	0.63	0.02	0.05	0.30
502S														
605D														
605S	0.46	0.02	0.51	0.35	0.03	0.02	0.60	0.87	0.00	0.13	0.89	0.00	0.02	0.08
606D														
606S	0.35	0.02	0.64	0.37	0.00	0.00	0.63	0.60	0.01	0.39	0.56	0.00	0.06	0.38
MW101	0.94	0.01	0.05	0.86	0.01	0.00	0.13	0.50	0.04	0.46	0.47	0.00	0.01	0.52
MW102D														
MW102S	0.59	0.00	0.41	0.81	0.00	0.01	0.19	0.70	0.00	0.30	0.72	0.00	0.02	0.26
MW103	0.51	0.03	0.46	0.81	0.00	0.03	0.15	0.30	0.19	0.51	0.50	0.01	0.01	0.48
MW104D														
MW104S	0.63	0.02	0.35	0.70	0.00	0.08	0.23	0.53	0.01	0.46	0.52	0.00	0.00	0.47
MW604	0.83	0.00	0.17	0.82	0.00	0.03	0.14	0.58	0.00	0.41	0.63	0.00	0.04	0.33
T2	0.26	0.03	0.71	0.22	0.01	0.01	0.77	0.94	0.01	0.05	0.93	0.00	0.02	0.06

Appendix A

Selected Results of PAM Statistical Analyses

Charlevoix Municipal Well									
Analyte Name	Well ID	Units*	Trend Test (80% Confidence)		Compare-to-Standard Test (95% Confidence)			Compare-to-Baseline Test (95% Confidence)	
			Result	Slope Estimate (Units*/Yr)	Result	UCL (Units*)	Standard (Units*)	Result	UPL (Units*)
Benzene	11	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	2	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	206	ug/l	No Trend	0#	**Exceedance**	**15.9366**	**5**	**Worse**	**19**
Benzene	209	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	210	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	212	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	3	ug/l	No Trend	0#	**Exceedance**	**18.3317**	**5**	No Change	516.7604
Benzene	316	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	320	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	402	ug/l	No Trend	0#	**Compliance**	**1.0979**	**5**	No Change	35
Benzene	405	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	406	ug/l	**Downward**	**-0.30638#**	**Compliance**	**0.05**	**5**	No Change	357.4298
Benzene	407	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	408	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	409	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	250
Benzene	501D	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	501S	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05

Charlevoix Municipal Well									
Analyte Name	Well ID	Units*	Trend Test (80% Confidence)		Compare-to-Standard Test (95% Confidence)			Compare-to-Baseline Test (95% Confidence)	
			Result	Slope Estimate (Units*/Yr)	Result	UCL (Units*)	Standard (Units*)	Result	UPL (Units*)
Benzene	502D	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	502S	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	601	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	602	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	603	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	605D	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	605S	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	606D	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	606S	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	607	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	608	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	609	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Benzene	610	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	380
Benzene	611	ug/l	**Downward**	**-0.37982#**	**Exceedance**	**5.3833**	**5**	No Change	282.8945
Benzene	T2	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
Ethylbenzene	11	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	2	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05

Charlevoix Municipal Well									
Analyte Name	Well ID	Units*	Trend Test (80% Confidence)		Compare-to-Standard Test (95% Confidence)			Compare-to-Baseline Test (95% Confidence)	
			Result	Slope Estimate (Units*/Yr)	Result	UCL (Units*)	Standard (Units*)	Result	UPL (Units*)
Ethylbenzene	206	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	209	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	210	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	212	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	3	ug/l	No Trend	0#	**Compliance**	**187.4146**	**700**	No Change	161.3174
Ethylbenzene	316	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	320	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	402	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	110
Ethylbenzene	405	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	406	ug/l	**Downward**	**-0.28523#**	**Compliance**	**0.84642**	**700**	No Change	229.8773
Ethylbenzene	407	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	408	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	409	ug/l	No Trend	0#	**Compliance**	**1.0141**	**700**	No Change	72
Ethylbenzene	501D	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	501S	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	502D	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	502S	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05

Charlevoix Municipal Well									
Analyte Name	Well ID	Units*	Trend Test (80% Confidence)		Compare-to-Standard Test (95% Confidence)			Compare-to-Baseline Test (95% Confidence)	
			Result	Slope Estimate (Units*/Yr)	Result	UCL (Units*)	Standard (Units*)	Result	UPL (Units*)
Ethylbenzene	601	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	1.9
Ethylbenzene	602	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	603	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	605D	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	605S	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	606D	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	606S	ug/l	No Trend	0#	**Compliance**	**0.84642**	**700**	No Change	1
Ethylbenzene	607	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	608	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	609	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
Ethylbenzene	610	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	88
Ethylbenzene	611	ug/l	No Trend	-0.076356#	**Compliance**	**46.6511**	**700**	No Change	292.11
Ethylbenzene	T2	ug/l	No Trend	0#	**Compliance**	**0.05**	**700**	No Change	0.05
PCE	11	ug/l	**Downward**	**-0.1441#**	**Exceedance**	**40.9637**	**5**	No Change	552.538
PCE	2	ug/l	No Trend	0.01748#	**Exceedance**	**61.3477**	**5**	No Change	123.7086
PCE	206	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
PCE	209	ug/l	**Downward**	**-0.13747#**	**Exceedance**	**126.0499**	**5**	No Change	640.0645

Charlevoix Municipal Well									
Analyte Name	Well ID	Units*	Trend Test (80% Confidence)		Compare-to-Standard Test (95% Confidence)			Compare-to-Baseline Test (95% Confidence)	
			Result	Slope Estimate (Units*/Yr)	Result	UCL (Units*)	Standard (Units*)	Result	UPL (Units*)
PCE	210	ug/l	**Downward**	**-0.13822#**	**Exceedance**	**24.1723**	**5**	No Change	140.4287
PCE	212	ug/l	No Trend	-0.09102#	**Compliance**	**0.05**	**5**	No Change	2.3
PCE	3	ug/l	**Downward**	**-0.11864#**	**Exceedance**	**12.2499**	**5**	No Change	44.1684
PCE	316	ug/l	No Trend	0.057163#	**Exceedance**	**20.6988**	**5**	No Change	30.3974
PCE	320	ug/l	**Downward**	**-0.021324#**	**Compliance**	**1.749**	**5**	No Change	3.4687
PCE	402	ug/l	No Trend	-0.19571#	**Exceedance**	**67.551**	**5**	No Change	906.5078
PCE	405	ug/l	No Trend	-0.038441#	**Exceedance**	**60.3389**	**5**	No Change	459.7808
PCE	406	ug/l	**Downward**	**-0.39588#**	**Compliance**	**3.1427**	**5**	No Change	1271.496
PCE	407	ug/l	**Downward**	**-0.29437#**	**Exceedance**	**163.1086**	**5**	No Change	7456.7277
PCE	408	ug/l	No Trend	-0.097364#	**Exceedance**	**10.2982**	**5**	No Change	31.9885
PCE	409	ug/l	No Trend	0#	**Exceedance**	**90.6096**	**5**	No Change	335.2541
PCE	501D	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	1.3
PCE	501S	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	1.3
PCE	502D	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
PCE	502S	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
PCE	601	ug/l	**Downward**	**-0.15369#**	**Exceedance**	**30.0905**	**5**	No Change	246.533
PCE	602	ug/l	**Downward**	**-0.15215#**	**Compliance**	**3.4203**	**5**	No Change	32.9857

Charlevoix Municipal Well									
Analyte Name	Well ID	Units*	Trend Test (80% Confidence)		Compare-to-Standard Test (95% Confidence)			Compare-to-Baseline Test (95% Confidence)	
			Result	Slope Estimate (Units*/Yr)	Result	UCL (Units*)	Standard (Units*)	Result	UPL (Units*)
PCE	603	ug/l	No Trend	-0.046334#	**Exceedance**	**12.807**	**5**	No Change	167.7889
PCE	605D	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
PCE	605S	ug/l	**Downward**	**-0.079324#**	**Exceedance**	**34.2265**	**5**	No Change	74.949
PCE	606D	ug/l	No Trend	0.078351#	**Compliance**	**0.84642**	**5**	No Change	1
PCE	606S	ug/l	No Trend	-0.0023532#	**Exceedance**	**34.6142**	**5**	No Change	56.079
PCE	607	ug/l	No Trend	-0.061886#	**Exceedance**	**5.4311**	**5**	No Change	19.8334
PCE	608	ug/l	**Upward**	**0.025993#**	**Exceedance**	**18.3591**	**5**	No Change	19.824
PCE	609	ug/l	No Trend	0#	**Compliance**	**1.3611**	**5**	No Change	7.8
PCE	610	ug/l	**Upward**	**0.12462#**	**Exceedance**	**17.8941**	**5**	No Change	63.086
PCE	611	ug/l	**Upward**	**0.4275#**	**Exceedance**	**62.2475**	**5**	No Change	115.6994
PCE	T2	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
TCE	11	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
TCE	2	ug/l	**Upward**	**0.062515#**	**Compliance**	**4.0787**	**5**	No Change	6.5137
TCE	206	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	2.4
TCE	209	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	1.1
TCE	210	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	1
TCE	212	ug/l	**Downward**	**-0.16215#**	**Exceedance**	**17.0372**	**5**	No Change	88.6134

Charlevoix Municipal Well									
Analyte Name	Well ID	Units*	Trend Test (80% Confidence)		Compare-to-Standard Test (95% Confidence)			Compare-to-Baseline Test (95% Confidence)	
			Result	Slope Estimate (Units*/Yr)	Result	UCL (Units*)	Standard (Units*)	Result	UPL (Units*)
TCE	3	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
TCE	316	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
TCE	320	ug/l	**Downward**	**-0.16455#**	**Exceedance**	**10.8739**	**5**	No Change	63.5349
TCE	402	ug/l	No Trend	-0.023775#	**Compliance**	**2.1178**	**5**	No Change	3.7598
TCE	405	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	1.3
TCE	406	ug/l	**Downward**	**-0.36174#**	**Compliance**	**0.05**	**5**	No Change	243.2842
TCE	407	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	1.5
TCE	408	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
TCE	409	ug/l	No Trend	-0.079298#	**Compliance**	**0.93026**	**5**	No Change	39.3571
TCE	501D	ug/l	**Downward**	**-0.1773#**	**Exceedance**	**15.4478**	**5**	No Change	76.6005
TCE	501S	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	1.4
TCE	502D	ug/l	**Downward**	**-0.18207#**	**Exceedance**	**47.7221**	**5**	No Change	67.4648
TCE	502S	ug/l	**Downward**	**-0.22987#**	**Exceedance**	**12.6149**	**5**	No Change	18.6901
TCE	601	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
TCE	602	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
TCE	603	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
TCE	605D	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	1.3

Charlevoix Municipal Well									
Analyte Name	Well ID	Units*	Trend Test (80% Confidence)		Compare-to-Standard Test (95% Confidence)			Compare-to-Baseline Test (95% Confidence)	
			Result	Slope Estimate (Units*/Yr)	Result	UCL (Units*)	Standard (Units*)	Result	UPL (Units*)
TCE	605S	ug/l	No Trend	-0.041698#	**Compliance**	**3.6714**	**5**	No Change	13.7937
TCE	606D	ug/l	No Trend	-0.082948#	**Compliance**	**2.1259**	**5**	No Change	3.3477
TCE	606S	ug/l	No Trend	0#	**Compliance**	**0.84642**	**5**	No Change	1
TCE	607	ug/l	No Trend	0#	**Compliance**	**0.84642**	**5**	No Change	1
TCE	608	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
TCE	609	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
TCE	610	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	0.05
TCE	611	ug/l	No Trend	0#	**Compliance**	**0.05**	**5**	No Change	13.316
TCE	T2	ug/l	**Downward**	**-0.14036#**	**Compliance**	**1.8933**	**5**	No Change	12.05
Toluene	11	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	2	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	206	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	209	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	210	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	212	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	3	ug/l	No Trend	0#	**Compliance**	**21.4461**	**1000**	No Change	19
Toluene	316	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05

Charlevoix Municipal Well									
Analyte Name	Well ID	Units*	Trend Test (80% Confidence)		Compare-to-Standard Test (95% Confidence)			Compare-to-Baseline Test (95% Confidence)	
			Result	Slope Estimate (Units*/Yr)	Result	UCL (Units*)	Standard (Units*)	Result	UPL (Units*)
Toluene	320	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	402	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	6.1
Toluene	405	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	406	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	15.2626
Toluene	407	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	408	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	409	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	25
Toluene	501D	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	501S	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	502D	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	502S	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	601	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	602	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	603	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	605D	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	605S	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	606D	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05

Charlevoix Municipal Well									
Analyte Name	Well ID	Units*	Trend Test (80% Confidence)		Compare-to-Standard Test (95% Confidence)			Compare-to-Baseline Test (95% Confidence)	
			Result	Slope Estimate (Units*/Yr)	Result	UCL (Units*)	Standard (Units*)	Result	UPL (Units*)
Toluene	606S	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	607	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	608	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	609	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Toluene	610	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	160
Toluene	611	ug/l	**Downward**	**-0.22178#**	**Compliance**	**3.5581**	**1000**	No Change	82.0373
Toluene	T2	ug/l	No Trend	0#	**Compliance**	**0.05**	**1000**	No Change	0.05
Xylenes	11	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	2	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	206	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	209	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	210	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	212	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	3	ug/l	No Trend	0#	**Compliance**	**80.0179**	**10000**	No Change	133
Xylenes	316	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	320	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	402	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	71

Charlevoix Municipal Well									
Analyte Name	Well ID	Units*	Trend Test (80% Confidence)		Compare-to-Standard Test (95% Confidence)			Compare-to-Baseline Test (95% Confidence)	
			Result	Slope Estimate (Units*/Yr)	Result	UCL (Units*)	Standard (Units*)	Result	UPL (Units*)
Xylenes	405	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	406	ug/l	No Trend	-0.24072#	**Compliance**	**0.05**	**10000**	No Change	386.2949
Xylenes	407	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	408	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	409	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	61.3
Xylenes	501D	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	501S	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	502D	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	502S	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	601	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	602	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	603	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	605D	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	605S	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	606D	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	606S	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	607	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05

Charlevoix Municipal Well									
Analyte Name	Well ID	Units*	Trend Test (80% Confidence)		Compare-to-Standard Test (95% Confidence)			Compare-to-Baseline Test (95% Confidence)	
			Result	Slope Estimate (Units*/Yr)	Result	UCL (Units*)	Standard (Units*)	Result	UPL (Units*)
Xylenes	608	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	609	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05
Xylenes	610	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	181
Xylenes	611	ug/l	**Downward**	**-0.32783#**	**Compliance**	**7.2446**	**10000**	No Change	367.8402
Xylenes	T2	ug/l	No Trend	0#	**Compliance**	**0.05**	**10000**	No Change	0.05

NOTES:

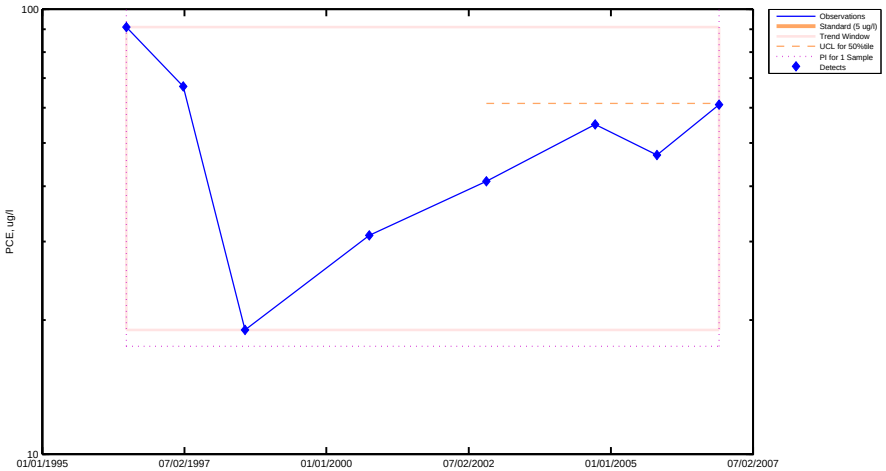
means trend coefficient of log-transformed data. Log(2) times its reciprocal is doubling(+)/halving(-) time.

Statistical Note: ND surrogate = 0.5 X Median of Nondetects' PQLs.

These results obtained on 06/05/2007.

2
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



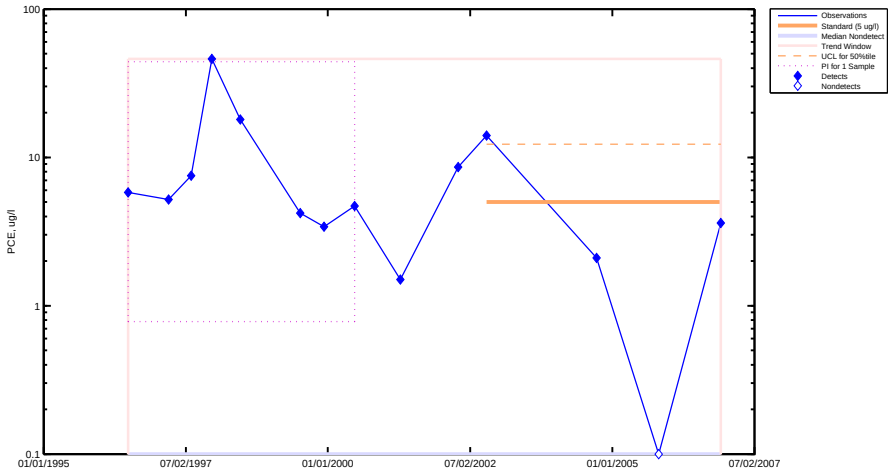
Standard Test (95%): Exceedance <UCL = 6.13e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.24e+002/1.74e+001 ug/l>
Trend Test (80%): No Trend <Slope = 1.75e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

3
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



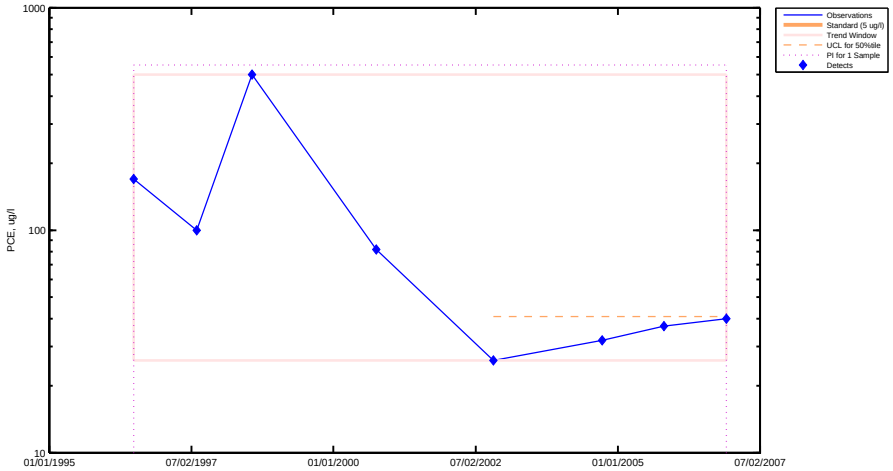
Standard Test (95%): Exceedance <UCL = 1.22e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 4.42e+001/7.79e−001 ug/l>
Trend Test (80%): Downward <Slope = −1.19e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

11
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



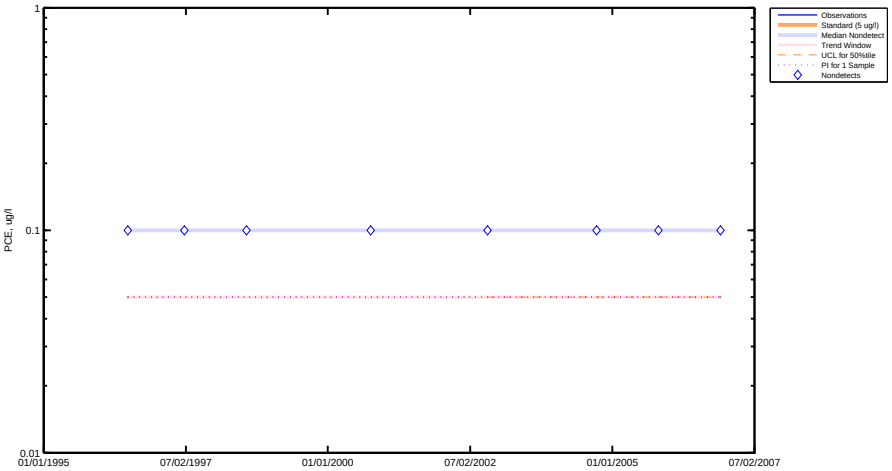
Standard Test (95%): Exceedance <UCL = 4.10e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.53e+002/9.15e+000 ug/l>
Trend Test (80%): Downward <Slope = −1.44e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

206
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend

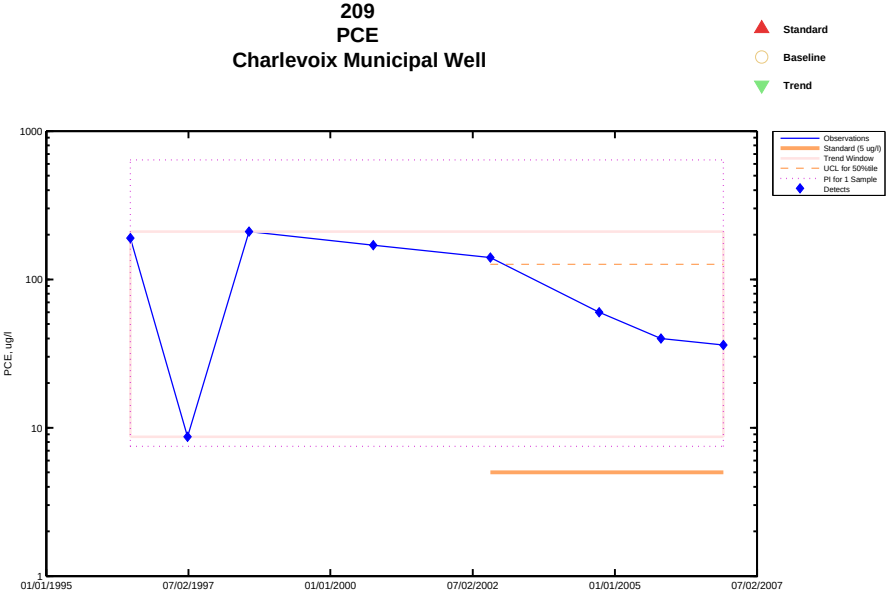


Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

209
PCE
Charlevoix Municipal Well

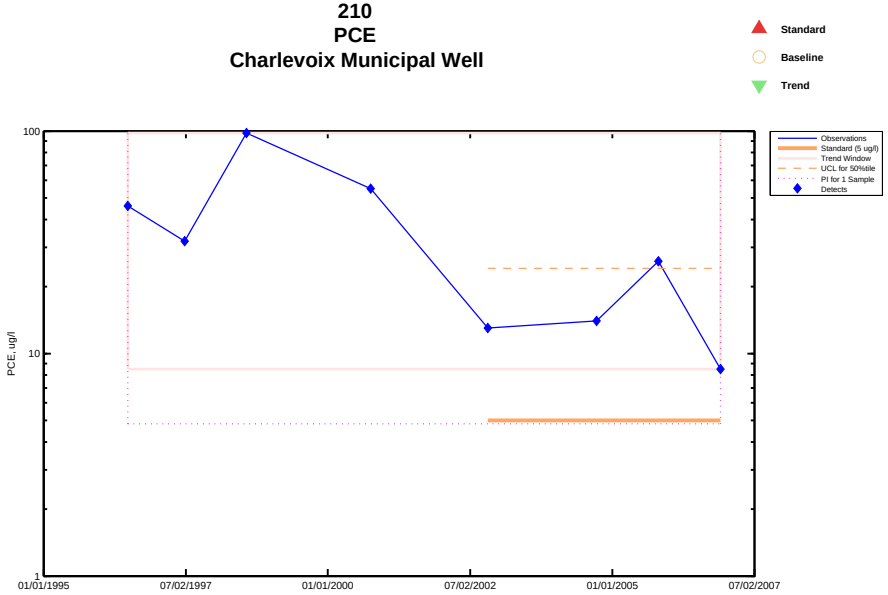


Standard Test (95%): Exceedance <UCL = 1.26e+002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 6.40e+002/7.48e+000 ug/l>
Trend Test (80%): Downward <Slope = −1.37e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

210
PCE
Charlevoix Municipal Well

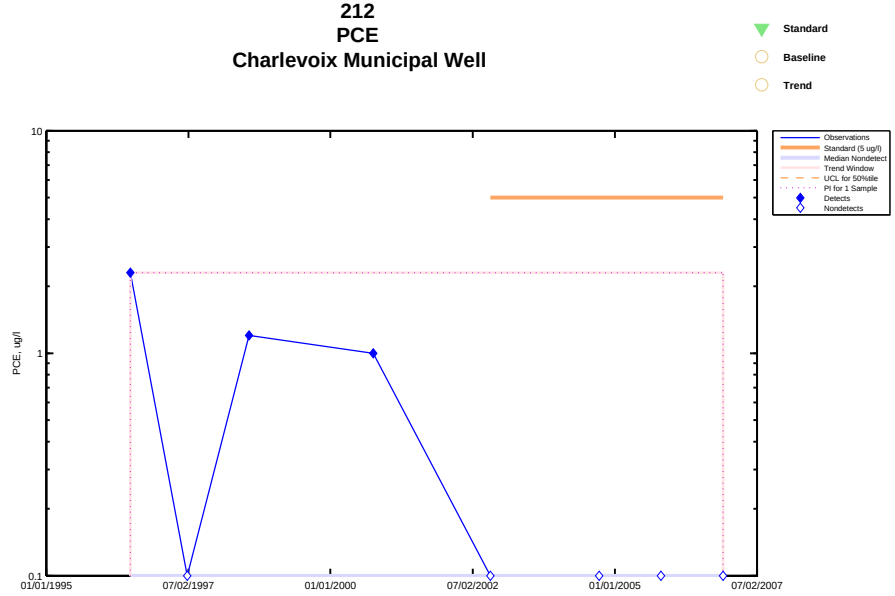


Standard Test (95%): Exceedance <UCL = 2.42e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.40e+002/4.83e+000 ug/l>
Trend Test (80%): Downward <Slope = −1.38e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

212
PCE
Charlevoix Municipal Well

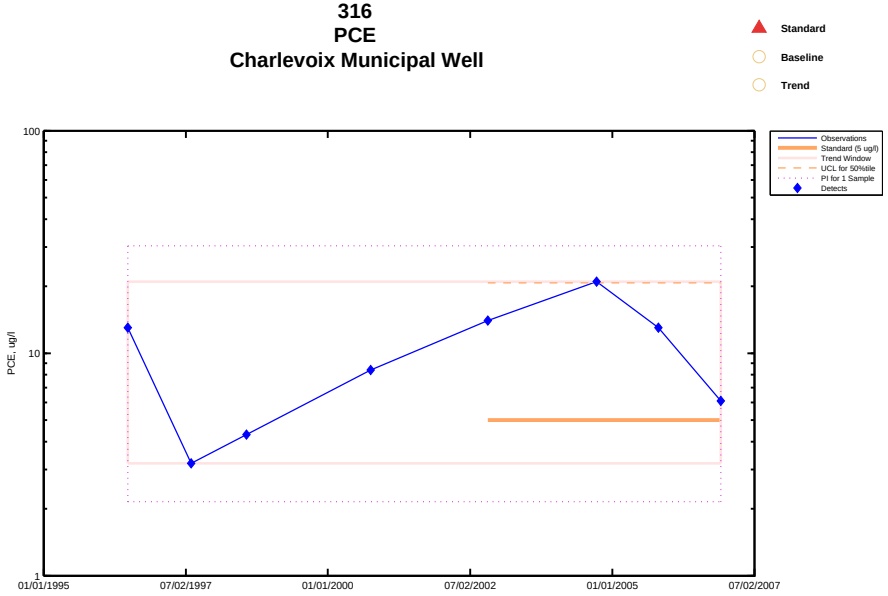


Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 2.30e+000/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = −9.10e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

316
PCE
Charlevoix Municipal Well



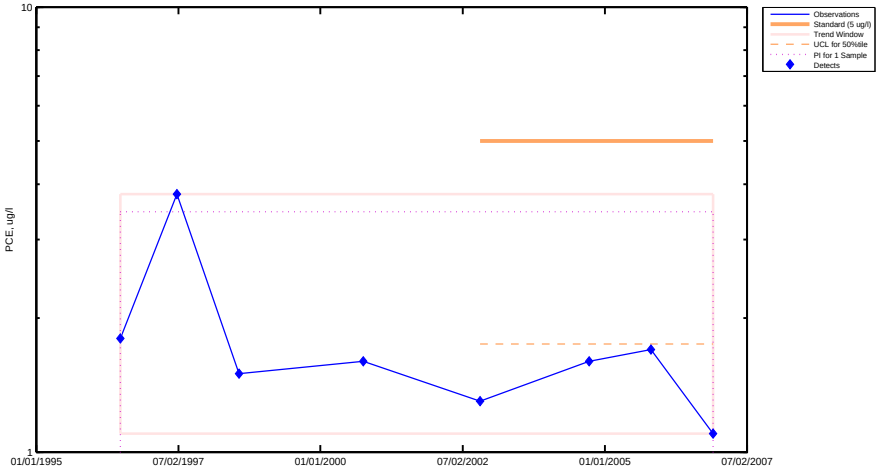
Standard Test (95%): Exceedance <UCL = 2.07e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 3.04e+001/2.15e+000 ug/l>
Trend Test (80%): No Trend <Slope = 5.72e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

320
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



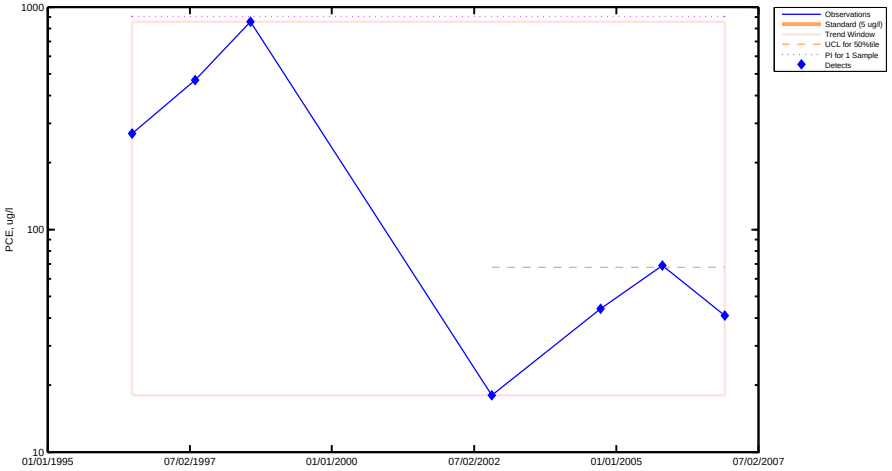
Standard Test (95%): Compliance <UCL = 1.75e+000 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 3.47e+000/6.51e−001 ug/l>
Trend Test (80%): Downward <Slope = −2.13e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

402
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



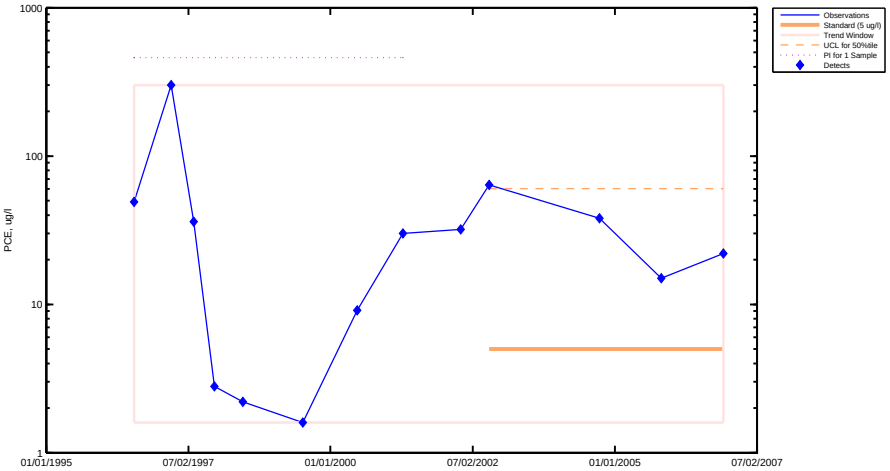
Standard Test (95%): Exceedance <UCL = 6.76e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 9.07e+002/0.00e+000 ug/l>
Trend Test (80%): No Trend <Slope = −1.96e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

405
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



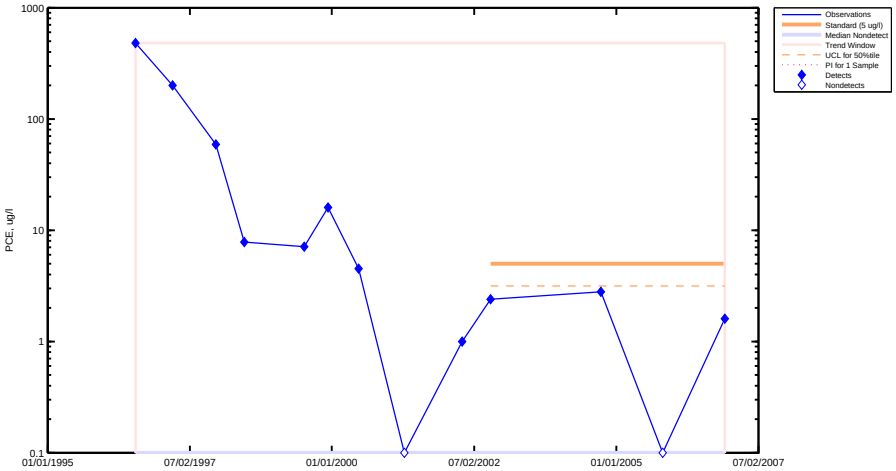
Standard Test (95%): Exceedance <UCL = 6.03e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 4.60e+002/0.00e+000 ug/l>
Trend Test (80%): No Trend <Slope = −3.84e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

406
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



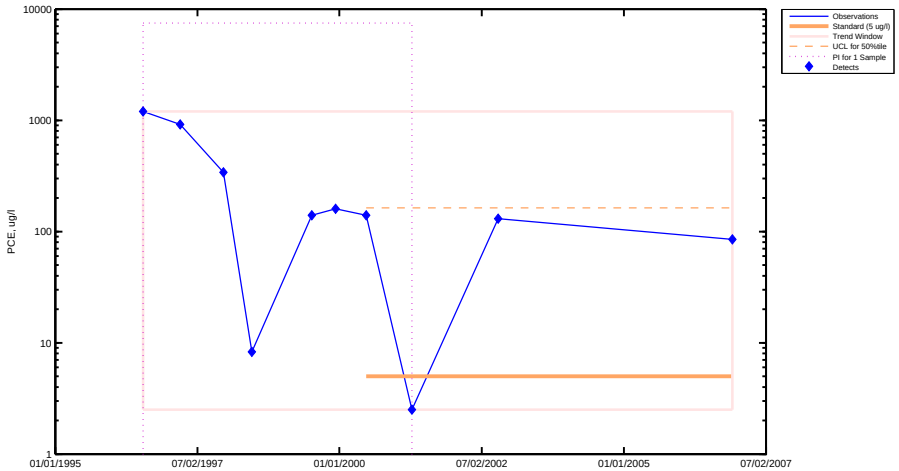
Standard Test (95%): Compliance <UCL = 3.14e+000 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.27e+003/0.00e+000 ug/l>
Trend Test (80%): Downward <Slope = −3.96e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

407
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



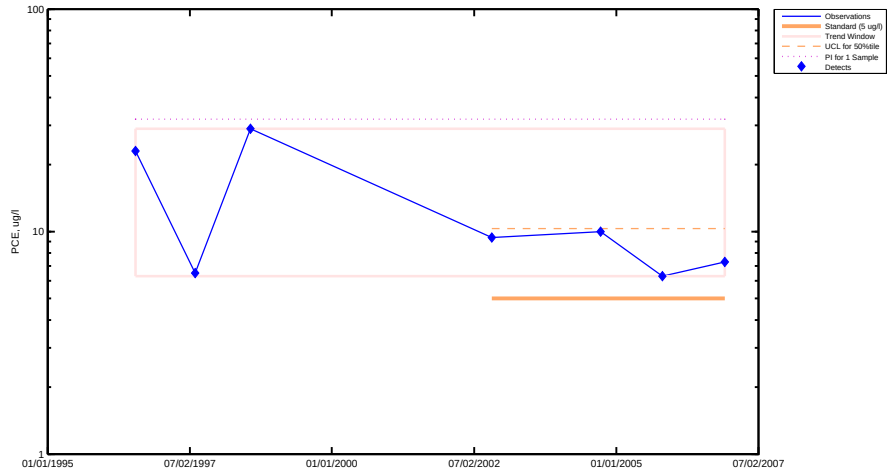
Standard Test (95%): Exceedance <UCL = 1.63e+002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 7.46e+003/8.88e−001 ug/l>
Trend Test (80%): Downward <Slope = −2.94e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

408
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



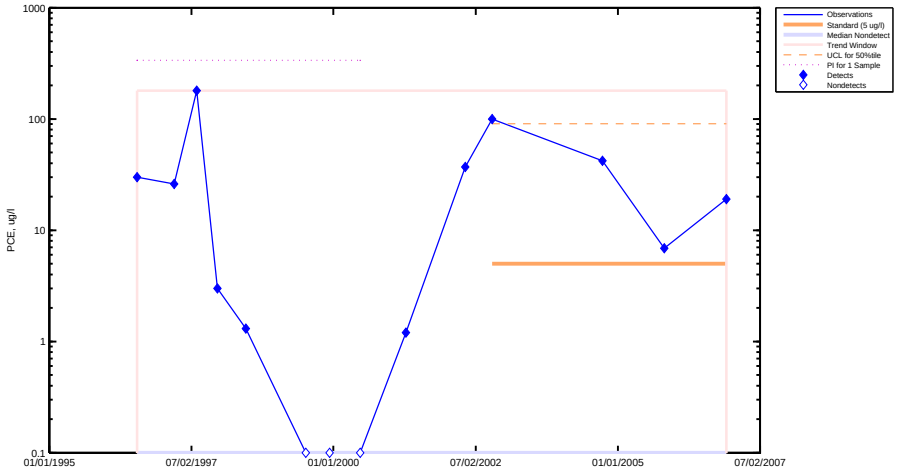
Standard Test (95%): Exceedance <UCL = 1.03e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 3.20e+001/0.00e+000 ug/l>
Trend Test (80%): No Trend <Slope = −9.74e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

409
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



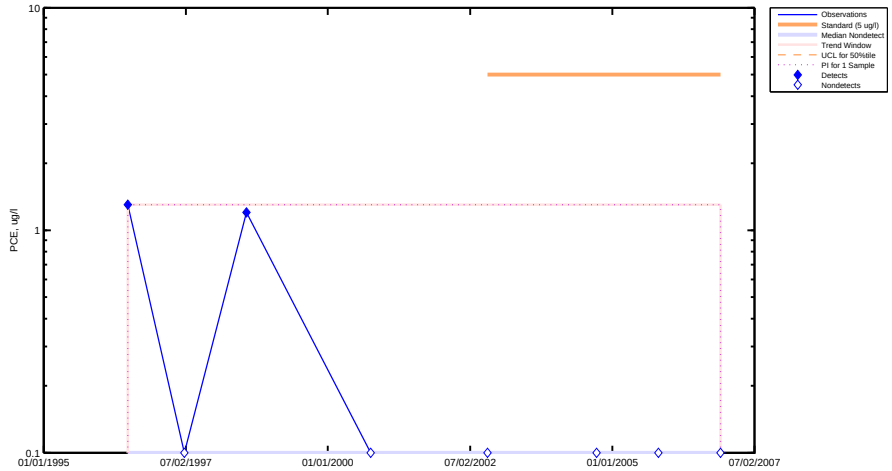
Standard Test (95%): Exceedance <UCL = 9.06e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 3.35e+002/0.00e+000 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

501D
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



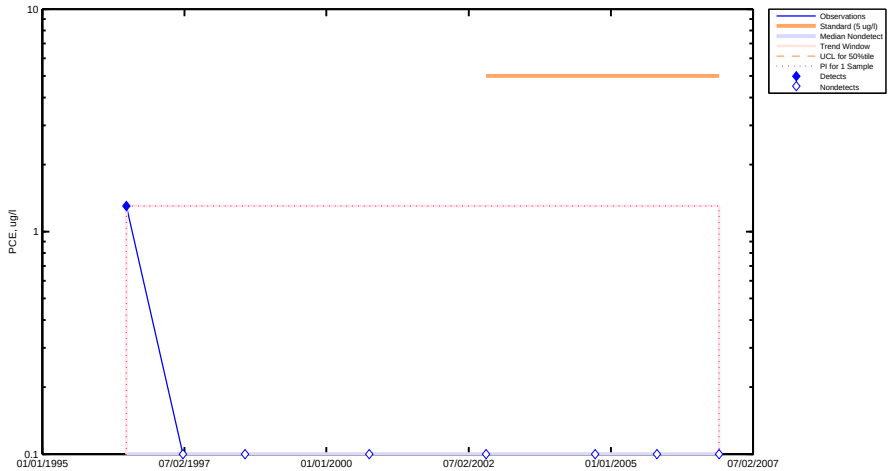
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.30e+000/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

501S
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



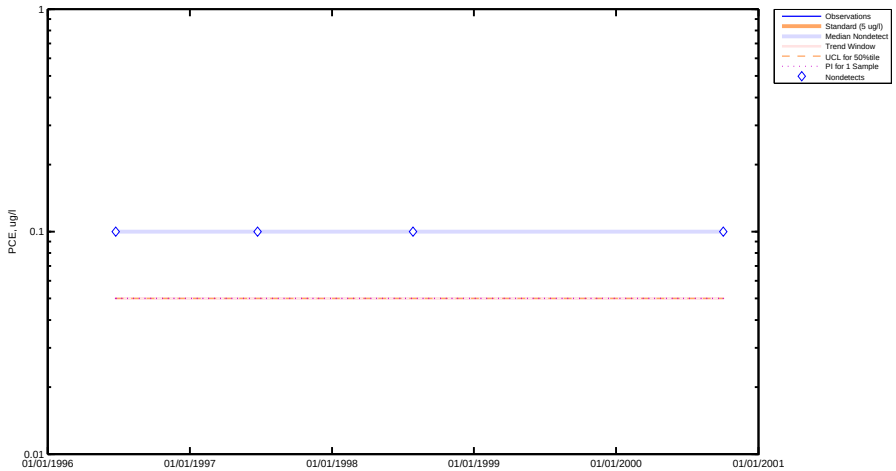
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.30e+000/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

502D
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



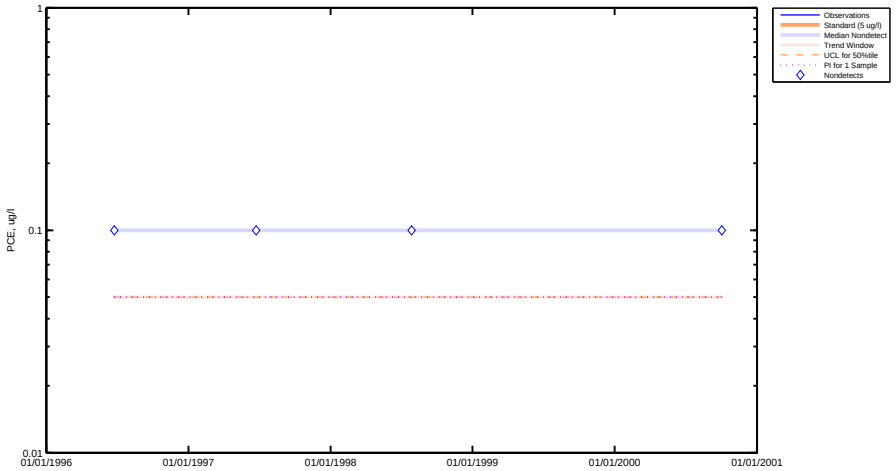
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

502S
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



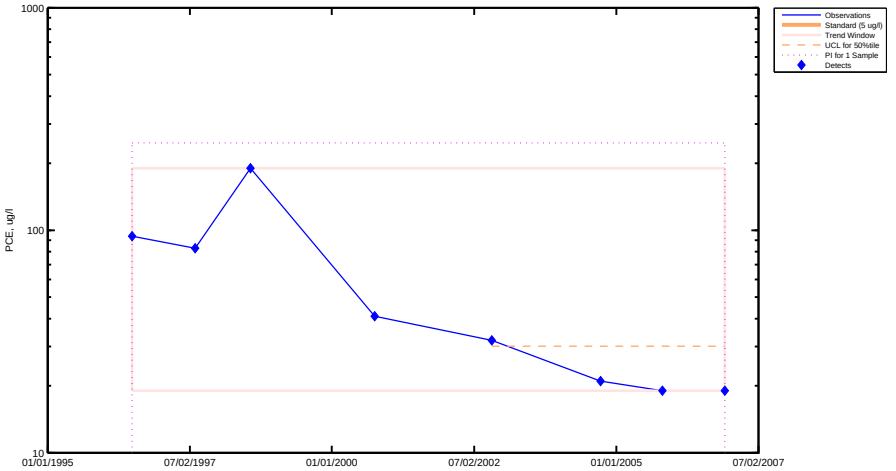
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

601
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



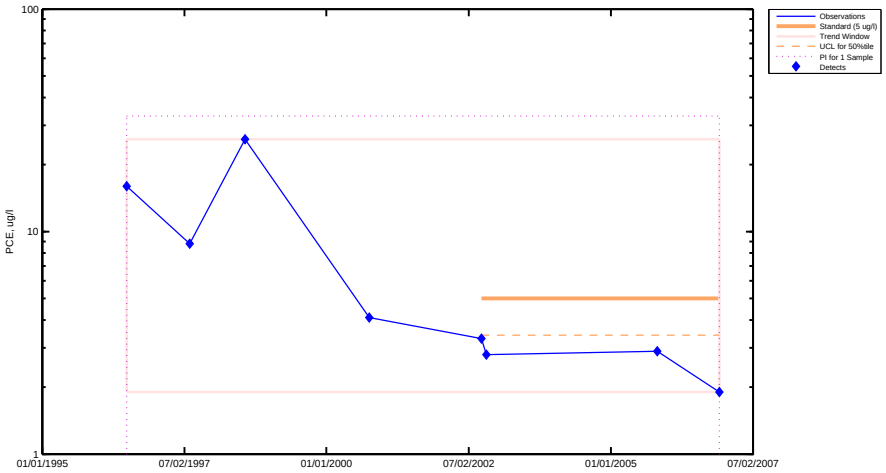
Standard Test (95%): Exceedance <UCL = 3.01e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 2.47e+002/7.39e+000 ug/l>
Trend Test (80%): Downward <Slope = −1.54e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

602
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



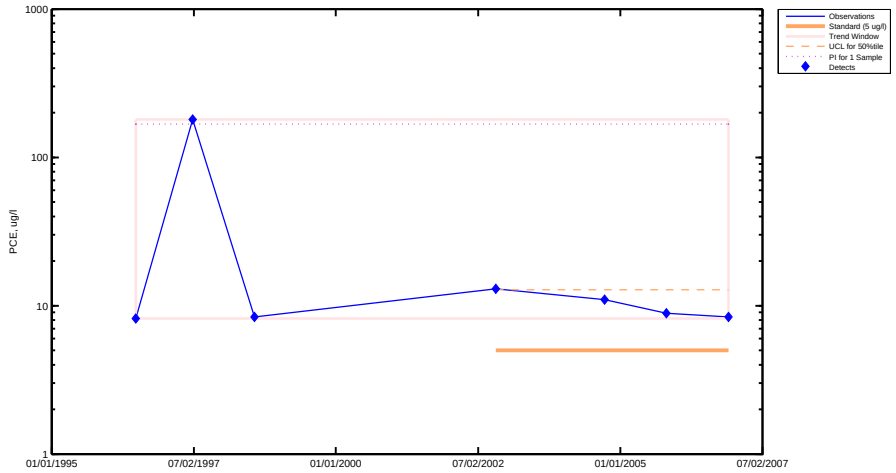
Standard Test (95%): Compliance <UCL = 3.42e+000 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 3.30e+001/3.35e−001 ug/l>
Trend Test (80%): Downward <Slope = −1.52e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

603
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



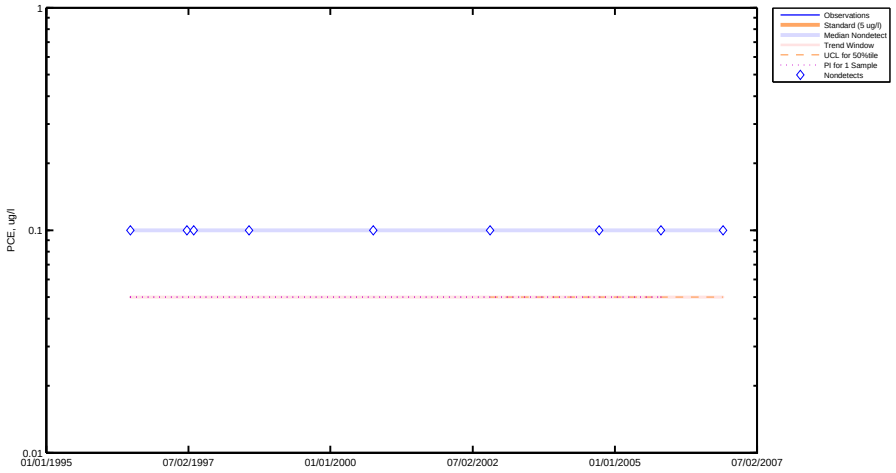
Standard Test (95%): Exceedance <UCL = 1.28e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.68e+002/0.00e+000 ug/l>
Trend Test (80%): No Trend <Slope = −4.63e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

605D
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



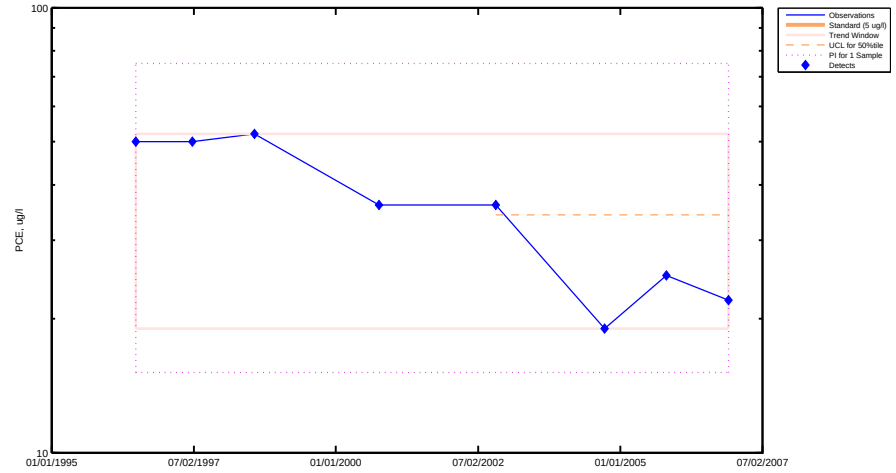
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

605S
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



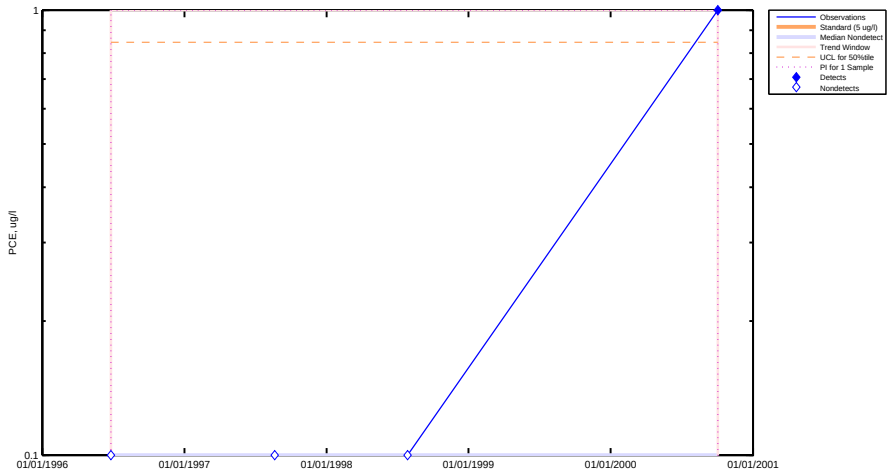
Standard Test (95%): Exceedance <UCL = 3.42e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 7.49e+001/1.51e+001 ug/l>
Trend Test (80%): Downward <Slope = −7.93e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

606D
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



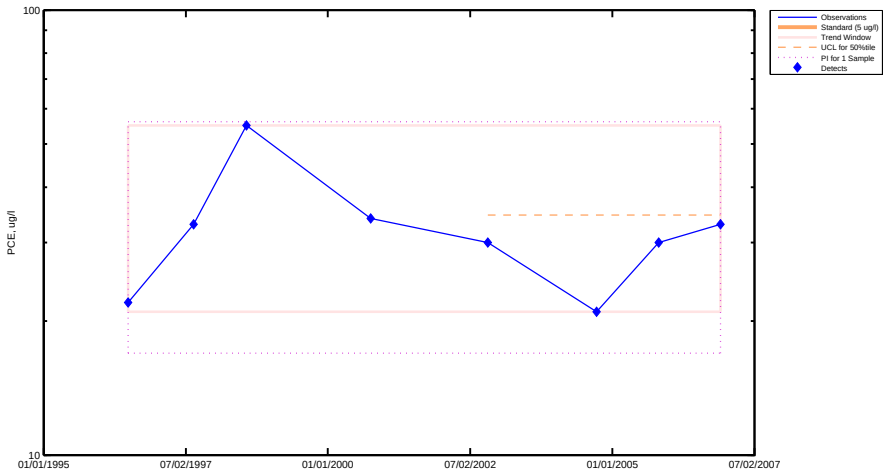
Standard Test (95%): Compliance <UCL = 8.46e−001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.00e+000/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 7.84e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

606S
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



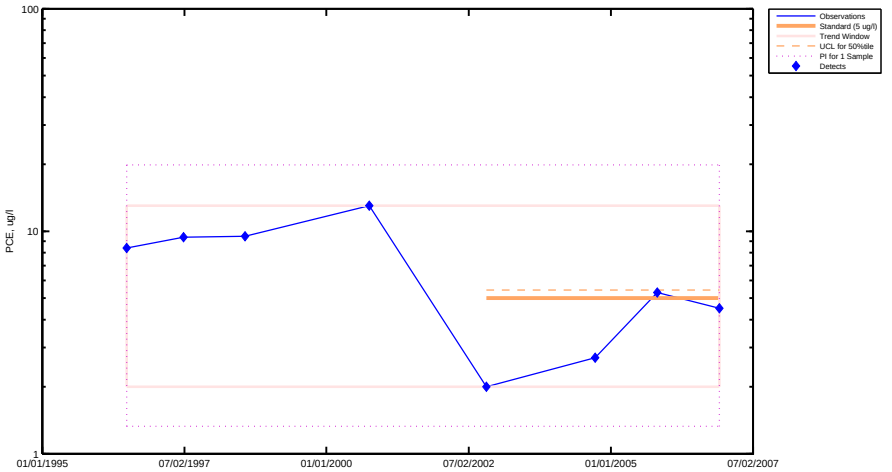
Standard Test (95%): Exceedance <UCL = 3.46e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.61e+001/1.69e+001 ug/l>
Trend Test (80%): No Trend <Slope = −2.35e−003 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

607
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



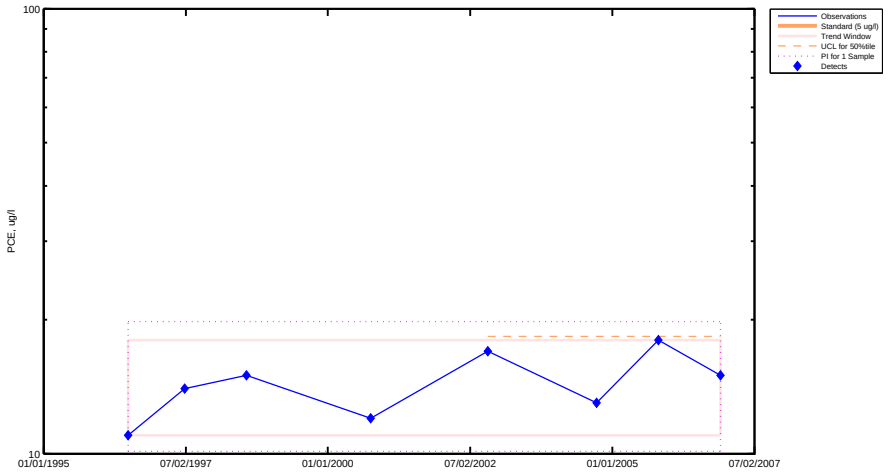
Standard Test (95%): Exceedance <UCL = 5.43e+000 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.98e+001/1.33e+000 ug/l>
Trend Test (80%): No Trend <Slope = −6.19e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

608
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



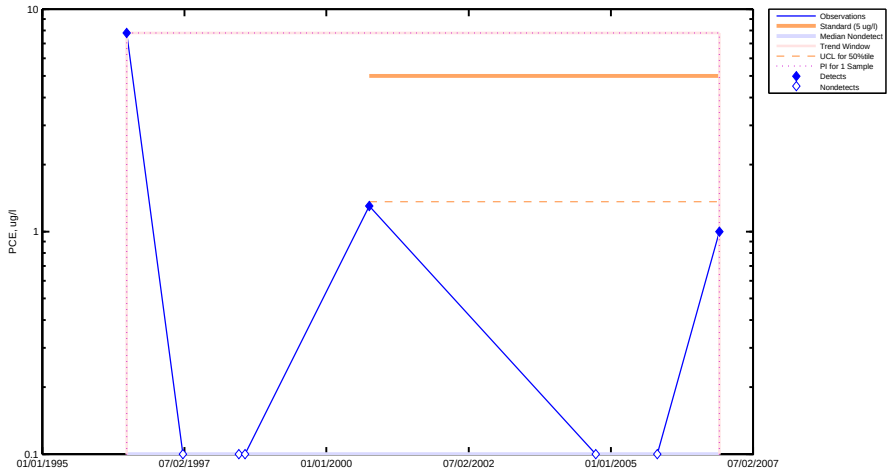
Standard Test (95%): Exceedance <UCL = 1.84e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.98e+001/1.01e+001 ug/l>
Trend Test (80%): Upward <Slope = 2.60e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

609
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



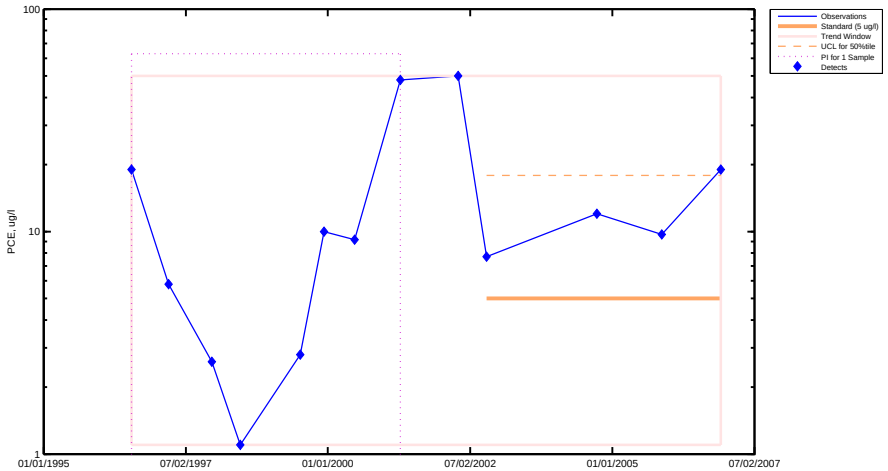
Standard Test (95%): Compliance <UCL = 1.36e+000 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 7.80e+000/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

610
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



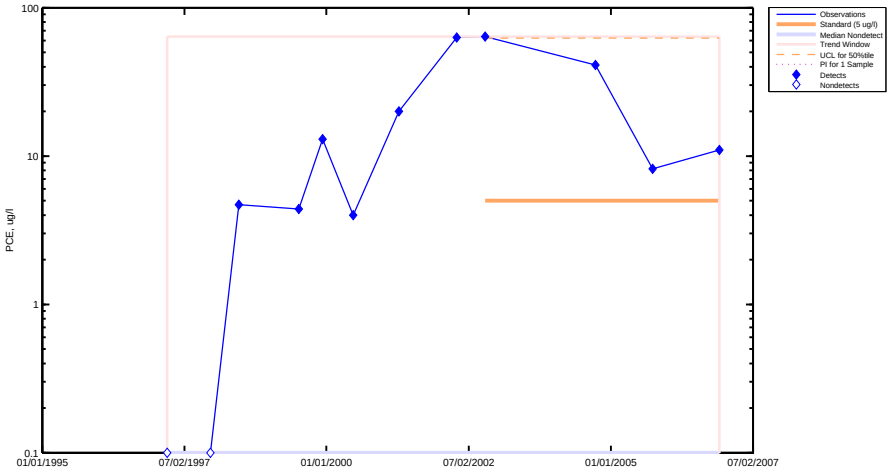
Standard Test (95%): Exceedance <UCL = 1.79e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 6.31e+001/6.23e−002 ug/l>
Trend Test (80%): Upward <Slope = 1.25e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

611
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



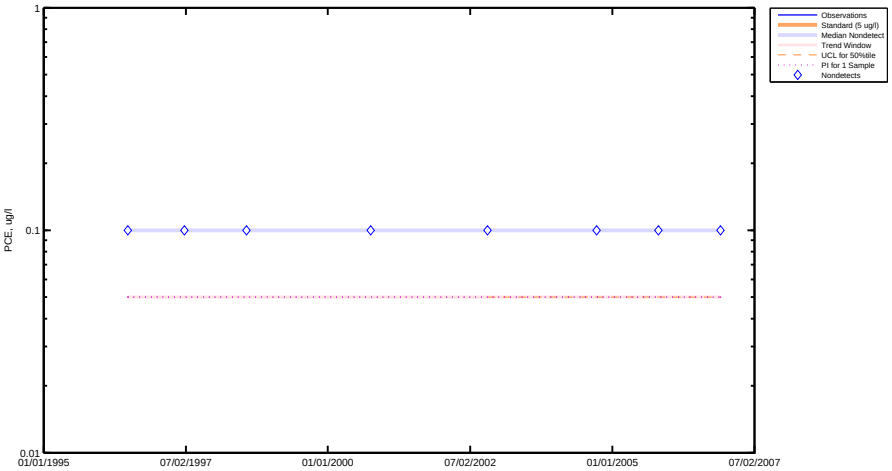
Standard Test (95%): Exceedance <UCL = 6.22e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.16e+002/0.00e+000 ug/l>
Trend Test (80%): Upward <Slope = 4.27e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

T2
PCE
Charlevoix Municipal Well

Standard
Baseline
Trend



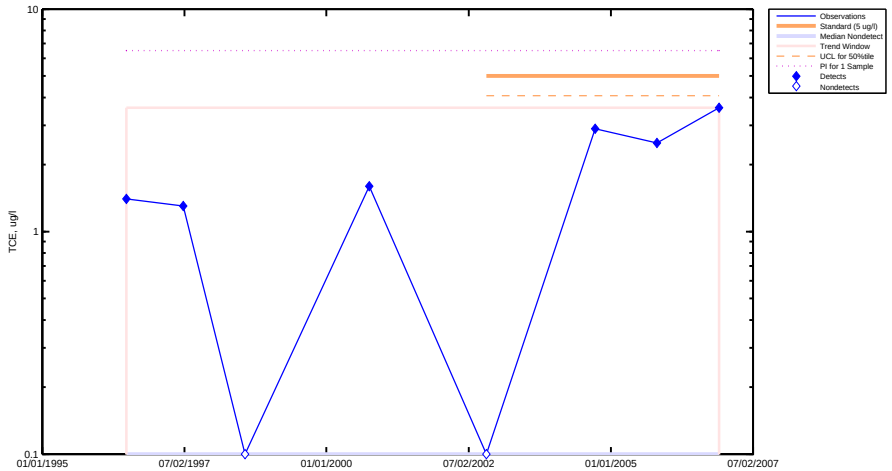
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

2
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



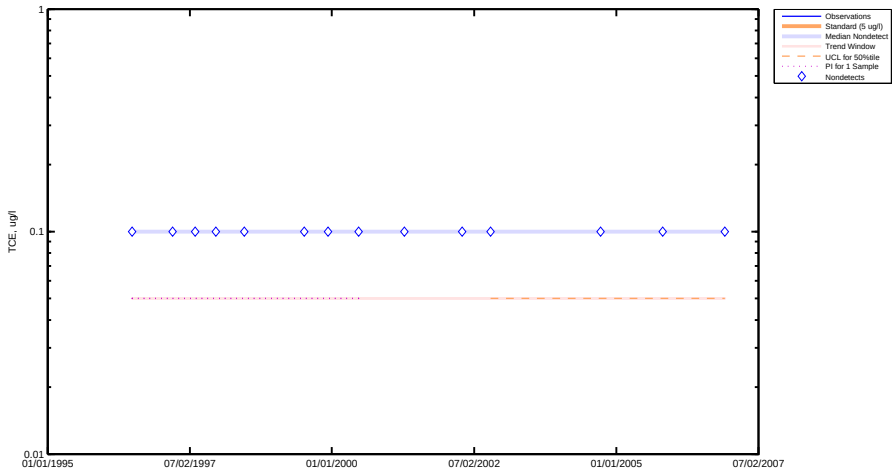
Standard Test (95%): Compliance <UCL = 4.08e+000 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 6.51e+000/0.00e+000 ug/l>
Trend Test (80%): Upward <Slope = 6.25e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

3
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



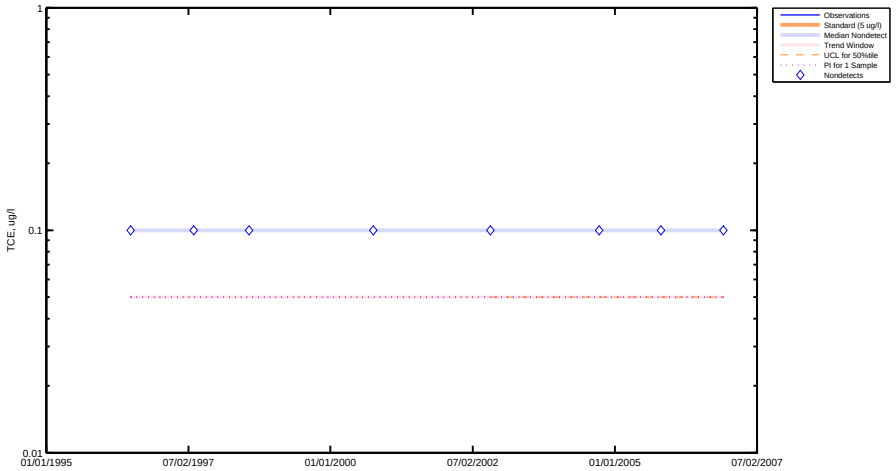
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

11
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



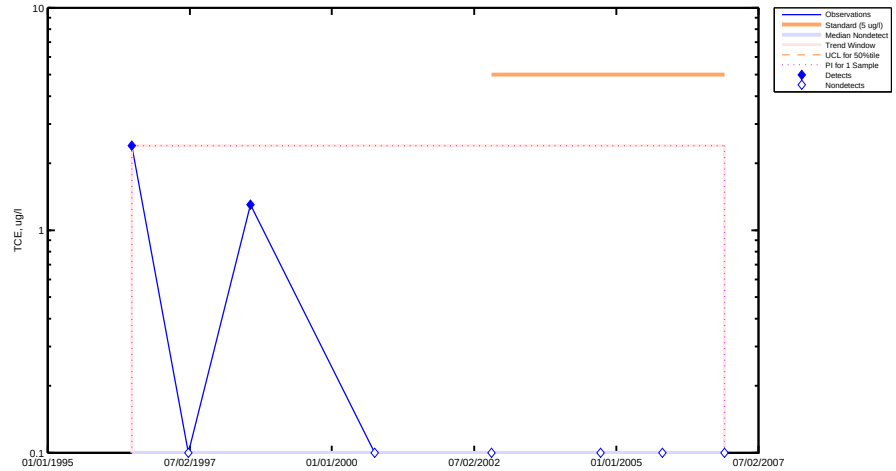
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

206
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



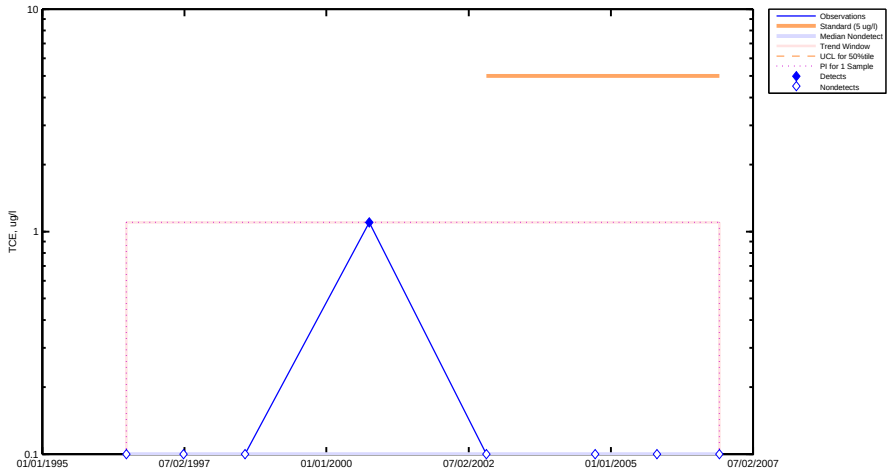
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 2.40e+000/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

209
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



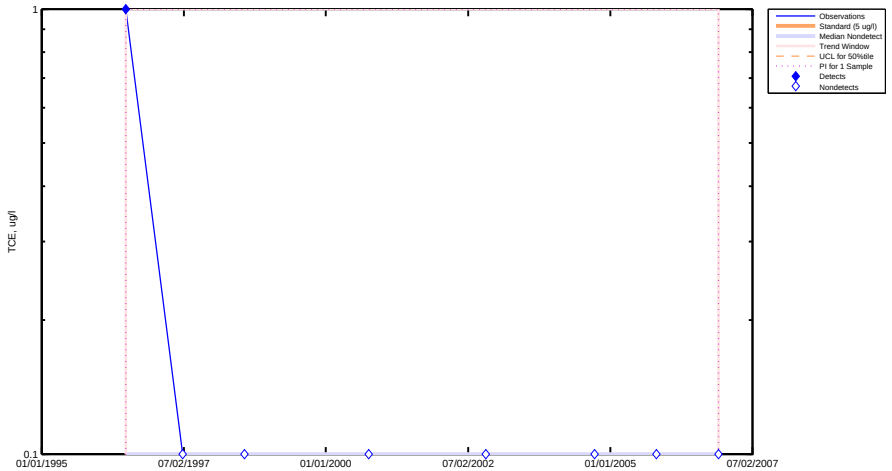
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.10e+000/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

210
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



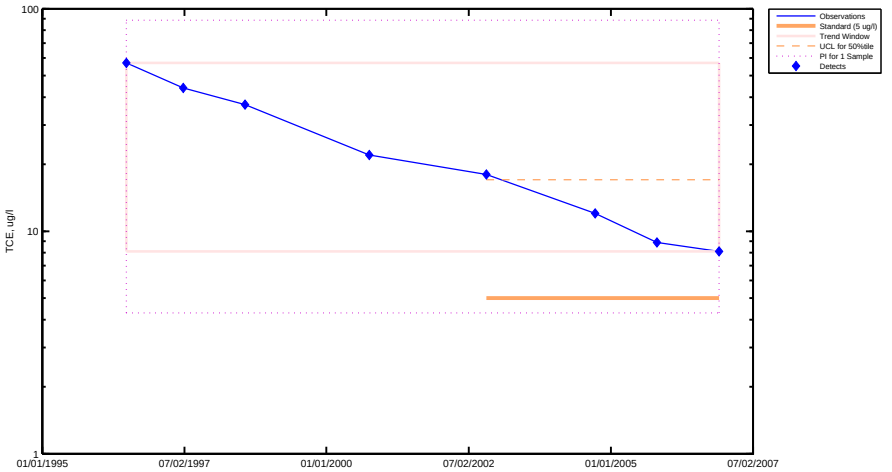
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.00e+000/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

212
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



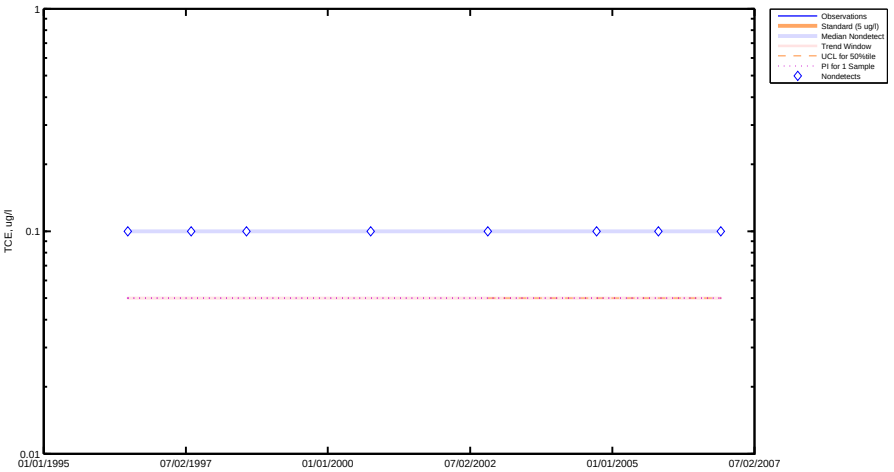
Standard Test (95%): Exceedance <UCL = 1.70e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 8.86e+001/4.30e+000 ug/l>
Trend Test (80%): Downward <Slope = −1.62e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

316
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



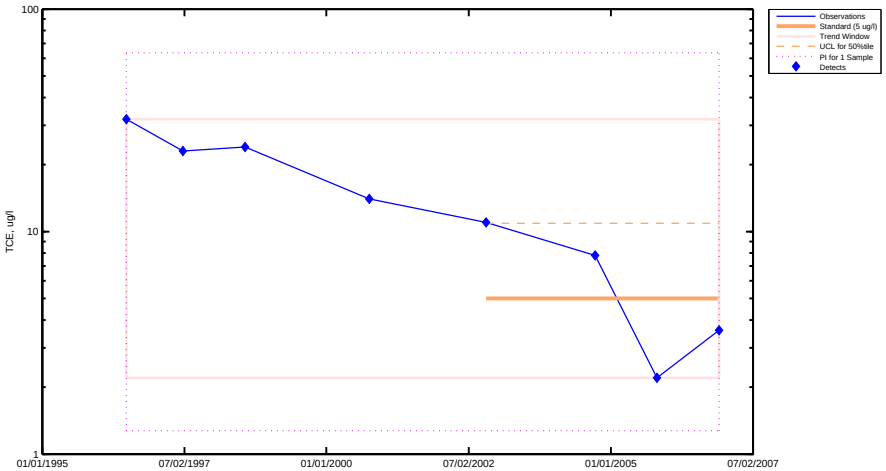
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

320
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



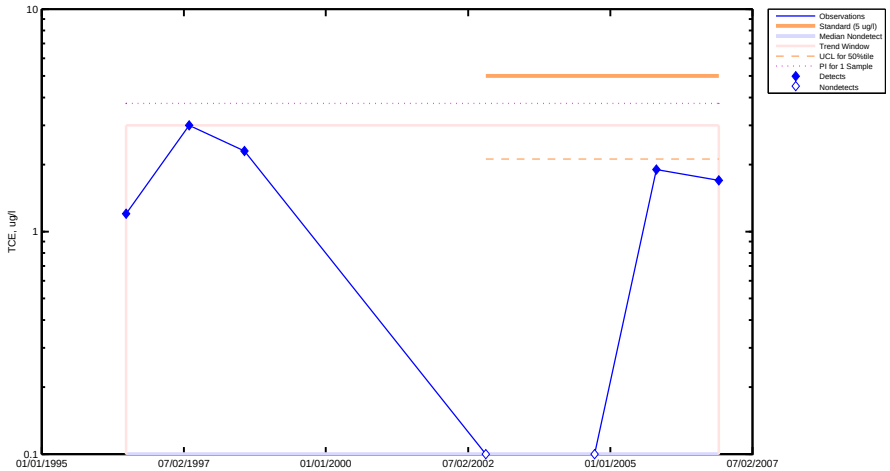
Standard Test (95%): Exceedance <UCL = 1.09e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 6.35e+001/1.27e+000 ug/l>
Trend Test (80%): Downward <Slope = −1.65e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

402
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



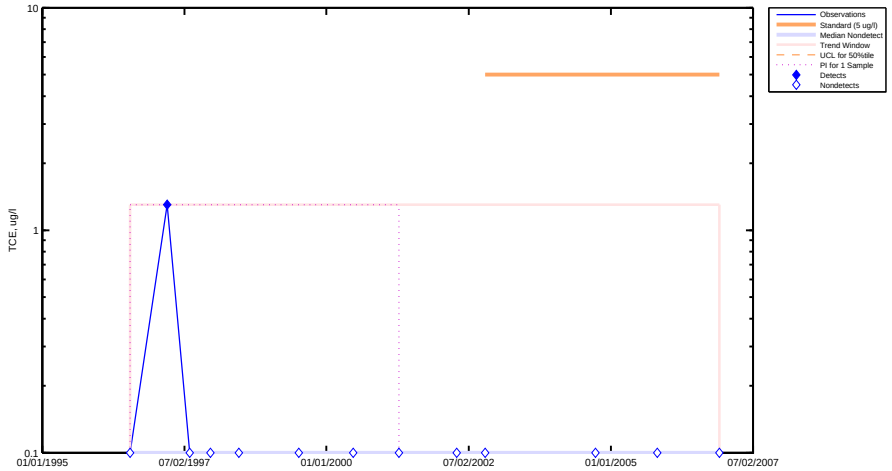
Standard Test (95%): Compliance <UCL = 2.12e+000 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 3.76e+000/0.00e+000 ug/l>
Trend Test (80%): No Trend <Slope = −2.38e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

405
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



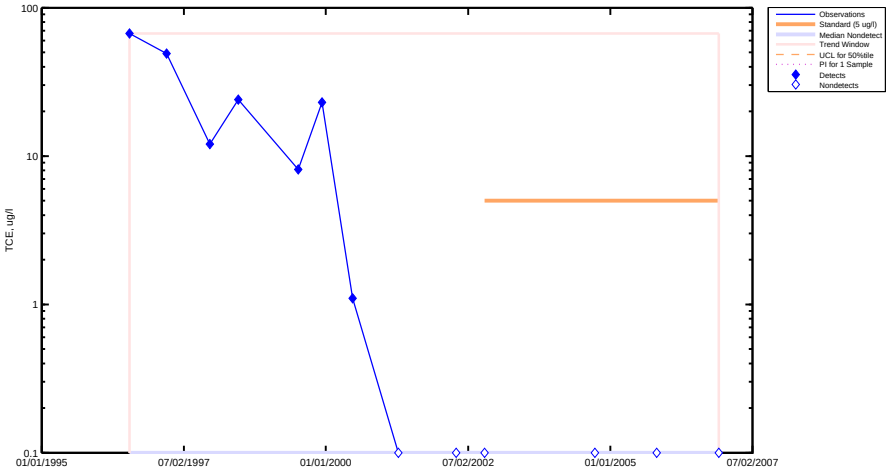
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.30e+000/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

406
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



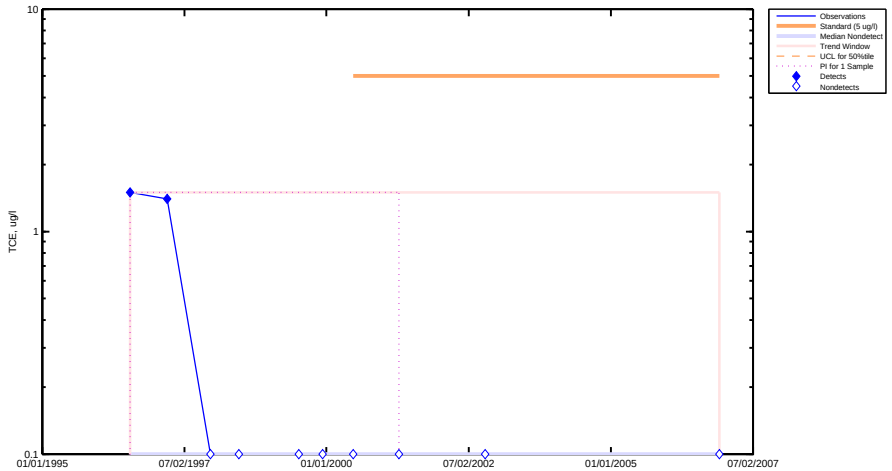
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 2.43e+002/0.00e+000 ug/l>
Trend Test (80%): Downward <Slope = −3.62e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

407
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



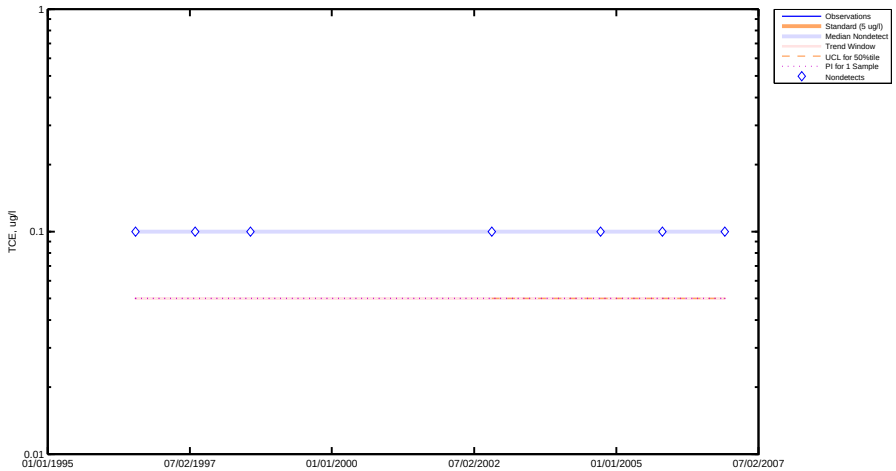
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.50e+000/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

408
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



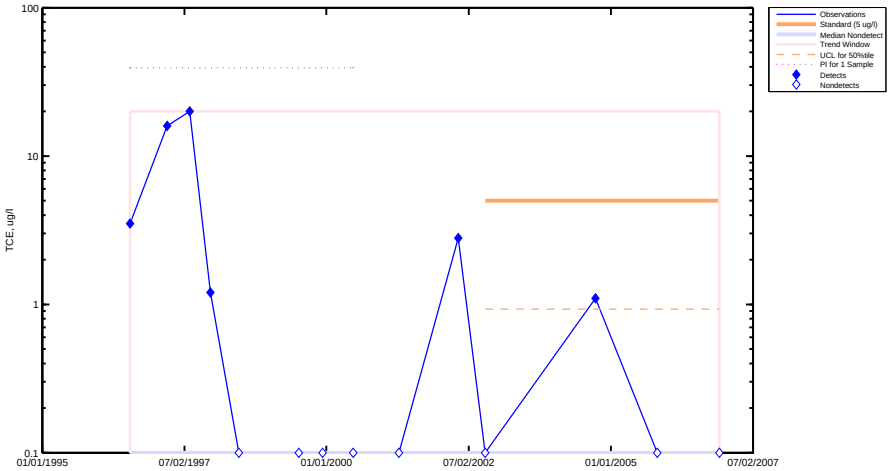
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

409
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



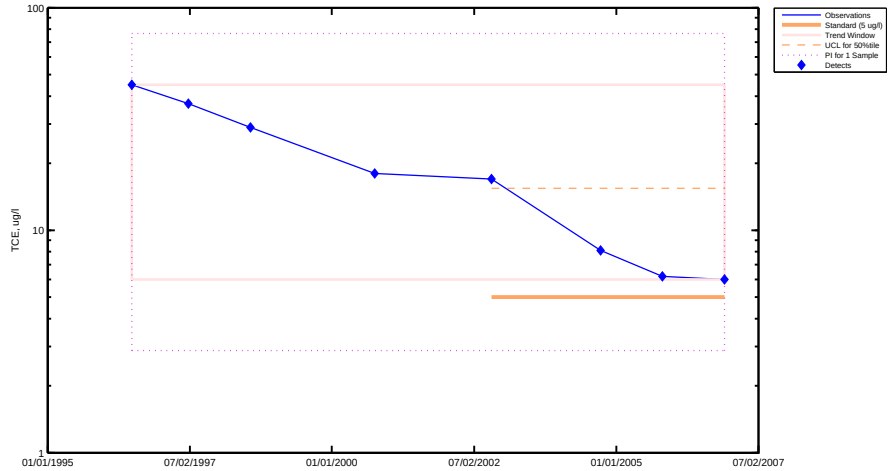
Standard Test (95%): Compliance <UCL = 9.30e−001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 3.94e+001/0.00e+000 ug/l>
Trend Test (80%): No Trend <Slope = −7.93e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

501D
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



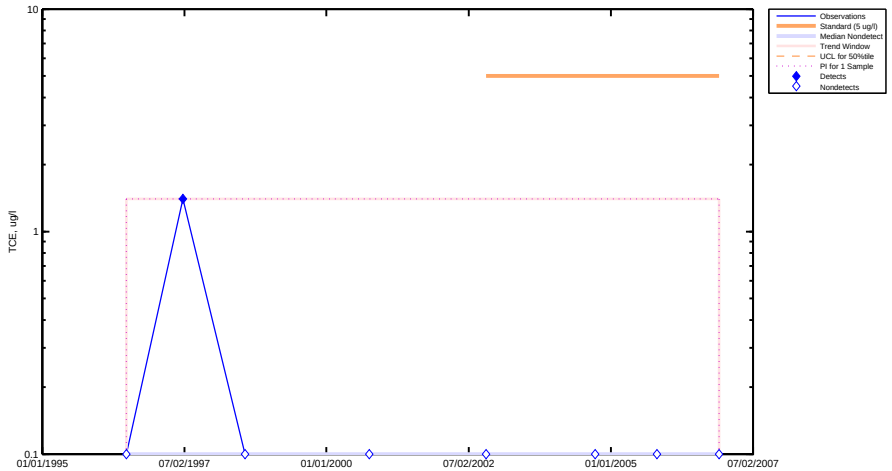
Standard Test (95%): Exceedance <UCL = 1.54e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 7.66e+001/2.88e+000 ug/l>
Trend Test (80%): Downward <Slope = −1.77e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

501S
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



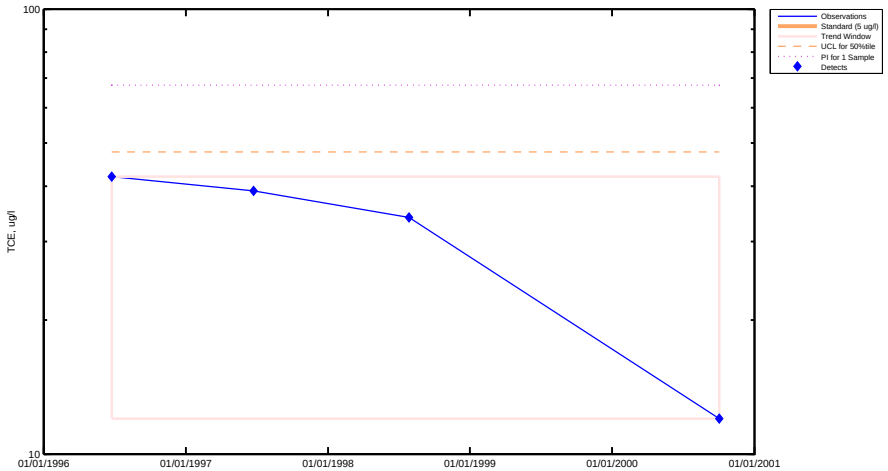
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.40e+000/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

502D
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



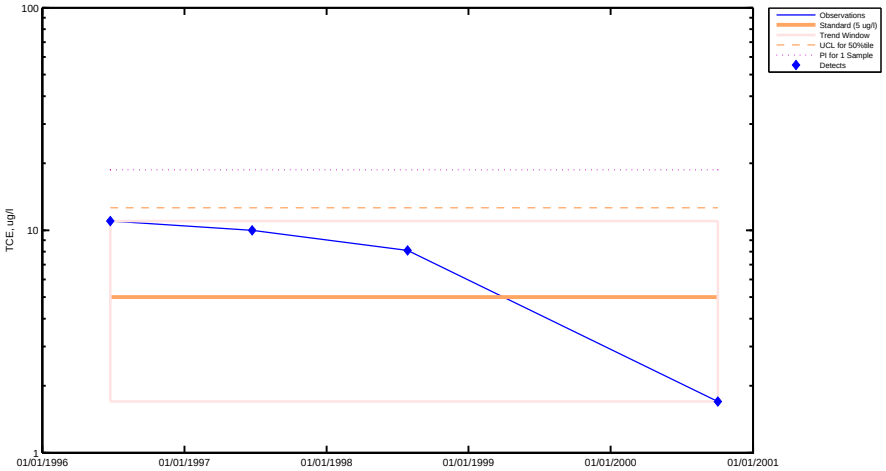
Standard Test (95%): Exceedance <UCL = 4.77e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 6.75e+001/0.00e+000 ug/l>
Trend Test (80%): Downward <Slope = −1.82e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

502S
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



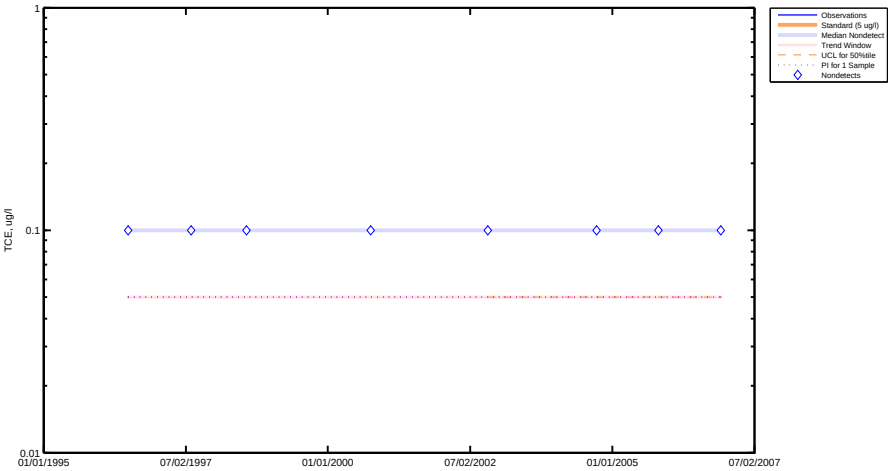
Standard Test (95%): Exceedance <UCL = 1.26e+001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.87e+001/0.00e+000 ug/l>
Trend Test (80%): Downward <Slope = −2.30e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

601
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend

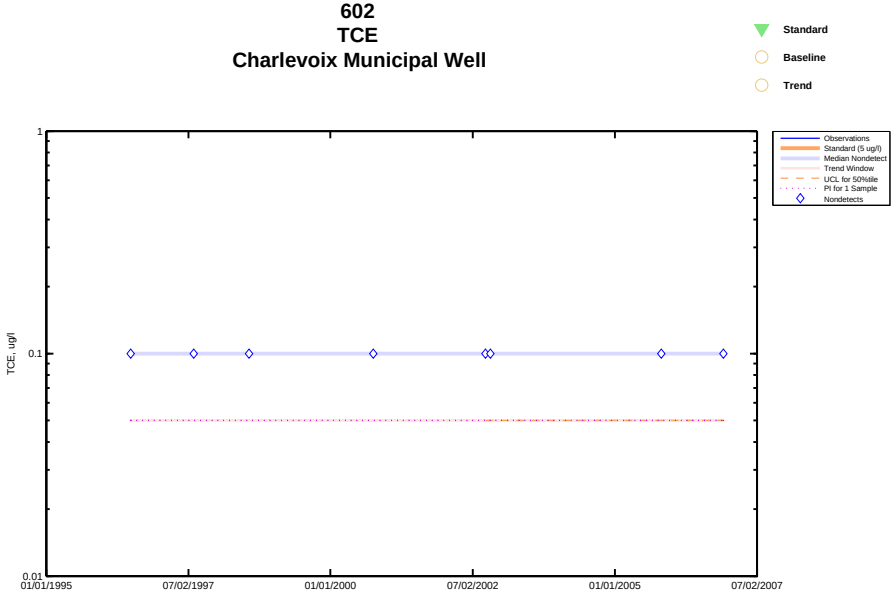


Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

602
TCE
Charlevoix Municipal Well

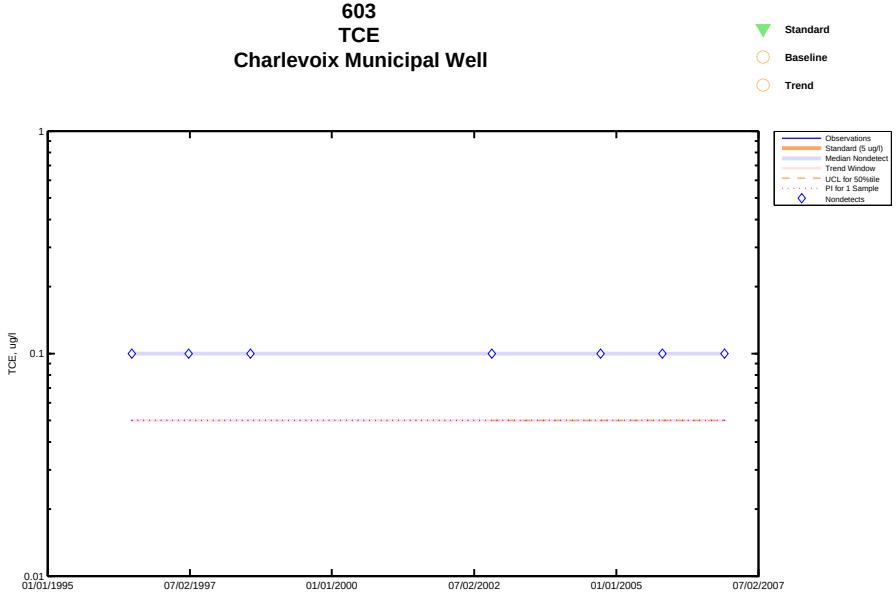


Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

603
TCE
Charlevoix Municipal Well

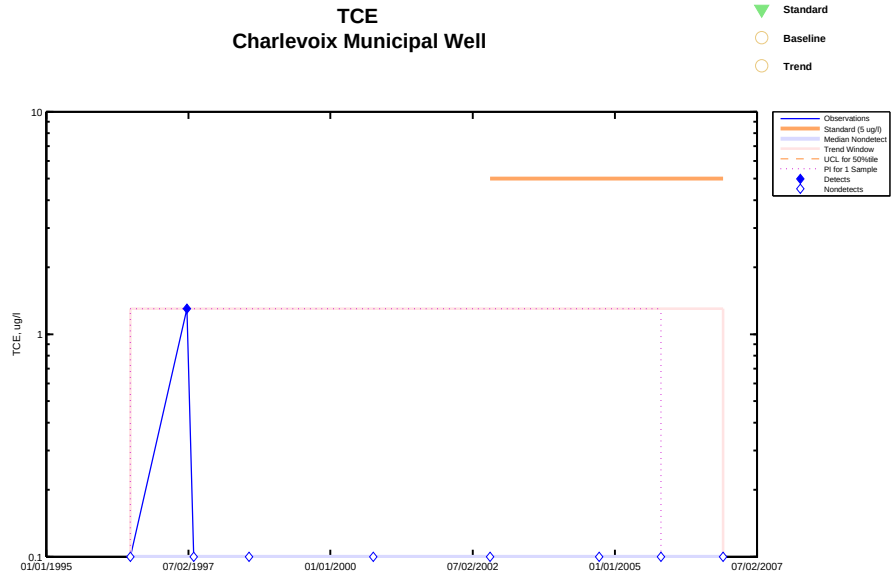


Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

605D
TCE
Charlevoix Municipal Well

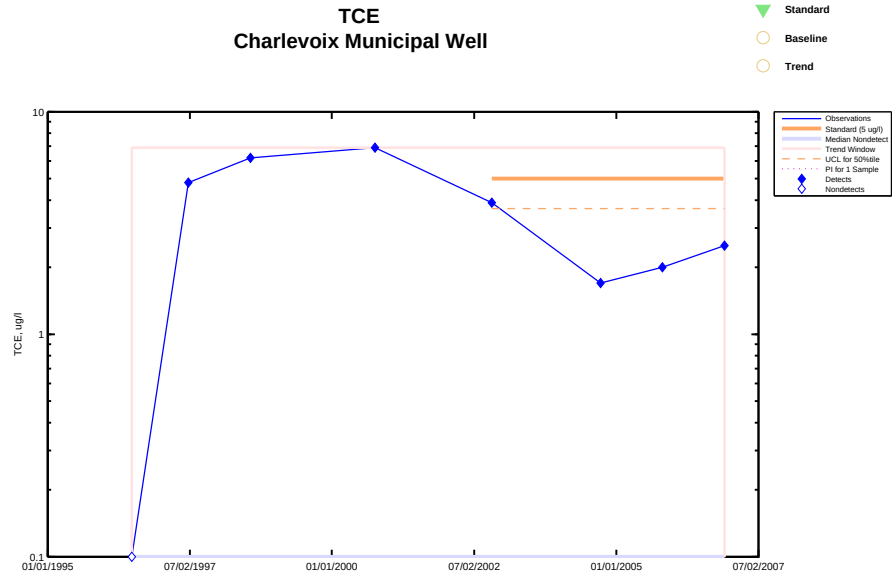


Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.30e+000/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

605S
TCE
Charlevoix Municipal Well

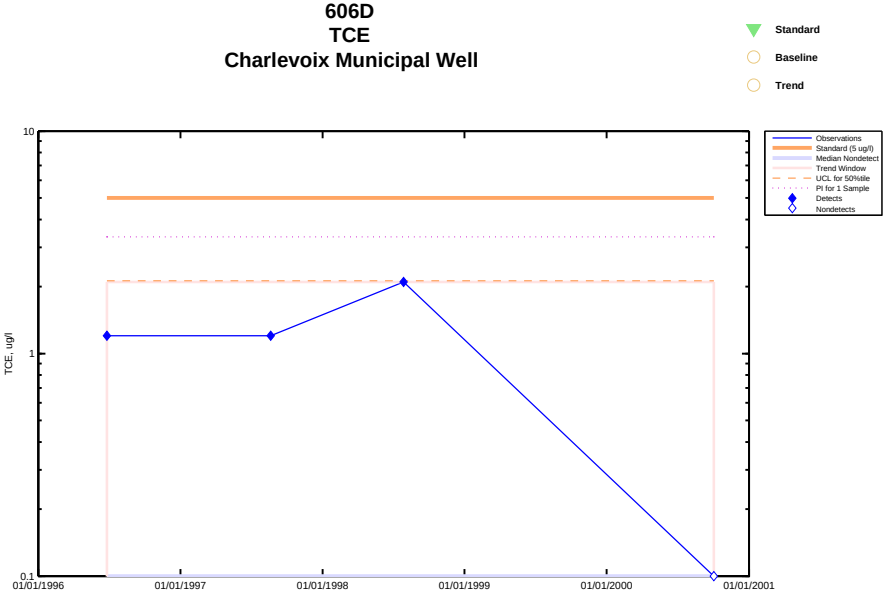


Standard Test (95%): Compliance <UCL = 3.67e+000 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.38e+001/0.00e+000 ug/l>
Trend Test (80%): No Trend <Slope = −4.17e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

606D
TCE
Charlevoix Municipal Well

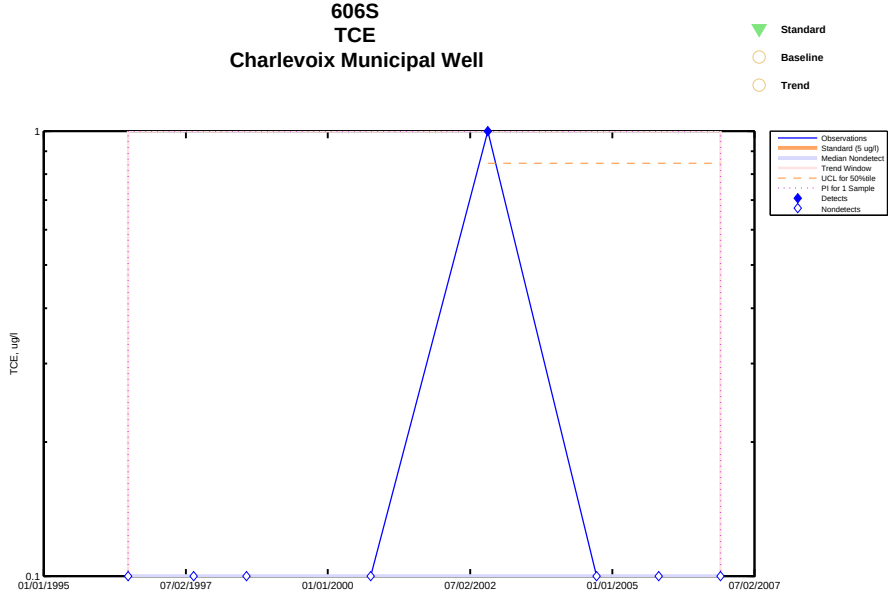


Standard Test (95%): Compliance <UCL = 2.13e+000 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 3.35e+000/0.00e+000 ug/l>
Trend Test (80%): No Trend <Slope = −8.29e−002 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

606S
TCE
Charlevoix Municipal Well

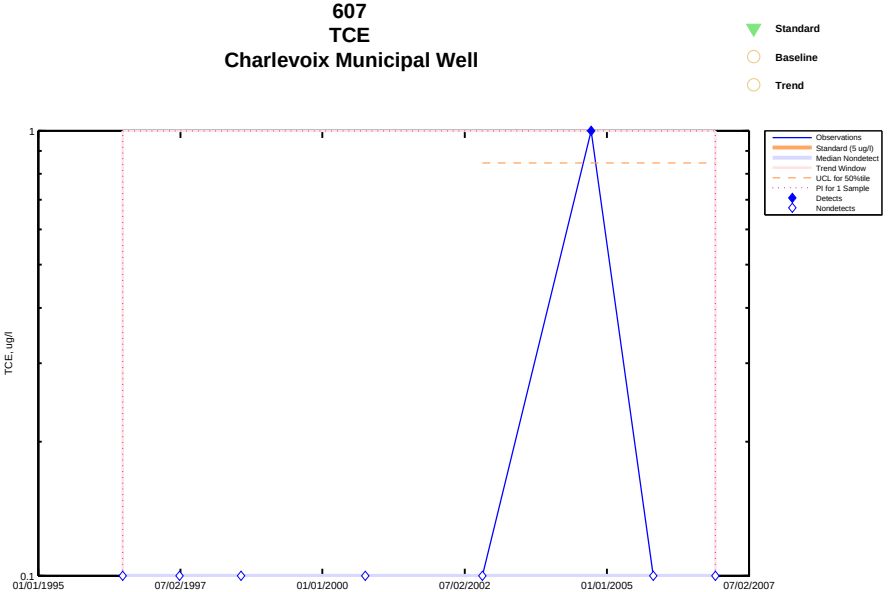


Standard Test (95%): Compliance <UCL = 8.46e−001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.00e+000/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

607
TCE
Charlevoix Municipal Well

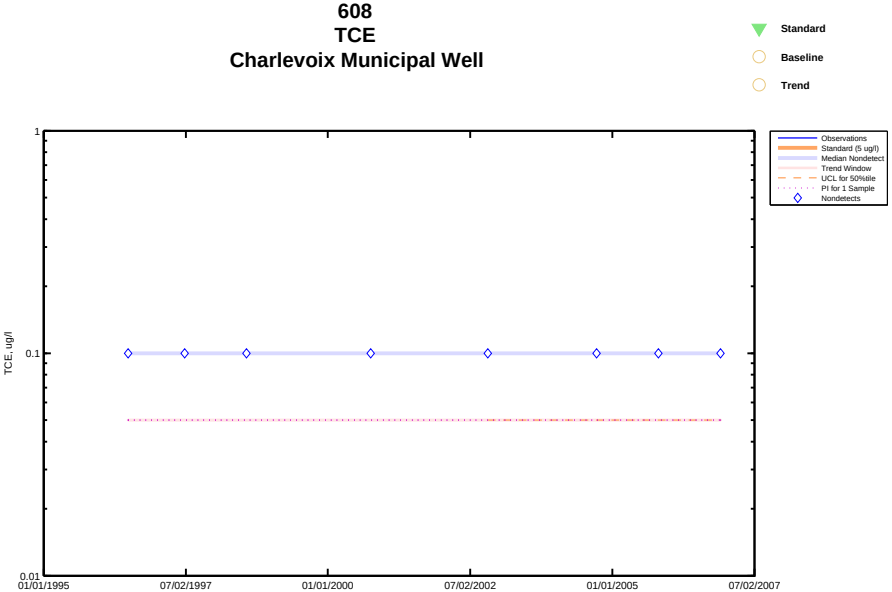


Standard Test (95%): Compliance <UCL = 8.46e−001 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.00e+000/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

608
TCE
Charlevoix Municipal Well



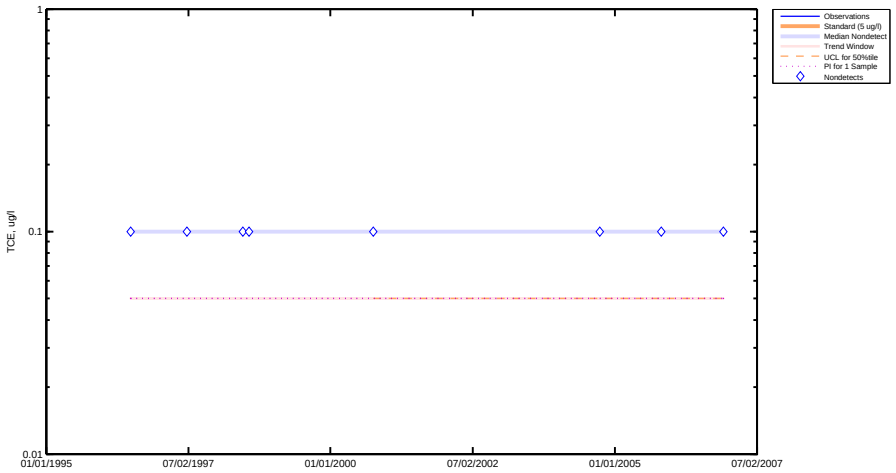
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

609
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



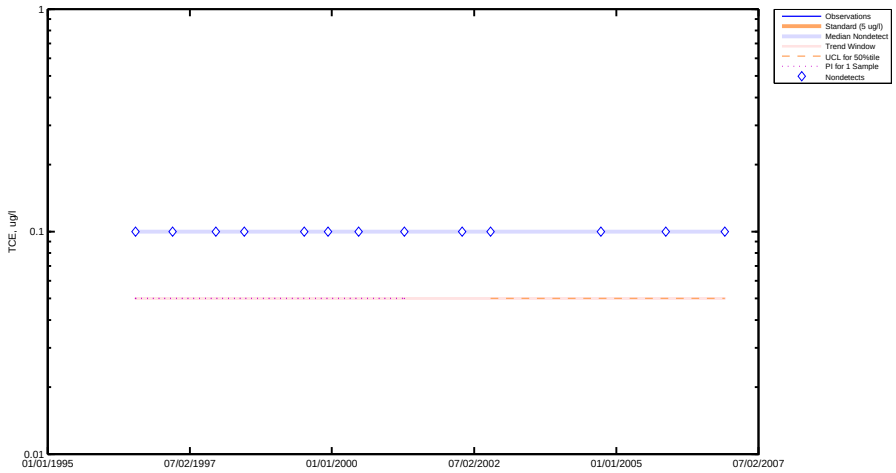
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

610
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



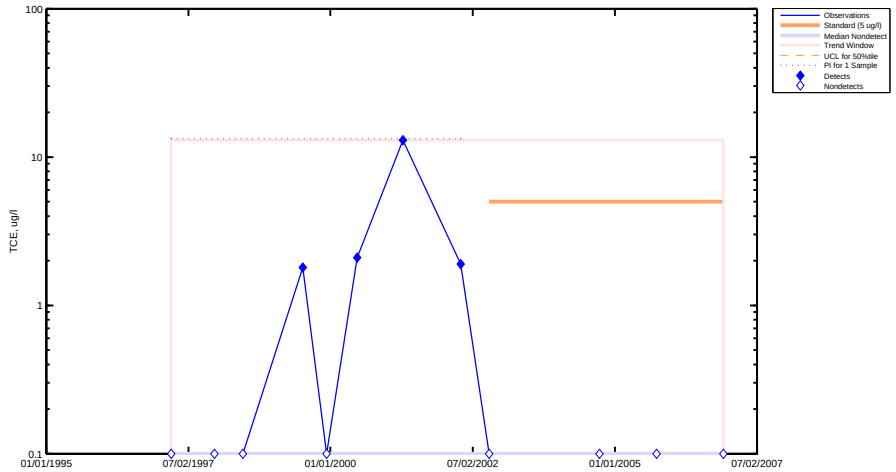
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 5.00e−002/5.00e−002 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

611
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



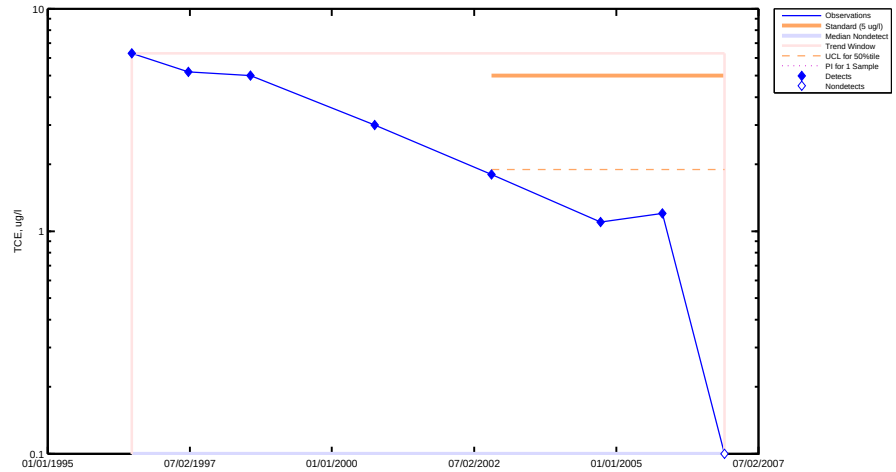
Standard Test (95%): Compliance <UCL = 5.00e−002 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.33e+001/0.00e+000 ug/l>
Trend Test (80%): No Trend <Slope = 0.00e+000 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

T2
TCE
Charlevoix Municipal Well

Standard
Baseline
Trend



Standard Test (95%): Compliance <UCL = 1.89e+000 ug/l>
Baseline Test (95%): No Change <UPL/LPL = 1.20e+001/0.00e+000 ug/l>
Trend Test (80%): Downward <Slope = −1.40e−001 log−ug/l/year>

Statistical Note: ND surrogate = 0.5 X Median of nondetects' PQLs

Run Date: 05−Jun−2007
Prepared by: US EPA, Region 5

APPENDIX F:
MAROS COMPARISON

Comparison of Summit Software and MAROS

(Based on Application at the Camp Allen Site

A secondary objective of this ESTCP project is to compare the recommendations and performance of the Summit Software with those of the Monitoring and Remediation Optimization Software (MAROS) for long-term monitoring optimization (LTMO) at one of the three Department of Defense (DoD) demonstration sites. Both software products were applied to the data for the Camp Allen site, which was the first site evaluated in the ESTCP project. This summary follows the structure of a White Paper that was submitted early in the project to describe the manner in which the two software products would be compared.

The italics represent the text from the White Paper, and the regular text represents the observations and comparisons. In places, the terms "Summit monitoring tools" and "Sampling Optimizer" used in the White Paper have been changed to "Summit Software".

How do the two software products compare with respect to user functionality and ability to implement site-specific objectives?

With respect to <u>user functionality</u>, comparisons will include the following:

- *GeoTrans, NAVFAC SE, and EPA will report on their perceptions of the comparative ease of learning the functionality of the two software products. GeoTrans personnel received training for both the Summit and MAROS tools by the software developers. The Summit training was a little more than a half-day, and the MAROS training was approximately a half-day hands-on training that was part of a two-day LTMO conference. The NAVFAC SE and EPA personnel received the same training as GeoTrans for the Summit Software, and the EPA personnel has previously received MAROS training.*

Both the Summit Software and MAROS were run without problems based on the training that was provided. However, understanding some of the options associated with the functionality required some limited further consultation with the software developers and/or some iteration running the software. Examples include the following:

- For the Summit Software, it took several iterations to determine which combination of interpolation technique and data transformation was better suited for the data. Similarly the default population size for the Sampling Optimizer module of the Summit Software was initially used, and no new monitoring plans with adequately small errors were identified. Consultation with the software developer led to using a larger population size, which gave monitoring plans with adequately small errors. Note that programming changes to the Summit Software that occurred subsequent to the analysis at the Camp Allen site would have likely mitigated this concern.

- For MAROS, there was some initial confusion with several issues. For instance, the units conversion created small decimal changes to the data. Also, it was not clear what to assign for flow parameters when actual conditions were non-uniform. Also, it was observed that the software produced false trends in some cases when all the data values were non-detect. These issues were resolved (or at least explained) by talking with the software developer. For instance, the software developer explained that the small decimal changes do not affect the statistical outcome, and the software developer confirmed that the software may report false trends (when all samples are non-detect) on the Linear Regression report and the Site Summary report, but that it would not impact the analysis of sampling frequency in the software.

Thus, we consider both software packages as "easy to use", but with the caveat that we had some questions regarding each software package that required quick clarification or guidance from the software developer. In addition, many modules are included in MAROS, and it may be a bit confusing for a "beginner" to understand the relationship between these modules and their use within the overall optimization process. On the contrary, Sampling Optimizer has fewer modules and has a more easily understood relationship between the modules.

Ultimately, the NAVFAC SE personnel did not apply either software due to personnel changes at NAVFAC SE, and USEPA did not apply the MAROS code to their example.

- *Relative ease of software installation on personal computers will be compared and reported.*

Both the Summit Software and MAROS are easy to install. In each case we were able to install the software within minutes. A license is required to activate and use the Summit Software. In addition, some Sun and Microsoft products (Sun Java JRE 6.0 or higher, Microsoft .net Framework Runtime 2.0 or higher) need to be installed prior to installing the Summit Software.

- *The relative ease of importing site data into the software, and the flexibility regarding input formats, will be compared and reported.*

The input formats for both software packages are very similar. MAROS uses Excel, MS Access, or ERPIMS file types for input. The Summit Software uses Excel CVS file type for input. In each case there were certain modifications to the data required, such as ordering the columns, changing the flags for NDs, adding detection limits, etc. These types of changes were no more significant for one software versus the other software; the user simply needs to follow the instructions regarding the input structure and requirements for that software product. These modifications took on the order of minutes to several hours for each software product for someone experienced with performing such operations in MS Excel or MS Access.

- *If either of the software products unexpectedly "freeze-up", or if any apparent software bugs are identified, it will be reported.*

MAROS did not have "freeze-up" issues. However, "freeze-up" was an issue during application of the Summit Software at the Camp Allen site, because the software frequently came to a near-halt with large population sizes, a parameter used in Sampling Optimizer module (e.g., population size of 800 for spatial analysis for the Camp Allen site). A simulation that usually took several minutes might end up running for several hours, and in some cases it was not clear whether the simulation would actually run to completion. It is important to note that the programming of the Summit Software was updated subsequent to the evaluation of the Camp Allen site, and the Summit Software "freeze-up" issue was not observed during analysis of the subsequent two demonstration sites. In the final testing version of each software, no bugs in either product were evident.

- *Features in each software package that are not available in the other will be identified and briefly described (e.g., MAROS includes a "Constituent of Concern (COC) Assessment", and Summit Software includes a "Data Tracker" module that identifies anomalous data in new sampling).*

MAROS and Summit Software are extremely different based on their logic, approach, and software structure. Key differences include the following:

  - The evaluation of data redundancy in the Sampling Optimizer (SO) module of the Summit Software uses mathematical optimization, which is unique relative to other LTMO software products. This allows sampling redundancy to be evaluated on a system-wide basis (e.g., best solution if one location is removed, if two locations are removed, if three locations are removed, etc.). A key benefit of this approach is that it allows the tradeoff between the number of samples and the accuracy of the resulting plume interpolation to be assessed. This is a significant improvement over the approach for evaluating data redundancy utilized in the Monitoring and Remediation Optimization Software (MAROS) software, which is not based on mathematical optimization. In MAROS, individual wells locations are evaluated for redundancy based on impacts of removing that well alone; consequently, the impact of removing groups of wells cannot be assessed and the aforementioned tradeoff cannot be evaluated.

  - The Summit Software approach to data redundancy evaluation provides plume visualizations for the baseline plan (i.e., all samples) versus improved plans (i.e., reduced numbers of samples) within the software. These comparative visualizations are quite effective for communication with stakeholders and regulators. MAROS does not include such plume visualizations.

- MAROS uses a Delaunay Triangulation approach for spatial interpolation and statistical trend analysis for temporal evaluation. Thus, analysis in MAROS is either spatial or temporal, but not spatiotemporal. The Summit Software includes the functionality to do spatiotemporal redundancy analysis, though it appears to yield more conservative results than the spatial redundancy analysis and our project team concluded that spatiotemporal analysis with the Summit Software may not be the preferred evaluation approach given the conservative results and increased computation time.

- MAROS results suggest an "optimal plan" which may be more straightforward for the user than the redundancy results from the Summit Software which consists of many potential sampling plans along a tradeoff curve. On the other hand, the EPA Region V team indicated that the trade-off curve approach implemented in the Summit Software was a major benefit. Thus, there is a tradeoff between the simplicity of one recommendation and the flexibility provided by a family of optimal solutions. Both software products require qualitative review of the results.

- MAROS includes a COC Assessment module and Data Consolidation module. The COC Assessment Module is to help the user choose COCs to evaluate based on risk evaluation, and choice of Preliminary Remediation Goals (PRGs). It has some regulatory criteria built in, but those may be superseded by user-defined values. It allows evaluation for up to 5 COCs based on their toxicity, prevalence, and mobility. The software generates a report regarding COC assessment. The software also allows the user to specify COCs to use for further MAROS evaluation regardless of the results of this module. The Data Consolidation Module provides the user options to handle non-detects. Options include using detection limits imported with the data, using half of the detection limits, using a fraction of the detection limits, or using user-specified detection limits that overrides the imported detection limits. This module can also consolidate duplicates (e.g., average, maximum, etc.) and perform time consolidation (e.g., quarterly, yearly, etc.). The Summit Software does not offer these types of functions.

- The Summit Software has a Data Tracker module that indicates if new data are "in-bounds" or "out-of-bounds" relative to expectations, based on previous data at that well. This functionality is not present in MAROS.

- MAROS indicates if the concentration trend at a well is increasing, decreasing, or stable. This functionality is not present in the Summit Software.

- *The degree to which trend analysis features in MAROS are similar to or different from the Data Tracker features in the Summit Software will be discussed. For instance, MAROS analyzes the whole data history for one COC at a well to see if there is evidence for a trend. The Summit Software, by contrast, evaluate new data to see if they are consistent with historical data, whether or not there is a trend.*

As mentioned above, the functions provided by each software package are different with respect to trend analysis. The Summit Software has a Data Tracker module that indicates if new data are "in-bounds" or "out-of-bounds" relative to expectations, based on previous data at that well, and that functionality is not present in MAROS. On the other hand, MAROS indicates if the concentration trend at a well is increasing, decreasing, or stable, and that functionality is not present in the Summit Software. Both functions are useful, and neither software offers both types of functionality. Also, in MAROS the quantitative sampling frequency analysis provides a recommendation based on the overall history and on the recent data. In the Summit Software, the user must select which data for each well to import as "historical data" for developing the prediction limits (i.e., the whole history or a portion of the history).

- *Any differences in the minimum sampling requirements (space and time) required to apply the software will be described.*

For MAROS, at least 4 sampling events for a specific well is required for the Trend Analysis module to determine a concentration trend result. Both the Sampling Frequency module and Data Sufficiency module recommend 6 or more sampling events.

Sampling Optimizer requires 15 or more (20 or more is preferable) sampling locations for a specific sampling event for spatial or spatiotemporal analysis. Data for at least 4 sampling events is required for a specific well for spatiotemporal analysis and for Data Tracker.

- *The degree to which the software use and functionality are described in a user's guide, and the quality of the user's guide, will be compared and reported.*

Both User's Guides are comprehensive and clearly presented.

- *The ability of the software to provide visualization of results and/or export results to other commonly used formats will be evaluated.*

MAROS provides nicely formatted and well organized graphs and reports pertaining to each step of the software application. However, very simple Excel figures are generated to indicate trend analysis results and uncertainty areas. No plume visualization is generated

within MAROS. To make a more presentable figure for a report, the user needs to have a base map, and software that can make category-type figures (e.g., Surfer, ArcGIS, AutoCAD). The MS Access-based output can easily be modified and exported so that it can be imported into a Surfer type of software (or equivalent) to make plume visualization figures.

The Summit Software provides plume distribution maps as well as uncertainty maps for direct visualization. It also provides side-by-side comparison of plume maps of potential plans with plume maps with all data points. For these maps, the software gives users the option to post either well names, measured concentrations, or "interpolated" concentrations at the removed locations. These are extremely useful maps. However, these maps can only be exported as image files, thus, it is difficult to directly import these image files into other software packages such as Surfer, ArcGIS, and AutoCAD.

The Summit Software allows the user to easily save a project and re-open it later, which is very useful. As for MAROS, after beginning a series of analyses, the user cannot archive any results until all those analyses are performed. However, reports can be saved after various steps, and an advanced MS Access user can presumably access the data tables and export them from within the code to another database at any time. Because all functions, code, queries and tables are open, the user can pull out data at any time and save them to another file.

With respect to <u>incorporating site-specific objectives</u> for evaluating data redundancy, the Summit Software allows for some flexibility to define and customize optimization objective functions and constraints. For instance, the Summit Software allow the user to specify multiple objectives (e.g., minimize number of wells sampled and minimize the error that results from removing sampling points) and provide a tradeoff curve that balances one objective versus the other. The Summit Software also allows the user to specify sampling locations that cannot be eliminated. The MAROS tools use a different approach based on a prescribed sequence of analyses (e.g., trend analysis, moment analysis), followed by reference to a "decision matrix" for suggested sampling locations and a "decision map" for suggested sampling frequency. The comparison between the two software products will include the following:

- *Can both software products reasonably address the site-specific goals and constraints expressed by the Installation and/or site personnel? If not, how do the required simplifications differ*?

Sampling Optimizer in the Summit Software incorporates site-specific goals and constraints. The site-specific goals are incorporated into the objective functions via parameters such as the "cutoff value" which delineates low concentration areas from high concentration areas. The error is calculated differently in each area, such that interpolation errors are weighted more heavily in areas of low concentration. Also, multiple COCs can be evaluated simultaneously if desired in Summit Sioftware. The "Well Constraint" feature allows the

user to specify the maximum sampling frequency and the minimum sampling frequency for each well. For spatial analysis, this feature allows the user to specify which wells cannot be removed from the system. This feature also allows the user to use abandoned wells in the development of the "model" but to make sure those abandoned wells are not included in the sampling plans that are recommended.

MAROS does not incorporate these types of site-specific goals, which instead are generally evaluated along with other information during qualitative review of the results from MAROS to determine the final recommendations. MAROS does not allow the user to specify the minimum and/or maximum sampling frequency for each monitoring well, but the user can specify that one or more wells not be completely eliminated from the sampling plan.

- *How much effort is required to implement the use of site-specific objectives and constraints in the Summit Software?*

The input for parameters associated for the multiple objectives within the Sampling Optimizer of the Summit Software is extremely straightforward and requires very little effort (i.e., minutes). Of course, assessing the sensitivity of the results to those inputs (e.g., changing the cutoff value for the error calculation) would require multiple optimization runs. The Summit Software currently evaluates the tradeoff between two general types of objectives (cost and error). This tradeoff was of interest at all three demonstration sites.

Given the same data and site objectives, do the two software products agree with respect to fitted models and uncertainties? If not, can one determine why differences exist?

With respect to fitted models, the Summit Software builds a statistical or geostatistical model using either kriging or inverse distance weighting, and these models can be spatial, temporal, or spatiotemporal. The MAROS tools use a Delaunay Triangulation approach for spatial interpolation and trend analysis for temporal evaluation. Thus, analysis in MAROS is either spatial or temporal, but not spatiotemporal. Therefore, with respect to fitted models, the two packages are not directly comparable. One potential method for comparing the representation of the data in each software product would be to compare "plume visualizations" that illustrate such items as peak concentration, plume boundary shape and location, and concentration gradients. However, whereas the Summit Software provides a visualization of the modeled plume, the MAROS tool does not. Another potential comparison would be based on mass metrics provided by both the Summit and MAROS tools. MAROS provides measures of total mass, center of mass, and distribution of mass. The Data Tracker in the Summit Software can track new data relative to user-specified mass metrics, e.g., expected reductions in mass (the mass calculations are provided by the Model Builder module of the Summit Software). Comparisons will include the following:

- *Do the two packages provide similar estimates of plume mass? If not, can the differences be explained*?

Tracking of plume mass was not performed with Summit Software for the Camp Allen site (it was not part of the optimization formulation). Thus, the comparison cannot be made. It is important to note that calculations of plume mass in the Summit Software, if performed, would vary based on the model selected in Model Builder (consisting of the selected interpolation technique and data transformation).

- *Are changes in estimated mass over time provided by two software packages similar? If not, can the differences be explained?*

Even though a direct comparison was not performed for the Camp Allen site, a key limitation regarding mass tracking over time with the Summit Software was identified during the analysis of one of the other demonstration sites. This limitation pertains to the case where different events have different sampling distribution, which will generally be the case at most sites (due to different sampling frequencies, new wells, abandoned wells, and wells that cannot be sampled in specific events due to logistics). The software does not perform temporal interpolation or extrapolation to fill in missing values in events where specific wells are not sampled. As a result, mass or mass flux results will have higher variability and uncertainty for events with fewer samples, different spatial distribution of samples, and/or for events where key wells (e.g., wells with high concentrations) are not sampled. However, the calculations of mass within MAROS have the same limitation, since MAROS also does not fill in missing values based on temporal interpolation. Thus, the two software products have the same limitations in this regard.

With respect to <u>uncertainty evaluation</u> (which pertains to potential need for additional sampling), both the Summit and MAROS tools provide information regarding areas of relatively high uncertainty that are potential candidates for additional sampling. However, the methodology differs between the two software packages. Comparisons will include the following:

- *Do the two packages provide similar regions with high uncertainty? If not, can the differences be explained*?

We compared the uncertainty maps from Model Builder of the Summit Software versus the labeled triangles of MAROS at the Camp Allen site. It seems like the MAROS results indicate a larger area of high uncertainty because the triangles are large due to the sparse data. It seems like Summit Software results show the higher uncertainty area limited more closely to the vicinity of the higher measured concentrations. This may be because MAROS normalizes the "error" by the maximum of the known or estimated concentrations.

- *Are there differences in the manner in which the software suggests lowering the uncertainty (i.e., are specific new sampling data locations/frequencies suggested by either software)? If so, do the suggestions differ*?

MAROS recommends general levels of monitoring and may specifically recommend more frequent sampling for some wells in areas of higher uncertainty. However, neither software package provides specific recommendations (i.e., number of new wells and/or locations of new wells) to reduce the uncertainty. It is left to the user to evaluate the uncertainty results provided and identify additional sampling locations. This is a substantial limitation of both software packages.

Given the same data and, to the extent possible, the same sets of site-specific optimization goals and constraints, do the two software products agree with respect to recommendations regarding monitoring optimization? If not, can one determine why differences exist?

With regard to data redundancy analysis, the following results from the implementation of the two software packages will be compared:

- *The number of sampling points (in space) suggested to be removed*.

This is quite complicated to compare. For spatial analysis in the Sampling Optimizer module of the Summit Software, it does not seem that a direct comparison to MAROS is appropriate because the underlying "baseline" dataset for the spatial analysis in Sampling Optimizer uses a larger dataset (e.g., 42 wells for the Shallow Aquifer) that is a combination of data points measured over 5 years (2001 to 2005). MAROS started with a longer data set (starting in 2000 rather than 2001) but many locations were excluded from the "baseline" set it considers because there were not enough samples for many individual wells (i.e., MAROS starts with 12 shallow A wells and 22 shallow B wells for a total of 34 shallow wells, versus 42 shallow wells used in Sampling Optimizer). So immediately comparisons are difficult.

We can compare the spatial results from Sampling Optimizer with the well redundancy results from MAROS for the deep aquifer since the same wells are used in both software packages. There are no wells recommended to be removed from the deep aquifer based on MAROS results. As for Sampling Optimizer, no new monitoring plans were identified within the acceptable range of errors (i.e., no wells recommended to be removed) when two abandoned wells are excluded from future plans. However, 7 to 10 wells were identified to be removed for the selected plans if these two abandoned wells are replaced. Again, comparing the results is difficult because we cannot evaluate the case where the abandoned wells are in the baseline, but definitely removed from the future plans, with MAROS (which is our base case with Sampling Optimizer).

When comparing the spatiotemporal results from Sampling Optimizer (which we could only do for the deep aquifer due to data limitations for the Summit Software for the shallow aquifer) with the well redundancy analysis results from MAROS for the deep aquifer, there are no wells recommended to be removed by MAROS, and there are either no wells or two wells

recommended to be removed by Sampling Optimizer depending on which plan we select. However, there are only 21 wells used for spatiotemporal analysis for Sampling Optimizer while there are 31 wells used for well redundancy analysis in MAROS (21 Area A deep wells and 10 Area B deep wells) due to data requirements of Sampling Optimizer. Again, this makes direct comparison of results very difficult.

- *The similarity or difference in the portions of the plume where reductions in sampling locations are suggested*.

There is only one well recommended to be removed from the shallow aquifer by MAROS, and the same well along with 24 other wells are recommended to be removed for the selected plan by spatial analysis with Sampling Optimizer. As discussed previously, we are not sure these results are easily compared. Adding to the complication is that other plans along the tradeoff curve provided by the Summit Software could also be evaluated.

For the deep aquifer, MAROS recommended no wells to be removed. Spatial analysis with Sampling Optimizer did not recommend any well to be removed. But if two abandoned wells are replaced, 7 to 10 wells were identified to be removed by Sampling Optimizer for the selected plan.

- *The recommended sampling frequency at each existing sampling location*.

The current sampling frequency at Camp Allen is annual sampling. MAROS can recommend more frequent sampling for wells with insufficient data or with increasing trends (e.g., semi-annual or quarterly sampling). In Sampling Optimizer, users can specify the minimum and maximum frequencies for each well (e.g., users can define quarterly sampling as the maximum frequency). But we decided to keep annual sampling as the maximum frequency for all the wells (i.e., no wells can sample more frequent than annually).

As discussed earlier, spatiotemporal analysis could not be performed for the shallow aquifer by Sampling Optimizer due to insufficient data to conduct a spatiotemporal simulation. For the deep aquifer, only 21 wells were analyzed spatiotemporally in Sampling Optimizer while 31 wells were analyzed for frequency in MAROS. The table below summarizes the recommended sampling frequency at each well for the deep aquifer.

Area A Wells	***MAROS***	***Sampling Optimizer***	***Area B Wells***	***MAROS***	***Sampling Optimizer***
A1-MW4B	*Annual*	*N/A*	*B-MW2B*	*Annual*	*N/A*
A1-MW6B	*Annual*	*Annual*	*B-MW2C*	*Biennial*	*N/A*
A1-MW9B	*Quarterly*	*Annual*	*B-MW3B*	*Annual*	*Annual*
A1-MW10B	*Annual*	*Annual*	*B-MW5B*	*Quarterly*	*N/A*
A1-MW24B	*Quarterly*	*N/A*	*B-MW11B*	*Annual*	*Annual*
A1-MW25B	*Semi-annual*	*Annual*	*B-MW15B*	*Biennial*	*Annual*
A1-MW31B	*Annual*	*Annual*	*B-MW19B*	*Quarterly*	*N/A*
A2-MW11B	*Annual*	*N/A*	*B-MW22B*	*Biennial*	*N/A*
A2-MW23B	*Annual*	*Annual*	*B-MW33B*	*Annual*	*Annual*
A2-MW28B	*Annual*	*Annual*	*B-MW34B*	*Annual*	*Annual*
A2-MW32B	*Annual*	*Biennial*			
A-MW1B	*Quarterly*	*N/A*			
A-MW13B	*Biennial*	*Annual*			
A-MW14B	*Biennial*	*Every 5 Year*			
A-MW15B	*Annual*	*Every 3 Year*			
A-MW16B	*Biennial*	*Annual*			
A-MW17B	*Annual*	*Annual*			
A-MW18B	*Annual*	*Annual*			
A-MW19B	*Semi-annual*	*N/A*			
A-MW30B	*Annual*	*Annual*			
A-MW31B	*Biennial*	*Biennial*			

There is some similarity in that, of the four wells recommended by Sampling Optimizer to have frequency greater than annual, two of them also had sampling frequency greater than annual recommended by MAROS. Also, annual frequency was suggested for a majority of the wells by both packages. However, it is really not easy to make any direct conclusions from the comparison.

- *How do the recommended sampling plans compare with respect to the site-specific objectives*?

No simple comparison is evident.

Based on these comparisons, an overall discussion will be presented regarding the extent to which the recommended sampling plans resulting from two software packages are similar or different. Feedback will be solicited from the site personnel to determine if recommendations from one software package are more satisfactory to them than the other.

GeoTrans personnel (and other members of the project team) will attempt to explain the causes of any differences in results obtained using the two software products. The causes might involve differences in modeling approaches or optimization algorithms. Ascertaining these causes may

require review of the software documentation and/or discussions with the developers. To the extent possible, GeoTrans personnel (and other members of the project team) will attempt to describe the strengths and limitations of the two software products and provide suggestions regarding applications appropriate for and not appropriate for each.

These issues have been addressed to the extent possible in the discussion provided above. Both software products are easy to use and are well documented. The primary advantage of the Summit Software is that the redundancy evaluation is based on mathematical optimization which allows sampling redundancy to be evaluated on a system-wide basis (e.g., best solution if one location is removed, if two locations are removed, if three locations are removed, etc.). A key benefit of this approach is that it allows the tradeoff between the number of samples and the accuracy of the resulting plume interpolation to be assessed. This is a significant improvement over the approach for evaluating data redundancy utilized in MAROS, which is not based on mathematical optimization. In MAROS, individual wells locations are evaluated for redundancy based on impacts of removing that well alone; consequently, the impact of removing groups of wells cannot be assessed and the aforementioned tradeoff cannot be evaluated.

Also, the Summit Software approach to data redundancy evaluation provides plume visualizations for the baseline plan (i.e., all samples) versus improved plans (i.e., reduced numbers of samples) within the software. These comparative visualizations are quite effective for communication with stakeholders and regulators. However, these maps can only be exported as image files, thus, it is difficult to directly import these image files into other software packages such as Surfer, ArcGIS, and AutoCAD. MAROS does not include such plume visualizations.

The Summit Software has a Data Tracker module that indicates if new data are “in-bounds” or “out-of-bounds” relative to expectations, based on previous data at that well. This functionality is useful but is not present in MAROS. MAROS indicates if the concentration trend at a well is increasing, decreasing, or stable. This functionality is also useful, but is not present in the Summit Software.

Both software products suffer from some similar limitations. With respect to areas of uncertainty, neither software package provides specific recommendations (i.e., number of new wells and/or locations of new wells) to reduce the uncertainty. With respect to mass calculations, neither software performs temporal interpolation or extrapolation to fill in missing values in events where specific wells are not sampled. As a result, mass or mass flux results will have higher variability and uncertainty for events with fewer samples, different spatial distribution of samples, and/or for events where key wells (e.g., wells with high concentrations) are not sampled.