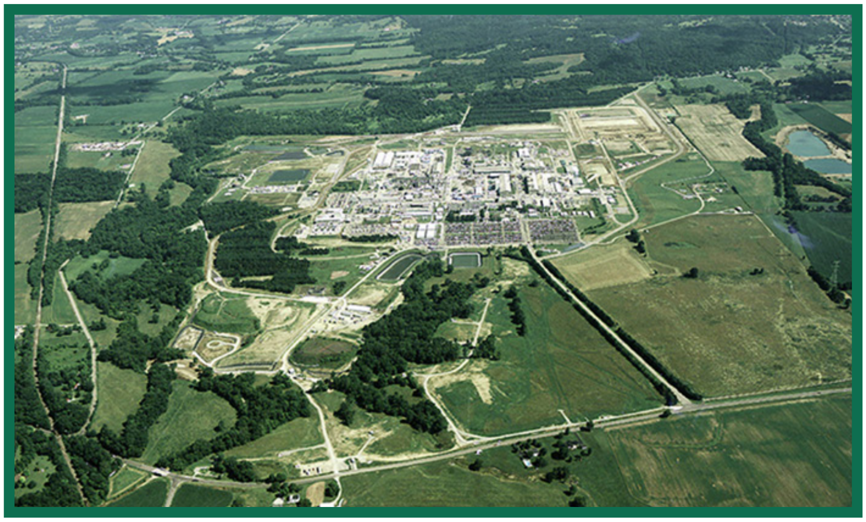
# ESTCP
# Cost and Performance Report

**(ER-200714)**



## Demonstration and Validation of the Geostatistical Temporal-Spatial Algorithm (GTS) for Optimization of Long-Term Monitoring (LTM) of Groundwater at Military and Government Sites

**August 2010**

ENVIRONMENTAL SECURITY
TECHNOLOGY CERTIFICATION PROGRAM

U.S. Department of Defense

# COST & PERFORMANCE REPORT
Project: ER-200714

## TABLE OF CONTENTS

Page

**TABLE OF CONTENTS (continued)**

**Page**

## LIST OF FIGURES

## LIST OF TABLES

**Page**

## ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| 1,1,1-DCE | 1,1-dichloroethene |
| 1,1,1-TCA | 1,1,1-trichloroethane |
| 2,4-DNT | 2,4-dinitrotoluene |
| | |
| AFB | Air Force Base |
| AFCEE | Air Force Center for Engineering and the Environment |
| AFP44 | Air Force Plant 44 |
| | |
| BRAC | base realignment and closure |
| | |
| CAS | Chemical Abstracts Service |
| CDF | cumulative distribution function |
| CERCLA | Comprehensive Environmental Response, Compensation, and Liability Act |
| CES | cost-effective sampling |
| COC | contaminant of concern |
| CSM | conceptual site model |
| CV | coefficient of variation |
| | |
| DCDF | declustered cumulative distribution function |
| DERP | Defense Environmental Restoration Program |
| DoD | Department of Defense |
| DOE | Department of Energy |
| DPT | direct punch technology |
| | |
| ESTCP | Environmental Security Technology Certification Program |
| | |
| FUDS | formerly used defense sites |
| | |
| GIS | geographic information system |
| GTS | Geostatistical Temporal-Spatial |
| GUI | graphical user interface |
| GWTP | groundwater treatment plant |
| | |
| HHRA | human health risk assessment |
| | |
| IDE | interface development environment |
| ITRC | Interstate Technology & Regulatory Council |
| | |
| LTM | long-term monitoring |
| LTMO | long-term monitoring optimization |
| LWQR | locally weighted quadratic regression |
| LZ | Lower Zone |
| | |
| MCL | maximum contaminant level |

## ACRONYMS AND ABBREVIATIONS (continued)

| | |
|---|---|
| MDL | method detection limit |
| NOP | Nebraska Ordnance Plant |
| QLR | quantile local regression |
| RCRA | Resource Conservation and Recovery Act |
| RDX | hexahydro-1,3,5-trinitro-1,3,5-triazine |
| RL | reporting limit |
| RMSE | root mean squared error |
| ROI | return on investment |
| SAIC | Science Applications International Corporation |
| SGZ | shallow groundwater zone |
| TCE | trichloroethene |
| TNB | 1,3,5-trinitrobenzene |
| TNT | 2,4,6- trinitrotoluene |
| USEPA | U.S. Environmental Protection Agency |
| UZLU | Upper Zone lower unit |
| UZUU | Upper Zone upper unit |
| VOC | volatile organic chemical |

# ACKNOWLEDGEMENTS

This work was supported by the Environmental Security Technology Certification Program (ESTCP) of the Department of Defense (DoD) as part of Project ER-200714 and through the Air Force Center for Engineering and the Environment (AFCEE). Mr. Phil Hunter (AFCEE) served as the Principal Investigator (PI). Dr. Kirk Cameron (MacStat Consulting, Ltd.) authored the Geostatistical Temporal-Spatial (GTS) Algorithm and served as the lead scientist and programmer. Project management and software development support was provided by Science Applications International Corporation (SAIC). SAIC personnel included Robert B. Stewart and Michael J. Kenny.

The PI wishes to acknowledge all parties who provided site data and testing of the GTS software including Robert Johnson, Dave J. Becker, and Jon Atkinson.

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

- Air Force Plant 44 Site, Tucson, AZ (AFP44 site)
- Former Nebraska Ordnance Plant Site, Mead, NE (NOP site)
- Fernald DOE Site, Ross, OH (Fernald site)

Their time and efforts are greatly appreciated.

*This page left blank intentionally.*

# 1.0 EXECUTIVE SUMMARY

## 1.1 OBJECTIVES OF THE DEMONSTRATION

The primary objective of this ESTCP project was to demonstrate and validate use of the Geostatistical Temporal-Spatial (GTS) groundwater optimization software, developed by MacStat Consulting and Science Applications International Corporation (SAIC) for, and under the auspices of, the Air Force Center for Engineering and the Environment (AFCEE), at three DoD and Department of Energy (DOE) sites. The three demonstration sites were as follows:

- Air Force Plant 44 Site, Tucson, AZ (AFP44 site)
- Former Nebraska Ordnance Plant Site, Mead, NE (NOP site)
- Fernald DOE Site, Ross, OH (Fernald site)

## 1.2 TECHNOLOGY DESCRIPTION

The GTS software demonstrated in this ESTCP project offers a set of tools for long-term monitoring optimization (LTMO) and consists of five major modules:

- **Prepare** imports analytical and water-level data, imports site boundaries and shape file overlays, and enables data management via (a) an internal SQLite database, (b) creation of analysis variables, and (c) identification of outliers.

- **Explore** allows for basic statistical exploration via data summaries and graphs, analysis and ranking of contaminants based on optimization potential, and identification and analysis of multiple vertical aquifer horizons.

- **Baseline** displays initial groundwater monitoring network status, fits nonlinear baseline trends via locally weighted quadratic regression (LWQR), displays trend maps, builds spatial models via bandwidth selection, computes and displays potentiometric surfaces, and constructs and displays concentration-based plume base maps using quantile local regression (QLR).

- **Optimize** allows for both temporal and spatial optimization. Temporal optimization in GTS consists of two components: (1) *temporal variograms* applied to groups of wells and (2) *iterative thinning* of individual wells. More than one temporal optimization method allows for flexible handling of the kinds of data available at different installations. Spatial optimization within GTS consists of (1) searching for statistical redundancy via mathematical optimization using the GTSmart algorithm; (2) determining optimal network size with the aid of cost-accuracy trade-off curves; and (3) assessing whether new wells should be added and where (i.e., *network adequacy*).

- **Predict** allows import and comparison of new sampling data against previously estimated trends and maps. Two options include trend flagging and plume flagging to identify potentially anomalous new values.

To support the Optimize module, GTS also includes a separate stand-alone Excel spreadsheet Cost Comparison Calculator to realistically calculate the financial benefits of implementing a GTS-optimized sampling program, as well as return on investment (ROI).

Some of the advantages of the v1.0 release of GTS demonstrated in this project include the following:

- Substantial projected annualized and life-of-project cost savings from implementing a GTS-optimized program, in the range of 30 to 60%. ROI for a GTS-optimized monitoring program is generally 1 to 2 years or less.
- Equally applicable to site-specific plumes and unit-wide or base-wide studies involving multiple source areas, plumes, and monitoring conditions. This is because GTS does not require or utilize plume-specific configuration data, fate-and-transport models, or other hydrogeologic modeling information.
- Innovative exploratory tools for assessing data characteristics, ranking contaminants of concern (COCs) for optimization potential, and analyzing multiple aquifer horizons. These tools can also assist in identifying and developing anthropogenic or background data sets.
- Sophisticated built-in graphics for data visualization, including contour mapping, complex trends, post-plots, and shape file annotation.
- Trend estimates derived from LWQR, allowing for fitting of complex and/or seasonal time series data. All other currently available LTMO tools only offer fitting of linear trends, an assumption that does not match the reality of most long-term monitoring (LTM) data sets.
- Semi-nonparametric surface map estimates made using QLR, a smoothing technique not bound by the constraints of kriging. By design, QLR is made to handle skewed data sets as well as significant proportions of non-detects, data features ubiquitous to LTM networks.
- Automated redundancy searches employing mathematical optimization, both during temporal and spatial analyses. Spatial optimization is performed with a quasi-genetic algorithm unique to GTS, known as GTSmart.
- Use of multiple cost-accuracy trade-off curves to gauge points of optimality. Defensible bias measures of statistical accuracy allow for rigorous analysis of potential trade-offs.

## 1.3 DEMONSTRATION RESULTS

Key results of the project are listed below:

- The GTS software was found to be easy to use and navigate by the testers and mid-level site analysts, even though none of these users was formally trained on the software. Because GTS v1.0 represents a major overhaul and upgrade to the previous beta version, with a software architecture that was completely redesigned, a significant number of software bugs, logic flaws, and glitches were encountered during both internal and external testing of the software. By the end of project, no significant bugs or software errors remained.

- Graphical outputs in GTS were found to be quite helpful and attractive to users. These, combined with the unique exploratory data tools built in the software, were rated as one of its strong points.
- GTS was found to be effective as an optimization tool. Significant degrees of redundancy were identified at each demonstration site. The iterative thinning function recommended reductions in sampling frequency ranging from 50 to 75% across the three demonstration sites, while the GTSmart algorithm found degrees of spatial redundancy ranging from 16 to 40%. Further, GTS was run successfully at larger sites having more than 200 distinct well locations.
- Of the temporal optimization tools, iterative thinning was found to be superior in performance to temporal variograms. The variograms were easily computed, but yielded poor to mixed results. Overall, the results did not enable reliable or replicable estimates of optimal sampling intervals, since few variogram ranges (denoting points of optimality) could be identified at the test sites.
- A goal of this project was to enable users to perform water-level-aided spatial mapping as an option in GTS. Internal testing of this feature led to mixed results and a decision not to include it in v1.0 of the software. However, as a by-product of this testing, GTS now includes the ability to create potentiometric surface maps of groundwater levels. Users found this to be a useful tool and visualization feature.
- When the input data sets were essentially equivalent, GTS optimization results were shown to be highly reproducible when comparing results from expert users and independent mid-level analysts. Except for the Fernald site, where the input data sets substantially differed, the optimized sampling intervals were identical on a site-wide basis at the other demonstration sites and differed only slightly when broken down by aquifer zone. Spatially, the levels of redundancy found using the same COCs were very similar at both the AFP44 and NOP sites. Further, a locational analysis of which wells were flagged as redundant showed statistically significant similarity in common locations and spatial proximity.
- The trend and plume flagging tools in GTS were shown to be reasonably effective in flagging potentially anomalous measurements from a reserved subset of data from each demonstration site. And, because the reserved data sets were collected “close in time” to the historical data—being observations from the next year of sampling—the projected (i.e., extrapolated) trends and plumes successfully predicted (i.e., bounded) over 90% of the new measurements. Nevertheless, the trend and plume flagging features may be too sensitive in flagging anomalies; further investigation indicated that perhaps only 30% of the trend anomalies and 65% of the plume anomalies were values actually deserving further investigation or verification.

- The network adequacy function successfully located areas of substantial mapping uncertainty at each demonstration site, and recommended coordinate locations for the siting of new wells. Because GTS cannot determine whether a suggested new location coincides with a physical obstruction or is unfeasible for other reasons, users were able to successfully override specific locations and to document those decisions visually on a post-plot of both existing and recommended locations.

## 1.4 IMPLEMENTATION ISSUES

Based on application of GTS v1.0 to the three demonstration sites during this project, the software has certain limitations that could be mitigated by future improvements. These include:

- GTS requires a number of input fields in ASCII text format in order to create a sufficient analysis database. Some users may find the directions for importing data and creating or augmenting databases within GTS more complicated than need be. The software would be improved if this process were streamlined and simplified.
- GTS does not offer sophisticated handling of radiochemical data, particularly measurements recorded with non-positive values (i.e., zeros or negatives). These data must first be converted to positive values, unless they represent non-detects with a known, positive detection or reporting limit. GTS could be improved by allowing a specific option for radiochemical data.
- Optimized sampling intervals from temporal variograms in GTS often do not match the optimized sampling intervals from iterative thinning using the same data. Further improvements to the temporal variogram algorithm may be needed, especially to account for sites with spatial trends that are actively changing over time.
- Cost-accuracy trade-off curves in GTS are not interactive. Although the bias limits can be adjusted by the user, the spatial optimization must be completely re-run each time those limits are changed, in order to see the impact of the revised limits and to generate a new optimal network. The software could be improved by combining the current trade-off curves into a single, weighted curve that would allow for interactive selection of different sampling plans by the user.
- There is no way in GTS v1.0 to batch print graphics. Since a GTS analysis typically generates a large number of statistical graphics, users may be frustrated with the inability to document graphical results outside the application. The software could be improved by enabling an option to do batch printing to popular image formats.
- The mathematical optimization algorithm in GTS is not a true genetic algorithm wherein portions of the binary string “DNA” representing alternate network configurations are allowed to “mate,” “mutate,” and create “offspring.” Instead, GTS does a “smart search” through the space of potential network configurations, selecting for testing only those strings with interwell spacing comparable to the

full network. The software might be improved by incorporating a true genetic algorithmic search.

- The Prepare module may identify too many data records as "outliers" at some sites, necessitating needless user review and override. GTS could be revised and streamlined by combining the temporal and spatial outlier searches into a single, improved algorithm that better accounts for local trend fluctuations.
- "Time slices" in GTS—discrete, non-overlapping periods of sampling—are computed automatically, but are not adjustable by the user. The software could be improved by allowing user input to define or adjust time slices to accord with site-specific remedial events or histories.
- The Predict module readily identifies anomalous future measurements but may be too sensitive in flagging anomalies. GTS could be revised with improved trend and plume flagging routines to better avoid flagging non-anomalous values.

The level of effort and computation time for applying GTS at the three demonstration sites are documented within this report, as well as a basis for estimating the costs of applying the software to other sites. Estimated cost-benefit analyses at each of the three sites are presented, along with projected ROI from implementing the GTS-optimized sampling plans. Estimated total cost savings compared to the baseline monitoring program ranged from 39 to 45%, with ROI ranging between 4 and 6 months. The specific well-by-well optimization recommendations computed by the ESTCP project team are listed in appendices to this report. A GTS users guide was finalized as part of this project and was submitted as a separate deliverable to ESTCP. The software and users guide are now available free for use by the public.

*This page left blank intentionally.*

# 2.0 INTRODUCTION

## 2.1 BACKGROUND

The Department of Defense (DoD) has invested over $20 billion in environmental restoration through the Defense Environmental Restoration Program (DERP) to address restoration needs at active installations, formerly used defense sites (FUDS), and in connection with base realignment and closure (BRAC). Across the agency, thousands of sites are engaged in long-term maintenance, remedial investigations, or groundwater cleanup.

Since groundwater contamination is common at DoD sites, large monitoring networks comprising dozens, hundreds, or even thousands of wells are in place at many facilities, as required for LTM by Resource Conservation and Recovery Act (RCRA) permits or under a Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) response. Frequently, the monitoring network has been installed either piecemeal or haphazardly over time, the result of changing goals and objectives, oversight by multiple contractors, changing subsurface conditions, and differing regulatory requirements. Relatively few sites have undergone a comprehensive optimization analysis, designed to identify an optimal network size and configuration, and to optimize the sampling plan and frequency of monitoring.

With moderate size or larger monitoring systems, there can be redundancy in the number and placement of wells (*spatial redundancy*) and inefficient frequency of monitoring (*temporal redundancy*). There is also a risk that portions of the site may be too sparsely sampled (*under-coverage*) to adequately assess or characterize subsurface conditions. Optimization of existing monitoring systems aims at improving their effectiveness and reducing overall site cleanup costs, without losing information *critical* to satisfying regulatory and monitoring objectives, site characterization, or to measuring remedial success.

Redundancy and optimality in this project are treated as *statistical* concepts. Redundancy is premised on what can be estimated with sufficient accuracy when existing data are removed from the current system. The remaining data (the *reduced-data set*) must be used to reconstruct features or characteristics that were estimated from the *full-data set*. This may include the reconstruction of temporal features such as trends when selected sampling events are eliminated, or spatial features like surface maps when selected wells are removed. Redundancy is defined as the ability of the reduced-data set to reconstruct the original trend or map within certain bounds on probable error. Forcing reproduction of the original trend or surface map guarantees that an overall characterization of the plume (and its rate of change) can likewise be reconstructed using the reduced data.

Of course, any measurement collected at a unique point in time and space provides some (statistical) information about the LTM network. Conversely, information is always lost when data are removed from the system. So judging an LTM network as “optimal” entails balancing a mathematical trade-off between this loss of information and the cost savings realized by not collecting, analyzing, and measuring the additional data. An optimized system is one that entails—compared to the current system—a minor loss of (statistical) information but a significant gain in cost savings.

Most current approaches to optimizing LTM network design typically rely on professional engineering judgment as opposed to statistical logic. Engineering-based approaches often involve "piece-wise" revamping of the monitoring network instead of a more objective statistical evaluation. Facilities may change subcontractors periodically, resulting in a patchwork quilt of LTM recommendations concerning well placement, network sufficiency, and sampling frequency. There can also be subtle pressure by contractors to justify and maintain LTM programs so as not to risk cuts in funding, as well as additional pressures by regulators not to reduce monitoring efforts for fear of losing vital data.

Due to these factors and the substantial costs associated with LTM, AFCEE has actively pursued testing of statistical optimization strategies for its LTM networks. The goal is to design a monitoring network able to capture necessary contaminant information—including the ability to meet DERP or regulatory objectives—but to do so at the lowest possible cost. One such strategy developed in coordination with AFCEE is the subject of this demonstration: the GTS statistical optimization software tool.

GTS is designed to mathematically optimize LTM groundwater networks. Version 1.0 of the software has five modular components linked together in a wizard-type user interface. These components enable the following key tasks:

1. Data summary and exploration, including identification of chemical constituents best suited for optimization, and analysis of multiple aquifer horizons (should they exist)
2. Estimation of nonlinear baseline trends and concentration-based surface maps
3. Temporal optimization of sampling frequencies and spatial optimization of the number and locations of wells
4. Identification of recommended locations for new wells, predicated on reducing mapping uncertainty
5. Tracking of new data against projected trends and concentration surfaces in order to flag potential anomalies, outliers, or recent plume changes.

GTS also includes a separate cost-benefit estimating tool designed to realistically quantify the potential savings and ROI achievable by implementing an optimized sampling program.

## 2.2 OBJECTIVE OF THE DEMONSTRATION

The primary objectives of this project included the following:

1. **To promote** widespread adoption of statistically based optimization efforts across DoD and government facilities involved in LTM, especially through the public release of GTS v1.0.
2. **To accelerate** the transfer and usage of GTS as a viable software technology to analysts and site managers desiring to physically optimize their LTM networks by improving and completing the user interface. This project will enhance the

functionality of GTS, improve performance, and make the tool more user-friendly for effective transition to potential users.

3. **To incorporate** as an automated feature simple, site-specific flow regime information into the GTS mapping capability by allowing the inclusion of water level data for one or more sampling events.

4. **To demonstrate** the applicability, usability, and effectiveness of an enhanced GTS software interface at sites representing multiple branches of DoD. The fully functional interface will be tested by the target audience—mid-level analysts with some statistical and geostatistical experience and a hydrogeologic background—to ensure that such analysts can arrive at similar optimization results to those generated when statistical or geostatistical experts evaluate the same data using the same software.

## 2.3 REGULATORY DRIVERS

There are no regulatory issues directly associated with this project, although the initial impetus for GTS was to more efficiently and cost-effectively meet regulatory requirements for LTM under both RCRA and CERCLA. 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 during LTM, which will ultimately address regulatory objectives and allow for improved communication between site stakeholders. Implementation of optimal sampling plans suggested for the demonstration sites is not within the scope of this project.

*This page left blank intentionally.*

# 3.0 TECHNOLOGY

## 3.1 TECHNOLOGY DESCRIPTION

GTS is a set of freeware, desktop software tools, designed to perform mathematical optimization of LTM groundwater networks. GTS allows any contaminated site with the minimum number of well locations (i.e., 20 or more for spatial optimization) and distinct sampling events (i.e., 6-8 per well location for temporal optimization) to quickly (i.e., within a few to several days after electronic data gathering and preparation) analyze and develop an optimal groundwater monitoring plan. Not only can these plans be periodically reviewed and updated over the life of the facility, but they also allow for efficient use of sampling resources, providing the necessary analytic and sampling data for good regulatory and remedial decisions, while simultaneously eliminating unnecessary, superfluous, or wasteful data collection and expense.

Given the minimal data requirements, any site undergoing LTM could potentially utilize the updated GTS software. This includes both larger and smaller sites due to the modular design of GTS and its ability to separately and independently optimize sampling frequencies and well locations.

The main GTS application (v1.0) consists of a set of five modules linked by a wizard-style graphical user interface (GUI). A schematic of the overall modular design is presented in Figure 1. The GTS distribution package also contains a separate Excel cost-benefit calculator spreadsheet for quantifying the resource savings achievable through implementation of a GTS-optimized sampling program.

The five modules in the main GTS application consist of Prepare, Explore, Baseline, Optimize, and Predict. All of these modules are built using open-source or license-free (to the user) runtime environments. R (www.project-r.org) is the statistical engine behind GTS, responsible for all statistical computations and estimates. The MatLab runtime environment (www.mathworks.com) is used to visually display maps, trends, and other statistical graphics. SQLite (www.sqlite.org) serves as an open-source database to house data imported into GTS and to store results. Finally, QT (http://en.wikipedia.org/wiki/Qt_(framework)) and C++ have been utilized to create the GUI with which users interact.

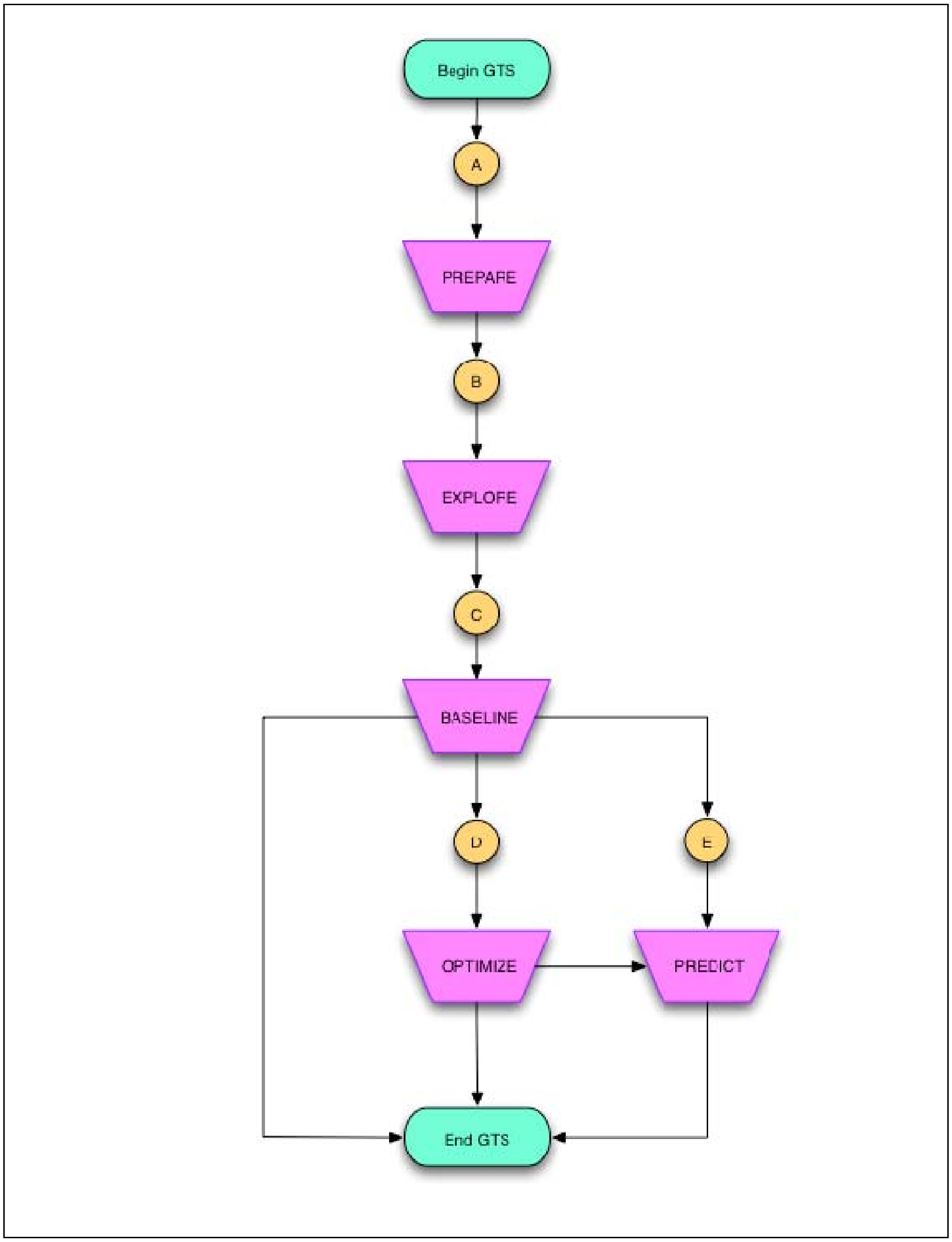



**Figure 1. Overall modular design of GTS.**

**Prepare Module**

The Prepare module enables data import and simple data checking. [More detail about the Prepare module and any other GTS functionality may be found in the GTS Users Guide, which has been provided as a separate deliverable for this project.] Users can view a simple map of the well network, import shapefiles as GIS-overlays for visual annotation, and check for outliers in the imported database (see Figure 2). Of some importance, GTS only uses existing site data for its analysis. No geophysical or hydrogeologic modeling is required or utilized. A spatial analysis usually requires at least 20 distinct wells to be useful, and a full temporal analysis requires at least 8 distinct sampling events of historical monitoring data per well. Other necessary information includes:

- Well ID and location
- Sample date
- COCs, concentration values, and reporting limits
- Screen depth, interval, aquifer zone
- Water level measurement data (optional)
- Geographic information system (GIS) data (Esri Shapefiles) to represent key features of the site (optional)

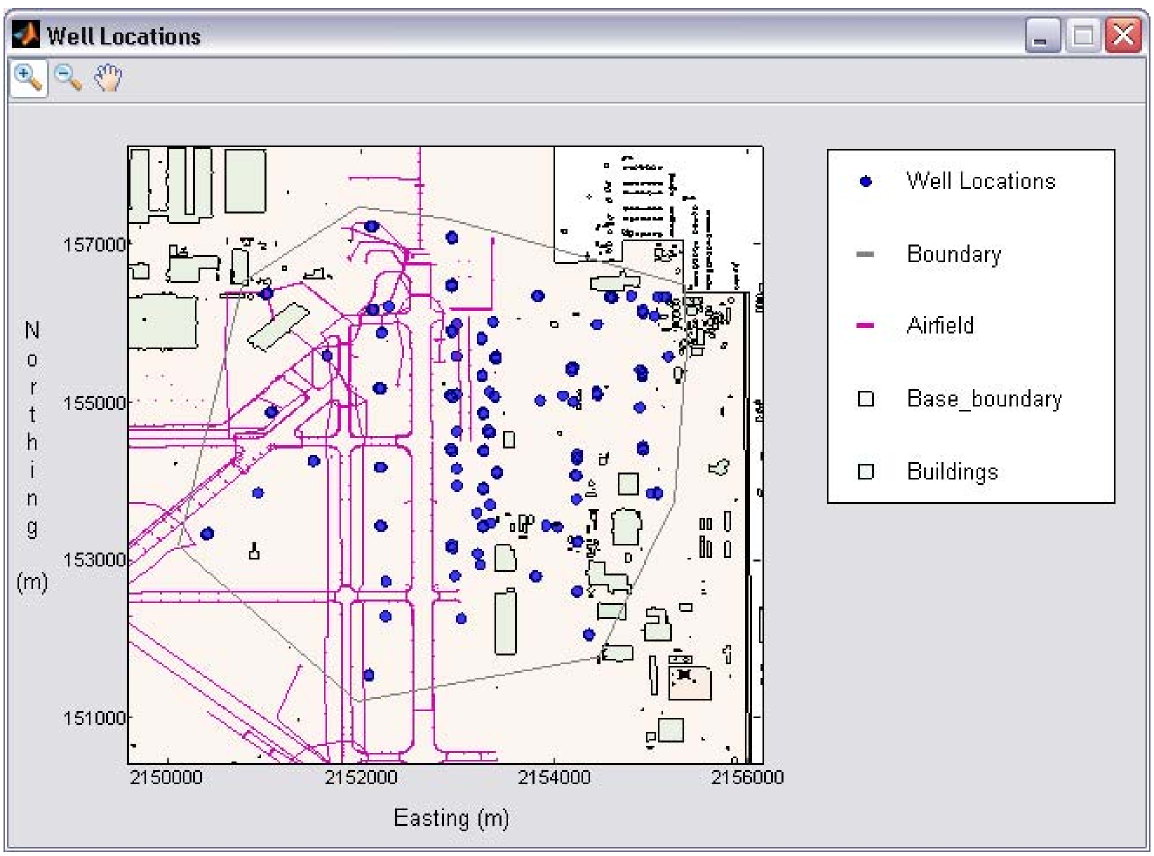



**Figure 2. Example of location map in GTS.**

GTS also creates a series of data-specific "time slices" in this module. Each time slice represents a kind of "snapshot" or "window of time" where, by default, a large majority of the distinct wells has been sampled. By analyzing a series of such snapshots, GTS assesses the degree of repeatability of its estimates of spatial redundancy; well locations are not classified as redundant unless they are redundant across a majority of the time slices, thus showing the results can be replicated over time. A schematic of logic and features of Module A is given in Figure 3.

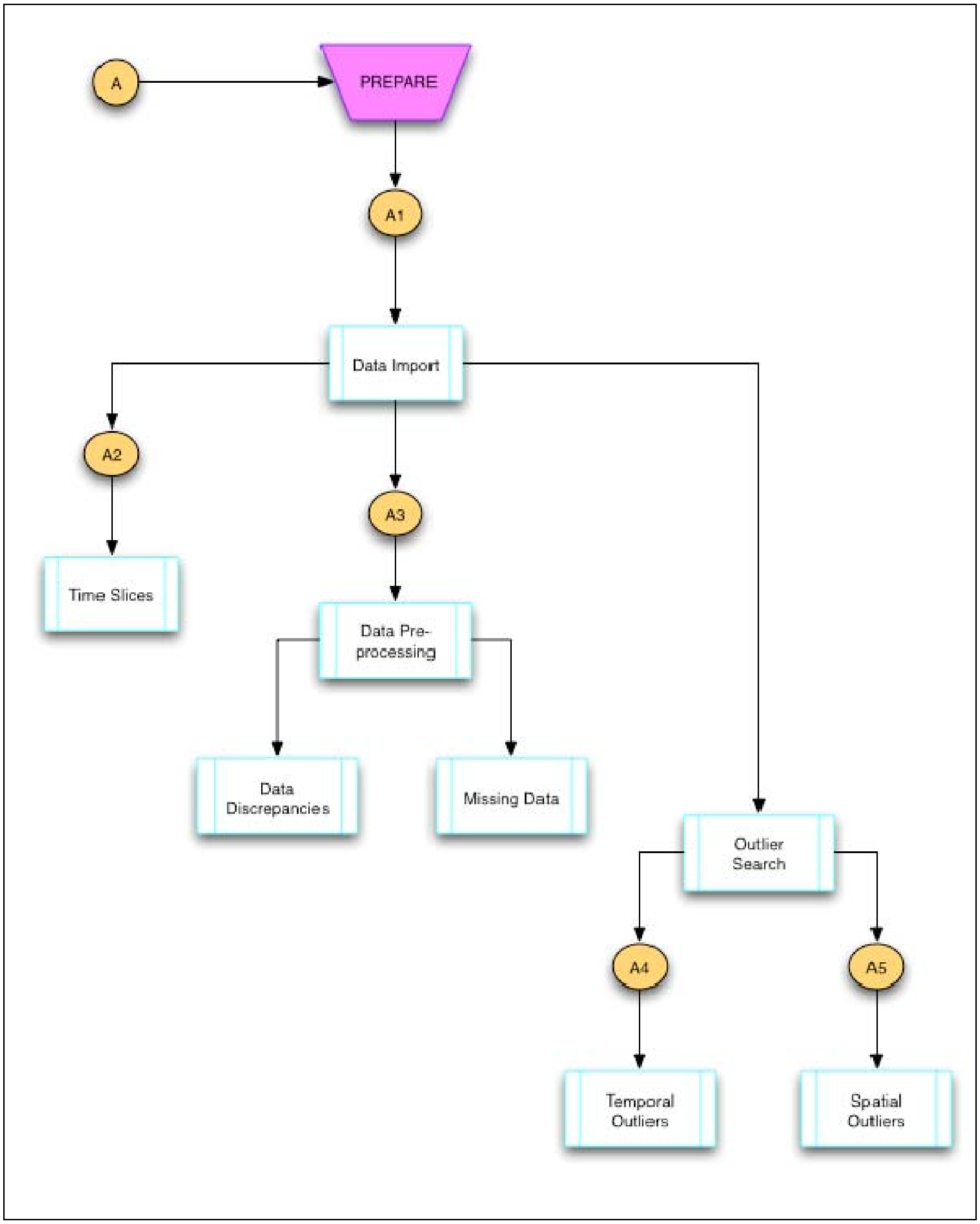



**Figure 3. Schematic of prepare (Module A) logic.**

## Explore Module

The second GTS module enables the user to prepare simple data summaries and to examine exploratory graphs. These tools can be used in their own right to gain a feel for data characteristics and/or data quality through visualization of time series plots of individual wells, side-by-side boxplots of COC-specific concentration levels, and post-plots of concentration hot spots or exceedances of regulatory levels (see Figure 4). An overview of the logic and features of Module B is given in Figure 5.

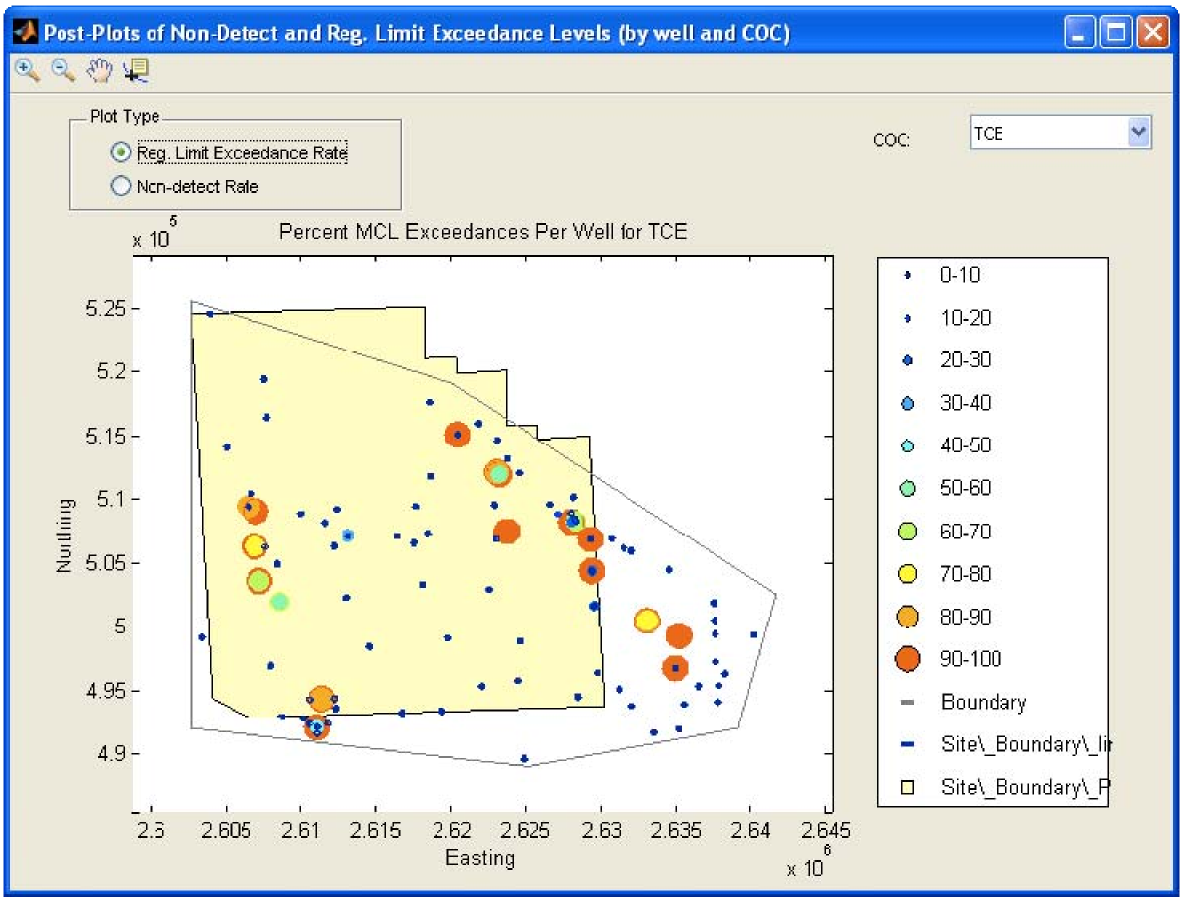



**Figure 4. Example post-plot of regulatory limit exceedances.**

The exploratory tools can also be used as part of a more extensive analysis to better prepare the data for optimization. GTS enables the user to rank COCs for optimization potential by examining frequency and location of detections and regulatory exceedances, toxicity and mobility factors, and key statistical indicators. Lower ranking COCs can then be excluded from further analysis. GTS also provides an analysis of vertical aquifer horizons. Horizon-specific variograms and boxplots can be examined to determine the degree of similarity in concentration levels and spatial correlation patterns. The user can decide to perform a simple 2D (i.e., two-dimensional) analysis, grouping all horizons into a single horizontal plane, or instead a 2.5D (i.e., "layer cake") approach, where each horizon is analyzed separately. Users can also delete or merge specific layers or horizons as needed.

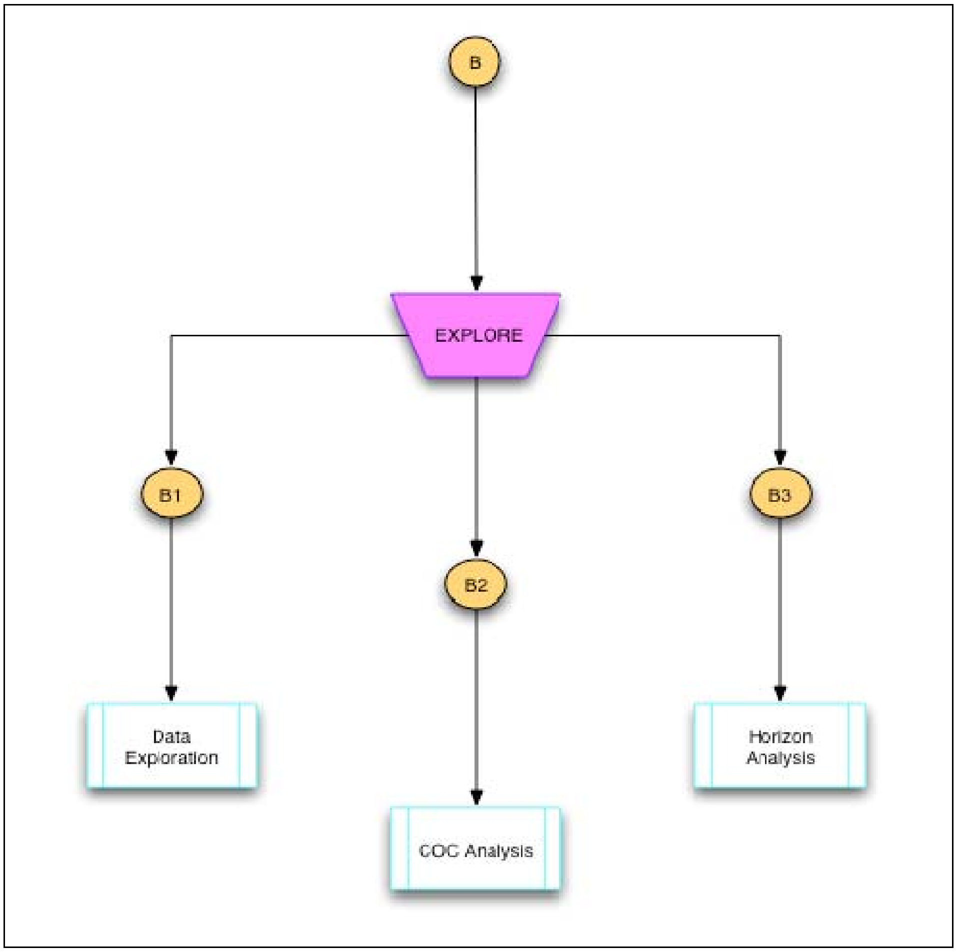


**Figure 5. Schematic of explore (Module B) logic.**

**Baseline Module**

As indicated in the introduction, GTS achieves optimization via an *empirical definition of redundancy*: sampling events and/or wells are redundant if trends and maps initially built with data from those locations or events can be accurately reconstructed without subsequently using them (that is, utilizing only more critical wells and events). To this end, a key step prior to any GTS optimization is to create baseline trends and/or base maps using the original data set in order to test the accuracy of reconstructions based on reduced-data subsets.

The Baseline module offers tools to construct such baseline trends and base maps. Like data exploration in GTS, these tools can be employed in their own right if a user does not necessarily need an optimization but merely wants documented estimates of temporal trends and/or maps of plume extent for each time slice. An overall schematic of the logic and features of Module C is given in Figure 6.

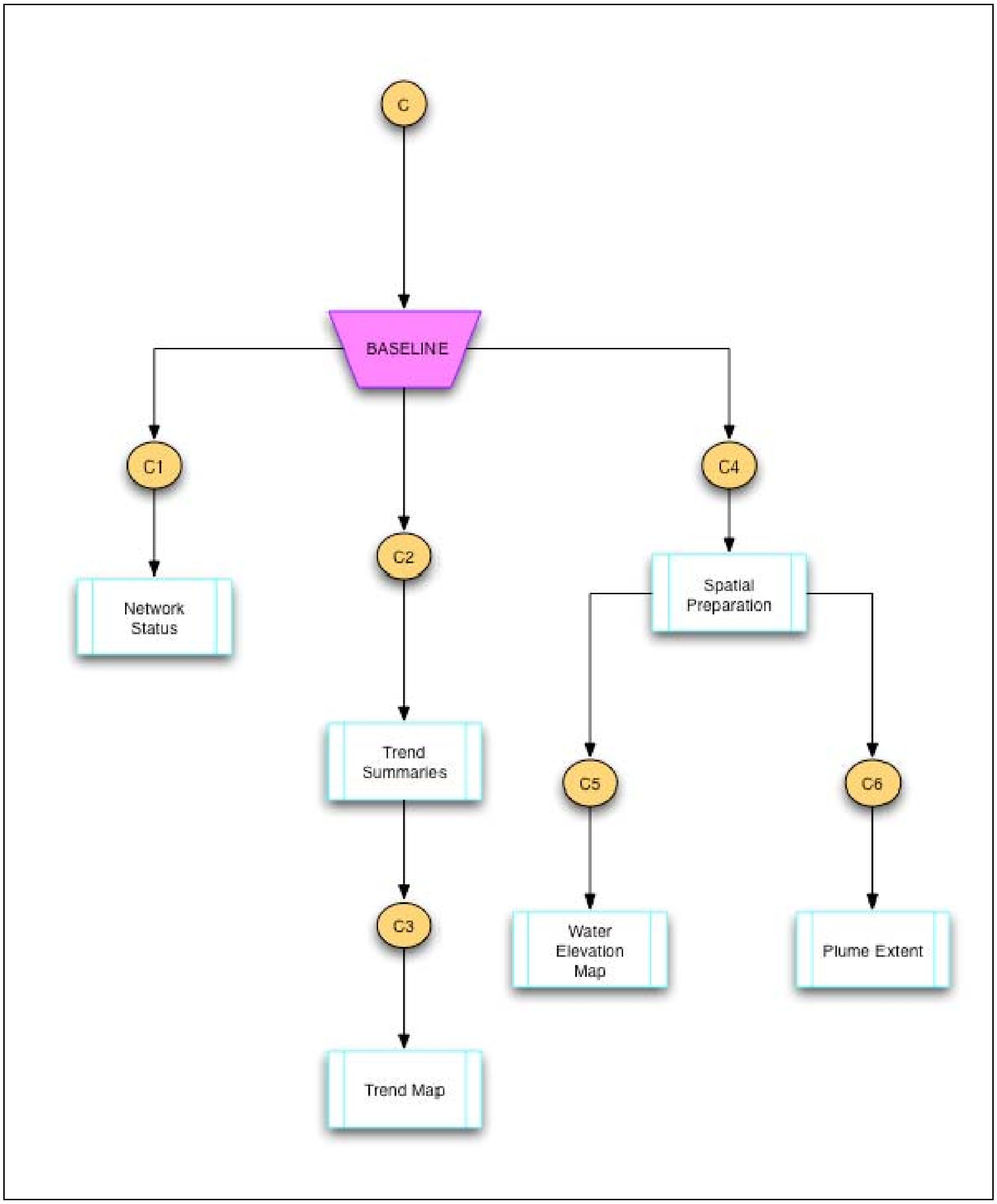



**Figure 6. Schematic of baseline (Module C) logic.**

Trends are estimated in GTS via a type of local regression known as LWQR, as it can readily fit complex and/or seasonal trends along with confidence bounds around those trends. LWQR constructs an estimate at any point (in time) *x* as a weighted average of the sample measurements in a local neighborhood surrounding *x*. Local regression enjoys several optimal properties as a statistical technique and several practical benefits: (1) it is inherently nonlinear and thus capable of describing trends that are actively changing; (2) it estimates the average trend and thus provides a smooth estimate of how the mean concentration is changing over time; and (3) a by-product of the fitting process is a series of local trend *slopes*, which can be used to gauge rates and directions of change at particular points or periods of time.

This last benefit is exploited by GTS in constructing trend maps, which spatially represent trend movement during a specific time period. These maps point to where different kinds of trends are occurring and how probable it is that the trends represent something real. They can also be used to flag or confirm changes in plume extent over time and to help identify areas of the site where additional sampling might be warranted (see Figure 7).

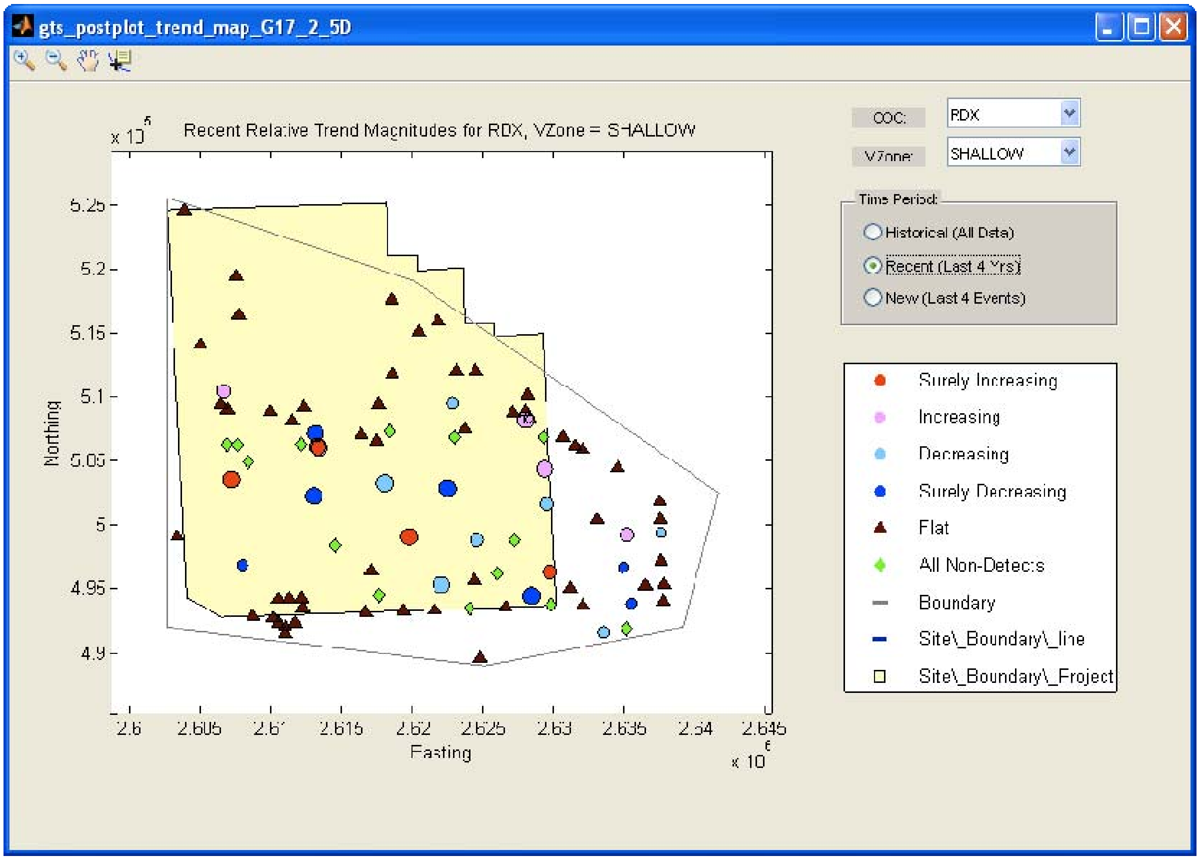



**Figure 7. Example trend map in GTS.**

**Plume maps** (e.g., base maps) are uniquely created in GTS using QLR, a quasi-nonparametric fitting and spatial estimation procedure designed specifically for GTS. QLR employs local regression instead of kriging, which, unlike the latter (1) does not require development of a spatial covariance model but still accounts for the presence of spatial correlation; (2) as a smoother, does not assume that sample data values have been measured without error; and (3) does not require only one measurement per sampling location or per sampling event.

Instead of requiring an a priori spatial covariance model, the user decides on a degree of smoothness of the map through adjustment of a bandwidth parameter. In practice, the process is mostly automated since GTS computes a default bandwidth for each map, which can be overridden when desired. As a smoother instead of an interpolator, local regression is akin to linear regression through a scatter cloud of points. The best-fitting line may not coincide with any specific point, yet it attempts to capture the overall trend. Similarly, a surface map fitted with local regression attempts to capture the best overall surface trend. The method explicitly assumes each data point is measured with some degree of error. It also explicitly allows for multiple data points at any given location.

Standard forms of kriging require that there be only one data point per location to avoid collinearity in the kriging equations. Given inconsistent sampling schedules across wells at most sites, choosing data from a given sampling event often does not include sampling information from all the wells of interest. But widening the snapshot of time to include more wells typically leads to multiple data points at some locations, necessitating perhaps an averaging of these measurements before input to kriging, even though this action tends to reduce the observed variability of the data set and violate the assumption of identically distributed measurements.

Mapping in GTS does not apply local regression *directly* to the concentration data. Like other regression techniques, it assumes that residuals around the local trend or surface are approximately normal in distribution. But in practice, essentially every LTM network has: (1) significant fractions of non-detect measurements among one or more COCs, and (2) high levels of skewness in the (univariate) concentration distributions (i.e., significant non-normality). Neither of these data features is adeptly handled by standard spatial mapping techniques without the use of special data transformations.

GTS accounts for these real-world difficulties by using QLR as a mapping engine. QLR first constructs an estimate of the overall observed (i.e., empirical) declustered cumulative distribution function (DCDF), based on recent concentration data from the site [“declustered” refers to adjusting the cumulative distribution function (CDF) for the preferential clustering of sampling locations in higher level concentration areas]. Then each concentration is converted to a value between 0 and 1 (i.e., the unit interval) using the DCDF and further converted to values along the real line via a second logit transformation. These logit-transformed values are then fitted using local regression and the resulting estimates back-transformed utilizing the same two-step transformation process in reverse to get concentration-domain map estimates. The name “quantile” in QLR comes from the fact that the first step of the transformation changes each concentration into an equivalent quantile from the DCDF.

The advantages of QLR include (1) non-detects can be handled without resorting to complicated imputation schemes; (2) the impact of extreme skewness is minimized since all estimation is done on the logit-transformed values and only afterwards back-transformed into concentration estimates; (3) plume detail and intensity can be reasonably captured since each logit-domain estimate is linked directly back to the observed concentration distribution at the site (i.e., DCDF); (4) a range of possible spatial models is fit to the observed data, with one model identified by GTS as the preliminary best choice; (5) the entire map building process is automated within the GTS software interface—except for choice of spatial bandwidth if the user decides to override the GTS-computed defaults—allowing an analyst to construct statistically sophisticated maps without the need for expert consultation or setup.

By design, GTS *does not* fully automate the process of fitting either spatial or temporal models. Although standard statistical techniques such as “residual checking” are employed to help guide the fitting process, it is well known (see [4]) that strict reliance on “black-box” modeling approaches can lead to poor-fitting models. In GTS, the user has the option to provide input at critical junctures in the model building exercise and override the GTS defaults.

In addition to the baseline trends, trend maps, and concentration base maps, the Baseline module also provides the user with a visual and tabular overview of the baseline network status. The status report includes estimates of the empirically derived baseline sampling frequency/interval associated with each well, as well as a graphical summary of which locations are "critical" to the network, "redundant," or "protected." Connected with this last feature, users can designate selected wells as protected, meaning that those particular locations are shielded from spatial optimization (i.e., always kept as critical wells and never classified as redundant). GTS also allows import of water level data and visualization of an estimated water table surface, along with how the water table changes across time slices (see Figure 8).

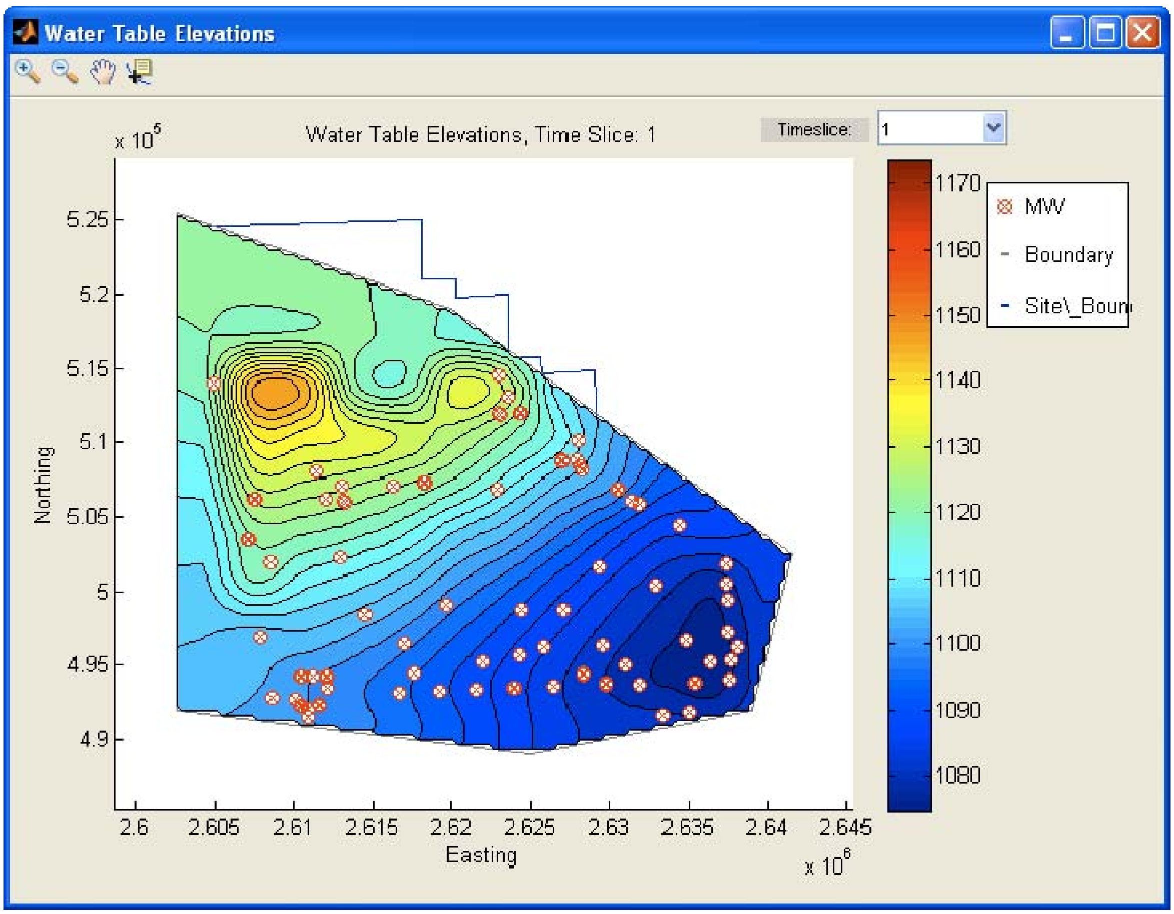



**Figure 8. Example water table map.**

## Optimize Module

Once baseline trends and base maps are constructed, users can begin optimization. GTS offers separate temporal and spatial optimization functions, depending on the needs and data availability of different sites. Temporal optimization in GTS consists of two components: (1) *temporal variograms* applied to groups of wells and (2) *iterative thinning* of individual wells. More than one temporal optimization method allows for flexible handling of the kinds of data available at different installations. Temporal variograms are most useful at sites with limited sampling histories and less historical data. Iterative thinning, by contrast, reconstructs the entire trend at each well, a more difficult statistical task requiring larger amounts of data (generally at least eight samples per well), but providing well-specific optimal sampling schedules and readily accounting for seasonal trends or fluctuations. Figures 9 through 11 provide an overview of the logic and features of Module D.

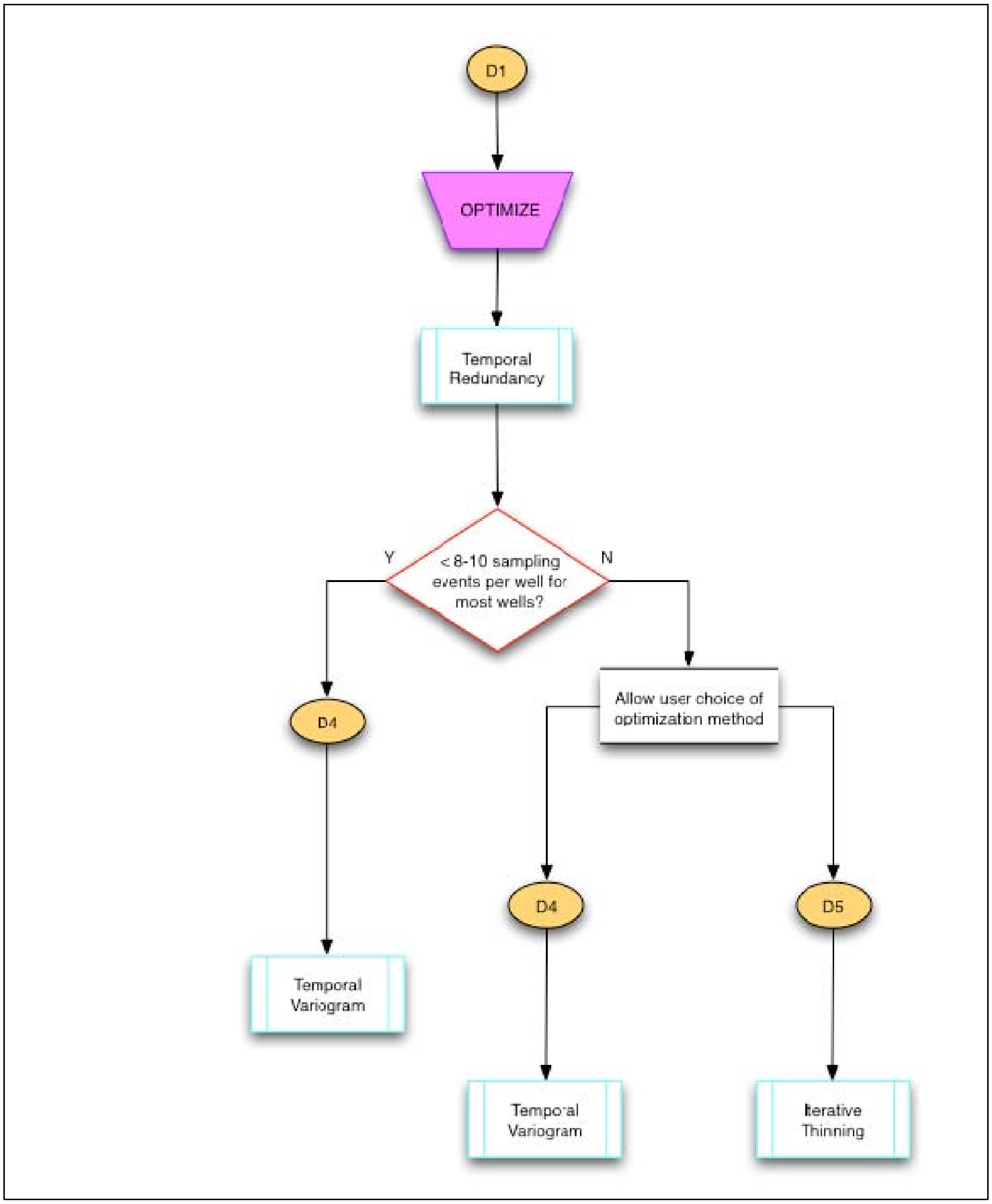



**Figure 9. Schematic of optimize (Module D) logic — temporal redundancy.**

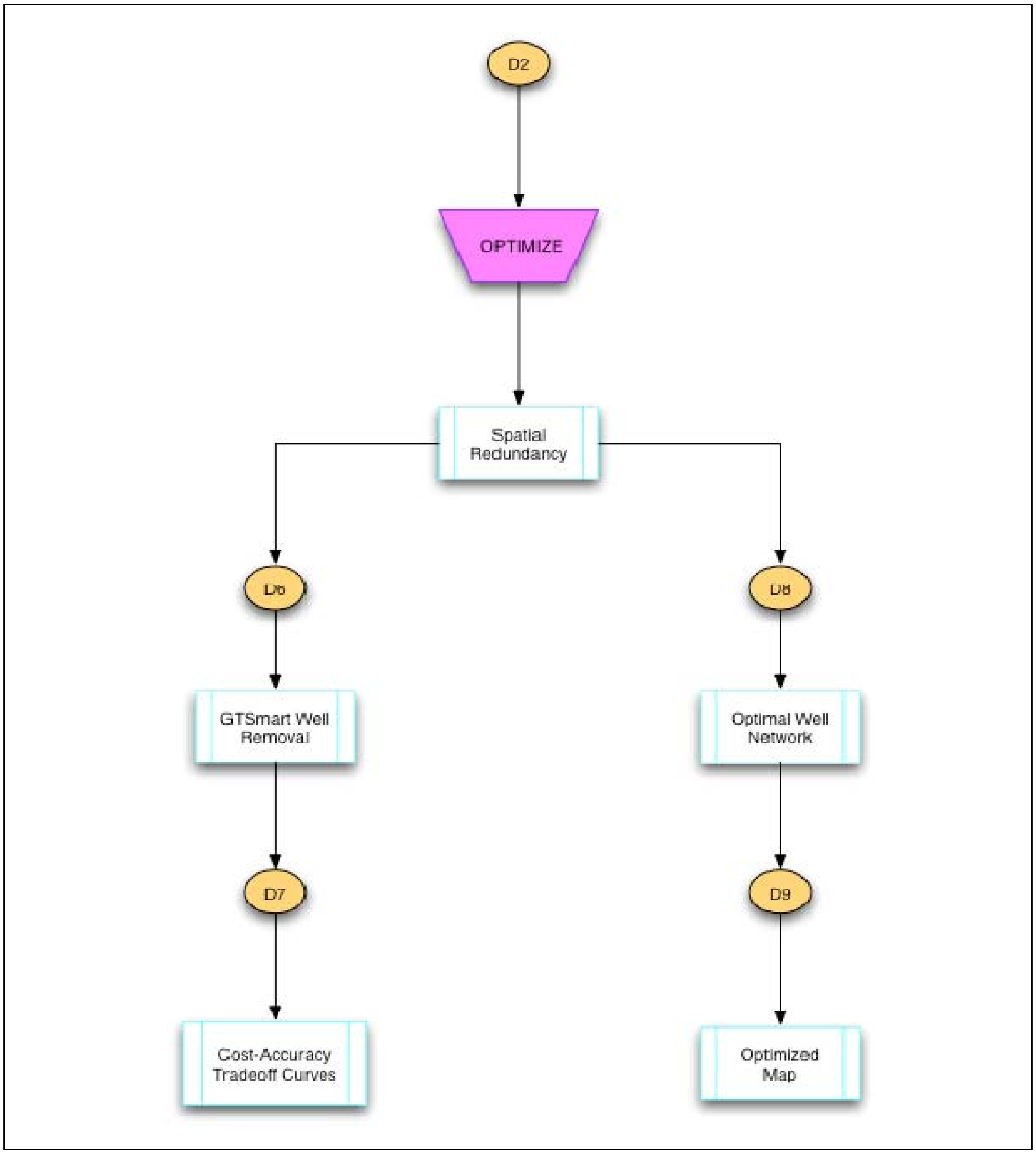



**Figure 10. Schematic of optimize (Module D) logic — spatial redundancy.**

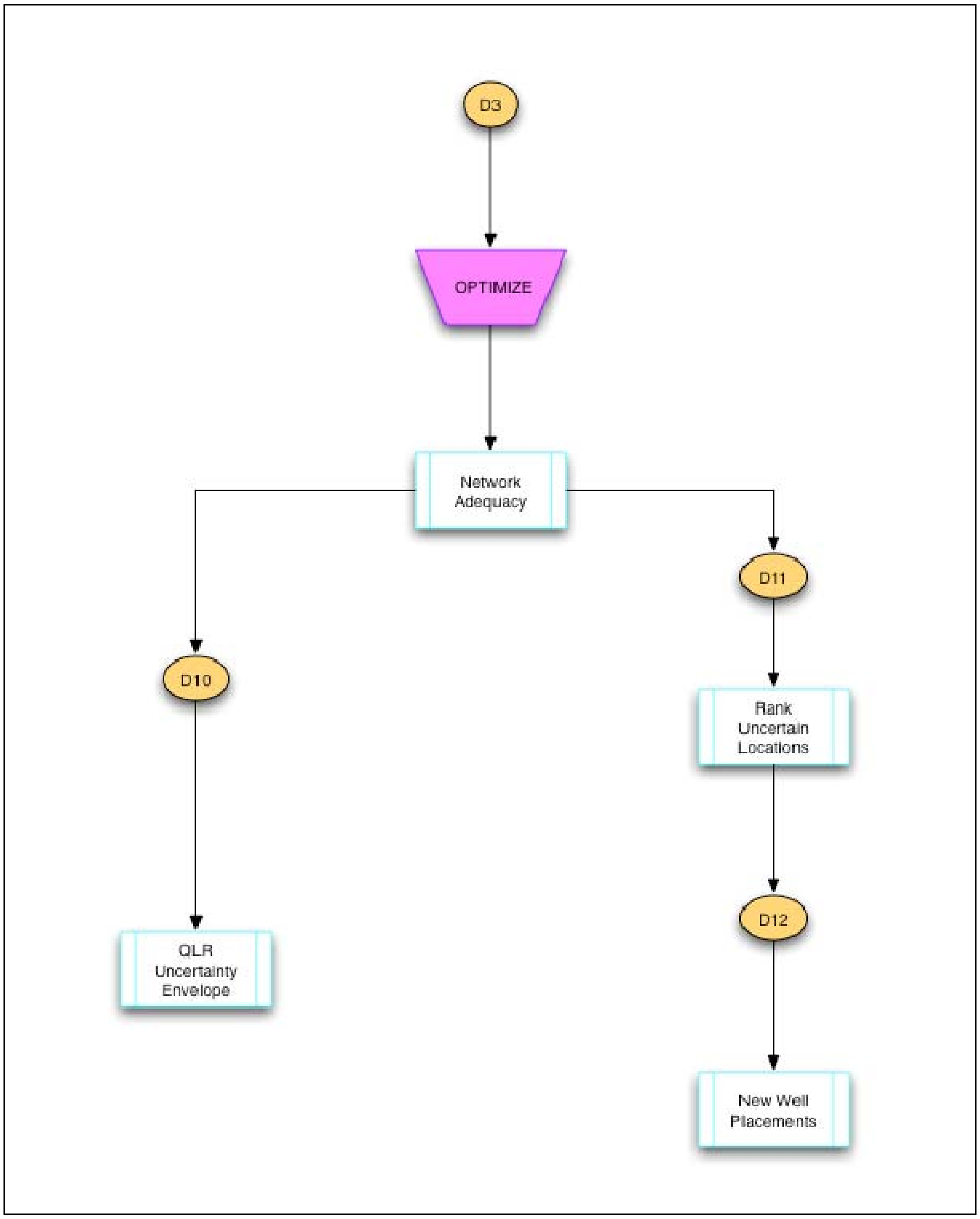



**Figure 11. Schematic of optimize (Module D) logic — network adequacy.**

The temporal variogram optimizes sampling frequencies simultaneously over a *group* of well locations (see Figure 12). These locations might represent all wells at a given site, those connected with a particular regulatory unit or part of a treatment system network. Whatever the grouping, the temporal variogram provides a single optimal sampling interval that can be applied to every well within the group. The temporal variogram itself is a smoothed curve, fit to a scatterplot of squared differences between all possible measurement pairs (y-values) versus the time lag between successive sampling events (x-values). The curve is estimated using LWQR.

After GTS constructs the temporal variogram, the user is prompted to identify an approximate range in its structure. Because the variogram assesses the correlation between the observed data and lag time between samples, positive temporal correlation is exhibited on the variogram by small values for small time lags and larger values for large time lags. Small values on a variogram indicate a high degree of correlation, while higher values represent a loss of correlation and greater statistical independence. The range is identified as the first lag at which the variogram begins to level off or plateau. GTS sets the optimal sampling interval to this chosen range of the temporal variogram, if it exists. Sampling intervals smaller than the range are associated with correlated, and therefore somewhat redundant, sampling results.

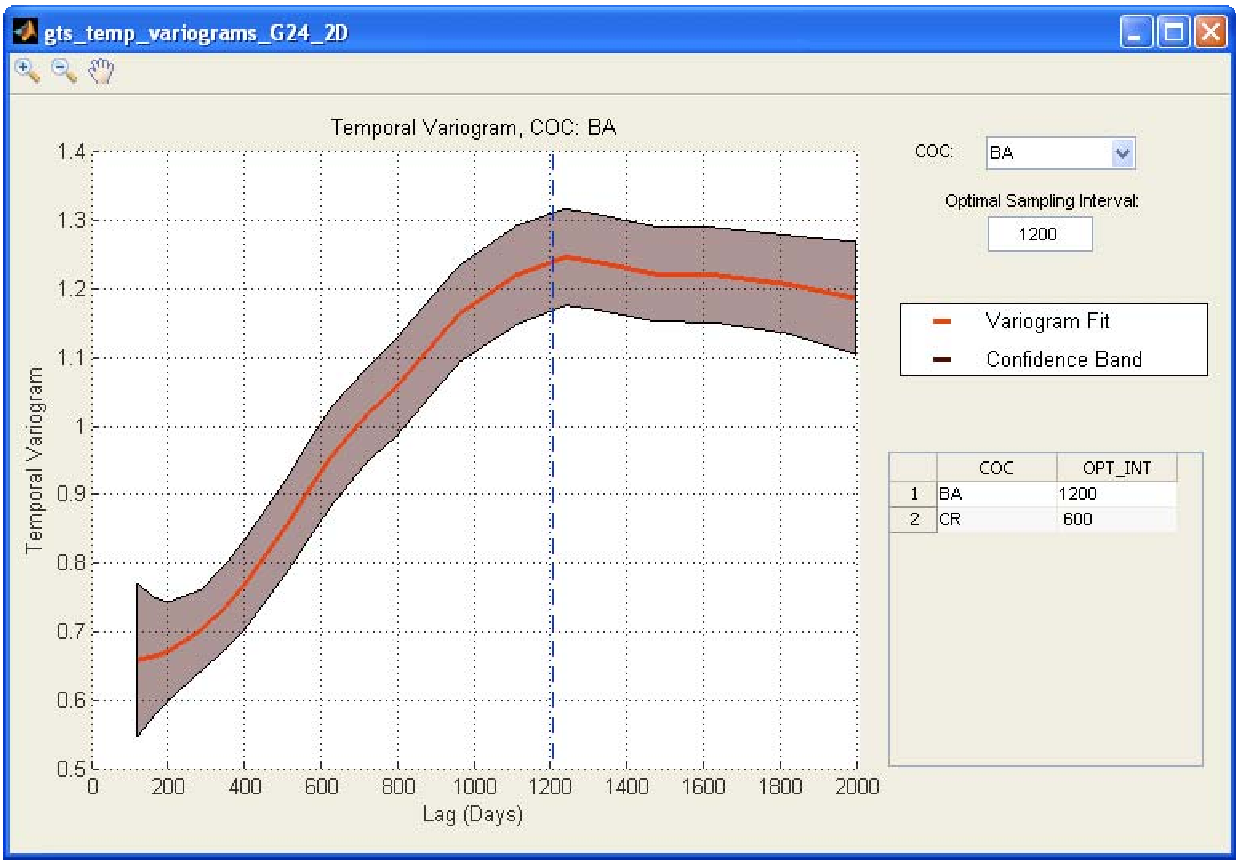



**Figure 12. Example of temporal variogram in GTS.**

Iterative thinning optimizes the sampling frequencies at *individual* wells. Because each location is analyzed separately, a different recommended sampling interval is generated for each well. GTS then combines these well-specific sampling intervals into a common operational sampling frequency for all the wells using the median optimal interval. Iterative thinning is based on a straightforward idea: (1) take the existing, historical data for a given well location and constituent; (2) determine the current average sampling interval; (3) fit a trend to these data along with statistical confidence bounds around the trend; (4) iteratively remove, at random, certain fractions of the original data; and (5) re-estimate the trend based on the reduced data set to determine whether or not the trend still lies within the original confidence bounds. If too much of the new trend falls outside the confidence limits, stop removing data and compute a new, optimized sampling interval using the remaining data.

The other optimization function within GTS—spatial optimization—consists of the following steps: (1) searching for statistical redundancy via mathematical optimization; (2) determining optimal network size with the aid of cost-accuracy trade-off curves; and (3) assessing whether new wells should be added and where (i.e., network adequacy).

To find spatial redundancy, GTS identifies optimal subsets of the existing monitoring network through mathematical optimization. This measures the degree of deterioration in GTS-estimated site maps by comparing site-maps made using a series of potentially "optimal" reduced-data networks against their corresponding base maps. GTS uses a quasi-genetic algorithm, GTSmart, to search through alternate network configurations, where every alternate configuration temporarily removes a certain percentage of the wells. For each such configuration, a tentative site-map is constructed. Then the relative residuals (or relative differences) between the tentative concentration estimates on the site map and the corresponding base map estimates are used to assess the degree of redundancy via three statistical measures: (1) trimmed mean absolute bias, (2) upper 90th percentile absolute bias, and (3) maximum absolute bias.

For each of these measures, bias is computed between the site map and base map estimates by taking the absolute value of the logged ratio between the site map and base map. The ratio of the two map estimates allows an estimate of the *relative* rather than *absolute* difference between the site map and base map; logging the ratio gives more statistical weight to mismatches between high areas of one map and corresponding low areas on the other (e.g., overestimating concentrations near boundaries of a plume). These necessarily positive-valued residuals are then plugged into standard formulas for computing the 95% trimmed mean, the upper 90th percentile, and the maximum. Thus, three measures of bias are computed for each alternative site map.

All three statistical measures are graphed against the degree of well removal, among the thousands of alternate configurations tested, to form cost-accuracy trade-off curves (see Figure 13). Default, user-adjustable limits on the acceptable levels of bias are also plotted. The trade-off curves display the relationship between well removal and map bias and identify at what point the bias measures exceed their limits. GTS designates a well configuration as optimal when it exhibits the largest degree of well removal among those configurations whose bias measures are still within the acceptable bias limits. In other words, an optimal well configuration balances reduction in cost (through the removal of wells) and consequent loss of map accuracy (as measured by bias). If many wells are statistically redundant, the trade-off curves will indicate a significant cost reduction without substantial information loss. If few wells are redundant, the loss of accuracy will be large even when a small number of wells are removed.

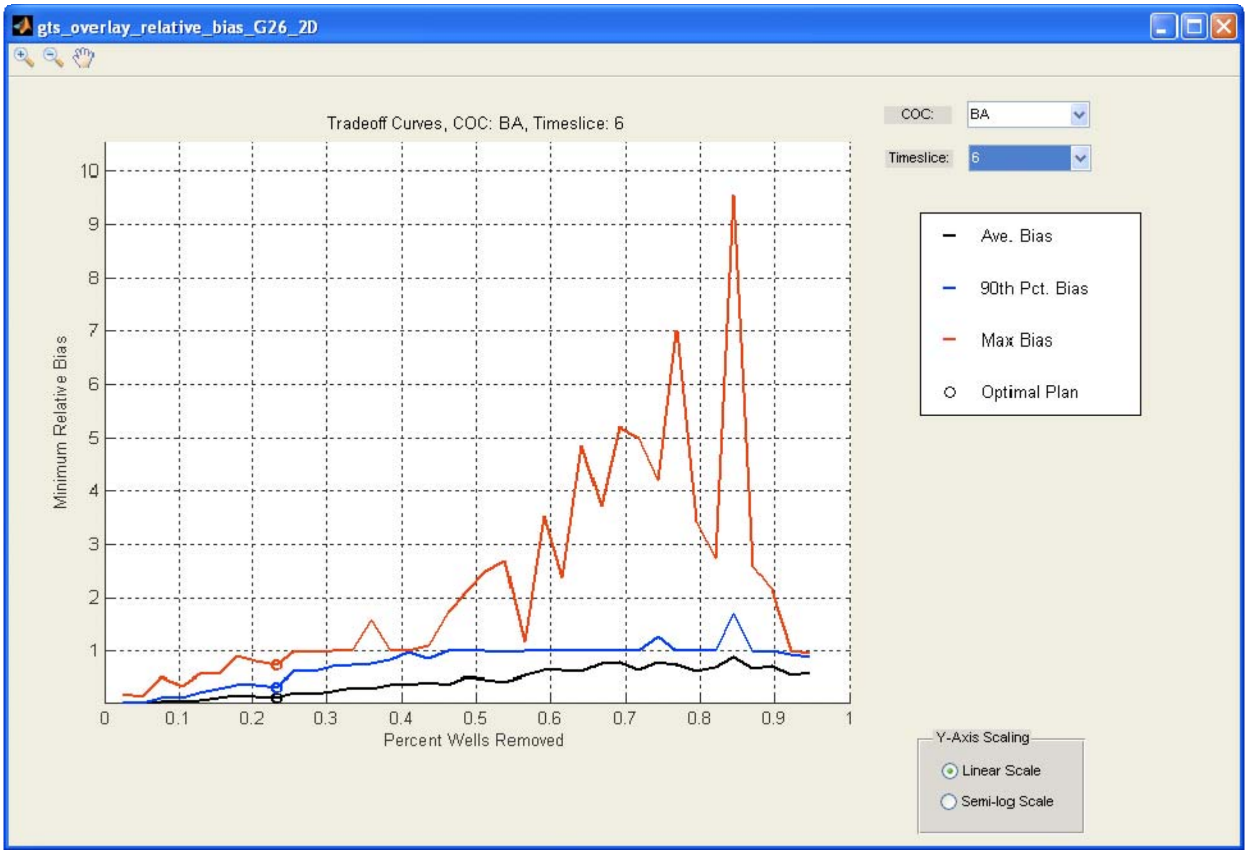



**Figure 13. Example of cost-accuracy trade-off curves in GTS.**

Once a point of optimality has been computed, GTS tags as redundant all wells that were not included in that configuration for a given COC and period of sampling (i.e., time slice). The remaining wells are deemed critical to the network. The same process is repeated for other time slices and COCs and then combined automatically to determine a ranked list of critical and redundant wells at the site. The user is presented with a list of wells and their optimization status, along with a post-plot of the well network showing which locations are redundant and which are critical. GTS also displays side-by-side before and after maps of the plume extent for each time slice and COC (and aquifer zone, if applicable), in order to document any differences due to the optimized network (see Figure 14).

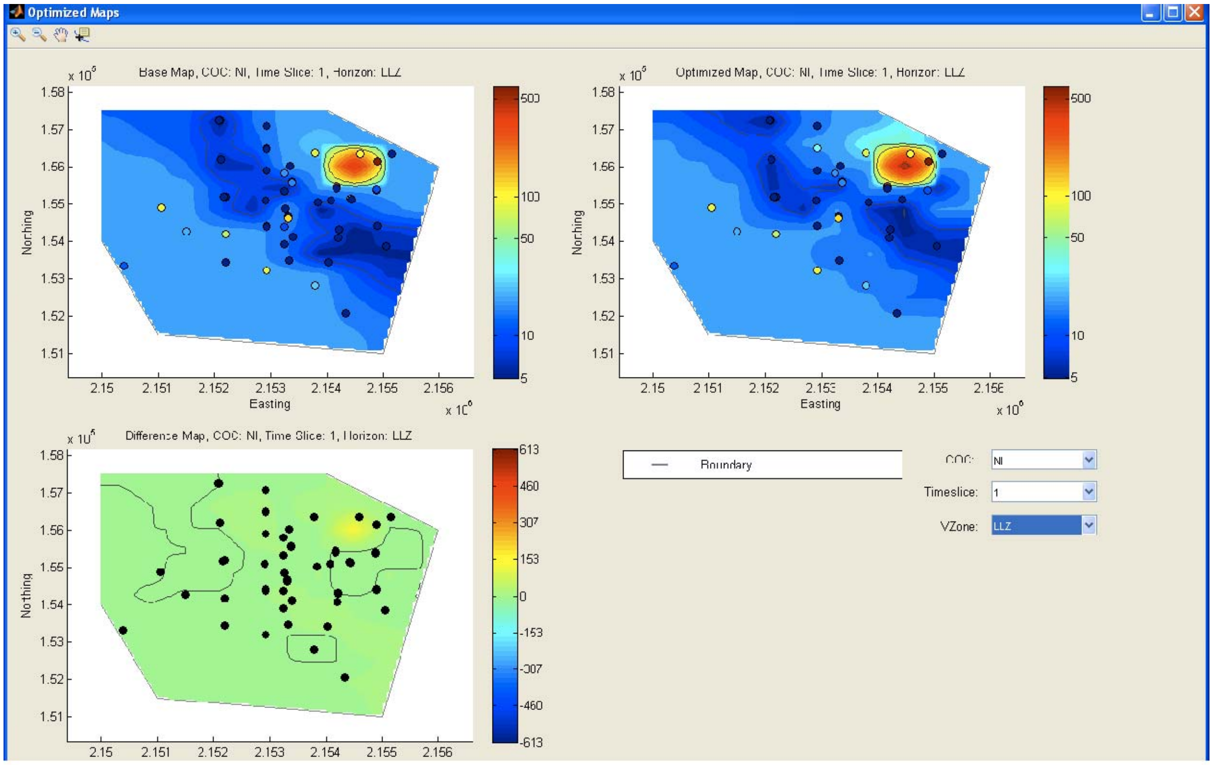



**Figure 14. Example baseline versus optimized maps in GTS.**

The last step of spatial optimization in GTS is the network adequacy analysis. This function determines whether any portions of the site warrant new sampling locations. To do this, GTS generates a risk envelope for each COC. The risk envelope is a map of estimated coefficients of variation (CVs), a result of applying QLR at each pixel on the map to estimate both a (mean) concentration and its associated standard deviation for each time slice. The CV is simply this standard deviation divided by its associated (mean) concentration estimate (and then averaged across time slices), providing a unitless measure of uncertainty at each pixel. By combining and ranking these uncertainty values across COCs, GTS flags good candidate locations for the placement of new wells, subject to user override (see Figure 15).

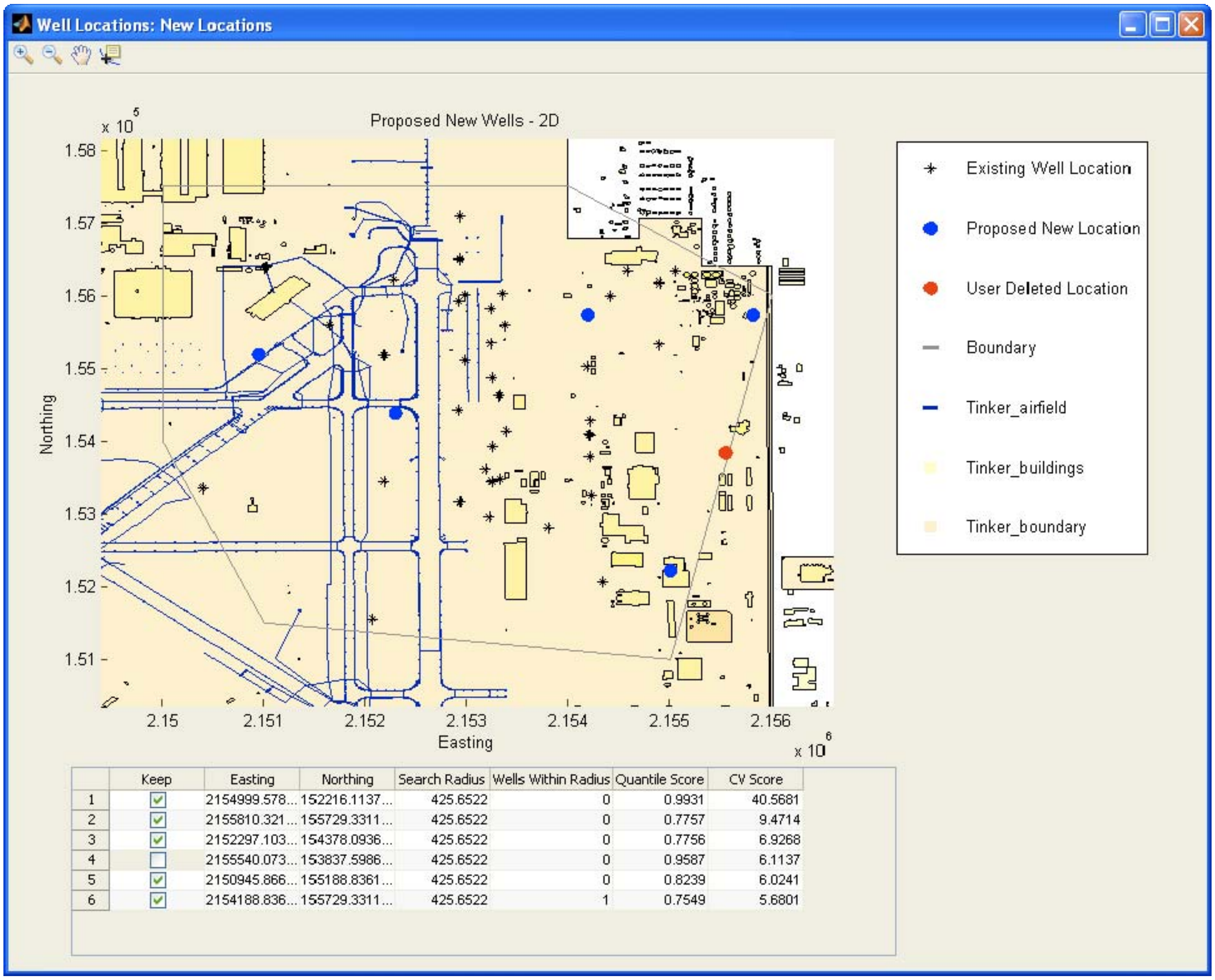



**Figure 15. Example of network adequacy post-plot.**

Once GTS optimization is completed, users can export tables of the results for use in the Excel-based GTS cost-comparison calculator. The calculator is designed to compute a realistic, site-specific ROI associated with a recommended optimized sampling program. In straightforward fashion, it builds two sets of cost estimates: a baseline set representing the original (non-optimized) monitoring program and an optimized set using the GTS recommendations concerning sampling frequency and network size. It then computes the difference between these two sets of costs to determine the potential savings realized from optimization and the ROI.

To make the cost accounting as realistic as possible, the cost-comparison calculator allows site-specific entry of such factors as constituent groups (including relative sampling rates to account for parameters that are collected only sporadically or in select portions of the site); field sampling and analytical method costs; management, reporting, mobilization, and labor costs; costs for drilling any new wells; and costs associated with performing the optimization study. All this information is combined with the GTS recommendations for which wells are critical or redundant, optimized sampling frequencies, and whether any new well locations are needed.

## Predict Module

The last module allows users to import and compare new sampling data against previously estimated trends and maps. A schematic of the logic of the Predict module is given in Figure 16. The goal of these features is to enable identification of potential outliers, anomalous values, or early warning changes in hydrogeologic conditions, plume intensity, or extent. The two available options within GTS v1.0 include trend flagging and plume flagging. In the first, a prediction band around the baseline trend at each well is linearly extended into the future to the newly imported sampling events. If any new measurement falls outside the prediction band, that sampling event and the associated well are flagged (see Figure 17). Users can then investigate explanations for the apparent anomalies.

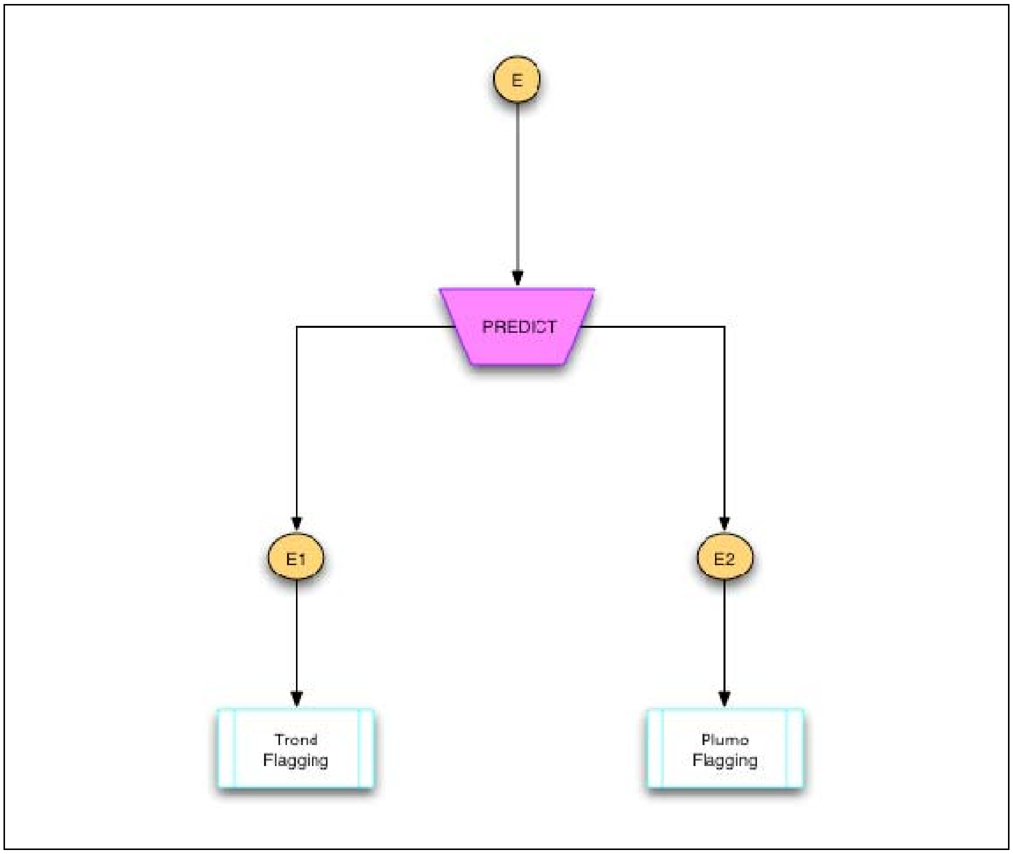



**Figure 16. Schematic of predict (Module E) logic.**

The second option—plume flagging—has a similar purpose but compares the new data against a prediction envelope constructed around the plume map. Data falling outside the envelope are flagged for additional follow-up. Of interest, unlike trend flagging, plume flagging can be utilized to check data sampled from new well locations that do not yet have a temporal history. It can also be utilized to periodically track abandoned wells, perhaps locations deemed redundant during optimization, to verify that the projected plume using the critical well network adequately reproduces concentration levels at locations no longer being regularly monitored.

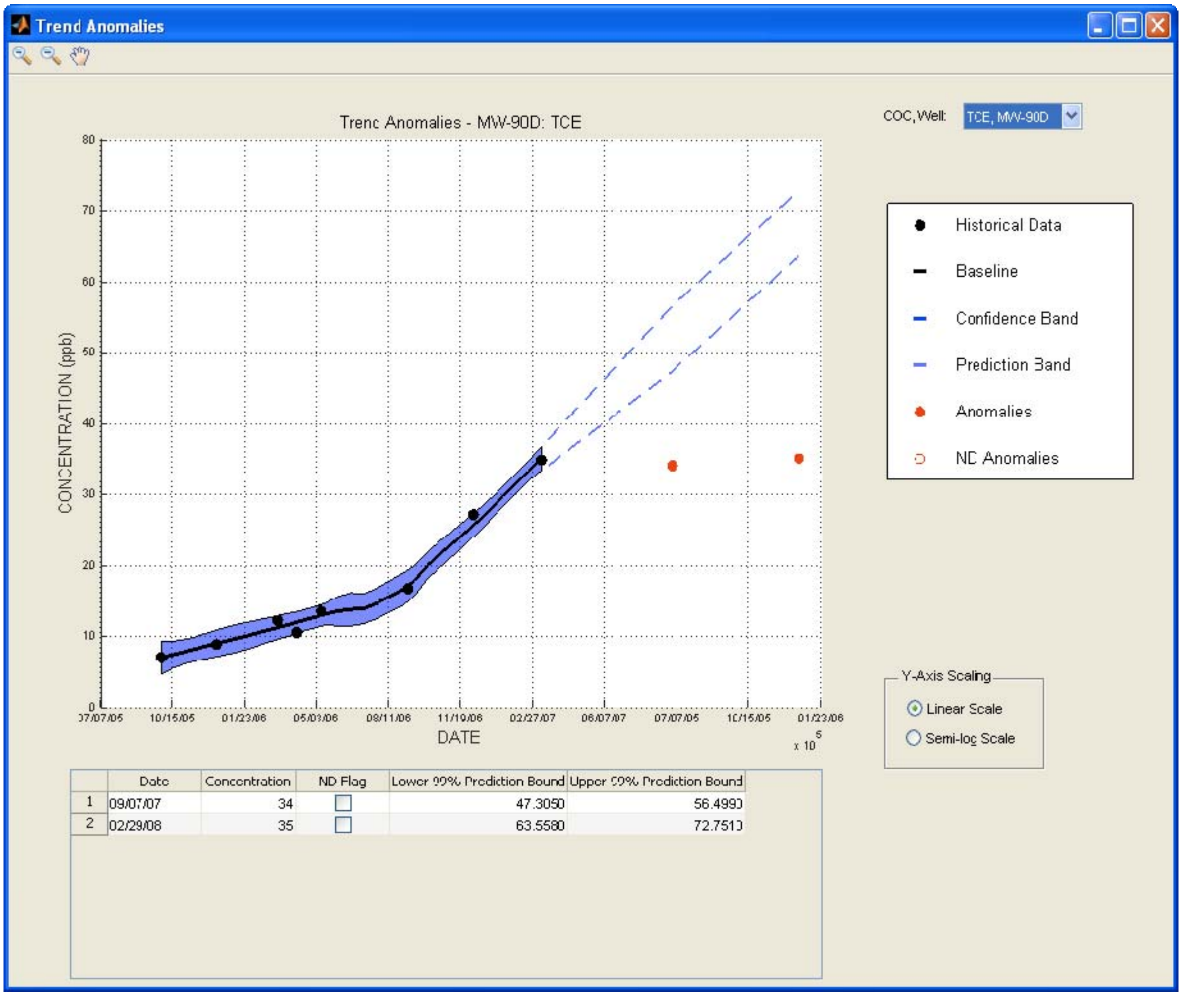



**Figure 17. Example of trend flagging.**

## 3.2 ADVANTAGES AND LIMITATIONS OF THE TECHNOLOGY

By way of overview, GTS is designed to balance the practical and scientific difficulties inherent in optimization schemes, namely, how to perform a scientifically defensible optimization analysis without requiring substantial involvement by statistical or mathematical experts. The software builds in several state-of-the-art statistical and geostatistical analytical routines, all tailored to LTMO yet woven into a user interface designed to smartly guide the user through a complex series of analyses. GTS is designed to be run by mid-level analysts with some—though not expert—level statistical and geostatistical background.

### Benefits of GTS

The first and most important benefit of GTS is that it offers a more resource-effective long-term groundwater monitoring program. This benefit is realized in three primary ways:

1. By reducing sampling frequency and minimizing spatial redundancy in existing networks

2. Through statistically defensible addition of new well locations to better characterize contaminant plumes
3. Via trend mapping and trend flagging to better monitor changes over time in site conditions and to identify anomalies or unexpected sampling results.

Several hundred and possibly thousands of DoD, DOE, and Environmental Protection Agency (USEPA) sites could benefit from the techniques within GTS. Projected annualized and life-of-project cost savings from implementing a GTS-optimized program at a given site can be significant, in the range of 30 to 60%. ROI for a GTS-optimized monitoring program is generally 1 to 2 years or less.

GTS is equally applicable to site-specific plumes, and unit-wide or base-wide studies involving multiple source areas, plumes, and monitoring conditions. This is because GTS does not require or utilize plume-specific configuration data, fate-and-transport models, or other hydrogeologic modeling information. Instead, it merely attempts to reconstruct maps and trends, based on the general extent of existing groundwater wells. GTS assumes that accurate reconstruction of these features will enable and assist continued regulatory, monitoring, and remedial decisions as needed, using the optimized network.

Operationally, GTS offers stand-alone spatial and temporal optimization modules. Even at sites that are poorly characterized or have insufficiently large well networks to warrant a spatial analysis, a temporal optimization can still be conducted, including trend mapping and trend flagging. Past applications of GTS have demonstrated that most of the projected cost savings is realized through the temporal analysis.

Technically, GTS also offers several additional benefits. These include:

- <u>Statistically-based, semi-objective LTM optimization</u>, *built to be run by non-experts.* Most currently available tools either place substantial reliance on qualitative review by expert hydrogeologists (in combination with statistical analysis) or offer less sophisticated and more heuristic statistical methods. GTS is designed to incorporate state-of-the-art statistical tools within a user interface negotiable and interpretable by mid-level analysts. GTS compliments and encourages professional judgment from stakeholders in negotiating an optimal monitoring plan.
- <u>Innovative exploratory tools for assessing data characteristics</u>, ranking COCs for optimization potential and analyzing multiple aquifer horizons. These tools can also assist in the identification and development of anthropogenic or background data sets, such as are needed to set defensible concentration limits when delineating contaminated versus uncontaminated wells.
- <u>Sophisticated built-in graphics</u> for data visualization, including contour mapping, complex trends, post-plots, and shapefile annotation.
- <u>Trend estimates derived from LWQR</u>, allowing for fitting of complex or seasonal time series data. All other currently available LTM optimization tools only offer

fitting of *linear* trends, an assumption that *does not match* the reality of most LTM data sets. Neither do other methods provide a rigorous and non-subjective way to assess redundancy in sampling frequencies.

- Semi-nonparametric surface map estimates made using QLR, a smoothing technique not bound by the constraints of kriging. By design, QLR is made to handle skewed datasets as well as significant proportions of non-detects, data features ubiquitous to LTM networks.
- Empirical, data-driven assessment of redundancy. GTS does not rely, as do some tools, on the kriging variance—known to be a poor absolute measure of variability—for judging spatial redundancy. Instead, a reduced-network is optimal if it can accurately reproduce the base map.
- Automated redundancy searches, both during temporal and spatial optimization. The most complicated computational tasks are transparent to the user within the GTS interface.
- Use of multiple cost-accuracy trade-off curves to gauge points of optimality. Defensible bias measures of statistical accuracy allow for rigorous analysis of potential trade-offs.
- A straightforward, realistic cost-comparison calculator that estimates cost savings to be realized from implementing the GTS-optimized monitoring program, using baseline cost data supplied by the user. The calculator also computes estimated ROI accrued from performing a GTS optimization.
- User-ready summary reports of the results of GTS optimization; these include lists of optimal sampling intervals by well; recommended operational sampling intervals by site/area, well group, or aquifer horizon; lists of redundant and non-redundant well locations; and areas recommended for new wells.

**Limitations of GTS**

Although extremely versatile and capable, v1.0 of GTS has certain limitations, some of which became apparent during this ESTCP demonstration:

- Effective spatial optimization in GTS requires a minimum of 20-25 wells and at least two sampling events per well; temporal optimization requires at least 1 well and 4-8 distinct sampling events per location.
- GTS requires a number of input fields in ASCII text format to create a sufficient analysis database. Some users may find the directions for importing data and creating or augmenting databases within GTS more complicated than need be.
- QLR, the GTS mapping engine, is by design a “smoother” rather than an interpolator, that is, it may not replicate or “honor” observed measurements when creating map estimates, unlike, for instance, kriging. To the extent that these observations are precisely known or fixed, users may find QLR-based maps less appealing than interpolated maps.

- GTS does not offer sophisticated handling of radiochemical data, particularly measurements recorded with non-positive values (i.e., zeros or negatives). These data must first be converted to positive values, unless they represent non-detects with a known, positive detection or reporting limit.
- Optimized sampling intervals from temporal variograms in GTS often do not match the optimized sampling intervals from iterative thinning using the same data. Further improvements to the temporal variogram algorithm may be needed, especially to account for sites with spatial trends that are actively changing over time.
- Cost-accuracy trade-off curves in GTS are not interactive. Although the bias limits can be adjusted by the user, the spatial optimization must be completely re-run each time those limits are changed, in order to see the impact of the revised limits and to generate a new optimal network.
- There is no way in GTS v1.0 to batch print graphics. Since a GTS analysis typically generates a large number of statistical graphics, users may be frustrated with the inability to document graphical results outside the application.
- The mathematical optimization algorithm in GTS is not a true genetic algorithm wherein portions of the binary string "DNA" representing alternate network configurations are allowed to "mate," "mutate," and create "offspring." Instead, GTS does a "smart search" through the space of potential network configurations, only selecting for testing those strings with interwell spacing comparable to the full network.
- GTS v1.0 does not track changes in contaminant or plume mass, nor does it allow users to specify contaminant mass as an optimization criterion.
- GTS may not give valid/accurate spatial results in subsurface environments that are highly fractured and discontinuous with poor hydraulic connection. Spatial mapping techniques in general (not just those in GTS) inherently assume that concentration patterns at known wells can be extended (e.g., interpolated, smoothed) to unsampled locations. This may be problematic at sites with large contrasts in hydraulic conductivity (preferential pathways).
- There is no current method to correctly handle distinct well screens at different depths possessing the *same* location name and identical easting/northing coordinates. This limitation can occur with either direct punch technology (DPT) samples that take multiple discrete measurements at different depths but along the same borehole, or possibly with cluster wells that have multiple screens at distinct depths. As long as the name of each well screen or discrete sampling point/depth is unique, GTS will analyze the data appropriately. If identical names are used for such locations, however, regardless of depth, the user must adjust the naming convention outside the program.

## Other Technologies

As of this writing, at least four other technologies fairly similar in aim and scope to GTS have been or are being developed. These include the Three-Tiered Monitoring Strategy developed by Parsons Engineering (www.parsons.com), Summit Tools developed by Summit Envirosolutions (www.sampleoptimizer.com), MAROS developed by GSI Environmental (www.gsi-net.com/software/free-software/maros.html), and the geostatistical optimization module of VSP developed by Battelle (http://vsp.pnl.gov/index.stm).

The Three-Tiered Monitoring Strategy has not been released as stand-alone software, but is currently under development. Until now, it has been a proprietary algorithm used on a consulting basis. Substantial emphasis is placed on expert qualitative review by a consulting hydrogeologist. Its spatial analysis relies on kriging and its known shortcomings. Previous versions of the Three-Tiered approach also did not use mathematical optimization to identify redundancy. The temporal analysis does only linear fitting of trends and uses a rule-based rather than empirical strategy to derive optimal sampling frequencies.

Summit Tools was developed under ESTCP grant ER-200629 and released in 2009. The ESTCP version is a proprietary software system that is free for use by government and DoD employees; commercial users must buy an annual license. All users must purchase upgrades if desired. It relies in part on kriging for spatial mapping but also incorporates other spatial modeling techniques as well as automated redundancy searches based on efficient genetic algorithms. Summit Tools utilizes an automated “black box” approach to spatial modeling, with its attendant risks, in order to simplify user input. Sampling frequency optimization is handled via a joint spatio-temporal redundancy search. This requires highly regular baseline sampling intervals to be effective. Summit Tools also includes a Data Tracker module designed to identify potential anomalies/outliers in new data, based on linear or exponential-decay projections of baseline trends.

MAROS was also developed under the auspices of AFCEE and is freely available. As an optimization software product, MAROS is the most mature of the competing technologies but lacks many of the advanced statistical features included within either GTS or Summit Tools. It fits only linear trends and offers a heuristic, rule-based approach for determining optimal sampling frequencies. MAROS does not perform spatial mapping, per se, but relies on Delauney triangulation and nearest neighbor analysis to assess spatial redundancy. Users desiring detailed site maps must employ third-party mapping software. Also, only one measurement per sampling event and location is allowed when conducting spatial evaluations. A new version of MAROS is currently under development and promises to add significant new capabilities.

VSP recently released a new geostatistically based optimization module for conducting spatial optimization of well locations and temporal optimization of sampling frequencies. New documentation of these capabilities was being prepared at the time of writing this report.

Although other optimization approaches exist (for instance [18-20]), they depend in large measure on coordinated use of numerical groundwater simulation models (e.g., fate and transport). Some utilize Kalman filters and/or simulated annealing to update the models and predict where in the network uncertainty might most be reduced. None of these methods has apparently been translated into stand-alone, public domain software. Furthermore, *numerical groundwater models are not available at a majority of potential sites where GTS might be utilized.*

To roughly compare the features offered by GTS, MAROS, the Three-Tiered approach, Summit Tools, and VSP, the following "measles chart" in Figure 18 gives a comparative overview.

**Figure 18. LTMO software feature comparison chart.**

| Feature/Capability | GTS | MAROS | Summit Tools | 3-Tiered | VSP |
|---|---|---|---|---|---|
| Built-in database | ● | | ● | | |
| Data filtering, manipulation | ● | | | | |
| Rich visualization, statistical graphics | ● | | ● | | ● |
| Data checking, outlier search | ● | | | | |
| Freeware | ● | ● | | | ● |
| Publicly released | ● | ● | ● | | ● |
| Print/save reports | ● | ● | ● | | |
| Exploratory data tools | ● | ● | | | |
| COC ranking analysis | ● | ● | | | |
| Multiple horizon analysis | ● | | | | |
| Linear trends | | ● | | ● | ● |
| Complex, nonlinear trends | ● | | | | |
| Trend analysis | ● | ● | | | |
| Trend maps | ● | ● | | | |
| Mapping engine | | | | | |
| Quantile local regression | ● | | | | |
| Kriging/quantile kriging | | | ● | ● | ● |
| Delauney triangulation | | ● | | | |
| Water table mapping | ● | | | | |
| Mass flux/moment analysis | | ● | ● | | |
| Temporal optimization | | | | | |
| Temporal variograms | ● | | | | |
| Iterative thinning | ● | | | | |
| Cost-effective sampling (CES) | | ● | | | |
| Spatio-temporal optimization | | | ● | | |
| Spatial optimization | | | | | |
| Mathematical optimization | ● | | ● | | |
| Optimize by multiple site objectives | | | ● | | |
| Steepest descent (i.e., sequential, well-by-well) | | ● | | ● | ● |
| GTSmart (quasi-genetic) search | ● | | | | |
| Genetic algorithm search | | | ● | | |
| Network adequacy analysis | ● | ● | | | |
| Cost-comparison calculator | ● | | | | |
| Spatio-temporal optimization | | | ● | | |
| Built-in qualitative analysis | | | | ● | |
| Data tracking | | | | | |
| Trend flagging/data tracker | ● | | ● | | |
| Plume flagging | ● | | | | |

*This page left blank intentionally.*

# 4.0 PERFORMANCE OBJECTIVES

This section provides a summary of the performance objectives stated in the Technical Demonstration Plan for evaluating GTS in this project, including a conclusion as to whether or not each performance objective was met. Table 1 summarizes these performance objectives. To avoid repetition, a detailed discussion of each performance objective is deferred until Section 7.0 that explains the criterion, how it was assessed, and the basis for the assessment.

**Table 1. Performance objectives.**

| Performance Objective | Data Requirements | Success Criteria | Criteria Met? |
|---|---|---|---|
| **Qualitative Performance Objectives** | | | |
| Ease of use, software (primary) | Feedback from independent site testers operating the software | Users find GTS easy to use as indicated by user feedback and by general lack of error or system crashes in installation and use. | YES |
| Ease of use, user manual (primary) | Feedback from independent site testers using the manual | Users find GTS manual easy to use and understand. | PARTIALLY |
| Graphical output requires limited explanation (secondary) | Feedback from independent site testers operating the software and interpreting results | Users find GTS graphical outputs require limited explanation. | YES |
| Software reliability (primary) | Feedback from software beta testers | By end of project, GTS does not have any significant bugs. | YES |
| Release GTS as fully functional, stand-alone freeware (primary) | Complete/upgrade GTS interface and computational engine using open source and license-free runtime coding tools | GTS is free-to-use, stand-alone desktop application with a single (.exe) installer. | YES*<br>*Separate cost-comparison calculator is currently an Excel spreadsheet |
| Accessible to non-experts (primary) | Design user interface so that GTS can be run and interpreted by those without expert statistical training | GTS can be successfully performed and interpreted by mid-level analysts. | YES |
| Robustness (primary) | GTS analyses from a cross-section of site conditions and COCs | Can be applied across sites with a variety of COCs, hydrogeologic terranes, remedial solutions, etc. | YES |
| Water level-aided mapping (secondary) | Develop spatial mapping option that utilizes both concentrations and water head-level data | GTS can create maps based on either concentrations or a combination of concentrations and water-level data. | NO/PARTIALLY |
| **Quantitative Performance Objectives** | | | |
| Ease of use (primary) | Log of number and type of operational difficulties encountered by independent site analysts | GTS users encounter few operational difficulties. | PARTIALLY |

**Table 1. Performance Objectives. (continued)**

| Performance Objective | Data Requirements | Success Criteria | Criteria Met? |
|---|---|---|---|
| Reproducibility of temporal optimization (primary) | Quantitative comparison of temporal optimization results between GTS design team and independent site analysts | Expert and new users arrive at similar reductions in monitoring frequency using same site data and information. | YES |
| Reproducibility of spatial optimization (primary) | Quantitative comparison of spatial optimization results between GTS design team and independent site analysts | Expert and new users arrive at similar optimized network configurations (i.e., placement of wells) using same site data and information. | YES |
| Predictability (secondary) | Quantitative assessment of reserved validation data from each demonstration site | GTS Predict module successfully projects trend and plume estimates to encompass >90% of near future measurements. | PARTIALLY |
| Optimization effectiveness (primary) | Numerical measures of degree of temporal and spatial redundancy identified at each demonstration site, along with associated cost savings | GTS is able to identify significant redundancy in larger groundwater monitoring networks and can generate optimized sampling programs. | YES |
| Accuracy (primary) | Numerical comparisons between GTS base maps/trends and site concentration data | There is a low degree of statistical difference between original site data and GTS-constructed base maps and trends. | YES/PARTIALLY |
| Versatility (primary) | GTS analyses from larger sites with more than 200 well locations | Revised software is able to perform optimization at sites with >200 wells. | YES |
| ROI (secondary) | Cost-benefit analyses from demonstration sites | Projected ROI is ≤ 3 years at each site. | YES |

## 5.0 SITE DESCRIPTION

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

- *Data history* — Full temporal optimization in GTS requires a minimum of eight distinct monitoring events for most groundwater wells.
- *Network size* — Spatial optimization in GTS requires at least 20-25 distinct well locations; to achieve the performance objective for versatility (see Table 1), some sites with more than 200 well locations were required.

In selecting the sites, the project team also 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 to select each site from a different federal agency, if possible. Furthermore, the project team looked for willingness on the part of the site team to participate in the effort and consider implementation of results. Table 2 provides a summary of the demonstration sites.

**Table 2. Characteristics of demonstration sites.**

| | Air Force Plant 44 | Fernald Site | Former NOP |
|---|---|---|---|
| **Agency** | Air Force | Dept of Energy | Army |
| **Location** | Tucson, AZ | Ross, OH | Mead, NE |
| **Geographic location** | West (arid) | Mideast (Ohio valley) | Midwest (plains) |
| **Remediation system** | Pump-&-treat with 25 extraction wells | Pump & Treat after extensive excavation of contaminated soils | Pump & Treat with 10 extraction wells |
| **Primary COCs** | TCE, chromium, 1,4-dioxane, 1,1-DCE | Uranium | TCE and RDX |
| **Aquifers evaluated** | SGZ, UZUU, UZLU, LZ | Single aquifer | SHALLOW, MEDIUM, and DEEP aquifers |
| **Sampling frequency** | Quarterly (most wells) | Quarterly (most wells) | Semi-annual, but varies by well |
| **Monitoring network** | 208 (206 at risk) | 467 wells and DPT locations (376 active) | 250 (177 at risk) |

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

### 5.1 SITE LOCATION AND HISTORY

#### Air Force Plant 44, Tucson, AZ

AFP44 is located in the northern portion of the Tucson Basin within the Sonoran Desert section of the basin and range physiographic province in southern Arizona (see Figure 19). The basin is bounded on the west and south by the Sierrita, Black, and Tucson Mountains, on the south and southeast by the Santa Rita Mountains, and on the east and north by the Empire, Rincon, Tanque

Verde, Santa Catalina, and Tortolita Mountains. Elevations range from 2500 feet above sea level in the center of the basin to 9400 feet above sea level in the Santa Rita Mountains.

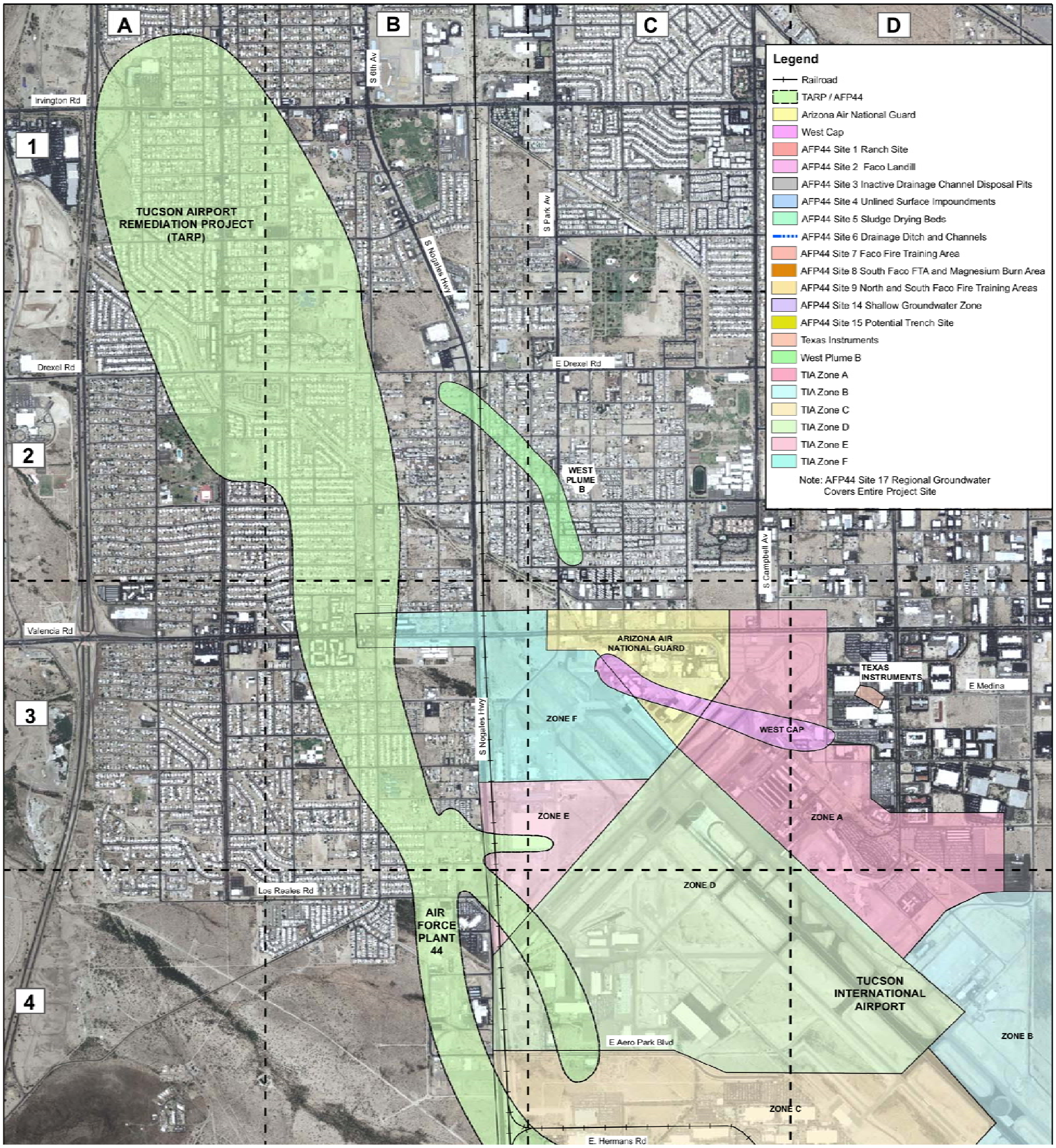



**Figure 19. Air Force Plant 44, Tucson, Arizona.**

Weapons manufacturing at AFP44 began in the 1950s and continues today at the government-owned, contractor-operated facility. From the 1950s through the mid 1970s, hazardous materials were stored, handled, and disposed in a manner consistent with widely accepted industry practices of the time. Releases to the environment occurred involving primarily chromium and chlorinated solvents, including trichloroethene (TCE) and 1,1,1-trichloroethane (1,1,1-TCA).

The primary known release sources included sludge drying beds, unlined lagoons, degreasers, and uncontrolled landfills. Chlorinated solvents associated with AFP44 are present in off-site groundwater to the northwest, commingled with the same compounds released from other nearby sites.

Groundwater impacts were discovered in the early 1980s at AFP44 and were investigated by the USAF to define the extent and magnitude of the contamination. An extensive drilling and sampling program, followed by a human health risk assessment (HHRA), led to the identification of several sites where contaminant concentrations were sufficiently elevated to warrant remediation.

Remedial actions at AFP44 were initiated in 1986 with the implementation of a site-wide groundwater extraction and injection system referred to as the Groundwater Reclamation System. The groundwater treatment plant (GWTP), which treats groundwater collected by the system, was designed to remove both chromium and chlorinated solvents from extracted groundwater at rates up to 5000 gal per minute. Chromium treatment was discontinued at the GWTP in 1994 when treatment switched to a well head system that targeted only those wells where chromium exceeded the maximum contaminant level (MCL). The Groundwater Reclamation System continues to treat chlorinated solvents in groundwater, with some modifications implemented in the 1990s to maximize contaminant mass removal. After 22 years of operation of the groundwater treatment system, as well as successful operation of five soil remediation systems, the chlorinated solvent plume in the regional groundwater has been significantly reduced.

Sampling was conducted for 1,4-dioxane at AFP44 in the early 1990s; however, no detections were noted in analytical results. An improved, more accurate method of sampling (USEPA Method 8270, Modified) was developed to analyze 1,4-dioxane at a lower detection limit. The new method allows 1,4-dioxane to be detected at 1-2 ppb detection levels as opposed to the older detection level of 100 ppb.

**Former NOP, Mead, NE**

The former NOP occupies approximately 17,250 acres located 0.5 miles south of the town Mead in Saunders County, NE (Figure 20). 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 assembly 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 USEPA National Priorities List of Superfund sites in August 1990 because contamination was identified in the groundwater and the soils at the site, and the release of contamination from this site is considered to be a potential threat to public health, welfare, and the environment.

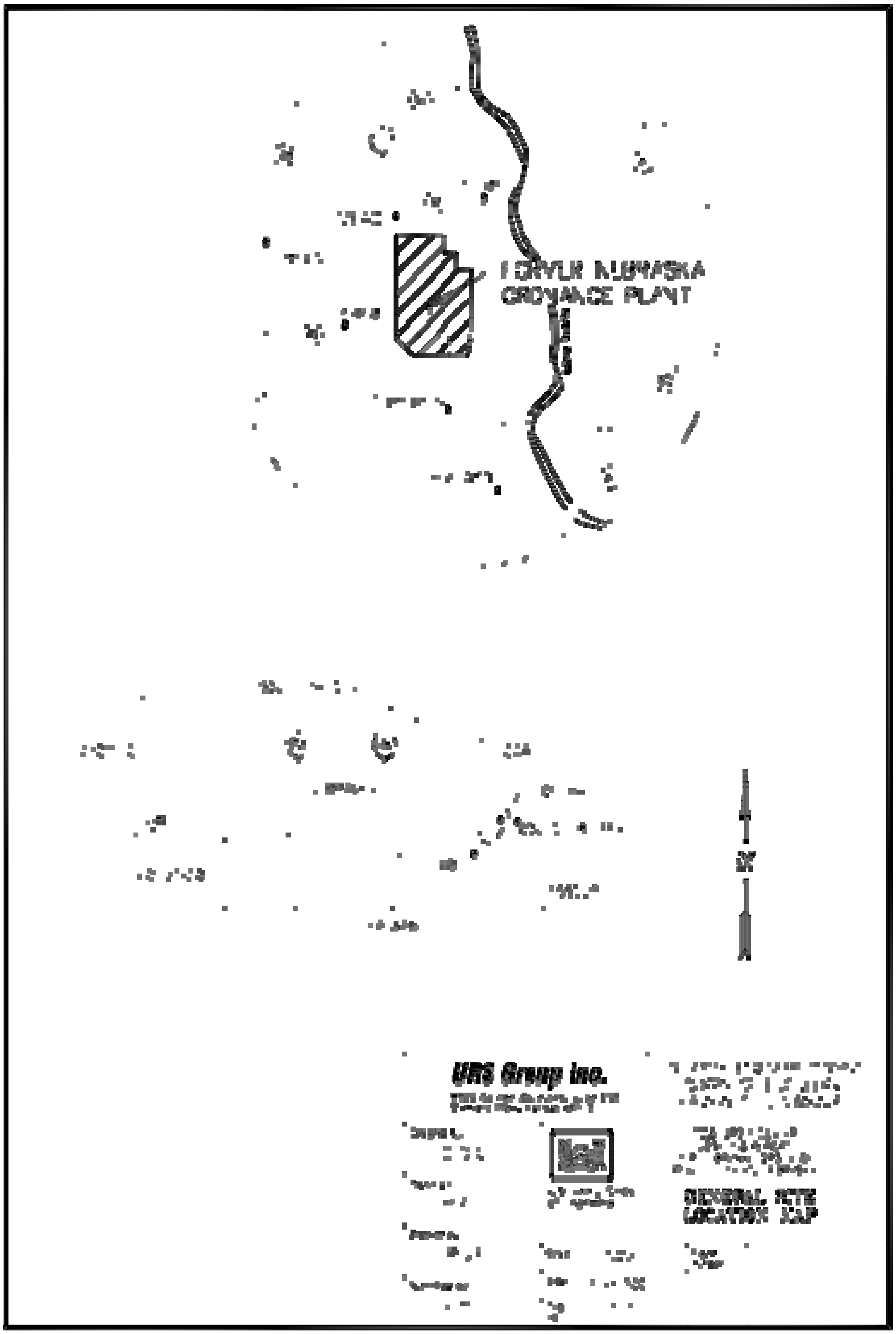



**Figure 20. Former Nebraska Ordnance Plant (NOP), Mead, NE.**

### Fernald DOE Site, Ross, OH

The Fernald Site is located near Ross, OH, about 18 miles northwest of Cincinnati (Figure 21). It occupies 1,050 acres of land, 136 of which were covered by buildings when DOE had active operations there. Its mission was to produce uranium metal for use as fuel in DOE nuclear reactors. The Fernald Site operated in this capacity for nearly 40 years, from 1952-1989, before being shut down. Altogether, 462 million pounds of high-purity uranium metal were produced, along with 2.5 pounds of waste per pound of refined uranium. Thus, approximately one billion pounds of waste materials were stored at the facility during its operational life.

After production activities at the site ceased in 1989, the 1990s were dedicated to site remediation activities, including the demolition and removal of buildings, the excavation of contaminated soils, and the construction of an on-site disposal facility as a repository for demolition debris and contaminated soils. In addition, historical site activities had resulted in groundwater contamination that migrated off-site, with uranium the primary contaminant of concern. Active remediation (pump and treat) was used to contain and treat contaminated groundwater. In the early 2000s, primary remediation activities at the site were completed, leaving only active groundwater remediation taking place, along with its associated groundwater monitoring network.

## 5.2 SITE GEOLOGY/HYDROGEOLOGY

### AF Plant 44, Tucson, AZ

The Tucson Basin is a broad, northwest-trending alluvial valley encompassing approximately 750 square miles in Pima County. AFP44 is situated at the western margin of the Tucson Basin. The Tucson Basin is located in the Alluvial Basin Hydrogeologic Province and the Basin and Range Geologic Province. These provinces are characterized by alluvial material that consists of clays, silts, sands, and gravels that eroded from the mountains and filled the basins. The coarser material is generally found near the mountains, while the finer material is found toward the center of the basins. Discontinuous layers of sand and gravel are encountered toward the center of the basins and probably represent ancient stream sedimentation.

The mountains bounding the Tucson Basin consist of crystalline igneous, metamorphic, and sedimentary rock. Geologists assume that AFP44 is underlain at great depths by crystalline rock consisting of granite, granite-gneiss, schist, andesite, basalt, and limestone that make up the mountains adjacent to the basin.

Several thousand feet of alluvial sediments deposited in the Tucson Basin are interbedded locally with volcanic flow, agglomerates, and tuffaceous sediments. The alluvial sediments that underlie the site have been characterized as belonging to four groups, which in descending stratigraphic order are surficial deposits, Fort Lowell Formation, Tinaja Beds, and the Pantano Formation.

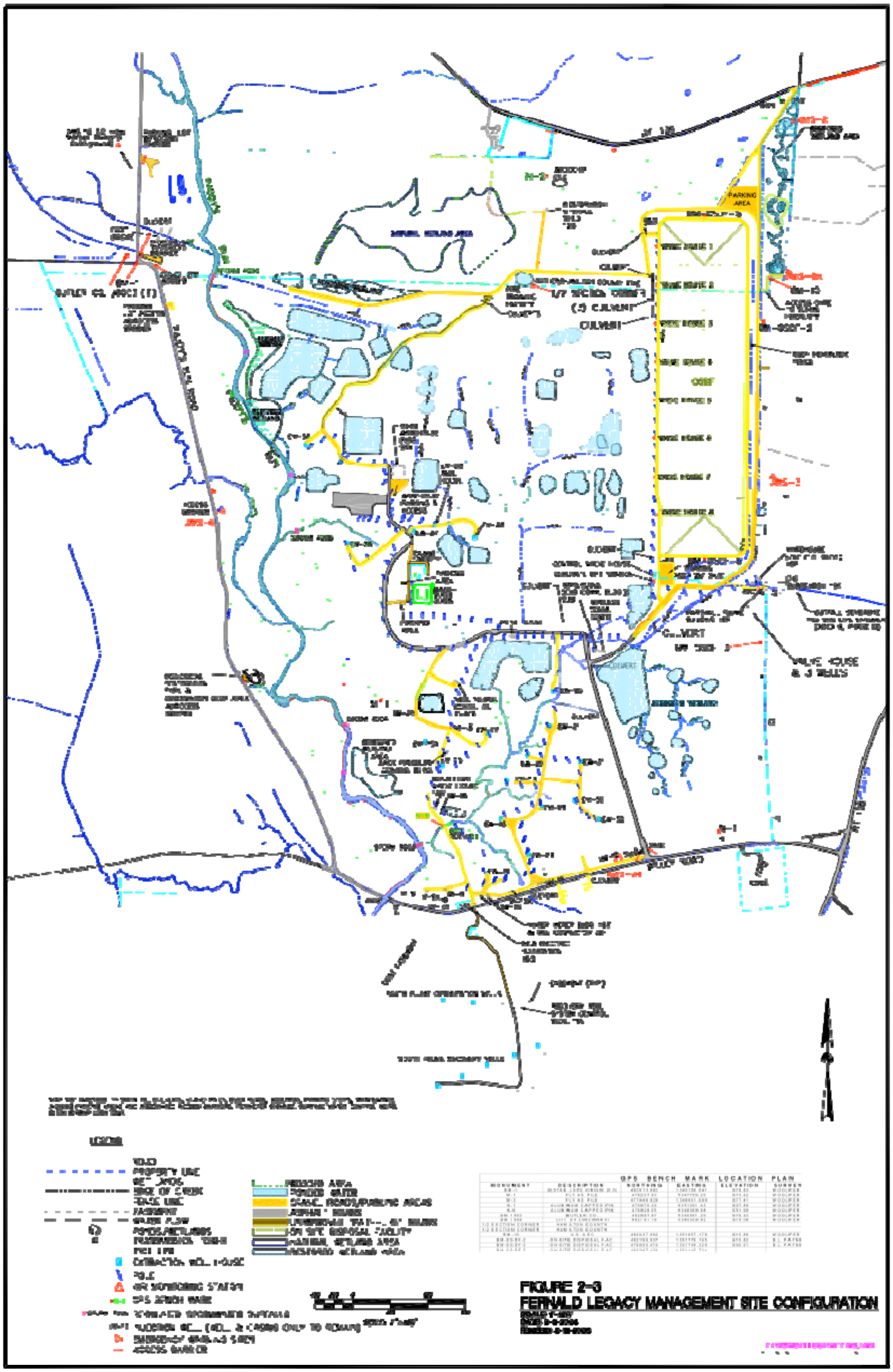



Figure 21. Fernald DOE site, Ross, OH.

The general hydrogeology beneath AFP44 includes a perched shallow groundwater zone (SGZ) and a regional aquifer (Figure 22). Within the regional aquifer at AFP44, an upper zone and a lower zone are separated by a clay aquitard. Within the upper zone, an upper unit and a lower unit are also separated by a clay aquitard. These units pinch out to the north and west and are therefore not hydrogeologically significant in the vicinity of AFP44.

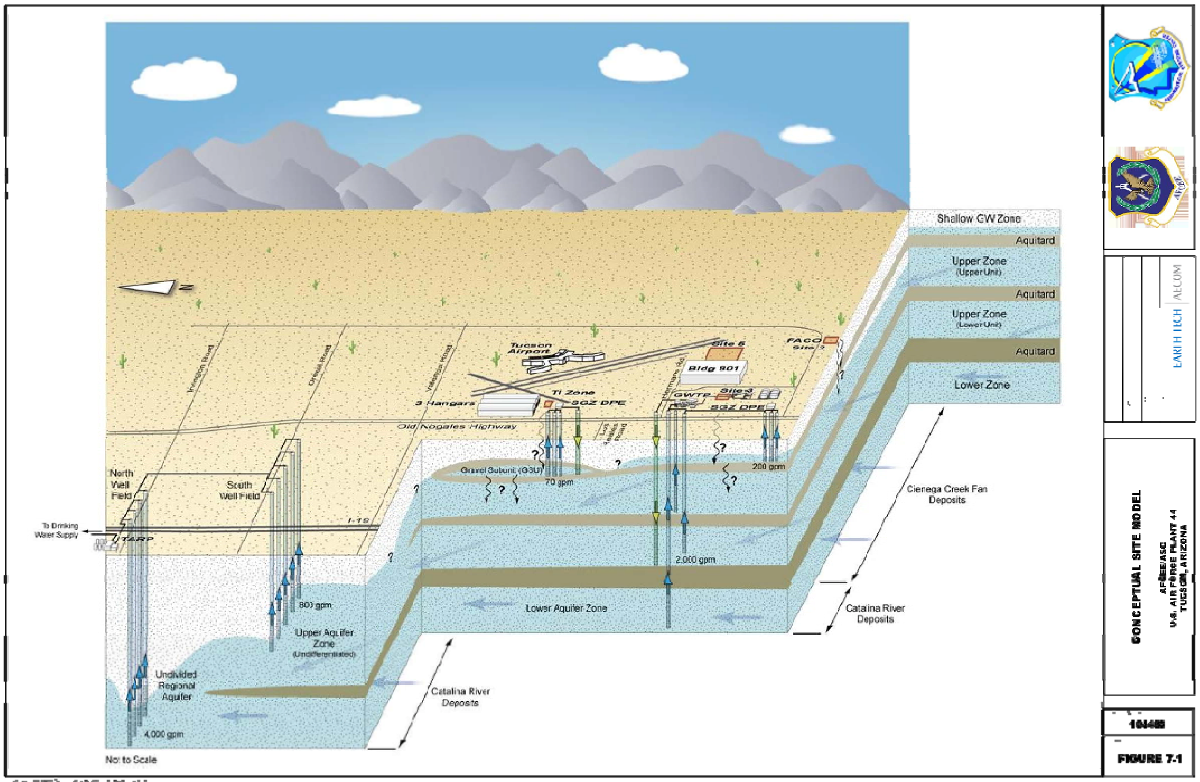



**Figure 22. Conceptual site model at AFP44.**

The SGZ consists of partially saturated silty clay, identified in the northwest portion of AFP44 and comprising an estimated 70 to 100 acres. The SGZ has a highly heterogeneous, complex region of inter-layered sandy clay and clay with numerous thin lenses of sand and gravel. Vertical migration of fluid is restricted by a distinct clay aquitard between the SGZ and underlying upper aquifer zone.

The upper aquifer zone, located in the Fort Lowell Formation, consists of gravelly sand with some clayey sand and sandy clay to a depth of 200 ft bgs and ranges in thickness from approximately 60 to 100 ft. This zone is underlain by a relatively impermeable layer of clay and sandy clay. The clay layer ranges in thickness from 100 to 160 ft and restricts the movement of groundwater between the upper and lower aquifer zones. Groundwater occurs in this upper zone under unconfined to semi-confined conditions.

The lower aquifer zone is located in the Pantano Formation and consists of clayey sand with lenses of gravelly sand and sandy clay. The top of the lower aquifer zone is approximately 300 ft bgs. Groundwater occurs in the lower zone under semi-confined conditions.

**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 ft. The unconsolidated material consists of topsoil, loess, 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 23).

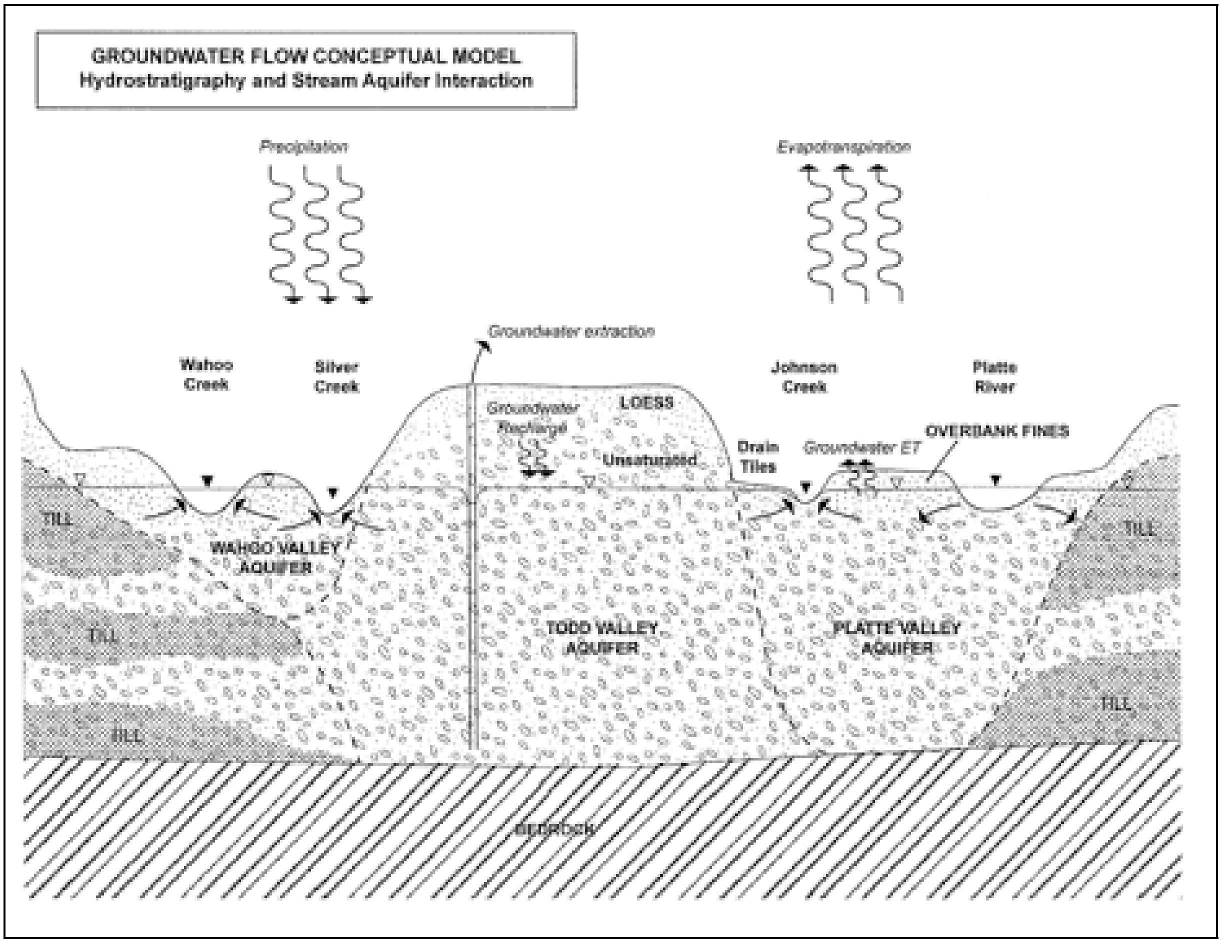



**Figure 23. NOP conceptual site model.**

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 aquifers. 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 ft thick) and a lower sand and gravel unit (17.5-72 ft thick). The upper sand unit is overlain by 4-23 ft of Peoria Loess. The unconsolidated material in the Platte River Valley (i.e., in the immediate vicinity of the Platte River) ranges in thickness from 39 to 49 ft. Overbank silts and clays ranging from 10-17 ft 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 ft to 58.0 ft. 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 ft.

### Fernald, Ross, OH

The Fernald site occupies approximately 1050 acres of land 18 miles northwest of Cincinnati (Figure 24). The former production area occupied approximately 136 acres in the center of the site. Paddy’s Run flows north to south along the western boundary of the site. The Great Miami River flows generally north to south to the east of the site before turning to the southwest south of the site. The site is situated on top of glacier overburden, consisting primarily of clay and silt with minor amounts of sand and gravel that overlies the Great Miami Aquifer. The Great Miami Aquifer itself contains a non-continuous clay interbed that separates the Great Miami Aquifer into an Upper and Lower portion.

The Great Miami Aquifer is underlain by shale inter-bedded with limestone. Paddy’s Run has eroded the glacial overburden, exposing the sand and gravel that make up the Great Miami Aquifer. Groundwater flow in the Great Miami Aquifer, in general, is to the east, southeast, and south across the facility, towards the Great Miami River.

The Fernald Site is located within a buried valley glacial outwash aquifer system, covered by younger glacial overburden. There is a perched groundwater system contained within this glacial overburden. The overburden is composed principally of clay-rich till having a sustainable groundwater yield of approximately 1 gal per minute. Horizontal flow is substantially greater than vertical flow, ranging from 1 to 58 ft per year horizontally but only 0.85 to 2.15 ft per year vertically.

The main aquifer consists primarily of well-sorted sand and gravel material. It has a sustainable yield of 400 gal per minute, with horizontal flow ranging from 400 to 1000 ft per year.

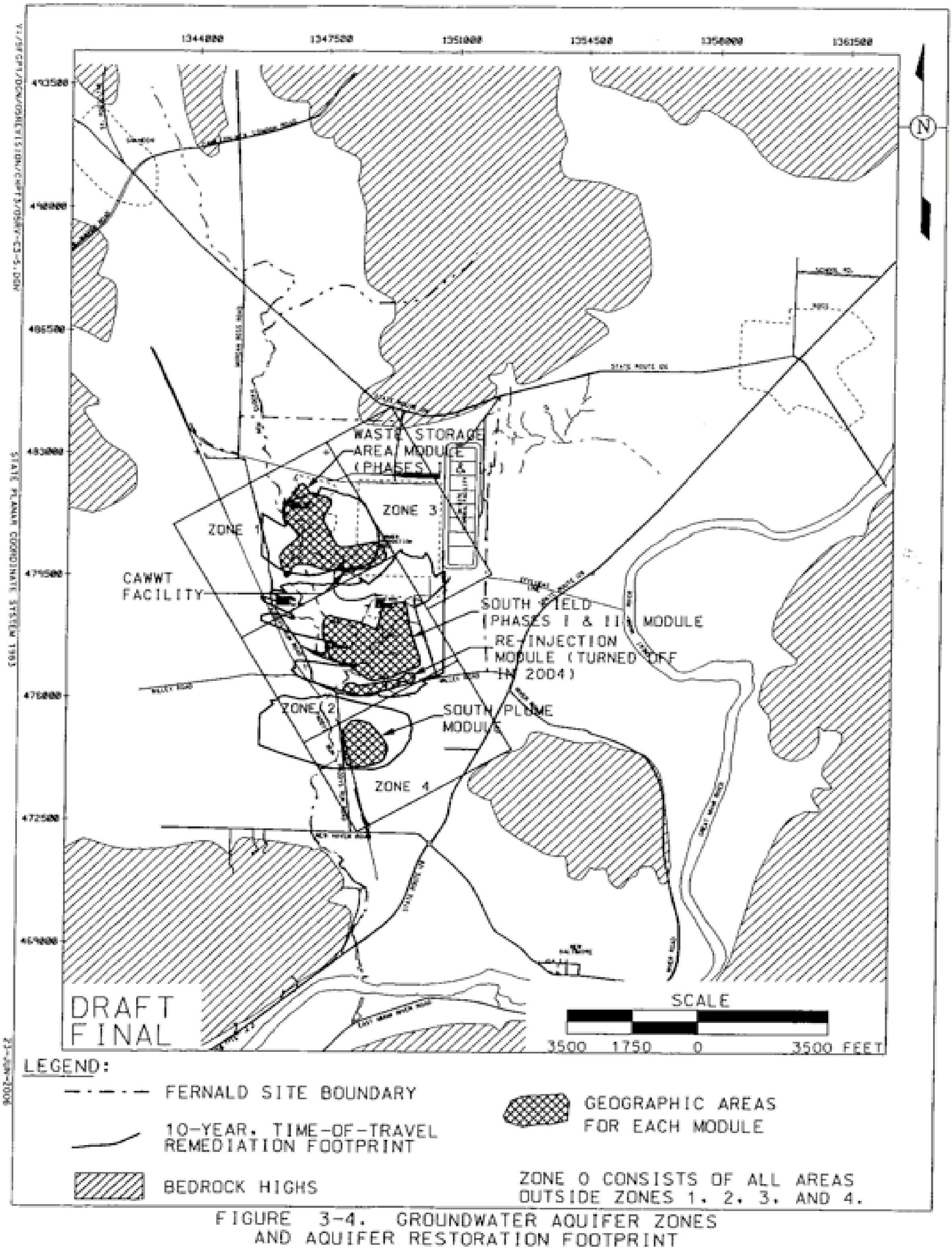



**Figure 24. Fernald groundwater aquifer zones.**

## 5.3 CONTAMINANT DISTRIBUTION

### AFP44, Tucson, AZ

The extent of contamination at AFP44 is described in the comprehensive HHRA for 1,4-dioxane in groundwater that was completed in 2004. It related to 1,4-dioxane at AFP44 but also addressed potential risks to receptors north of AFP44 within the footprint of the 1,4-dioxane plume in the regional groundwater. See Figure 25 for a map of plume extent.

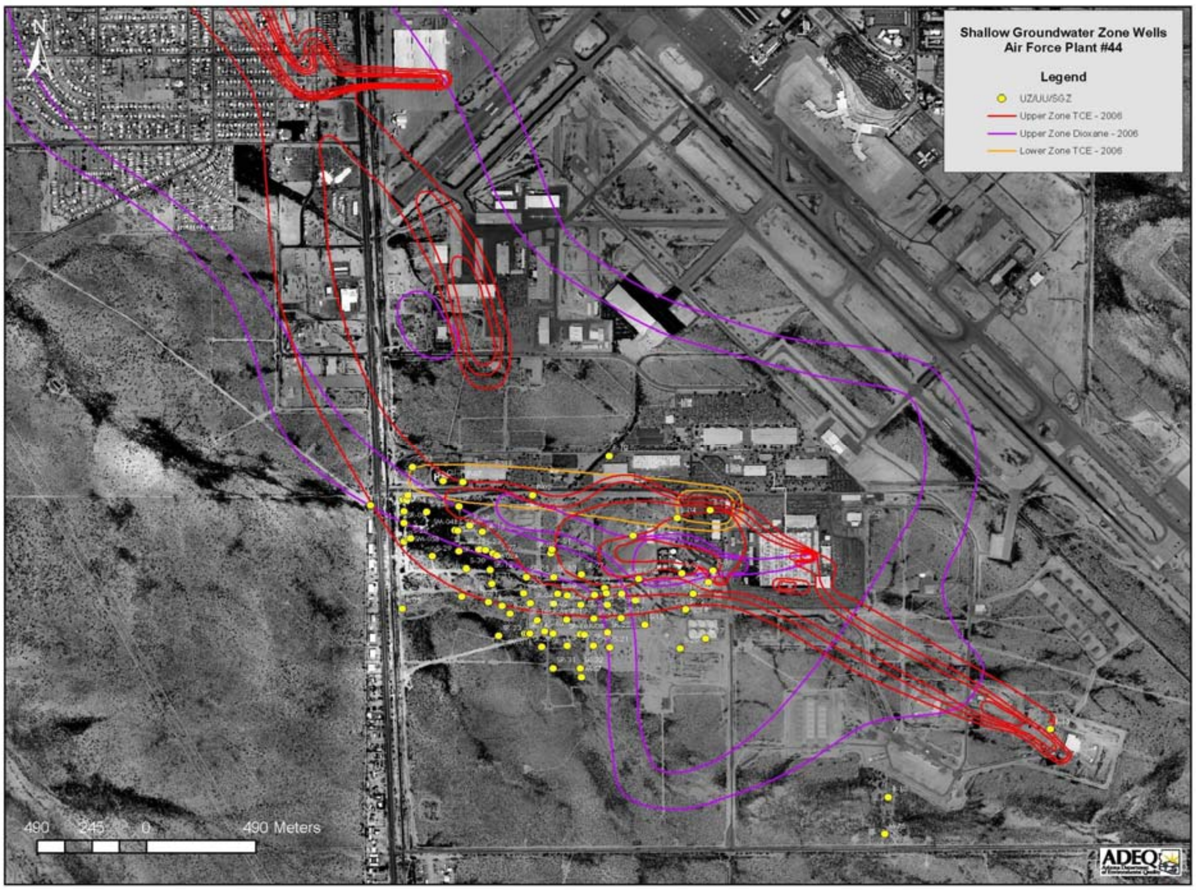



**Figure 25. Plume extent at AFP44.**

Prior to detection of 1,4-dioxane in groundwater, three contaminants had been detected in groundwater at levels that exceeded either promulgated groundwater standards or human health risk-based criteria—these included TCE, 1,1-dichloroethene (1,1-DCE), and chromium (total). Concentrations of other chemicals, including degradation products of TCE, 1,1-DCE, and 1,1,1-TCA, were infrequently detected at concentrations below respective screening criteria. The area downgradient of AFP44 also has TCE and 1,1-DCE contamination in regional groundwater above 5 and 7 ppb, respectively, that covers the area north-northwest to approximately Irvington Road. A groundwater containment system is already in place at AFP44 to reduce or eliminate off-site migration, thereby managing these COCs.

1,4-dioxane, a stabilizer for 1,1,1-TCA, has also been identified in groundwater in the vicinity and downgradient of AFP44. Drinking water extraction wells operated by the City of Tucson are located within the downgradient area of contamination. Groundwater is treated through an air stripping system prior to its distribution in the City of Tucson water supply. The City of Tucson has stated that all water supplied to the community through their water system will be at or below 3 ppb for 1,4-dioxane.

As an emerging contaminant, since the completion of the HHRA, additional investigations of 1,4-dioxane in the vicinity of AFP44 and downgradient of AFP44 have taken place and have found the levels ranging from non-detect to 11 ppb in 2006, from non-detect to 16 ppb in 2007, and from non-detect to 8.8 ppb in the spring of 2008. At AFP44 itself, a 2008 round of groundwater monitoring yielded 1,4-dioxane results from 144 wells that ranged from non-detect to 1400 ppb.

**NOP, Mead, NE**

The following volatile organic chemicals (VOCs) and explosive compounds were identified at the site (primary COCs are indicated with an asterisk):

VOCs —

- TCE*
- Methylene chloride
- 1,2-dichloropropane

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)

The site generally distinguishes plumes based on TCE and RDX (Figure 26). The four plumes (or “lobes”) of groundwater contamination identified at the site include:

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

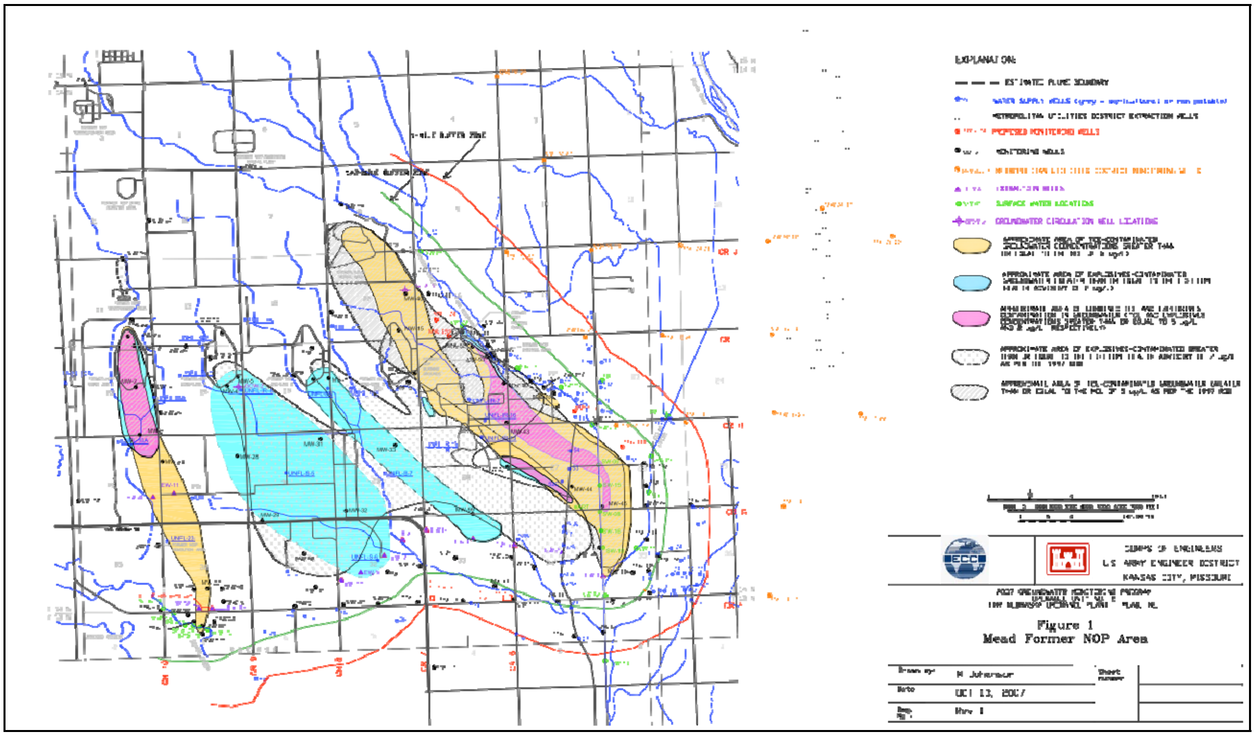



**Figure 26. NOP plume extent.**

**Fernald Site, Ross, OH**

The primary contaminant (COC) at the site is dissolved uranium, consistent with historic operations at Fernald. As noted, the site produced high purity uranium metal from 1952 through 1989. During that time period a significant amount of uranium was released to the environment, resulting in contamination of soil, surface water, sediments, and groundwater on and around the site. While there were other COCs besides uranium, uranium was by far the most significant and extensive contaminant of concern in environmental media, including groundwater.

During the 1990s and early 2000s, site remediation took place. High-level wastes were shipped off-site for disposal. Low-level contaminated material including building debris and soils were placed in an on-site disposal facility constructed for that purpose. The remediation process included deep and extensive excavations to remove soils contaminated with uranium that were believed to be sources for observed uranium groundwater contamination.

Groundwater contamination of the Great Miami Aquifer is believed to have resulted from infiltration of contaminated surface water through the bed of Paddy's Run, the storm sewer outfall ditch, the Pilot Plant drainage ditch, and the waste storage area ditch. In addition, groundwater contamination resulted from the emplacement of uranium-contaminated wastes in disposal areas such as the South Fields, and subsequent uranium leaching. There is no significant groundwater contamination of the underlying bedrock. Uranium contamination is not uniformly distributed over the vertical profile of the Great Miami Aquifer. In general, contamination levels are highest in groundwater associated with the water table in the vicinity of original source areas, with the center of mass of uranium contamination becoming deeper as one moves downgradient with the plume, reflecting vertical gradients in groundwater flow and recharge of clean groundwater from infiltration through uncontaminated soils downgradient of old source areas (Figure 27).

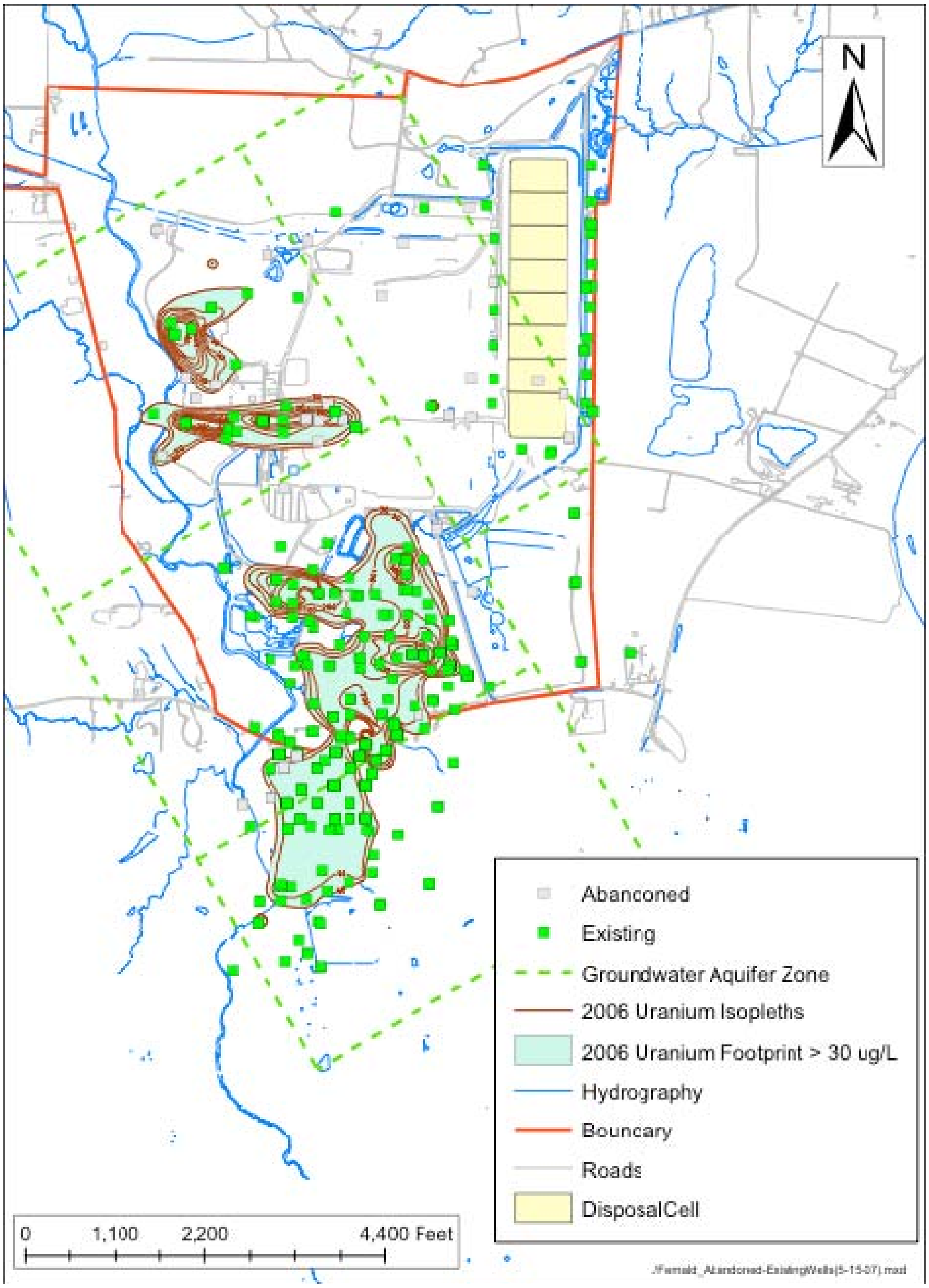



**Figure 27. Uranium extent at Fernald.**

# 6.0 TEST DESIGN

## 6.1 CONCEPTUAL EXPERIMENTAL DESIGN

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

- The ESTCP project team obtained preliminary approval and information from the site for review prior to site visit (including relevant descriptive reports and preliminary electronic data if available).
- The ESTCP project team conducted a site visit to present an overview of the GTS software and the project and to receive input from the site on specific issues and characteristics that might impact the optimization strategy. This input included overviews of the conceptual site model (CSM), data availability and format, contaminant drivers, and a tour of the site area.
- After discussion of the types of data needed to run a GTS analysis, the site and its contractors provided the ESTCP project team with the most updated version of historical sampling data in electronic format. This included not just analytical concentration data but also site boundary information and available water level data.
- Upon receipt of the electronic data, the ESTCP project team prepared the data for use in GTS. This preparation required the following steps:
    - Data screening and exploration — all historical concentration and water level data were examined for inconsistencies and obvious data quality issues. Significant questions or issues with the data were addressed to the site for possible resolution.
    - Data standardization — field names in the site data were standardized and matched to the expected GTS field name inputs.
    - Reserving the most recent year of sampling data for use in the GTS Predict module, in order to test the flagging of newer anomalous data against GTS baseline trend and plume estimates using the Trend and Plume Flagging features.
    - Creating tab-delimited (text) versions of the analytical data file, boundary file, and water level file (if separate from the analytical data) that could be directly imported into GTS.
- The prepared and standardized site historical data was supplied to both the ESTCP project team and the mid-level analyst responsible for performing an independent GTS optimization at that site.
- Two independent GTS analyses were performed using the same standardized data package: one by the (expert) ESTCP project team and one by the (non-expert) mid-level site analyst. The mid-level analyst supplied the ESTCP project team with a write-up of their results and the GTS project files they generated.

- The ESTCP project team analyzed the reserved last year's data by feeding it into the GTS Predict module. This was done to assess the functionality of the GTS trend flagging and plume flagging features. Summary reports were prepared of any anomalous data and the effectiveness of these techniques.
- Preliminary results of the optimization were communicated to representatives of the site, either via e-mail, phone, or in-person presentation (AFP44).
- Detailed comparison was made between the independent analyses conducted by the ESTCP project team and the mid-level site analyst in order to assess GTS usability, functionality, and reproducibility. These comparisons are incorporated into this final report.

In addition to the general experimental design described above, the following activities were also performed:

1. To perform a "layered" or 2.5D optimization analysis in GTS, each well location must have an aquifer zone designation. At AFP44, a number of wells either had uncertain designations or long well screens that traversed two aquifer zones as specified in the CSM. After consultation with site hydrogeologists, two versions of the AFP44 data package were prepared—one in which the uncertain wells were assigned to the uppermost of the possible zone designations and another in which these wells were assigned to the lowermost zone. Both variations of the data were analyzed by the ESTCP project team, while only one variation was supplied to the mid-level site analyst.
2. At the NOP site, a comparative study was performed by applying the Summit Tools and MAROS software applications, using the standardized data package for NOP. This was done to prepare a white paper comparison between GTS, Summit Tools, and MAROS.
3. At the NOP site, the standardized data package had to be subsequently revised when the ESTCP project team discovered that approximately 2000 of the analytical data records were essentially duplicates of other records. These duplicates were removed, a revised data package was sent to the ESTCP project team and mid-level site analyst, and the expert optimization analyses at NOP were re-done using the revised data.
4. At the Fernald site, a substantial number of the historical sampling locations involved DPT, as opposed to other locations that were more permanent monitoring wells. To apply GTS to these data, closely-spaced DPT sampling events were relabeled as single "wells" in order to create an approximate data history for each such location.
5. DOE arranged for its contractor to apply the GTS software to an additional site (Paducah, KY), and provided feedback to the ESTCP project team by preparing and submitting a summary report (see Appendix E). In addition, AFCEE arranged for the AFP44 data to be analyzed by two independent site analysts with differing levels of experience. Both of their summary reports are included in Appendix B.

## 6.2 BASELINE CHARACTERIZATION

At each demonstration site, optimization with GTS could only occur after first establishing a set of baseline conditions, especially since the redundancy analysis is predicated on comparing alternate and potentially optimal sampling programs against the initial baseline conditions. To establish an appropriate baseline, the following steps were conducted:

- Historical data acquisition and preparation
- Developing an optimization strategy
- Creating a set of estimated baseline trends and plume maps within GTS
- Estimating costs of the baseline monitoring program.

Each of these steps is discussed in more detail below.

### Historical Data Acquisition and Preparation

The first critical step was to obtain historical data in electronic format from each site and to then prepare that data for import into GTS. This was done prior to actual testing of the revised software. Significant results or observations stemming from this process include:

- *Data Quality Review.* An initial review of data quality was imperative. The ESTCP project team found substantial numbers of missing or unavailable pieces of information in its initial requests for historical analytical measurements and water level data. Follow-up questions/requests for clarification and additional data were forwarded to each site representative. Data review included items such as consistency of well names, availability and consistency of x-y coordinates in a consistent coordinate system, consistency of reporting limits and method detection limits for non-detects, completeness of the electronic data, and the presence of duplicate records. The review also looked at consistency of screened depth intervals, aquifer zone designations, surface elevations, and the amount of available water level data. Furthermore, time series plots of the concentration data were made to determine if any wells exhibited unusual data histories that might reflect data quality problems. Although this step took several days of manual labor per site, it is necessary for application of any kind of LTMO software.
- *Input File Format.* Sampling data imported into GTS can have a variety of possible text delimiters separating the fields. However, tab-delimited format is recommended. The order of fields within a text data file is not important, but the field names must exactly match the list of acceptable names in the GTS users guide. Not all fields listed in the user's guide are critical to GTS analysis, though fields that help locate each measurement within a Cartesian coordinate grid or that identify a measurement's magnitude and type (i.e., detected, trace, non-detect) are. Also critical is the standard Chemical Abstracts Service (CAS) number for each chemical contaminant. GTS matches the CAS number against its internal database to determine chemical-specific information such as standardized name, toxicity, mobility, and common regulatory limits. GTS also assumes, except for radiologic parameters, that all units have been standardized to parts per billion

(ppb or $\mu$g/L) concentration, and that this designation is consistent across records for a given chemical.

- *Sampling Event Constraints.* Although a full optimization analysis in GTS requires at least 8 distinct sampling events, there is no requirement that these events be either evenly spaced or spaced at least, say, quarterly. It is also possible to have multiple sample measurements on the same chemical at the same well with the same sampling date (e.g., field or lab duplicates). Due to properties of the local regression mapping engine utilized by GTS, users are not forced to have only one measurement per location per sampling event, or to perform averaging or random selection of such data records. Furthermore, GTS automatically groups irregularly spaced measurements into discrete subsets representing non-overlapping periods of time. These discrete time intervals are the time slices discussed in Section 3.1.

- *Rules for Non-Detects.* Non-detects are a persistent feature of groundwater monitoring data. To reasonably account for non-detect sample records, GTS requires the user to supply four fields: a strictly numeric measurement/concentration field (PARVAL), a PARVQ field designating whether the sample is detected, non-detect, or a trace value, and fields for the method detection limit (MDL) and reporting limit (RL). Each of these fields is typically present within ERPMS-consistent databases, so the user does not need to further manipulate the data outside GTS. Within the program, a set of rules is followed in order to impute a value for each non-detect. Broadly speaking, non-detects with positive values in the PARVAL field are set to half that value on the assumption that PARVAL contains a sample-specific reporting limit, while non-detects with zero or missing PARVAL are set to half the RL or MDL, whichever is present. Note: other laboratory or data quality flags can be imported into GTS but are not used directly to impute non-detects. Instead, these flags can be examined by the user to help validate other information within the sample record.

- *Outliers.* During data preparation, the ESTCP project team screened each dataset for obvious data inconsistencies, something each user is encouraged to do prior to GTS import. However, within the program, GTS v1.0 provides two different algorithms for flagging potential outliers: temporal outliers and spatial outliers. Using these screening tools, users are able to tag and eliminate statistical discrepancies from subsequent analysis and optimization (including, for instance, "dilution outliers" where a non-detect has an unrealistically high RL due to multiple dilutions in the lab). The sample records flagged as outliers are not removed from the database, but simply removed from analysis. The user can also generate outlier reports to document which specific data records were not utilized.

- *Data Filtering.* To maximize user convenience during data preparation and to account for electronic "data dumps" that tend to be inherently messy from the perspective of data screening, GTS provides a filtering mechanism within its internal database once a dataset has been imported. Although the viewing and sorting options within the database are somewhat limited, users can create complex, multilevel filters to significantly pare the data to be used during analysis. For this ESTCP project, almost all the initial screening was conducted outside the program, primarily to ensure that both the ESTCP project team and the mid-level site analysts would begin working with the same datasets. In more typical applications, filtering can provide a very valuable tool for winnowing data to a desired subset.

**Developing an Optimization Strategy**

The strategy for performing GTS optimization varied somewhat at each demonstration site, based on site-specific characteristics and contaminant drivers. However, GTS utilizes one guiding principle and one over-arching assumption in optimization. The critical assumption is that GTS will be applied to sites with potentially too many sampling measurements rather than too few. With the exception of its network adequacy analysis and temporal variograms, GTS establishes optimality by removing data from the current monitoring system and identifying some portion of this data as redundant. It is therefore primarily designed to establish optimality by eliminating analytical data redundancy.

The related guiding principle is that redundancy can best be discovered by comparing concentration trends and maps estimated from the full (non-optimized) data against corresponding trends and maps constructed from reductions in the data (i.e., reduced-data sets). Reduced-data trends and maps that are identical or very similar to their full-data counterparts indicate the presence of redundancy, while significantly different trends and maps suggest that critical data has been lost during reduction.

Significant results and observations about this process include the following:

- *Numbers of Contaminant Drivers.* The number of critical COCs varied by site, based on the input and feedback of site personnel. At Fernald, the only key driver was uranium; this COC constituted by far the bulk of the raw dataset. No other parameters were sampled more than sporadically or at more than a few wells. At AFP44, the database was preselected by site contractors to include four key COCs: chromium, TCE, 1,4-dioxane, and 1,1-DCE. All four were considered to have widespread presence in groundwater and to thus be contaminant drivers, though 1,4-dioxane was not sampled in every aquifer zone. At NOP, seven contaminants were part of the database, including three explosives and four VOCs. NOP site representatives asserted that only two of these COCs were actual contaminant drivers: TCE and RDX. Results of the GTS COC ranking analysis at NOP bore out this assertion. TCE and RDX were judged by GTS to have the best optimization potential of any of the chemicals.
- *COC Ranking and Optimization Constraints.* To minimize overall computing time and resources, GTS currently sets an upper bound to four on the number of COCs that can be simultaneously optimized. Obviously, this maximum is arbitrary, but reflects the fact that most sites have only a handful of key contaminant drivers. Contaminants in datasets with larger numbers of COCs are screened and ranked using the GTS Explore module, specifically the COC ranking analysis. This analysis develops a ranking of optimization potential for each COC, based on

factors such as the areal extent and frequency of sampling, rates and areal extent of both detections and exceedances above regulatory limits, sample sizes in the database, and mobility and toxicity factors. In practice, the COC ranking analysis can be used to identify those contaminant drivers that are most useful for optimization. During the ESTCP demonstration, the COCs to be optimized were already pre-set by site personnel at Fernald and AFP44. At NOP, however, the ranking analysis was applied to all seven database contaminants; TCE and RDX not only emerged as the highest ranked COCs, but their ranks were substantially higher than any of the other contaminants. Consequently, only these two drivers were optimized (see Table 3) by the ESTCP project team. Note, however, that the NOP independent site analyst also optimized both TNT and methylene chloride in his final analysis (due to a software glitch that has since been corrected).

**Table 3. COCs used during GTS optimization by ESTCP project team.**

| Site | COCs Optimized |
|---|---|
| **AFP44** | TCE, chromium, 1,4-dioxane, 1,1-DCE |
| **NOP** | TCE, RDX |
| **Fernald** | Uranium |

- *Evaluation of Multiple COCs.* Because multiple contaminant drivers may be present, GTS can optimize multiple COCs (up to a maximum of four) simultaneously, either during redundancy analysis or when assessing network adequacy (i.e., need for new well locations). To accomplish this during temporal optimization, GTS computes an optimal sampling frequency for each COC (either per-well, per-aquifer zone, or per-site) and then computes the median optimal sampling frequency across the COCs. In spatial optimization, a critical index is computed for each distinct well location by computing the fraction of COC-time slice pairs in which that well was deemed critical to the network (i.e., non-redundant). If the overall critical index is less than 0.5 after all COCs have been analyzed, that well is flagged as redundant. When analyzing network adequacy, GTS computes and maps a unitless uncertainty index for each COC across the site based on coefficients of variation. New wells are suggested only at locations where multiple COCs exhibit high levels of uncertainty.
- *Evaluation of Multiple Aquifer Zones.* Because distinct aquifer zones may exhibit very different concentration patterns and thus distinct plume maps, GTS can analyze multiple aquifers or aquifer zones simultaneously during a given spatial optimization run. To do this, the user must either select a 2D (i.e., two-dimensional) or 2.5D (i.e., two-and-a-half dimensional) approach at the end of the Explore module and prior to creating base maps. The 2D option treats all well locations as if screened in a single aquifer or layer. Plume maps generated under this option thus approximate the concentration distribution across a single, horizontal plane. The 2.5D option by contrast allows for multiple, distinct aquifer layers to be analyzed sequentially, with separate maps and optimization results generated for each layer. The user does not need to segregate the data by aquifer layer or go outside the program to perform a full analysis; rather, the sorting,

analysis, and concatenation of results across layers is done automatically within GTS.

- *Multiple Zones at Demonstration Sites.* The same optimization strategy was pursued by the ESTCP project team and mid-level site analyst at the AFP44 and NOP sites. In each case, the 2.5D option was selected, due to the presence of multiple, distinct aquifer zones (note: the second site analyst at AFP44 selected a 2D analysis for comparative purposes). At the NOP site, each of the SHALLOW, MEDIUM, and DEEP aquifers was analyzed, with substantially different levels of spatial redundancy. At AFP44, the layering was more complex and less distinct. The topmost layer (SGZ) extends across only part of the site, while the next layer (Upper Zone) is divided into an upper and lower unit, Upper Zone upper unit (UZUU) and Upper Zone lower unit (UZLU). Furthermore, the deep Lower Zone (LZ) only contains a small number of screened intervals, making it difficult to perform a GTS spatial analysis on just that layer. As a consequence, both the mid-level site analyst and the ESTCP project team chose to combine the Lower Zone and the Upper Zone Lower Unit into a single aquifer layer for purposes of the analysis (LZ-UZLU). GTS includes a feature that allows such merging of aquifer horizons (as well as deletion of certain layers or unmerging of combined layers) within the program, without any alteration to the raw data. In sum, both of these sites were optimized using a 2.5D (layered) analysis, each with three distinct aquifer zones.

  The Fernald site was exceptional in two ways: (1) Based on initial input from site representatives, the hydrogeology at various depths was not considered distinct enough to warrant a 2.5D analysis. Indeed, within the raw electronic data, only a small percentage of the records was distinguished by aquifer zone; the vast majority did not contain an aquifer designation. Consequently, the ESTCP project team analyzed Fernald as a 2D, single layer optimization. (2) Unknown to the ESTCP project team, the mid-level site analyst at Fernald retrieved additional information from the site and subsequently filled in the missing aquifer zone designations, thus editing and altering the standardized data package that had been prepared. The analyst then proceeded to run both a 2D analysis and a 2.5D optimization using the filled-in zone designations in order to perform a sensitivity analysis. Of interest, the site analyst's report (Appendix D) indicates that concentration levels of uranium in the three most populated aquifer layers were quite similar, somewhat buttressing the choice of a 2D analysis. More discussion of these differences can be found in Section 6.5.5.

- *Multiple Plumes within an Aquifer.* Unlike MAROS and similar software, GTS does not use or require plume-specific information such as locations of source areas, or designations as to which wells monitor the source or the tail of the plume. Instead, GTS is designed to estimate a concentration map across the entire site area of interest (as indicated by either the convex hull around the observed well locations or a separate boundary file imported by the user). GTS is thus able to estimate multiple plumes (and hot spots, source areas, etc) within a bounded region. This feature was needed at the demonstration sites since, in each case for at least one of the contaminant drivers, there were either multiple plumes (uranium at Fernald; TCE and RDX at NOP) or multiple lobes off the same plume (TCE and 1,4-dioxane at AFP44).

- *Measuring Plume Error.* With any spatial mapping algorithm, discrepancies or errors will occur between the actual concentrations at unmeasured locations and the corresponding map estimates. The goal, of course, is to minimize this error, but inevitable trade-offs occur depending on how error is measured or weighted. With QLR—the mapping engine in GTS—an additional source of potential error occurs at measured locations since QLR is a smoother and not an interpolator like kriging. To gauge the accuracy of a base map, GTS considers the weighted errors or residuals between map estimates at known locations and the observed concentrations. However, GTS also assumes that the absolute magnitudes of errors in high-concentration areas (e.g., plume interiors) are not as critical as similarly sized errors in low-concentration areas (e.g., near plume or site boundaries). Therefore, GTS computes by default a kind of relative residual, in particular, the logarithm of the ratio between the map estimate and the corresponding known concentration. By computing residuals in this manner, less statistical weight is placed on larger discrepancies in high-valued areas, while more weight is given to significant discrepancies in low-valued regions.

  GTS also differentially weights the relative residuals according to the spatial density of the measured observations. Observations in more sparsely sampled areas are given greater statistical weight due to the fact that they inform a relatively larger share of the site areal extent, while observations in clustered locations individually receive lesser weight. Computation of these weights is achieved by computing the ratio of the area of the Voronoi polygon associated with each measured location divided by the total site area.

- *Protected Wells.* In developing an optimization strategy for each site, the ESTCP project team requested input from site personnel as to whether any well locations should be protected (i.e., excluded) from a redundancy search. These protected wells are always kept in the optimized sampling program, regardless of what happens to other locations. The NOP site requested that 77 locations, mainly site boundary wells, be protected from GTS optimization. At AFP44, only two wells were so designated. None were suggested by Fernald personnel; however, in reviewing information provided by the site, 91 of the 467 distinct locations (mostly monitoring wells) had been abandoned by the time of the ESTCP demonstration, yet still had valuable historical data. To account for this status and to avoid flagging an already abandoned well as redundant, those 91 locations were labeled as protected for purposes of GTS analysis.

  To protect wells in GTS, there are two possible methods: (1) the user can add a binary field to the data file outside the program (PROTECT_FLAG) and prior to data import; well locations with value 1 in this field are then treated as protected while those with value 0 are eligible for optimization; or (2) the user can designate selected wells as protected *within* GTS via a series of checkboxes when viewing the baseline network status display. The first method was utilized for all three sites during data preparation and standardization to ensure that the same data structure was utilized both by the ESTCP project team and the mid-level site analysts.

- *Temporal Optimization Strategy.* GTS offers two different temporal strategies to accommodate varying monitoring networks and data configurations. Temporal variograms identify the sampling lag associated with a lack of event-to-event correlation. Samples collected at smaller (shorter) lags exhibit correlation and hence some statistical redundancy. Despite this straightforward idea, accurately estimating the inter-event correlation at a single well generally requires a significant amount of sampling data. To get around this limitation, GTS pools data from multiple wells into a single, average per-well event-to-event correlation estimate. In practice, this estimate is sensitive to fractions and patterns of non-detect measurements, so that temporal variograms do not always clearly identify a range (i.e., the smallest sampling lag associated with zero inter-event correlation). Because of this difficulty, users are encouraged where possible to first consider the other GTS temporal strategy, iterative thinning. Iterative thinning is performed by necessity on each individual well; it also requires at least eight distinct sampling events per location in order to estimate the baseline trend. From that baseline, data are "thinned" (i.e., reduced) at random to assess the degree of redundancy and ultimately an optimal sampling interval.

  For this ESTCP project, sites were purposely sought with enough historical data to allow a temporal redundancy search by either of the two methods within GTS. This requirement, along with the GTS recommendation to use iterative thinning where feasible, led each site analyst to perform and report iterative thinning as the primary temporal optimization tool. For its part, the ESTCP project team ran both methods at each site to compare the results. More generally, some sites using GTS may not have enough historical sampling data to make iterative thinning feasible. In these cases, temporal variograms can often still be calculated (due to the pooling of data across multiple well locations), though there is no guarantee that a clear range will be identified.

**Creating a Set of Estimated Baseline Trends and Plume Maps within GTS**

The third step in baseline characterization was to create the baseline trends and base maps by which GTS gauges redundancy. Since almost all redundancy and, hence, optimality in GTS is assessed by numerical comparisons against the baseline trends and base maps, it is critical that the baseline estimates be consistent with the temporal and spatial patterns observed within the measured data. To ensure this, GTS utilizes nonlinear local regression as its fundamental estimation engine: 1-dimensional regression for trends and 2-dimensional regression for maps. Nonlinear local regression can generate realistic (concentration) estimates for a variety of complex data patterns, both temporal and spatial, including such examples as seasonality and local hot spots. GTS also attempts to make good default choices in order to parameterize each local regression model. In the event the defaults do not lead to reasonable models, the software provides diagnostic tools to enable the user to adjust the model for a better fit. Significant results or observations stemming from this process include:

- *Removal of Data Gaps in Trend Estimation.* One of most significant challenges for local regression is fitting a reasonable trend during periods of time when there are large gaps between measured sampling events, e.g., when a well is not sampled for a few years prior to new sampling. Attempts to extrapolate a local trend to these gaps may result in wildly inaccurate estimates. To avoid these difficulties, GTS attempts to identify any substantial data gaps and to then exclude data prior to such a gap from trend estimation. Significant data gaps were identified for certain wells at each of the three test sites, suggesting that irregularly spaced sampling is the norm rather than the exception in groundwater monitoring networks. Users are also encouraged within GTS to examine time series plots of contaminant-well pairs with potential gaps to make sure the gaps are visually substantial; any inconsequential gaps can be easily overridden prior to trend estimation.

- *Classification of Trend Types.* Unlike simple linear regression, building an accurate model using nonlinear local regression requires additional data. To ensure that only those contaminant-well pairs with sufficient data are fit by local regression, GTS classifies each possible trend as either LWQR (local regression), Theil-Sen (nonparametric linear trend), FLAT (all measured values constant), FLAT-ND (all sampled values non-detects), or INSUFFICIENT (not enough data). No trends are fit to FLAT or FLAT-ND cases (due to lack of data variability), or in cases with less than four sampled values (INSUFFICIENT). For contaminant-well pairs with four to seven measurements, nonparametric linear trends are constructed using the Theil-Sen method, and for all the rest with eight or more measurements, nonlinear local regression is utilized. Table 4 below lists the number of trends at each site classified by type.

**Table 4. Numbers of trends classified by type at demonstration sites by ESTCP team.**

| Site | # Insufficient | # Flat or Flat-ND | # Theil-Sen | # LWQR | Total |
|---|---|---|---|---|---|
| **AFP44** | 99 | 113 | 97 | 342 | 651 |
| **NOP** | 57 | 295 | 53 | 57 | 462 |
| **Fernald** | 209 | 13 | 28 | 217 | 467 |

- *Trend Bandwidth Selection.* Any local regression model requires selection of a bandwidth parameter prior to fitting. GTS computes a default bandwidth value for each model based on internal checking of the residuals resulting from a range of alternate bandwidths. Despite this, perhaps due to unusual data clustering or general data sparseness, the default bandwidth may lead to highly inaccurate trend estimates over one or more portions of the date range. The bandwidth parameter also controls the degree of local smoothing in the trend estimate: larger bandwidths tend to give smoother, less variable trends, while smaller bandwidths react more nimbly to quickly changing local concentration patterns. To ensure that a reasonable model is fit, GTS allows the user to visually check the bandwidth alternatives and to override, if necessary, the default bandwidth. Some of the mid-level site analysts spent considerable time checking and tweaking the local regression trend models, especially at NOP and Fernald, while others tended to stick with the default bandwidth selections (AFP44).

- *Estimation of Confidence Bands.* Besides the local trend estimate, GTS also computes an approximate 90% confidence band around the trend. This band is useful in its own right as an indication of whether or not the mean concentration level exceeds a regulatory standard at any given point in time. It is also used in temporal optimization during iterative thinning as the numerical demarcation identifying when a reduced-data trend no longer reflects the baseline pattern. This occurs when the reduced-data trend falls substantially outside or beyond the confidence band surrounding the baseline trend. GTS utilizes one of two methods for constructing confidence bands. If the trend type is LWQR, the trend analogue to a standard confidence interval is used which properly accounts for the differential weighting of points in each local neighborhood where a trend estimate is made. If instead the trend type is Theil-Sen, the linear trend is then bootstrapped to estimate the confidence band. Currently, GTS does not use Theil-Sen trend cases when executing iterative thinning. At the test sites, since 178 (22%) of 794 non-flat trends with more than four observations were classified as Theil-Sen, more complete estimates of the optimal sampling intervals might have been made had these trends also been utilized.
- *Estimation Mesh for Maps.* In building concentration maps across a site area, the area must be discretized and estimates computed at each of a mesh of points. This is done to limit computational time, since interpolation between mesh points is typically much faster than computation of the local regression estimate at a mesh point. GTS currently employs a default mesh of approximately 100 evenly spaced points but allows the user to override this value by either increasing or decreasing the target number of mesh points (Figure 28). All of the site analysts opted to retain the default mesh spacing in their analyses. More generally, there are other spatial regression schemes that utilize unequally spaced meshes, whereby areas with clustered sample points receive tighter mesh coverage, while areas with sparse sample points receive fewer (i.e., looser) mesh points. Such schemes may more effectively map local areas where the plume is highly variable than the current GTS implementation.

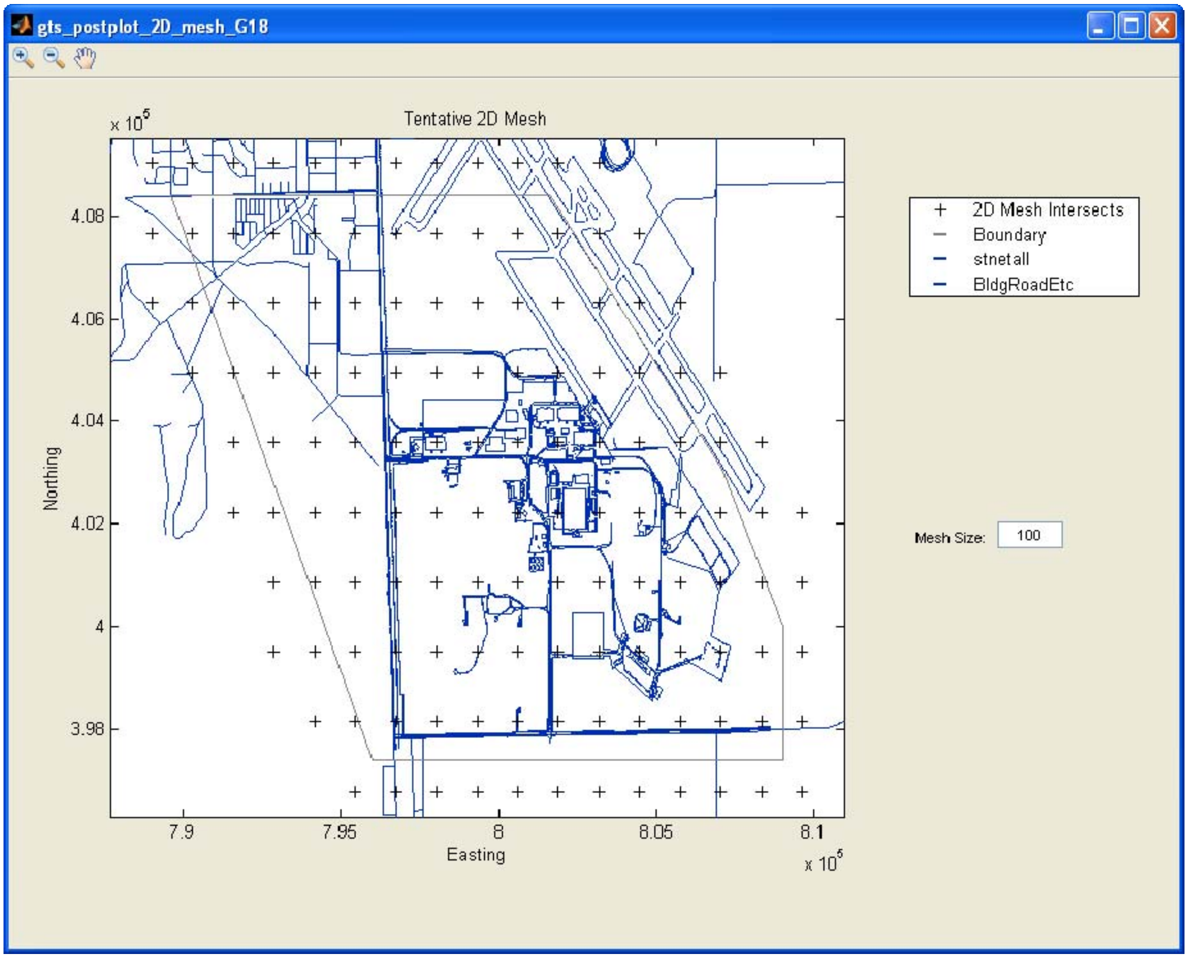



**Figure 28. Default GTS estimation mesh at AFP44.**

- *Declustered Cumulative Distribution Function.* To ensure that map estimates are consistent with the range of observed concentrations, GTS computes an empirical CDF to represent the statistical distribution of recent concentration levels. Each analytic observation sampled during one of the recent time slices is included in the CDF but weighted according to spatial density (Figure 29), that is, individual observations in clustered areas receive less weight than observations in more sparsely sampled locations to better reflect what proportion of the site is represented or informed by those concentration values. The net effect is that the weighting works to decluster the CDF estimate, resulting in a DCDF. The DCDF is used in turn by the QLR mapping engine to ensure that plume maps in GTS closely reflect the known concentration distribution and thus provide a more accurate baseline.

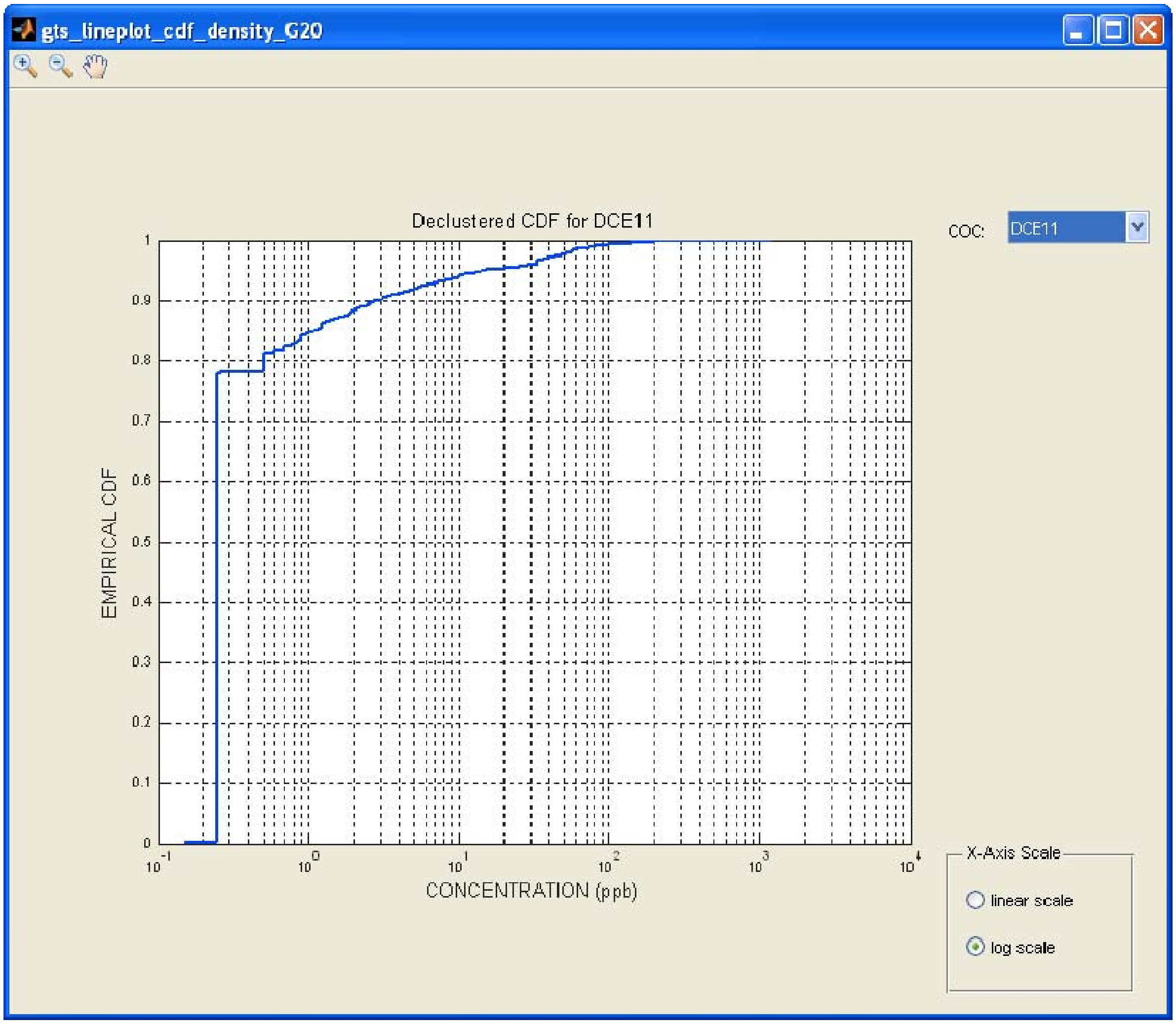



**Figure 29. Example declustered cumulative distribution function in GTS.**

- *Spatial Bandwidth Selection*. Like the local regression trend models, a bandwidth parameter must be chosen for each spatial regression model prior to constructing a base map. Using the weighted relative residuals described in **Developing an Optimization Strategy**, GTS computes a default bandwidth value for each map based on minimizing a series of diagnostic residual statistics across a range of possible bandwidths. If the default bandwidth does not result in an accurate or reasonable model, the user can override the default with a different bandwidth choice using a diagnostic interface within the program. The interface plots the relative residuals associated with each possible bandwidth as a color-coded post-plot (Figure 30). Residuals on the red end of the scale represent overestimates, blue residuals represent underestimates, and green residuals are close to the observed target.

  Although easy to use, some GTS testers suggested that the color-coded residuals did not provide enough diagnostic information to clearly identify superior regression models (i.e., base maps). At least three issues may have contributed to this impression: (1) The default regularly spaced mesh may not have allowed for fine enough interpolation around local hot spots, regardless of choice of bandwidth, leading to ill-fitting base maps. This could be improved by changing the mesh-building scheme within GTS to put more mesh points in the vicinity of clustered observations. (2) Plots of color-coded residuals are not the only useful diagnostic for selecting good bandwidths. GTS could be improved with additional spatial bandwidth diagnostic tools. (3) Because QLR is a *smoother* and not an *interpolator*, when low-valued and high-valued measurements are tightly clustered, the map estimate will necessarily be somewhere between, leading to the presence of both red residuals (overestimates) and blue residuals (underestimates) no matter what choice of bandwidth.

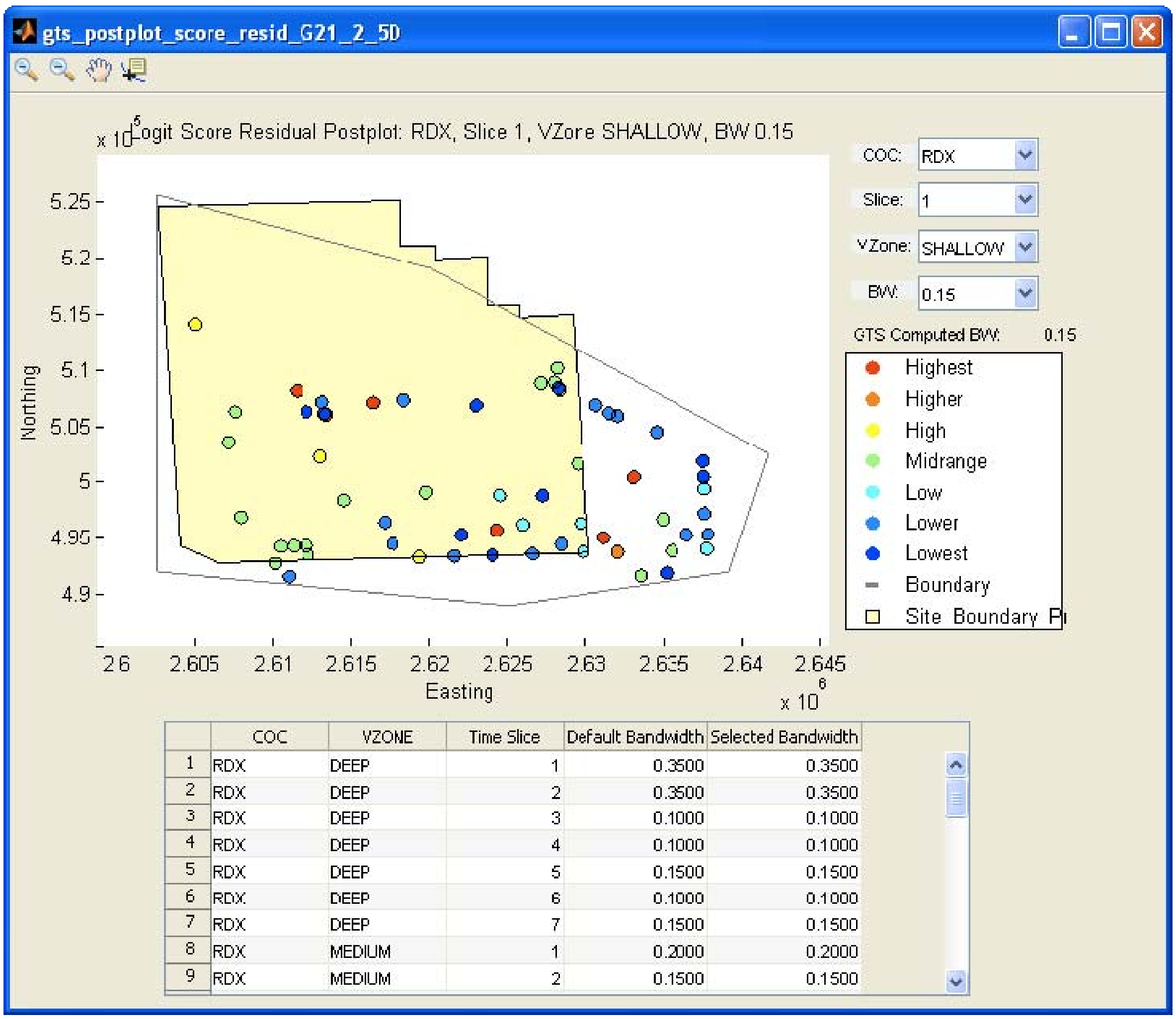



**Figure 30. Example residual post-plot.**

- *Multiple Time Slices, Multiple Zones.* GTS constructs a concentration base map of every contaminant for each time slice for which there is sufficient data as well as for each aquifer zone should multiple zones exist and when a 2.5D analysis has been selected. Having a base map for each time slice is, of course, useful for examining changes in plume extent and intensity over time, but the primary reason is to ensure that the optimization results in GTS are repeatable. That is, a given well can only be tagged as redundant for a particular contaminant if it is redundant across more than half the time slices. In this sense, GTS is fairly conservative when it comes to identifying spatial redundancy since the redundancy must exist relative to a majority of the base maps across the range of time slices.

Different aquifer zones are mapped separately to account for the possibility that groundwater concentration patterns may differ significantly by zone. This could be accomplished by performing multiple runs of the software, each run with a different subset of the data corresponding to a distinct aquifer zone. However, the GTS implementation adds significant ease of use by automatically mapping each aquifer zone separately when a 2.5D analysis is selected. Further, GTS allows the user to merge or delete specific zones for analysis purposes, a task that would be much more cumbersome outside the program. At AFP44, due to the small number of wells in the deepest aquifer zone and the somewhat fuzzy hydrogeologic distinction between aquifers at the site, both the site analyst and the ESTCP project team merged these wells into the UZLU to form a combined layer coded by GTS as LZ-UZLU.

**Estimating Costs of the Baseline Monitoring Program**

The final step in baseline characterization was to estimate the costs associated with the monitoring program at each test site prior to optimization. Site personnel and analysts were asked to provide site-specific estimates for laboratory and field sampling costs, as well as costs for factors such as mobilization, equipment, shipping, and labor rates. The current version of GTS includes a separate Excel spreadsheet into which results of an optimization run can be imported, and which guides the user in inputting baseline cost assumptions. The output of this spreadsheet is a realistic cost-benefits tally of the resources likely to be saved by implementing a GTS-optimized sampling program, including the ROI.

More detail on the baseline costs estimated at each site is provided in Section 8.3. Significant results or observations stemming from this process include:

- *Filled-In Cost Estimates.* Not every test site provided the full range of baseline cost estimates requested by the ESTCP project team. To generate cost savings at these sites, the GTS cost comparison calculator comes pre-loaded with costing assumptions that are fairly typical across the industry. These assumed costs were imputed to the missing values on the cost spreadsheet where necessary but are noted as estimates in Section 8.3.
- *Ease of Use Issues.* None of the independent site analysts completed or returned the GTS cost calculator spreadsheet. This was apparently because (1) the GTS cost calculator is a separate spreadsheet and not part of the main GTS application and therefore requires additional export of data from GTS and subsequent import and manipulation within the cost spreadsheet; (2) some of the site analysts did not

have access to the baseline cost assumptions for their site and therefore decided they could not complete the cost spreadsheet; and (3) time constraints. Ideally, the cost spreadsheet should be part of the main GTS application to encourage and ease its use (however, one site analyst opined that it should be kept as a separate application). Once in the spreadsheet, the process to complete a cost analysis is fairly straightforward but does require some user input and data manipulation. However, since none of the users completed this task, no direct comparison between the site analysts and the ESTCP project team could be made of the cost savings or ROI estimates. Instead, the cost savings reported in this report represent estimates made solely by the ESTCP project team.

## 6.3 TREATABILITY OR LABORATORY STUDY RESULTS

These items do not apply to this ESTCP project.

## 6.4 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 3.1 and illustrated on a flowchart in Figures 1, 3, 5, 6, 9, 10, 11, and 16. Further details are provided in the GTS software users guide, which has been provided as a separate deliverable for this project.

## 6.5 FIELD TESTING

Figure 3.1 charts the GTS v1.0 project and software testing schedule.

A summary of key results from testing of the GTS v1.0 software is presented in the following sections:

6.5.1. Schedule for Software Testing
6.5.2. Ease of Use, Installation
6.5.3. Software Bugs, Software Changes
6.5.4. Summary of Temporal Redundancy Evaluations
6.5.5. Summary of Spatial Redundancy Evaluations
6.5.6. Summary of Network Adequacy Evaluations
6.5.7. Summary of Trend and Plume Flagging Results
6.5.8. Import/Export Features
6.5.9. Computation Time/Level of Effort

### 6.5.1 Schedule for Software Testing

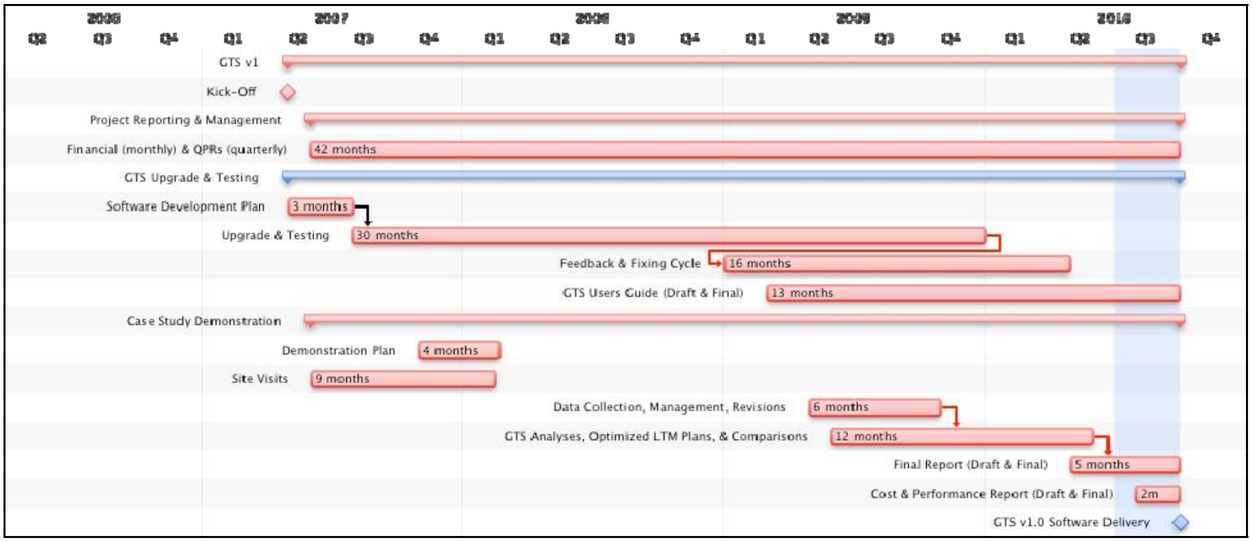



**Figure 31. GTS v1.0 project and software testing schedule.**

### 6.5.2 Ease of Use, Installation

Overall, the GTS software was found to be easy to use by the testers and mid-level site analysts. None of these users was formally trained on the software; questions regarding usage (and other project matters, including software bugs and development) were fielded in weekly conference calls sponsored by the ESTCP project team. Experience with other LTMO software varied among the testers; most had some previous experience running MAROS. Representative comments offered by testers concerning ease of use included the following:

> This tester rates the general usability of GTS as very good considering it is in beta form. Its modular structure is logical and relatively easy for the minimally experienced geostatistical practitioner to use. Installation and security and administrative rights elements of set up were performed by AFCEE/OSS personnel so the tester cannot adequately evaluate this component of the software. (AFP44)

> The five major modules coupled with Windows menu and dialog boxes allow an environmental professional with limited statistical training and expertise to navigate successfully through the many spatial and temporal elements of GTS. The GUI appears to be highly functional and user friendly. The ability on output graphs to change from linear to logarithmic units and to pan comprises a notable graphical robustness. (AFP44)

> The software is quite user-friendly. The screens are easy to navigate and read. The screen sequence is logical and appears to be structured to prevent a novice user from by-passing necessary steps. On the other hand, the ability to jump to other steps that have either already been conducted or that can be conducted based on the steps already completed make the program easy to navigate. (NOP)

Apart from bugs encountered during the Fernald application, GTS was easily used. The interface made sense and was clear. There are some relatively minor suggestions on improving the user experience described below. Based on my experience, GTS's major benefits are the exploration that can be done with data sets once loaded (outlier searches, data gaps, time series plots, etc.) The major impediments to its use will likely be the following: (1) difficulty in setting up the software and acceptable input files, (2) run times for some of the steps, (3) 'bugs' encountered during application, if my experience turns out to be representative, and (4) interpretation/reasonableness/defensibility of results." (Fernald)

The overall ease of use is good, as familiarity with the 5 main modules and their underlying windows comes fairly quickly. (Paducah)

The most consistent problems cited by users during this project related to ease of installation, data import/export (discussed in Section 6.5.8), and the level of interpretive detail offered in the users manual. GTS was certified to run only under a (32 bit) Windows XP or equivalent operating system environment. Users who attempted to install GTS under Windows Vista or Windows 7 were mostly successful but occasionally encountered glitches that prevented completion of the installation process. Additionally, several government users had to obtain special permissions and/or assistance from IT personnel in order to circumvent security firewalls. Frequently, the user had to install GTS on their computer as a system administrator in order for GTS to run properly.

Comments were received from some testers regarding the lengthy time required to initially install GTS. Updates to the software install fairly quickly. However, the first go-around requires installation of several separate software components, all related to the open source, freeware architecture of GTS. Once these components are installed, they do not have to be installed again except when a particular component has been upgraded. Specific comments related to installation included:

Installation should be easy for users with administrative privileges on their computers. For users without administrative privileges, installation can require significant intervention by a network administrator. Installation of multiple builds may cause problems. In my situation, two versions of the supporting program R were present (2.9.1 and 2.10.1). I deleted the older version (required administrator intervention), but then when GTS was opened, it couldn't find R. A deletion and re-install by the administrator was then needed. (Paducah)

Set up was a significant issue, primarily because we do not have administrative rights on our machines. In my case I was able, with the assistance of our system administrator, to install on my desktop but was unable to get GTS operational on my laptop (and abandoned trying once it was running on my desktop). (Fernald)

The installation process was somewhat lengthy, but relatively easy. The fact that the software uses a couple of proprietary run-time software means there are several steps to the installation that may be a bit confusing for novice computer users. This should not be an issue for the intended users, though, since they are likely to be quite computer literate. The biggest hurdle for DoD users will likely

> be that the software will require installation by IT staff with administrator rights. This is a problem for most software, although MAROS can be used without an installation, provided the user has Microsoft Access. (NOP)

As to the GTS users guide, testers found it straightforward but concise. Some comments indicated the manual should include more detailed help for interpreting GTS output and results. Representative comments included:

> The user's guide is well written and concise. There are a number of items and parameters that are not adequately explained, however. In some cases, the ramifications of making certain changes or parameter choices are also not explained. For example, "bandwidth" is not really explained before or at its first use in a way a new user would likely understand (I think my geophysics background helped me). The manual could more fully explain the ramifications of unflagging data points as outliers. Are they or are they not used? It seems they are not used. What happens to the later calculations if you don't change them? What happens if you do? The manual is silent on the genetic algorithm settings for the spatial optimization work. What are the trade-offs in changing the settings? Other questions for the manual: (1) What are the Logit scores? What are expansion factors? (NOP)

> The user's guide provides a good introduction to the GTS algorithm and helpful instructions in preparing input data files and navigating through the five modules and numerous submodules. (AFP44)

> The User's Guide was, in general, easy to understand and follow. However there were many times when I found the brief description of what GTS was doing inadequate. I would strongly suggest adding appendices that provide technical detail and references, when appropriate, for the various analysis methods and approaches embedded within GTS. (Fernald)

> The manual has been refined over the last half year and is in good shape. It is light on details, however. A companion guide that documents the math/stats involved in the various steps is recommended. (Paducah)

### 6.5.3 Software Bugs, Software Changes

GTS v1.0 represents a major overhaul and upgrade to the previous beta-version GTS v0.6. The software architecture was completely redesigned and all new software components/tools were utilized to build the new version, including a fundamental switch in the statistical/computational environment from Fortran to R, as well as a brand new interface and data housing structure. As such, a significant number of software bugs, logic flaws, and glitches were encountered during both the early internal testing of GTS, as well as in the external testing by the mid-level site analysts. Due to the project schedule, it was necessary to have most of the site testers begin their analysis prior to the final software release. While this caused some significant frustrations on their part, it had the beneficial side effect of identifying additional GTS bugs and flaws, issues that were addressed during the project. Apart from software design changes or suggestions that fell outside the scope of original project proposal, the ESTCP project team addressed each reproducible bug and flaw, resulting in the current final GTS v1.0 release. Tester comments related to software glitches included:

> Bugs and crashes were common in earlier builds, but the only known problem while analyzing with GTS using the 15March2010 version is the map legend issue described above. (Paducah)
>
> I encountered a number of problems as I worked through GTS, some of which were resolved by the GTS team, others of which are still outstanding. (Fernald)
>
> Given the difficulty in getting IT support for installation of various subsequent builds of GTS, I encountered a number of problems that potentially were related to the version I was using. In some cases it was related to the dataset I was using. I had reported a number of problems to the GTS team and either my mistake was identified or the code was updated. Due to time constraints and early bugs, I was not able to evaluate the Predict module to assess new data. I understand that the software has been used with the Mead dataset through this step by others. One problem I found with the March 2010 version was that I could not go back and reduce the number CoCs once I passed the CoC selection step. (NOP)
>
> The software tester encountered numerous bugs and runtime errors while running the GTS 29 Oct and 11 Nov builds, some of which were fatal, causing shutdown of GTS. These problems occurred both in the XP environment as well as in Vista. These runtime errors are described in detail in the next section. The 15 Mar 2010 version was run on Windows XP utilizing the input file used for the 2009 testing. No runtime errors or 'bugs' were encountered. (AFP44)

In addressing either internal or tester-identified issues, several modifications were made to GTS beyond the software development plan. In all, a total of 34 separate alpha or beta builds of GTS v1.0 were generated. Among the more significant changes:

- Modified the SQLite database structure to allow for data filtering and limited editing. Now within GTS, users can specify complex filtering criteria for creating specific subsets of the database with which to analyze. Immediately after data import, users can also edit individual records and/or fields.
- Improved the usefulness of GTS graphics by adding zooming and panning controls to each plot. Also added the ability on time series plots and other 2D line plots to switch between concentration and semi-log scales.
- Improved the utility of post-plots and maps by adding "tool tips" to allow the user to identify key information about specific well locations directly from the plot using the cursor, including well name, easting and northing coordinates, and relevant summary statistics.
- Improved the default identification and viewing of potential outliers in multiple ways. Early versions of GTS flagged far too many samples as outliers, requiring more work for a user to override non-outliers. The internal GTS logic for identifying both temporal and spatial outliers was made more conservative and accurate. Non-detects were visually identified on outlier plots to better distinguish

true outliers from non-outliers. Also, the user interface for examining spatial outliers was redesigned to allow the user to examine all measurements in the local neighborhood of a potential outlier. Finally, only cases with potential outliers are displayed, significantly reducing the number of plots a user must navigate to finalize the outlier list.

- Added the dot ranking chart for visually ranking and identifying contaminants most suitable for further optimization.
- Added an interface allowing users to merge and/or delete specific aquifer zones for purposes of analysis without having to manipulate the data outside GTS.
- Vastly improved the ability to save results in GTS. In the current version, users can request that their project be saved at almost any point in the program. Additionally, GTS internally stores the results of lengthy calculations and large batches of graphics so that those results/plots do not have to be recomputed unless other data has specifically been changed. This internal saving dramatically cuts down on run time.
- Changed the spatial mapping engine from multiple indicator local regression to QLR in order to substantially improve base map accuracy and also to dramatically speed map computation. In turn, this change speeds the lengthiest step in spatial optimization.
- Improved the method by which spatial residuals are computed and displayed when checking possible spatial bandwidths. Residuals are now computed on a logit-scale, in parallel with how the local regression estimates are generated. Calculation of residuals also now gives equal relative weight to underestimates and overestimates. Improved the internal method for computing default spatial bandwidths.
- Added an option for the user to easily change and visualize the spatial mesh at which map estimates are made.
- Further tested and improved the default parameters used to run the GTSmart spatial redundancy search, including the size of the network subset search space and the error criteria for identifying optimal networks.
- Added the critical index to the spatial optimization results to better identify redundant wells and to allow users to perform further graduated ranking of wells within the classifications of “redundant” or “essential.”
- Improved the utility of certain post-plots and water elevation maps by distinguishing locations by well type (e.g., monitoring well, extraction well, injection well, piezometer, etc.).
- Improved the utility of the trend flagging and plume flagging tools by allowing users to easily override suggested anomalies.

### 6.5.4 Summary of Temporal Redundancy Evaluations

GTS provides two tools to assess temporal redundancy—temporal variograms and iterative thinning. As discussed in **Developing an Optimization Strategy**, iterative thinning has proven to be a more reliable technique at many of the sites (both ESTCP and otherwise) at which GTS has been applied. However, it requires longer data histories at individual wells than temporal variograms and so is not always applicable. At all three test sites, enough historical sampling data was available to run (and compare) both tools. Presented below are the key results from those analyses, as well as a comparison between results obtained by the ESTCP project team versus the independent site analysts.

***Sampling Frequency Optimization Using Temporal Variograms***

Successful use of the temporal variogram requires that the variogram exhibit a distinct and easily recognized pattern, namely a continuous (and smooth) increase in variogram level as the lag time between sampling events increases, followed by a plateau or constant level when the variogram reaches its 'sill.' The sampling lag at which the sill is first achieved is known as the "range" and designates the point of zero correlation in concentration levels between pairs of sampling events spaced in time as much or more than the range.

Finding this kind of pattern can be difficult. Variograms with well-established sills usually require that (1) sample pairs exist in sufficient quantity at a variety of different lags in order to populate a significant range of possible sampling intervals; (2) concentration levels at most wells are reasonably stable (but not constant) over time so that trends do not overly influence the estimates of intra-pair correlations; (3) not too many wells included in the temporal variogram have non-detect or "flat" data histories (i.e., all or almost all measurements are non-detect or constant in value). Lack of variation in concentration levels precludes the ability to correlate sampling lags with concentration patterns.

At the ESTCP test sites, temporal variograms were easily computed but yielded poor to mixed results. Table 5 lists the number of approximate ranges identified by the ESTCP project team for each test site, against the number of temporal variograms computed. Overall, the results did not enable reliable or replicable estimates of optimal sampling intervals. At AFP44, a sill was evident at only three of 11 combinations of COC and vertical zone, including no cases for either TCE or 1,4-dioxane and no cases for the UZUU aquifer zone. On the plus side, both ranges identified in zone LU-UZLU for different COCs were close to 1200 days or slightly more than a 3-year recommended sampling interval.

**Table 5. Summary of temporal variogram results obtained by ESTCP team.**

| Site | Aquifer Zone | # COCs | # Sills Found | Median Sampling Interval | Range of Sampling Intervals |
|---|---|---|---|---|---|
| **AFP44** | LZ-UZLU | 4 | 2 | 1225 days | 1200–1250 days |
| | UZUU | 4 | 0 | — | — |
| | SGZ | 3 | 1 | 200 days | — |
| **NOP** | DEEP | 2 | 1 | 1500 days | — |
| | MEDIUM | 2 | 1 | 1500 days | — |
| | SHALLOW | 2 | 1 | 1250 days | — |
| **Fernald** | — | 1 | 0 | — | — |

The independent site analyst at AFP44 identified ranges for each combination of COC and aquifer zone, unlike the ESTCP project team. Further comparison of the respective results revealed that the independent analyst attempted to identify the range associated with a "secondary sill," so termed because it depicts a temporary plateauing of the variogram, followed by a further increase at larger sampling lags. This discrepancy between the AFP44 site analyst and the ESTCP project team underscores three important points:

1. Estimating optimal sampling intervals using temporal variograms is somewhat subjective, since the analyst must visually identify the sill (if it exists) and then flag the approximate range at which the sill begins. Although GTS documents whatever choice the user makes, multiple users may arrive at different estimates using the same data.
2. A secondary sill may or may not provide a nearly optimal sampling interval. On the down side, there will still be some correlation between sample pairs with lags longer than the range of the secondary sill and hence some statistical redundancy. On the other hand, a secondary sill usually represents a significant decrease in that correlation, leading to measurements that are often nearly independent with respect to sampling lag.
3. Description of the use and interpretation of temporal variograms in the GTS users guide may need to be more extensive. It is possible users may get the impression that they should pick a range regardless, whether or not a clear sill is evident.

Two COCs—RDX and TCE—were analyzed at the NOP site. Of these, only RDX resulted in variograms with identifiable sills, each with a range on the order of 3-4 years, depending on the aquifer. None of the TCE variograms reached a plateau. The independent site analyst at NOP did not find any identifiable sills, either for RDX or TCE. Upon further inspection, it was determined that his results were computed using a version of GTS that incorrectly limited the maximum range of sampling dates displayed by the temporal variogram. Thus, the sills for RDX were not evident on the variograms he examined. The final release version 1.0 of GTS has fixed this issue.

At Fernald, neither the ESTCP project team or the independent site analyst identified a sill for uranium, the only COC. Both analysts found the temporal variogram to be uniformly increasing over the possible range of sampling lags. As the site analyst put it:

> In the case of the Fernald data set, no sill was apparent (Figure 18), a result consistent with the fact that uranium concentrations have been gradually falling across the site over time. Whenever consistent temporal trends are present, one would not expect variogram sills to be evident.

***Sampling Frequency Optimization Using Iterative Thinning***

Iterative thinning is predicated on the notion of trend reconstruction. If a baseline trend can be accurately reconstructed using fewer and, hence, more infrequent measurements, an optimized sampling interval can be obtained by determining what level of sampling is still necessary to do an accurate reconstruction. As a corollary, the ability to generate the same trends should lead to equivalent decisions concerning whether regulatory standards have been exceeded, remedial action is necessary, or what kinds of temporal changes are occurring. Thus, although GTS v1.0 does not directly compute optimized sampling frequencies on the basis of probable regulatory exceedances or the pace and direction of concentration change over time (i.e., slope), such questions can be answered by the GTS approach. Further, unlike other existing LTMO methods for temporal optimization, the combination of using iterative thinning and local regression for trend fitting accounts for two ubiquitous features of groundwater monitoring: nonlinear temporal patterns, including complex and/or seasonal trends, and irregularly spaced sampling events.

In the current implementation, GTS attempts to optimize any contaminant-well pair with at least eight distinct sampling events and for which the measurement levels vary with time (i.e., not uniformly non-detect or flat). Many sites, including the ESTCP test sites, have such data histories. However, the number of eligible contaminant-well pairs can vary significantly, usually by contaminant, depending on general contaminant levels and sampling schedules (e.g., COCs may be sampled on differential schedules leading to different accumulated data histories). Table 6 lists the number of contaminant-well pairs analyzed by iterative thinning at each site, along with the basic results generated by the ESTCP project team. Important observations from this table include:

- At AFP44, 1,4-dioxane had not been sampled frequently enough to enable iterative thinning at contaminant-well pairs involving this COC. As such, the optimization results at this site are based on 1,1-DCE, TCE, and chromium.
- At AFP44, many wells were still being sampled quarterly (1Q) at the time of the demonstration, so much so that the median baseline sampling frequency was quarterly in each aquifer zone except for SGZ, where the baseline frequency was semi-annual. Iterative thinning suggested that most trends could be adequately reconstructed using an annual sampling effort instead, an overall 75% reduction in the current schedule.
- At NOP, relatively few contaminant-well pairs were eligible for iterative thinning. Although data existed for 462 contaminant-well pairs, 295 (64%) of these were always non-detect, 57 (12%) had an insufficient number of sampling events to fit any trend, and 53 (11%) had only enough data to fit a Theil-Sen nonparametric linear trend (but not the eight events required to do iterative thinning). That left 57 (12%) eligible pairs. On one hand, the small number of pairs might seem to provide a weak justification for recommending a change in sampling frequency. However, the vast majority of pairs that were always non-detect could conceivably be sampled at any frequency and still give the same result. So the key to temporal optimization are the contaminant-well pairs with variable trends, even if fewer of those exist.
- At NOP, the majority of wells in each aquifer zone were sampled semi-annually (2Q) at time of the demonstration. Iterative thinning suggested that adequate trend reconstruction could be done based on annual (4Q) sampling in two of the three aquifer zones, and every three quarters (3Q) in the remaining SHALLOW zone. Overall, the GTS analysis recommended roughly half the level of sampling effort as was currently being conducted.

- At Fernald, since the only COC analyzed was uranium, there was a 1-1 correspondence between the total number of wells and the total possible number of contaminant-well pairs. However, at 209 (45%) of the 467 locations, the data were insufficient to fit any trend, primarily because most of this group of wells was in fact DPT-type geoprobes, and thus temporary sampling locations rather than permanent wells. Another 13 (3%) locations were always non-detect for uranium, while 28 (6%) only had enough distinct sampling events to be fit via a nonparametric linear trend (Theil-Sen). The remaining 217 (46%) were analyzed with iterative thinning.
- At Fernald, a large majority of the wells with sufficient data were being sampled, on average, quarterly (1Q) at the time of the demonstration. The GTS analysis recommended an overall reduction in sampling frequency to once every three quarters (3Q), based on the median optimal sampling interval, a reduction in sampling effort of roughly 67%. However, at this site (and more so than the other two) there was significant variation in the well-by-well iterative thinning results (see Figure 32). In fact, 100 (46%) of the optimal intervals were either every two quarters (2Q) or quarterly (1Q). Closer examination of the results showed that 30 of these wells were being sampled weekly at the time of the demonstration. So a reduction in sampling frequency to quarterly at these locations was fairly substantial.

**Table 6. Summary of iterative thinning results obtained by ESTCP team.**

| Site | Aquifer Zone | Total # Wells | Eligible Pairs | Base Median Sampling Interval | Optimal Median Sampling Interval |
|---|---|---|---|---|---|
| **AFP44** | All | 208 | 342 | 1Q | 4Q |
| | LZ-UZLU | 69 | 133 | 1Q | 4Q |
| | UZUU | 85 | 136 | 1Q | 4Q |
| | SGZ | 54 | 73 | 2Q | 5Q |
| **NOP** | All | 250 | 57 | 2Q | 4Q |
| | DEEP | 58 | 16 | 2Q | 4Q |
| | MEDIUM | 96 | 21 | 2Q | 4Q |
| | SHALLOW | 96 | 20 | 2Q | 3Q |
| **Fernald** | — | 467 | 217 | 1Q | 3Q |

***Iterative Thinning Comparison between ESTCP Project Team and Site Analysts***

A comparison was also performed between iterative thinning results generated by the ESTCP project team versus those submitted by the independent site analysts. Key results of this comparison are shown in Table 7 and Figure 32. In general, both sets of analysts at AFP44 and NOP computed fairly similar results using GTS on the same data, underscoring the reliability of GTS as a computational tool. More significant differences were found at Fernald, as discussed below. Important observations include:

- The recommended site-wide optimal sampling intervals were identical for both the expert and independent site analyses at AFP44 and NOP. The only differences occurred in aquifer zone-specific recommendations—once at AFP44 and once at NOP. In each case, the median optimal intervals differed by one quarter in length.

- At Fernald, the data sets imported into GTS differed significantly between the ESTCP project team and independent site analyst (see **Developing an Optimization Strategy**). In particular, the Fernald analyst eliminated most of the geoprobe locations and any wells outside a fairly central and smaller area than that delineated by the site boundary utilized by the ESTCP project team. As a consequence, the Fernald analyst employed a total of 172 well locations in his analysis, contrasted with the 467 locations used by the ESTCP project team. Due to the difference in input data, it is somewhat difficult to make a direct comparison in results. Even the baseline frequencies differ—in the commonly supplied data set, 164 (76%) of 217 eligible wells had baseline sampling frequencies that were either weekly or quarterly (1Q). In the data set used by the Fernald analyst, 93 (77%) of 121 eligible wells had semi-annual (2Q) baseline frequencies, while only 22 (18%) were quarterly or weekly.
- Despite these obvious differences in the two Fernald analyses, both teams computed a lengthening of the optimal sampling interval by two quarters on average, and a typical reduction in sampling effort of at least 50%.
- At Fernald, the site analyst performed additional follow-up analyses of the iterative thinning results. He found that:

  > There was a correlation noted between base sampling frequency and the GTS-recommended frequency. The longer the base sampling frequency, the longer was the GTS-recommended sampling frequency. Ideally one would want the 'optimal' sampling frequency to be independent of the original sampling frequency.

  Actually, the correlation is entirely consistent with the fundamental assumption that GTS is appropriate only for sites with too much sampling data, rather than too little. Iterative thinning always attempts to remove data prior to trend reconstruction. This guarantees that the optimal sampling interval will never be shorter than the baseline interval; hence, the longer the baseline interval, the longer the optimal interval.
- The Fernald site analyst also noted that:

  > There was no correlation between the GTS-recommended sampling frequency and the average concentration for a well. One might expect that wells that are significantly and consistently elevated above cleanup guidelines, or significantly and consistently below, might be of lesser interest from a sampling frequency perspective than wells that have concentrations around the action level.

  This finding underscores how GTS is primarily concerned with trend reconstruction, regardless of concentration level. Other strategies for temporal optimization clearly exist, but it is also true that if a historical trend can be accurately reconstructed, the same regulatory or remedial decisions—one way or the other—will likewise tend to be made.

- The comparative histograms in Figure 32 for AFP44 of the individual contaminant-well, pair-specific optimal sampling intervals are very similar in shape and magnitude. A Kolmogorov-Smirnov comparative test of the two distributions found a highly non-significant p-value of 0.994, underscoring the visual similarity. Greater differences are seen in the comparative histograms for NOP, though the two distributions still exhibit similar patterns, enough so that the Kolmogorov-Smirnov test gave a non-significant p-value of 0.526.
- The comparative histograms in Figure 32 for Fernald of the individual contaminant-well, pair-specific optimal sampling intervals are fairly distinct, apparently due to the differing data sets that were analyzed. The Kolmogorov-Smirnov comparative test of the two distributions is highly significant ($p<0.0001$), confirming the visual differences. It is also clear that more of the optimal sampling intervals computed by the Fernald analyst are longer than those calculated by the ESTCP project team, much of this due to the longer average baseline intervals within the data set utilized in the independent analysis.
- When exactly the same data is analyzed (unlike the Fernald case), it can lead to differing individual optimal sampling intervals for at least four reasons: (1) Choice of outliers—the user is responsible for selecting a list of outliers to exclude from analysis. The choice here may impact which trends have sufficient data for iterative thinning. (2) Choice of COCs—the user must select which COCs to analyze. At NOP, the site analyst included in his final run methylene chloride and TNT along with RDX and TCE as contaminants to be optimized. The ESTCP project team only included RDX and TCE, since the other contaminants were ranked as having much poorer optimization potential. During iterative thinning, this difference in COC choice led the NOP analyst to optimize 80 contaminant-well pairs, as opposed to the 57 analyzed by the ESTCP project team. (3) Choice of temporal bandwidth—the user must review and finalize a temporal bandwidth for each contaminant-well pair that will be subjected to iterative thinning. Different bandwidths impact the smoothness of the trend and sometimes how much data is needed to reconstruct it accurately. (4) Thinning process—iterative thinning involves drawing subsets at random from the data history of a given contaminant-well pair. Although this process is repeated many times and the results averaged, the same pair might occasionally yield different results on different runs through the iterative thinning routine.

**Table 7. Comparison of iterative thinning results.**

| Site | Aquifer Zone | ESTCP project team Base Interval | Independent Site Analyst Base Interval | ESTCP project team Optimal Interval | Independent Site Analyst Optimal Interval |
|---|---|---|---|---|---|
| **AFP44** | All | 1Q | 1Q | 4Q | 4Q |
| | LZ-UZLU | 1Q | 1Q | 4Q | 4Q |
| | UZUU | 1Q | 1Q | 4Q | 4Q |
| | SGZ | 2Q | 2Q | 5Q | 4Q |
| **NOP** | All | 2Q | 2Q | 4Q | 4Q |
| | DEEP | 2Q | 2Q | 4Q | 5Q |
| | MEDIUM | 2Q | 2Q | 4Q | 4Q |
| | SHALLOW | 2Q | 2Q | 3Q | 3Q |
| **Fernald** | — | 1Q | 2Q | 3Q | 4Q |

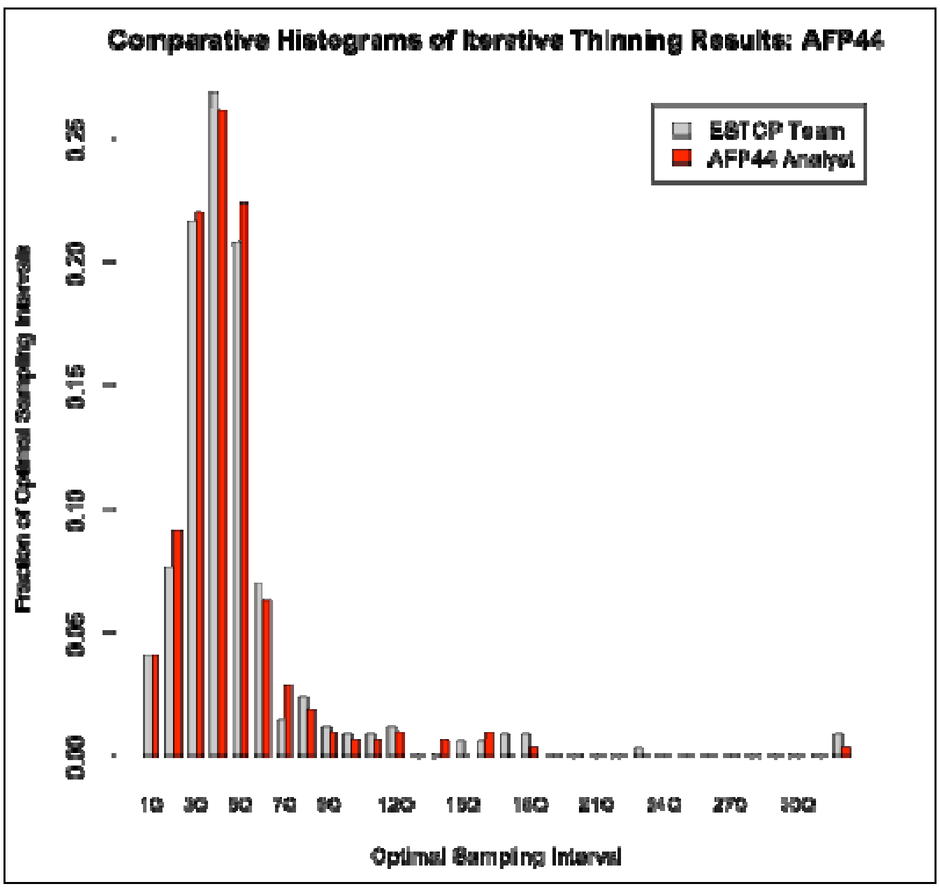



**Figure 32. Comparative histograms of individual optimal sampling intervals.**

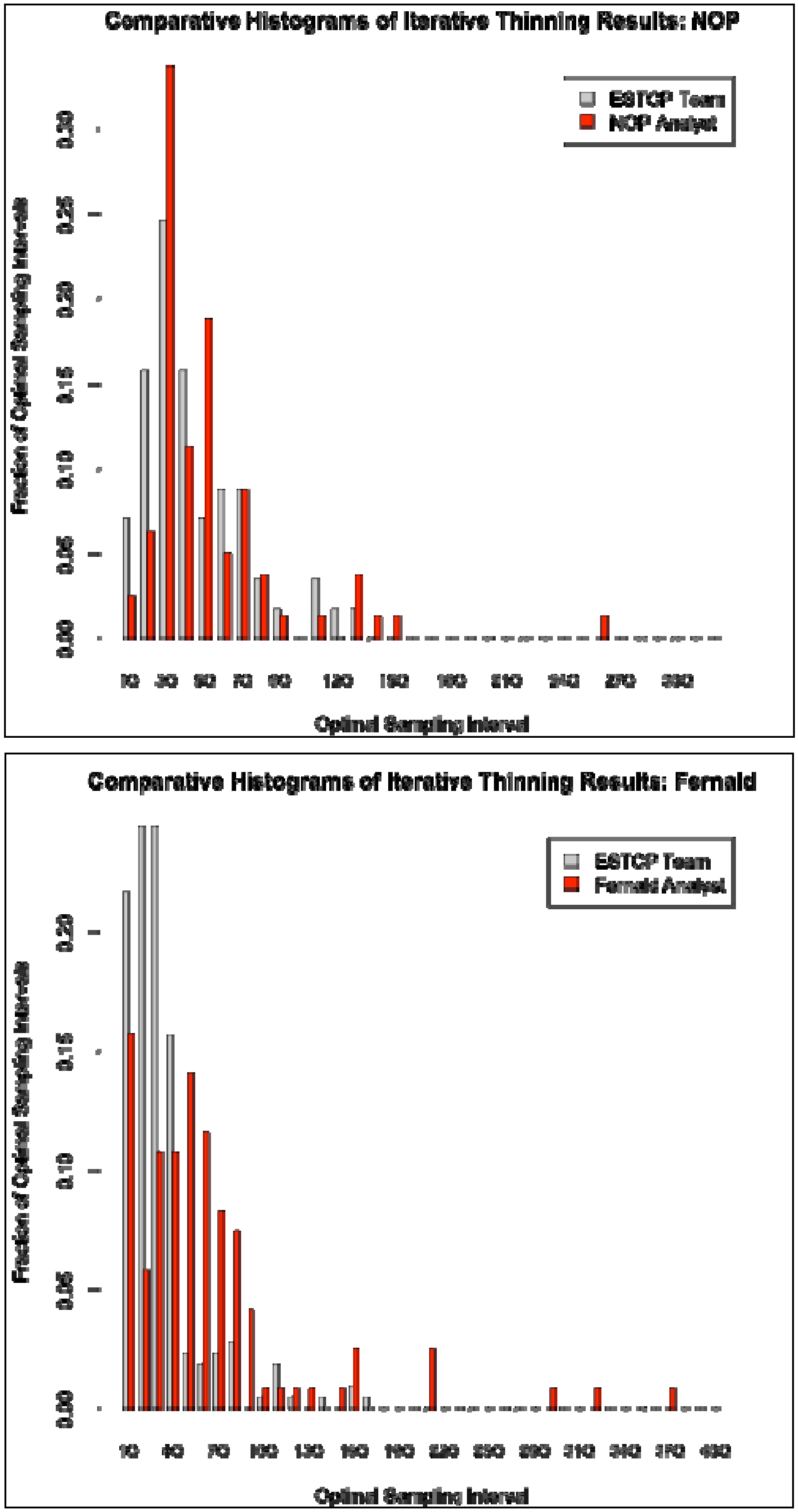



**Figure 32. Comparative histograms of individual optimal sampling intervals.** (continued)

### 6.5.5 Summary of Spatial Redundancy Evaluations

GTS v1.0 evaluates spatial redundancy using the same general philosophy as iterative thinning, but applied to maps instead of trends. A base map is created utilizing all applicable data, subsets of the data are randomly generated, and each subset is tested to determine how accurately the base map is reconstructed. Then, based on the degree of estimation error, a subset is deemed as optimal if it is the smallest network configuration that adequately recreates the base map.

Since the number of possible well subsets is prohibitively large for all but fairly small sites, a search procedure is required to intelligently winnow through possible subsets. One option in this regard is a genetic search algorithm, such as that employed by the Summit Tools LTMO software. The current version of GTS utilizes a quasi-genetic search strategy known as GTSmart. Like a true genetic algorithm, each possible network configuration (i.e., subset of well locations) is coded as a binary string, and a large initial population of such strings is generated for testing against the base map. On the other hand, the strings in GTS are not mated or mutated to form new strings as in a formal genetic algorithm. Rather, since QLR-based maps are computationally expensive, GTS picks only an optimal subset from the initial population of strings.

To ensure that the initial population of strings reasonably covers the search space of possible subsets, the search strings are formed smartly:

- The practical range of possible fractions of total number of wells included in a given subset (i.e., 0.05 to 0.96) is evenly divided into 13 bins (e.g., 0.05–0.12, 0.12–0.19, etc.). Then an equal number of unique strings is targeted for selection from each bin, that is, a randomly-generated string from a given bin is included only in the initial population if the fraction of kept wells falls within the range defined for that bin. The net effect is to force the initial population of strings to include a wide variety of possible well configurations, from subsets with only a few wells to those with nearly the full complement.
- Strings are also screened according to average interwell distance between pairs of locations. Based on a fixed percentile of the distribution of interwell distances in the full well configuration, strings are accepted for testing only if the average interwell distance in the string is at least as great as this percentile distance. This ensures that subsets in the initial population spatially cover the site area in a similar manner as the full well configuration, and strongly discourages strings that are tightly clustered in only a portion of the site.
- Protected wells—wells designated as ineligible for optimization—are always included in every string within the initial population.

Once the population of strings is formed, QLR is used to form a map for each string—based on data from wells included in that subset—and tested against the base map for absolute statistical bias. The optimal string is the subset that includes the least number of well locations, yet the map based on that string differs from the base map by no more than the bias constraints described in Section 3.1 (Optimize module). The same process is repeated for each COC, time slice, and aquifer zone (if a 2.5D analysis has been selected). Then the optimal strings are compared across time slices and COCs for each vertical zone (if any). A given location is tagged as redundant if it is missing from the optimal strings at more than half the COC-time slice pairs. All other locations are tagged as critical.

In the ESTCP demonstration, GTSmart was applied to each test site by the ESTCP project team in the configurations listed in Table 8. In addition, as discussed in Section 6.1, two versions of the AFP44 data package were prepared, given the uncertain aquifer zone designations for certain wells. This impacted the number of wells in the LZ-UZLU and UZUU zones but was otherwise the only difference between the two data sets. Table 9 summarizes the level of spatial redundancy found at each site, stratified by aquifer zone.

**Table 8. Data configurations used in spatial optimization by ESTCP team.**

| Site | Analysis Type | COCs | Aquifer Zones | # Time Slices |
|---|---|---|---|---|
| **AFP44** | 2.5D | TCE, chromium, 1,4-dioxane, 1,1-DCE | LZ-UZLU, UZUU, SGZ | 6 |
| **NOP** | 2.5D | TCE, RDX | DEEP, MEDIUM, SHALLOW | 7 |
| **Fernald** | 2D | Uranium | Single layer | 4 |

**Table 9. Summary of spatial redundancy computed by ESTCP team.**

| Site | Aquifer Zone | Total # Unprotected Wells | # Redundant Wells | Percentage Redundant |
|---|---|---|---|---|
| **AFP44 – Vers 1** | LZ-UZLU | 36 | 4 | 11% |
| | UZUU | 117 | 21 | 18% |
| | SGZ | 53 | 25 | 47% |
| | All | 206 | 50 | 24% |
| **AFP44 – Vers 2** | LZ-UZLU | 68 | 11 | 16% |
| | UZUU | 85 | 22 | 26% |
| | SGZ | 53 | 20 | 38% |
| | All | 206 | 53 | 26% |
| **NOP** | DEEP | 35 | 16 | 46% |
| | MEDIUM | 71 | 9 | 13% |
| | SHALLOW | 71 | 3 | 4% |
| | All | 177 | 28 | 16% |
| **Fernald** | — | 376 | 149 | 40% |

Important observations and results stemming from the spatial redundancy analysis include the following for each site, where comparisons of results with the independent site analysts are also noted:

***<u>Spatial Optimization at AFP44 Including Comparison with Site Analyst</u>***

- Despite the reclassification of 32 wells from zone UZUU to LZ-UZLU in creating version 2 of the database, roughly a quarter of the wells were found to be redundant using both versions. Similarly, both runs of the analysis found greater levels of redundancy in the uppermost aquifer zones and less in the deepest layers. This suggests a rough level of repeatability in the GTS results. Note, however, that there was greater redundancy found among the SGZ wells in the first run

(Version 1) than in the second run (Version 2), even though the same wells and data were available to both runs for this aquifer zone. Despite the "smart search" performed by GTSmart, the possible well subsets considered in any given optimization differ from run to run, leading to some variation in the results.

- The two versions of the database were compared to determine a) how many wells were found to be redundant in both optimization runs (i.e., overlap), and b) how close spatially were the two sets of redundant wells. Ostensibly, if clusters of wells are providing redundant statistical information (in terms of informing plume maps) and the concentration patterns are spatially continuous, there may not be a single "right" well to delete within a given cluster. Rather, more than one choice of redundant well might be possible and still allow accurate reconstruction of the base map. Under this supposition, if there exist specific areas of the site with redundant well clusters, different optimization runs on the same data ought to yield sets of redundant wells that either substantially overlap and/or are reasonably similar in spatial placement.

- To test this idea more concretely, the redundant wells (n=50) from version 1 of the database were compared against the redundant wells from version 2 (n=53). It was determined that 25 locations were the same in both runs. Further, based on extensive Monte Carlo sampling (N=10,000 runs) of same-sized sets of locations from the full list of 206 unprotected (i.e., eligible) AFP44 wells, it was found that a randomly picked set of 53 wells would only average about 13 locations in common with the version 1 redundant wells. Indeed, none of the Monte Carlo well sets had more than 24 locations in common, indicating that an overlap of 25 wells was highly statistically significant and that the separate GTS runs were consistently locating similar sets of redundant wells.

- The ESTCP project team also examined the spatial placement of both sets of redundant wells (see Figure 33). The two sets of locations are visually similar. To quantify the proximity, the average distance was computed between each well in the second set and its nearest neighbor in the first set. This mean distance was 170 ft, compared to a typical mean interwell distance of 521 ft between nearest pairs in a randomly selected test set of locations matched against the AFP44 version 1 redundant well set. Again, *none* of the Monte Carlo-generated well sets had a mean interwell pair distance less than 194 ft, suggesting that GTS was identifying redundant wells from the same areas of the site in both optimization runs.

- The site analyst optimized Version 1 of the database, as per the test design. As documented in Table 10, the site analyst identified 2-3 more wells as redundant per aquifer zone than the ESTCP project team, for an overall redundancy result of 28% (versus 24% for the ESTCP project team). The results seem quite similar, especially when viewed as a pattern across aquifer zones. Like the ESTCP project team, greater redundancy was identified at shallower depths than in the deeper aquifer zones, mostly attributable to the far greater density and clustering of wells in the SGZ layer.

- To compare the similarity between the results of the independent site analyst and those of the ESTCP project team, the same Monte Carlo testing was employed to measure the overlap and spatial proximity of the two sets of redundant wells. The site analyst matched 26 locations found by the ESTCP project team (out of 50 target redundant wells), and had a mean pairwise interwell distance of 243 feet. Thus, both the number of redundant locations in common and the mean interwell distance were slightly greater than the AFP44 Version 2 optimization run, but quite unlike the distribution of common locations or mean interwell distances exhibited by a randomly chosen set of wells. None of the Monte Carlo-generated well sets (n = 57 per set) had more than 25 wells in common with Version 1 of the ESTCP project team optimization run, and the typical number in common was only 14. Likewise, all of the random well sets had a mean interwell pair distance of at least 245 ft, with a mean value of 530 ft.

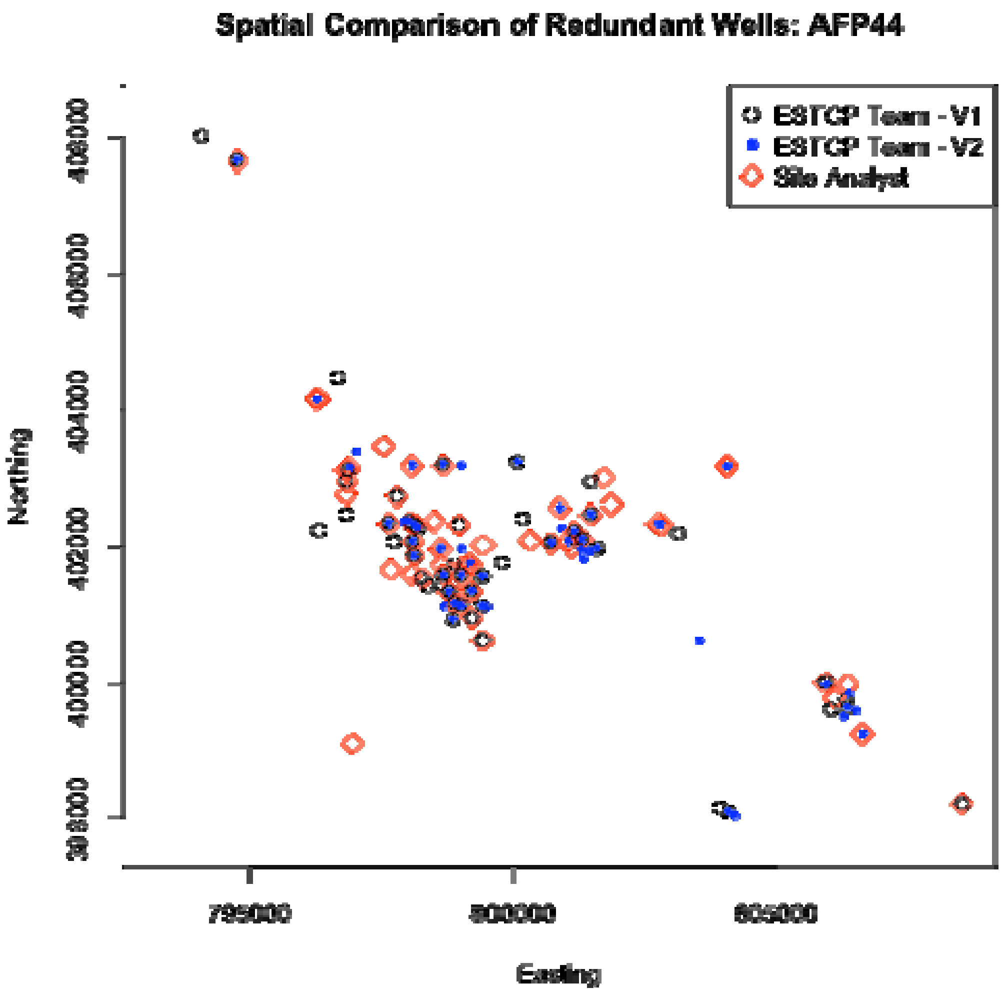



**Figure 33. Spatial comparison of redundant wells — AFP44.**

**Table 10. Comparison of spatial redundancy results.**

| Site | Aquifer Zone | Total # Eligible Wells | # Redundant Wells (% Redundant) — ESTCP project team | # Redundant Wells (% Redundant) — Independent Site Analyst |
|---|---|---|---|---|
| **AFP44 – Vers 1** | LZ-UZLU | 36 | 4 (11%) | 6 (17%) |
| | UZUU | 117 | 21 (18%) | 24 (21%) |
| | SGZ | 53 | 25 (47%) | 27 (51%) |
| | All | 206 | 50 (24%) | 57 (28%) |
| **NOP** | DEEP | 35 | 16 (46%) | 25 (71%) |
| | MEDIUM | 71 | 9 (13%) | 39 (55%) |
| | SHALLOW | 71 | 3 (4%) | 15 (21%) |
| | All | 177 | 28 (16%) | 79 (45%) |
| **Fernald** | — | 376 | 149 (40%) | 31 of 153 (20%)*<br>84 of 153 (55%)** |

* As summarized in written report submitted by Fernald site analyst
** As tabulated from GTS spatial optimization report submitted by Fernald site analyst

***Spatial Optimization at NOP, Including Comparison with Site Analyst***

- Only 16% of the unprotected wells were deemed redundant in the ESTCP project team analysis, including only 4% of the shallowest locations. However, the results varied substantially by aquifer zone, underscoring the importance of a 2.5D analysis at this site. The DEEP layer exhibited the smallest range of variation in concentration levels and much greater redundancy as a consequence (46%).
- By contrast, the independent site analyst found much higher levels of redundancy than the ESTCP project team (45% versus 16%), including greater redundancy within each aquifer zone (see Table 10). Upon further investigation, the differences are probably attributable to two factors: 1) outlier removal and 2) choice of COCs, discussed in more detail below.
  - o Outlier removal — Given the large fractions of non-detects in many of the analytes at the NOP site, and the variation in reporting limits, GTS identified a particularly large number of apparently spurious outliers at NOP. Most of these were weeded out (i.e., overridden) by the ESTCP project team prior to spatial optimization. The same was done by the NOP site analyst in his initial run through the data. However, when he re-ran the analysis on a newer version of GTS, the site analyst utilized the default set of outliers, resulting in the removal of a larger number of data points compared to the ESTCP project team. This had the impact of lessening the degree of observed variation at the site, particularly among COCs that already had very high non-detect levels (see below).
  - o Choice of COCs — Given the very high non-detect levels associated with both methylene chloride (86%) and TNT (96%) at NOP, the ESTCP project team chose not to optimize on these contaminants (or three others that were very similar) due to their poor optimization potential. Instead, only RDX and TCE were optimized, consistent with the persistent presence and extent of these chemicals at the site, and also consistent with

a comment from the NOP representative that remedial decisions at the site were made on the basis of those two COCs. By contrast, in the optimization run submitted to the ESTCP project team, the NOP site analyst also optimized on methylene chloride and TNT in addition to RDX and TCE.

Given the much smaller degree of variation in concentration levels for both methylene chloride and TNT (also exacerbated in the larger number of outliers removed by the independent site analyst), it was easier for GTS to remove additional wells and still accurately reconstruct a less variable base map. (At the extreme end, one could remove all but one well from a map consisting entirely of non-detects with a constant reporting limit.) Thus, the optimal network for monitoring methylene chloride and TNT was much smaller than the optimal network for monitoring RDX and TCE.

The net effect of this choice was therefore to increase the probability that a given well would be flagged as "redundant." Currently in GTS, each COC-time slice pair is given equal weight when forming the critical index used to distinguish essential from redundant wells. Any well tagged as "essential" in less than half the COC-time slice pairs is then flagged as redundant overall. By including methylene chloride and TNT in his analysis, the independent site analyst gave roughly half the spatial optimization weight to these COCs, at the expense of the two main contaminant drivers.

- As an aside, the independent site analyst generated two spatial optimization runs, one on an earlier beta version of GTS (not submitted to the ESTCP project team) and one on a more stable later release. In his earlier run, the site analyst utilized only RDX and TCE as contaminant drivers and commented that he found very similar levels of redundancy compared to the ESTCP project team (~20%). However, the ESTCP project team did not have access to the earlier run in order to make a detailed comparison of the results. The site analyst also noted that he apparently included methylene chloride and TNT in his second optimization run by mistake and attempted to deselect these COCs without success (a software glitch in GTS).
- To further parse out similarities and differences between the analyses of the ESTCP project team and site analyst, a post-plot of the two sets of redundant wells is presented in Figure 34. Although there are clearly more redundant wells identified by the site analyst, for reasons explained above, it is also evident that almost all the ESTCP project team redundant locations were also matched by the site analyst.
- To quantify the degree of overlap, the redundant wells (n=28) from the ESTCP project team analysis were compared against the redundant wells from the site analyst (n=79). Twenty-three locations were the same in both optimization runs. Further, based on extensive Monte Carlo sampling of same-sized sets of locations from the full list of 173 eligible NOP wells, it was found that a randomly picked

set of 79 wells would only average about 12 locations in common with the ESTCP project team redundant wells. Indeed, none of the Monte Carlo well sets had more than 23 locations in common, indicating that an overlap of 23 wells was highly statistically significant and that the independent GTS analyses were consistently locating many of the same redundant wells.

- The ESTCP project team also quantified the spatial placement of both sets of redundant wells. The mean interwell distance between nearest neighbor pairs from the two sets was 1348 ft, compared to a typical mean interwell distance of 1885 ft between nearest pairs in a randomly-selected test set of locations matched against the ESTCP project team redundant well set. Further, fewer than 0.5% of the Monte Carlo-generated well sets had a mean interwell pair distance less than 1348 ft, suggesting that GTS was identifying redundant wells generally from the same areas of the site in both optimization runs, despite the difference in total numbers of redundant wells.

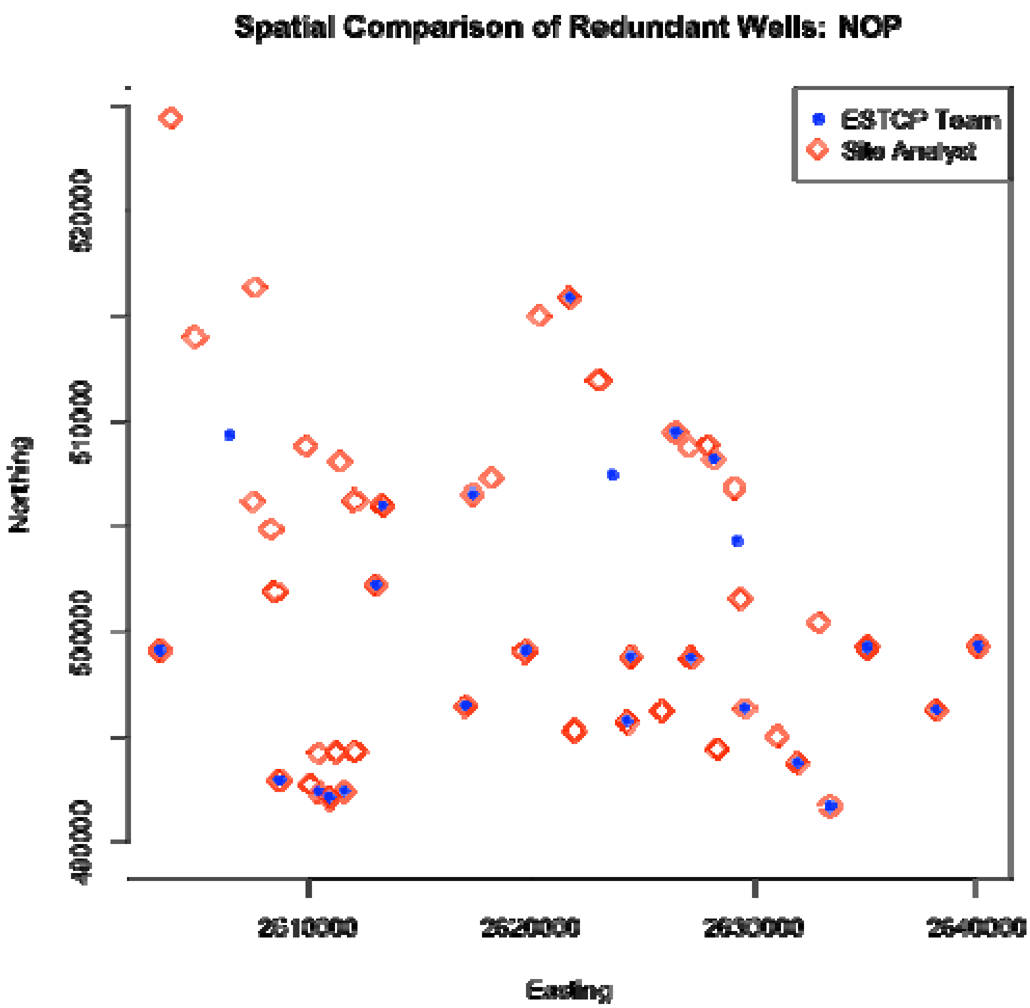



**Figure 34. Spatial comparison of redundant wells — NOP.**

***Spatial Optimization at Fernald Including Comparison With Site Analyst***

- At Fernald, when the results were stratified by well type, 49% of the DPT locations were found to be redundant, as opposed to 34% of the permanent wells (i.e., monitoring wells, extraction/injection wells, etc.). Optimization of the DPT locations reflected the following assumption: any location deemed redundant need not be mobilized for a direct push sample in the future, while those deemed critical should be resampled periodically within the same local subarea.

- In his sensitivity analysis comparing the impact of choice of bandwidth on the spatial optimization results at Fernald, the independent site analyst found significant differences depending on the bandwidths selected. As the analyst noted:

  > With the smallest spatial bandwidth selected, GTS identified 35 wells as redundant, not a significantly different number than for the base case when GTS self-selected well-specific bandwidths. However of these 35, only five were in common with the 31 wells GTS had selected for the base case. With the largest spatial bandwidth selected, GTS identified 84 wells as redundant; of these 84 eighteen were in common with the 31 wells selected as the base case. Clearly the selection of spatial bandwidths can have a significant impact on GTS results when evaluating monitoring well redundancy.

  These results underscore two points: (1) the importance of starting any GTS optimization analysis with an accurate base map, and (2) the fact that larger bandwidths lead to greater smoothing and less variation in concentration levels. Less variable maps tend to be easier to reproduce with fewer wells than maps with greater variation.

- In his sensitivity analysis considering the impact of 2D versus 2.5D optimization, the Fernald analyst remarked that while the numbers of redundant wells in the two runs were similar (31 versus 25 respectively of 153 eligible locations), "the specific wells selected as redundant [in the 2.5D case] were very different from the 2D analysis—only ten wells were identified by both the 2D and 2.5D analyses as redundant." While he did not provide the kind of comparative locational analysis discussed above, the result may point to nothing more than distinctly different spatial concentration patterns by aquifer zone. In that event, it would be surprising if GTS found nearly the same wells as redundant when treated as informing separate and distinct aquifers versus being treated as informing a single two-dimensional plane.

- Although the Fernald site analyst noted in his written report that using the default GTS spatial bandwidths in his 2D analysis produced 31 redundant wells (out of 153 eligible locations), the GTS-generated spatial optimization report he submitted listed 84 redundant locations (Figure 35). Apparently this corresponded to the case when the analyst set all the spatial bandwidths to their maximum value, lessening the degree of variation in the Fernald base maps. The analyst suggested that there seemed to be a remaining bug in the software, since when he re-ran several optimizations using different parameter choices (including bandwidth values)—switching back and forth within the same project file—the optimized network status post-plots did not always seem to match the locations listed in the text report. In any event, the ESTCP project team could not do a detailed locational analysis using what the Fernald analyst called his "base case" (i.e., 31 redundant wells, default GTS bandwidths), but instead had to analyze the submitted report.

- To tease out any similarities/differences between the analyses of the ESTCP project team and site analyst, a post-plot of the two sets of redundant wells is presented in Figure 35. Given the choice of the maximum spatial bandwidth in each case by the Fernald analyst, it is not surprising that he found a higher proportion of redundant wells than did the ESTCP project team. Furthermore, while there are some location matches (n=18), there are many more non-matches.
- To quantify the degree of overlap, the redundant wells (n=149) from the ESTCP project team analysis were compared against the redundant wells from the site analyst (n=84). Only 18 locations were the same in both optimization runs. Further, based on extensive Monte Carlo sampling of same-sized sets of locations from the full list of 153 unprotected Fernald wells (as employed by the site analyst), it was found that a randomly-picked set of 84 wells would average about 19 locations in common with the ESTCP project team redundant wells. The Monte Carlo well sets ranged from 10 to 28 wells in common, with a distribution indicating that an overlap of 18 wells was not at all statistically significant and no better than chance.
- The ESTCP project team also quantified the spatial placement of both sets of redundant wells. The mean interwell distance between nearest neighbor pairs from the two sets was 113 ft, compared to a typical mean interwell distance of 138 ft between nearest pairs in a randomly selected test set of locations matched against the ESTCP project team redundant well set. In this case, fewer than 2.5% of the Monte Carlo-generated well sets had a mean interwell pair distance less than 113 ft, suggesting that—even when a) the second data set was a partially overlapping subset of the first, b) the bandwidths were artificially inflated, and c) the total numbers of redundant wells were quite different—GTS still tended to identify redundant wells from the same general areas of the site in both optimization runs.

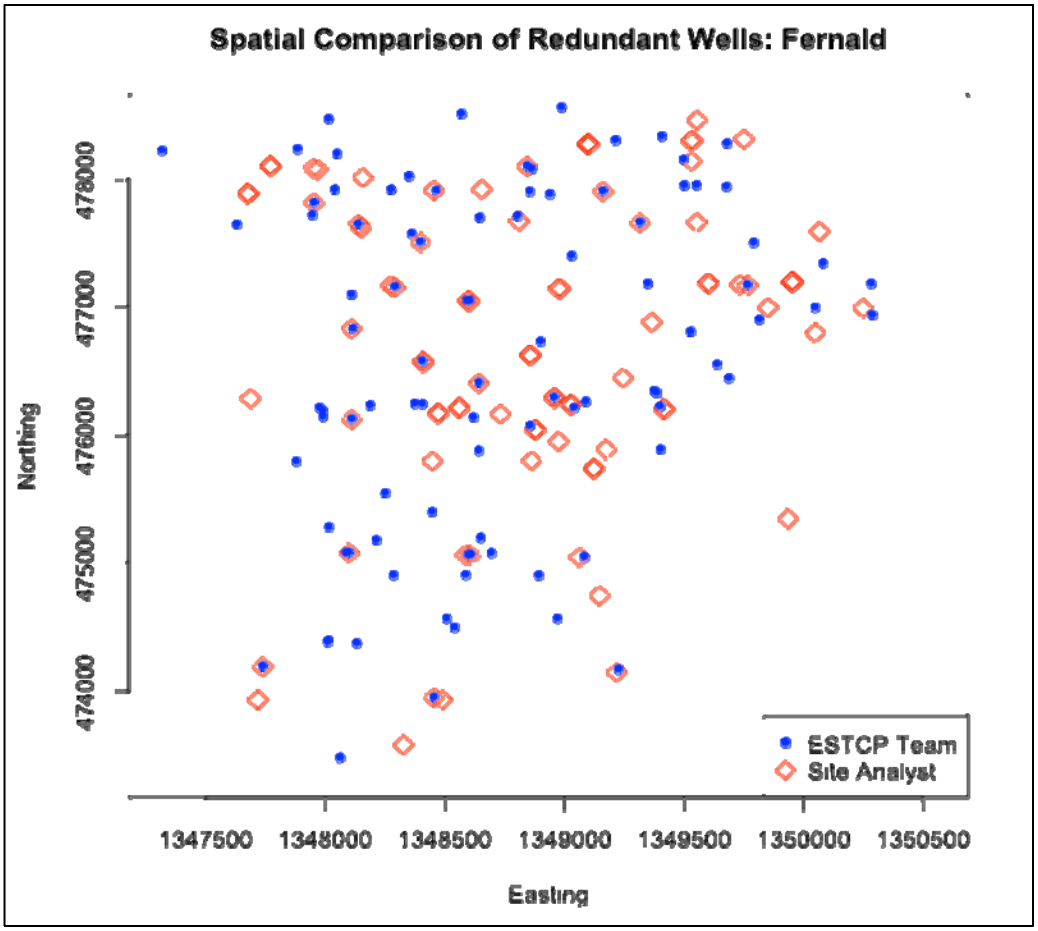



**Figure 35. Spatial comparison of redundant wells — Fernald.**

### 6.5.6 Summary of Network Adequacy Evaluations

As an option in spatial optimization, GTS determines if any new well locations are warranted, known as the network adequacy analysis. This is done by locating areas within the site boundary exhibiting both high relative uncertainty (as indicated by large coefficients of variation) and higher average concentration levels. GTS then searches the site over a fine grid to identify suggested coordinates for new wells within these subareas of higher uncertainty. To ensure reproducibility, a new location must exhibit high relative uncertainty across multiple COCs (assuming more than one COC is being analyzed).

At each of the demonstration sites, the network adequacy results were correctly and easily computed. The default list of suggested locations, however, varied in usability. GTS cannot determine whether a new location might be sited at a physical obstruction or perhaps in an inaccessible area. GTS also does not account for available construction and monitoring budgets. For these reasons, the user is allowed to override any or all of the GTS recommended well locations. This feature allows GTS to be utilized flexibly in site planning. Post-plots of the new location results designate user-accepted locations in a different color than overridden locations, thus documenting both what was computed and what was deemed useful.

Table 11 presents a summary of the numbers of suggested new wells, broken down by site and aquifer zone. At AFP44, GTS computed 24 recommended locations initially. Examination of the new well post-plots indicated that perhaps 13 of these locations should be eliminated, leaving 11 recommended new wells. Many of the eliminated wells were in close proximity either to other suggested wells or clusters of existing wells, and so represented probable redundancies. The case of aquifer zone SGZ was different: here it is known that the aquifer is only present over a small fraction of the boundary area. Any suggested wells placed outside the known extent of SGZ were overridden. The remaining two locations were kept, even though each was proximate to a cluster of existing wells. In practice, a knowledgeable site hydrogeologist might have overridden them also.

At NOP, 10 of 14 suggested wells were accepted by the ESTCP project team. Eliminated locations were in close proximity to existing wells. By contrast, when also including methylene chloride and TNT as COCs, along with RDX and TCE, the NOP site analyst found that GTS suggested 36 new well locations, the majority of which—especially for the SHALLOW layer—were located in the more sparsely-sampled northwestern section of the site. Some of these proposed wells were quite close to existing wells or even other newly proposed spots. Still, the addition of two highly non-detect COCs substantially changed the results. More detailed analysis suggested that two interdependent factors accounted for the differences:

1. In his final submitted analysis, due to time constraints, the NOP analyst did not override any of the suggested outliers identified by GTS. As discussed elsewhere, the variation in detection limits and high proportion of non-detects among some of the NOP analytes led GTS to flag way too many values as suspected outliers. Of almost 600 flagged records, the ESTCP project team decided that nine were probable outliers, including one value each for methylene chloride and TNT. By excluding all the default outliers—a large number of which were non-detect values for TNT and methylene chloride—the NOP analyst increased the relative level of uncertainty in areas of the site with generally low concentrations of these chemicals, and thus the likelihood of GTS suggesting additional new wells.
2. By including TNT and methylene chloride in the optimization, despite their high proportions of non-detects and poorer optimization potential, the overall relative uncertainty across all the optimized chemicals—particularly in the northwestern quadrant—was increased relative to an analysis based solely on RDX and TCE. This coupled with factor (1) led to the larger number of new wells reported by the NOP analyst.

At Fernald, 4 new well locations were suggested in the single aquifer layer that was analyzed, and all were considered reasonable choices by the ESTCP project team. Similar to the redundancy analyses, the independent site analyst at Fernald arrived at somewhat different results. When running a 2D analysis similar to the ESTCP project team, the independent analyst found that no new well locations were suggested. However, the Fernald analyst used only 172 well locations in a more limited and central portion of the site, compared to the 376 (active) wells and more extensive site area analyzed by the ESTCP project team. Comparing the same areas, GTS did not recommend any new wells for either team, so the general results were consistent.

Interestingly, the Fernald analyst also did a network adequacy run as part of the 2.5D analysis he conducted, after supplying the missing aquifer zone information. In that case, eight new well locations were suggested, five in the middle layer and three in the bottom layer. Note that this result underscores the importance of carefully deciding between a 2D and 2.5D approach within GTS. The map of relative uncertainty generated for each layer in a 2.5D analysis is based on the number, configuration, and concentration levels of wells in that layer. Comparing 2D to 2.5D on the same data will almost always give different results, as each layer in the 2.5D case will have fewer wells and often a different concentration pattern, generally leading to greater relative uncertainty (all other things being the same) and an increased need for new well locations.

Thus, it is not a flaw in GTS that the network adequacy results for the 2D and 2.5D cases at Fernald were different. Rather, a) if separate aquifer layers exist, b) that information is available to the database, and c) the concentration patterns in each layer differ, a 2.5D analysis should generally be utilized, especially to target new wells to the depth and aquifer layer where they are most needed. Note, however, that the Fernald analyst also expressed surprise that many of the suggested new locations were proximate to existing wells. This can occur within GTS v1.0 for at least two reasons:

1. The algorithm utilized by the site analysts did not force new wells to be located only in unsampled areas of the site. Instead, new locations were suggested in any area with sufficient relative uncertainty and high enough concentration levels. Users were encouraged to review and, if necessary, override the suggested placements. GTS also indicated how many existing wells were in the local vicinity of each newly suggested well, both numerically and visually (on the post-plot) to aid these decisions.
2. GTS uses QLR in spatial mapping rather than, say, kriging. As a smoother rather than an interpolator like kriging, there can be significant variability and hence uncertainty regarding average concentration levels even near existing well locations. This happens especially when the concentrations at closely spaced wells differ significantly (e.g., one high, one low). Contaminant levels in groundwater may not be spatially continuous (or at least smoothly so), depending on the complexity of the subsurface, preferential flowpaths, geochemical interactions with the subsurface soils, and so on. All of these factors can increase variability and caused the previous algorithm in GTS to sometimes suggest new wells close to existing wells or well clusters in order to better characterize the contaminant patterns.

Despite these factors, the experience of the software testers led the ESTCP project team to slightly alter the computation of new wells so that—in the future—none would be suggested near existing locations. The current release version of GTS includes these changes.

**Table 11. Summary of network adequacy results.**

| Site | Aquifer Zone | Number of GTS-Suggested Wells | Number of Accepted New Wells |
|---|---|---|---|
| **AFP44** | LZ-UZLU | 4 | 3 |
| | UZUU | 9 | 6 |
| | SGZ | 11 | 2 |
| **NOP** | DEEP | 4 | 3 |
| | MEDIUM | 4 | 3 |
| | SHALLOW | 6 | 4 |
| **Fernald** | — | 4 | 4 |

### 6.5.7 Summary of Trend and Plume Flagging Results

GTS v1.0 provides an interface for importing new data into the program that can then be checked for possible anomalies relative to previously constructed baseline trends and base maps. This import feature is distinct from the ability to incrementally append new data onto an existing database. The data imported for trend and plume flagging is also kept separate from the existing database.

To test the trend and plume flagging features (Predict: Module E), the most recent year's worth of sampling data was reserved from each test site, to be analyzed by the ESTCP project team. The goal was to determine whether the newer data was consistent with the older data, both temporally and spatially, and how well GTS would identify inconsistencies. To accomplish this goal at a temporal level, GTS constructs prediction bands around the baseline trends at contaminant-well pairs containing new data, linearly projects (i.e., extrapolates) these bands to the new sampling dates, and then compares the newer measurements against the projected prediction band. Spatially, GTS computes an approximate prediction envelope around the base map plume, and then interpolates the envelope to the coordinates of the new data to compare against the new concentration levels.

The independent site analysts were not asked to analyze this reserved data or to evaluate the trend and plume flagging features of GTS, though one tester at AFP44 did anyway. In general, both that tester and the ESTCP project team found the GTS algorithms for flagging anomalies to be somewhat too sensitive, resulting in more anomalies than made sense. According to the AFP44 tester:

> Criteria to identify anomalies may be too sensitive; many of the flagged values when viewed in time series seemed reasonable and didn't merit attention in the context of flagrant violation of prediction bands.

Table 12 offers a summary of the anomalies flagged by the ESTCP project team at each demonstration site. Despite the overly sensitive nature of the current GTS feature-set, users have the option to override any flagged anomaly, whether from trend flagging or plume flagging. So the final results of an analysis can be adjusted to better reflect the set of visually apparent anomalies. The principal reasons for too many flagged anomalies include:

- *Method of trend projection* — GTS v1.0 projects the baseline trend and associated prediction band linearly, based on the direction of the most recent baseline slope. In fact, many of the trends flattened out rather than continuing in the direction predicted by the baseline slope. A more conservative implementation of trend flagging would account for the possibility of a flat future trend, in addition to the directional projection currently employed.

- *Extrapolation is inherently difficult* — Any trend or plume extrapolation into the future is inherently uncertain, more so the farther the extrapolation. GTS will fail at this task some fraction of the time, no matter what projection method is utilized. For this reason, users are encouraged to review and override suggested anomalies whenever appropriate.

- *Lower bounds of the plume envelopes were often not quite low enough* — A number of essentially non-detect spatial anomalies fell just barely below the lower bound of the plume prediction envelope. An adjustment to the algorithm for constructing the prediction envelope may be needed.

- *Anomalies are more than just outliers* — The flagging algorithm in GTS is designed to identify not just obvious outliers, but also indications of temporal changes in trends or plumes, and even changes in detection/reporting limits for non-detects. To this end, some of the flagged anomalies may not be cause for alarm, but rather measurements to further investigate or document.

- *Plume envelope is approximate* — Due to transformation bias in back-transforming from logit-space to concentration scale when constructing the plume envelope, its nominal confidence level of 99% is only approximate. This might account for a higher than expected number of spatial anomalies in some cases.

**Table 12. Summary of trend and plume anomalies identified by GTS.**

| Site | # New Data Records Imported | # Default Trend Anomalies | # Probable Trend Anomalies | # Default Plume Anomalies | # Probable Plume Anomalies |
|---|---|---|---|---|---|
| **AFP44** | 1154 | 126* | 48 | 198** | 128 |
| **NOP** | 1786 | 108 | 62 | 25 | 19 |
| **Fernald** | 2099 | 174 | 13 | 33 | 17 |
| **Total flagged/total probable (%)** | | 408 | 123 (30%) | 254 | 164 (65%) |

*The AFP44 tester found 141 trend anomalies based on an analysis that eliminated a larger default number of outliers during the outlier screening; the ESTCP project team eliminated many fewer outliers prior to screening for anomalies in Module E.
**The AFP44 tester found 186 plume anomalies.

### 6.5.8 Import/Export Features

GTS v1.0 allows the import of ASCII text files, with one of several possible delimiters between fields (e.g., tabs, commas, spaces, etc.). GTS also allows separate import of water level (i.e., hydraulic head or depth to water) files for the purpose of creating potentiometric surface maps. In addition, the data import function can be used to build incremental databases; that is, new data in the same format can be added onto an existing database through successive use of the import command. So existing data are not deleted; rather, new data are appended into the data structure. This enables rich data sets to be accumulated over time and analyzed at periodic intervals.

For the purposes of annotating maps and post-plots, GTS allows the user to import Esri Shapefiles to be used as (static) graphic layers underneath a given plot or map. The number of Shapefiles that can be imported is only limited by system memory. Note here that Shapefiles cannot be manipulated within GTS, as say, within a GIS application.

Users can also import a simple site boundary text file, which delineates the vertices of a polygonal site boundary. In the current version of GTS, such a boundary is used not only to annotate the graphics but also to determine where map estimates should be made and what constitutes the analysis area of interest.

The most significant drawback to GTS import is the number and type of fields that are required to run an optimization. Given that GTS was originally developed for the Air Force, its input structure is based on standard ERPIMS conventions and field names. Any user must therefore ensure that his or her data is formatted according to these conventions. Altogether, 22 different data fields are required in GTS; some of these may have missing entries if complementary fields are populated (e.g., only one pair of the well screen depth fields SBD/SED and IBDEPTH/IEDEPTH need by populated; some databases tend to use the first pair, some the second). If potentiometric surface maps are desired, another three fields are required as part of either the main analytic database, or as part of a separate water level file.

Despite the large, required data structure, there is no requirement for data fields to be listed in any particular order. As long as the field names in the data file header match the GTS field names, the data are slotted into the right places within the internal SQLite database. Still, the experience of GTS testers during this project with data import varied considerably, with some having significant difficulties in getting GTS to correctly import their data. Relevant comments included:

> Data import is very involved and could be simplified; this is the single issue that could limit application to a wide audience. (AFP44)
>
> My initial attempts at loading data files failed — no error messages were thrown, there was no indication that something was wrong with the files, but GTS did not allow me to work with the data. After much experimentation I found that if I completely filled all blank fields, the load would be successful. (Fernald)
>
> I struggled with data import. My struggles were two-fold: manipulating the Fernald data so that it satisfied GTS's data paradigm, and producing input files that GTS would accept. (Fernald)

As a footnote, the tester at Fernald decided to manipulate the prepared input data well beyond the common data package that was supplied to both the site testers and the ESTCP project team. Much of this manipulation related to two factors: (1) the lack of adequate aquifer zone designations within the original data, and (2) the attempt to properly account for temporary DPT sampling locations within the context of LTM.

GTS has particular export capabilities but also drawbacks in this regard. On the plus side, each report in GTS (covering the results of a significant step in the analysis) can be exported to HTML and viewed in any standard web browser. These reports can also be easily sorted according to the report field headers. GTS also exports two text files of optimization results that are critical to completing the cost-benefit analysis using the GTS cost comparison calculator: the first provides a location-by-location listing of the temporary and spatial redundancy analyses (i.e., whether that well was flagged as redundant and the recommended sampling frequency if optimized temporally), while the second gives a listing of new wells recommended by GTS and their approximate coordinates. Both of these results files can be imported into Excel or another spreadsheet application for further summarizing or manipulation; they also must be imported into the GTS cost comparison calculator to derive the overall ROI associated with GTS optimization.

At the end of a project, users can document the database used in their analysis by exporting it to a tab-delimited text file. Note that this file contains not only the imported data but also several derived fields constructed by GTS internally to aid the analysis.

Unfortunately, GTS does not currently allow for graphics to be exported to image files. Initially, this capability had to be skipped due to the rather large number of graphics associated with a given analysis and the need to incorporate batch exporting of related graphics. The GTS project files were also designed to be somewhat self-contained, so that all the graphics from an analysis could be revisited by reloading the project. While the project files work as planned, users desiring to export graphics for other purposes must perform a screen capture and paste the graphic into an image-editing program. Relevant comments concerning graphical export included:

> There is not a way to save some of the graphics output, other than to do a screen capture, pasting the object into Paint or similar program and then saving as a JPEG file. The ability to save graphics would be very helpful for documenting and reporting the analysis results. (NOP)
>
> Reporting, in particular, the numerous graphics generated as output should be wholesale exported into a file for viewing and analysis; not sure what format would be best or universal. (AFP44)

### 6.5.9 Computation Time/Level of Effort

A summary of the amount of time it takes to apply GTS v1.0 is indicated in Table 13. This includes computation time primarily, though data preparation mostly encompasses manual labor. The amount of time required to run the optimization steps in GTS (temporal and spatial) varies considerably, according to the size of the network, amount of historical sampling data per well, and the hydrogeologic configuration of the site (i.e., number of separate aquifers and number of critical contaminants). Additional time is required to interpret and export results, as well as import results into the GTS cost comparison calculator to generate ROI.

**Table 13. General summary of time required to run GTS v1.0.**

| Task | Time | Comments |
|---|---|---|
| Data cleanup, screening, formatting | One to several days | Similar effort needed with other LTMO software; effort is primarily manual labor |
| Outlier screening (Module A) | Minutes to hours | Minutes to compute; review of a large number of outliers may require significant time |
| COC ranking, horizon analysis (Module B) | Minutes | |
| Baseline trends, base maps (Module C) | Minutes to hours | Minutes to < 1 hour to compute; more time may be needed for user to review/select temporal & spatial bandwidths |
| Temporal optimization — temporal variogram (Module D) | Seconds to minutes | |
| Temporal optimization — iterative thinning (Module D) | Minutes to hours | Wells with long data histories take more time; time increases linearly with number of wells being analyzed |
| Spatial optimization — redundancy search (Module D) | Minutes to hours | Time varies ~linearly with number of wells, number of contaminants, number of time slices, and number of separate aquifers; very large sites could require days of computing time |
| Spatial optimization — network adequacy (Module D) | Minutes | |
| Trend flagging (Module E) | Minutes | Time increases linearly with number of new records being analyzed |
| Plume flagging (Module E) | Minutes | Time increases linearly with number of new records being analyzed |

The two most computationally intensive steps in any GTS evaluation are temporal optimization by iterative thinning and the spatial redundancy search using the GTSmart algorithm. Table 14 provides a rough indication of the level of computational effort needed by the ESTCP project team to accomplish each of these steps at the three demonstration sites.

**Table 14. GTS computational time at three test sites.**

| Site | Data Configuration | Iterative Thinning | | GTSmart Redundancy Search | |
|---|---|---|---|---|---|
| | | Computation Time | Comments | Computation Time | Comments |
| **AFP44** | 3 aquifers, 208 wells, 4 COCs, 6 time slices | ~4 hrs | 342 COC-well pairs; <1 minute per pair | 14-15 hrs | 57 COC-zone-time slice triples; ~49 eligible wells per triple; ~15 minutes per optimization problem |
| **NOP** | 3 aquifers, 250 wells, 2 COCs, 7 time slices | 35-40 minutes | 57 COC-well pairs; <1 minute per pair | 10-11 hrs | 42 COC-zone-time slice triples; ~39 eligible wells per triple; ~15 minutes per optimization problem |
| **Fernald** | 1 aquifer, 467 wells, 1 COC, 4 time slices | 2.5 hrs | 217 COC-well pairs; <1 minute per pair | 6 hrs | 4 COC-time slice pairs; 209 eligible wells per pair; ~90 minutes per optimization problem |

## 6.6 SAMPLING METHODS

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

## 6.7 SAMPLING RESULTS

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

*This page left blank intentionally.*

# 7.0 PERFORMANCE ASSESSMENT

## 7.1 QUALITATIVE PERFORMANCE OBJECTIVES

### 7.1.1 Software Ease of Use

The expected performance metric is that GTS is easy to use and navigate by prospective users and that the GTS interface is well-designed and readily understood. The purpose of this performance objective is to indicate whether a mid-level analyst (i.e., one with some statistical and hydrogeological background) will be able to apply GTS to their site. During the demonstration, this objective was evaluated by having independent site analysts use GTS at the three test sites and report on their findings and experiences with the software. Although most of the site analysts had some previous exposure to MAROS, none had ever used the upgraded version of GTS nor was any user training on GTS provided, other than weekly phone support for questions. As documented in Section 6.5.2, navigation and use of the software was found to be straightforward and quickly understood. Installation was also generally straightforward, once proper administrative privileges were granted. Based on application of GTS by these independent analysts at the three demonstration sites, this performance objective was met.

### 7.1.2 Users Guide Ease of Use

The expected performance metric is that prospective users find the GTS user's guide/manual easy to utilize and understand and helpful in directing them on how to operate GTS and interpret its output. The purpose of this performance objective is to ensure that the software documentation for GTS is adequate and helpful in performing optimization analyses. The objective was assessed by gathering feedback on the user's guide from software testers and the independent site analysts who used GTS at the three demonstration sites. In general, users reported that the manual was well-written and straightforward in explaining how to operate each of the GTS modules. Comments were made by some testers that the GTS manual did not provide as much desired information on technical details regarding the GTS computational algorithms or how GTS derived certain results. Some users also desired additional guidance on how to correctly interpret GTS output/results. Based on this feedback, the performance objective was partially met.

### 7.1.3 Interpretation of Graphical Output

The expected performance metric is that prospective users will readily understand and correctly interpret GTS graphics and plots, perhaps in conjunction with consulting the GTS users guide. Since GTS incorporates a heavy dose of statistical graphics to convey optimization results, the purpose of this objective is to ensure that the graphics are both helpful and readily understood by the typical user. Direct feedback from software testers and the independent site analysts was solicited in order to evaluate this objective. In general, users found the graphics to be well-executed and helpful in conveying results. Some users suggested specific improvements to the program's graphics capabilities, such as improved legends or greater user control over symbols and colors. However, all users indicated good ability to use and interpret the existing graphics. Based on this feedback, the performance objective was met.

### 7.1.4 Software Reliability

The expected performance metric is that the final public release of GTS v1.0 does not exhibit any significant bugs or software glitches that impact/impede its ability to perform useful optimization analyses. The purpose of this objective is to identify whether there are any reliability issues associated with future use of the software. This objective was evaluated by testing the upgraded GTS software at three distinct sites, representing a variety of different conditions and data configurations, and by gathering direct feedback on software performance from the independent site analysts, as well as other interested software testers who participated in the ESTCP project team weekly conference calls.

Since GTS v1.0 represents a major upgrade and overhaul of the previous GTS beta software, many (i.e., hundreds) bugs, glitches, and crashes were encountered and reported by testers during this project. In all, 34 distinct alpha and beta builds of GTS were tested over the 3-year period, including seven in 2008, 19 in 2009, and another eight in 2010. Each build addressed multiple issues that were identified by testers. However, users also noted that by the final release in summer 2010, there were no significant bugs remaining. All testers were able to complete a start-to-finish optimization analysis without any crashes, bugs, or analysis-impeding issues. Thus, this performance objective was met.

### 7.1.5 Release GTS as Stand-Alone, Public Freeware

The expected performance metric is that GTS will be completely free to use and that it will be a stand-alone desktop application installed using a single executable file (.exe). The purpose of this objective is to ensure that GTS—funded by public moneys—can be used free of charge by the public. And further, the distribution and installation of GTS will be as uncomplicated as possible. This objective was evaluated by observing the characteristics of the GTS v1.0 end product. The design requirements for GTS mandated that free-to-use or open source software components be utilized in building the software. Many ideas were considered before settling on an architecture consisting of four major software technologies: (1) the open-source R statistical computing environment (www.r-project.org); (2) the open-source SQLite database tool; (3) the open-source QT interface development environment (IDE); and (4) the license-free MatLab runtime environment. Each of these pieces was critical to some aspect of GTS performance or functionality—R for statistical computing and optimization, SQLite for data housing and manipulation, QT for building the user interface, and MatLab for statistical graphics.

Because existing software technologies were leveraged in constructing GTS, a single installer was desired to avoid users having to install multiple, separate components with differing requirements. To this end, all the GTS component technologies were bundled together into a single executable file (.exe), with the exception of the Excel-based cost comparison calculator spreadsheet. The installer loads each component of GTS, including the GTS application itself, onto a desktop computer running Windows XP, with minimal input from the user. Although first-time installation can take up to an hour, updates are much more rapid as components that are already present do not need to be re-installed.

All of the major components used in GTS are open-source freeware, with the exception of MatLab. Because SAIC, part of the ESTCP project team, owns a MatLab developers license, it can freely distribute a license-free, cost-free executable of the MatLab runtime environment. This runtime environment is bundled into GTS v1.0. As far as the cost comparison calculator, Microsoft's Excel is, of course, not freeware, and so could not be bundled into the GTS executable. However, it is practically ubiquitous within the enterprise software arena. Any user with Excel on their computer can therefore access and run the GTS cost comparison calculator spreadsheet without any additional charge. In a future version of GTS, it is planned for the cost calculator to be coded directly into the interface with no need for Excel. However, even at present, almost no, if any, prospective users will need to pay anything to run GTS. Based on this architecture and design, the performance objective is met.

### 7.1.6 Accessible to Non-Experts

The expected performance metric is that GTS can be successfully run and interpreted by mid-level analysts. A mid-level analyst was defined for purposes of this demonstration as someone with some college-level background or professional experience in statistics, geostatistics, and hydrogeology, but who was not an expert in statistics or geostatistics. The purpose of the performance objective is to ensure that GTS can be successfully run by likely prospective users and that the labor costs associated with its use are not prohibitive. This was evaluated by having independent, non-expert testers run the software at the three demonstration sites and directly soliciting their feedback. Overall, none of the independent software testers were professional statisticians or geostatisticians, although the Fernald tester had previous professional experience in doing statistical analyses. All of the testers were likewise able to successfully complete one or more optimization analyses of their site data. Further, three testers commented in their evaluations that GTS could be reasonably navigated and applied by a professional with hydrogeological experience and some statistics background. Based on this feedback and their successful analyses of the demonstration site data sets, this objective is met.

### 7.1.7 Robustness of Software

The expected performance metric is that GTS can be applied across sites with a variety of COCs, hydrogeologic terranes, remedial solutions, etc. The purpose of this objective is to ensure that GTS is applicable to a large number of potential sites and conditions. This was evaluated by applying GTS to three different test sites, representing different branches of the government or DoD and covering a range of differing conditions. In addition, two versions of the AFP44 database were tested by the ESTCP project team and multiple data configurations were tested at each site by the independent analysts. Further, GTS was applied during the demonstration period by other interested software testers to several other sites, including Paducah, KY (DOE), Cape Canaveral (Air Force), Andrews Air Force Base (AFB), Tinker AFB, and Fort Dix (Army).

Regarding the three ESTCP demonstration sites, Table 2 and Section 5.0 document the variety of contaminants, numbers of wells, and aquifers optimized by GTS, including metals, organics, and radiologic parameters embedded within either alluvial valleys or buried valley glacial outwash aquifer systems, and with well sets ranging from 200+ to over 400. All of the test sites were undergoing or had undergone some type of remedial activity. Since spatial optimization in GTS is not plume specific, it does not require that the plumes be stable over time, only that maps can be estimated over a series of temporal snapshots (i.e., time slices). This allows for optimization at sites where concentration levels and patterns are actively changing, as indeed seen at the three demonstration sites.

The most important assumption (and limitation) of GTS is common to any geospatial mapping tool: each aquifer or aquifer layer is assumed to be spatially and hydraulically connected, leading to spatially continuous concentration patterns. Subsurface environments that are highly fractured or with strongly preferential pathways may not be good candidates for a GTS spatial analysis. On the other hand, GTS temporal optimization—particularly the well-specific iterative thinning feature—was shown to be applicable in any hydrogeologic environment, since it does not depend on spatial continuity and is especially useful at sites with complex or seasonal trends. And, since GTS is modular by design, users can flexibly apply either or both of the spatial and temporal optimization features, depending on site-specific conditions. All in all, the successful application of GTS to three very distinct test sites shows that the performance objective is met.

### 7.1.8 Water Level-Aided Mapping

The expected performance metric is that GTS can optionally estimate concentration maps using water level data as a covariate (and proxy for groundwater flow direction and potential). The purpose of this objective is to identify whether GTS can build more accurate and useful base maps by simultaneously utilizing both analytic concentration data and water level measurements. Unfortunately, internal development and testing of this feature on some of the test site data led to inconclusive results. Available resources and the project timetable did not allow for the development of additional improvements or deployment within the GTS interface. Thus the stated performance objective was not met. However, this work led to GTS incorporating a fairly robust mapping of the potentiometric surface as an added feature, something of a by-product of the original objective. Users commented that these water level maps—displayed in a temporal series by time slice—are quite useful as characterization tools in and of themselves.

## 7.2 QUANTITATIVE PERFORMANCE OBJECTIVES

### 7.2.1 Software Ease of Use

The expected performance metric is that GTS is easy to operate by prospective users, and testers will encounter few operational difficulties. The purpose of this objective is to ensure that GTS is set up in a manner that is conducive to use by prospective analysts. This was evaluated quantitatively by cataloging the number and types of operational problems and issues encountered by the independent site analysts. Table 15 lists the issues reported by type and number of similar reports.

The biggest operational issues included installation of GTS on government-owned computers and the variety of software bugs and crashes encountered while operating early beta versions of GTS. Installation of new desktop software on DoD or other government computers often requires specific administrator privileges. This difficulty is not unique to GTS but was reported by each of the testers. A more serious difficulty was the fact that due to the lengthy period of development needed to overhaul GTS and eliminate bugs from the software, there was not enough calendar time during the ESTCP project to wait to begin the case studies at the three demonstration sites until a completely stable version of GTS had been built. Instead, the case study analyses overlapped the GTS development phase, with two important consequences:

- The independent site analysts were given beta versions of GTS to perform their analyses. Since each beta version still possessed a number of unknown bugs, the testers all encountered new problems or bugs that sometimes crashed the software. In addition, as identified bugs were fixed and new versions of GTS built, testers were forced to install updates to the software and sometimes re-do portions of their analysis. At NOP, this became a significant issue, since the independent analyst had to wait for his IT staff to be able to schedule a GTS update, given the administrator privileges needed.
- Beta testing of GTS was more extensive than it would have been had not the development and demonstration phases of the project overlapped. While this posed an operational difficulty for the site analysts, it also allowed a larger number of testers to bang on the software before final release.

Four other issues were reported by more than one tester:

- *Data importing* — The process for importing data was considered too complicated by some users, requiring too many fields or too specific a format. One user was not clear as to which fields were required versus optional. One had difficulty loading a boundary file, though this was apparently due to insufficient guidance in the user's manual as to the type of boundary file that GTS accepts.
- *Graphics* — Some users commented on the inability in GTS to export plots and maps to common graphical formats, either singly or in batches. Instead, users are currently forced to capture individual screenshots of desired graphics and then import or modify those screenshots in other programs.
- *Optimization* — Users commented on the lengthy times needed for iterative thinning and especially for spatial optimization in GTS, perhaps requiring overnight computer runs. This limited their ability to test different variations of an optimization, such as by changing input parameters.
- *Outliers* — Some users found the GTS criteria for identifying potential outliers to be too sensitive, thus generating more outliers than reasonably existed. At large sites, this in turn entailed significant effort for user review and possible override of data points that were really non-outliers.

Despite these operational issues and difficulties, all testers rated the GTS interface as highly usable, easy to navigate, and readily understood. Based on this feedback, this objective was partially met.

**Table 15. Summary of operational difficulties encountered by software testers.**

| Type of Operational Difficulty | Description of Difficulty | # of Reports* |
|---|---|---|
| Installation | Lack of administrator privileges made installation difficult or lengthy | +++ |
| Bugs in beta testing | Several bugs and/or crashes encountered while operating beta versions of GTS | +++ |
| Data importing | Importing data is very involved/too complicated | ++ |
| | Zero/negative (radionuclide) data not handled by GTS without user adjustment | + |
| | Trouble loading boundary file | + |
| Graphics | No way to export graphics into other programs without creating screenshots | ++ |
| | Legends do not display correctly on 64-bit machine | + |
| User interface | Difficulties in switching back and forth (i.e., navigating the interface) during an analysis when changing parameters/settings or re-doing computations | + |
| | Keyboard shortcuts (e.g., Control-X) do not work with highlighted material | + |
| Optimization | Optimization runs took a long time | ++ |
| | Trouble deselecting COCs for optimization | + |
| Outlier analysis | Tedious to review outliers at sites with many wells | + |
| | Criteria for identifying outliers too sensitive | ++ |
| Trend/plume flagging | Criteria for identifying anomalies too sensitive | + |

* Each '+' symbol represents one distinct report

### 7.2.2 Reproducibility of Temporal Optimization

The expected performance metric is that GTS produces consistent, repeatable results during temporal optimization, such that different users analyzing the same data should generate substantially similar optimal sampling frequencies. The purpose of this objective is to determine whether the temporal optimization algorithms and features in GTS give valid results that can be replicated across multiple runs of the software or across multiple users. As detailed in Section 6.5.4, the optimized sampling intervals derived using iterative thinning at two of the sites were very similar when comparing the ESTCP project team's results with those of the independent site analysts. At both AFP44 and NOP, identical recommendations were computed for the overall, site-wide sampling interval, while the aquifer zone-specific intervals were identical in four of six cases, only differing by one quarter (1Q=90 days) in the other two. At Fernald, the independent analyst computed both the baseline sampling interval and the optimized sampling interval as longer by a quarter than the ESTCP project team did. This did not reflect a lack of validity in the GTS results but rather that the Fernald analyst used a fairly different subset of the original data package supplied to each site and that that subset exhibited longer average baseline sampling intervals.

Additional evidence of the repeatability of GTS temporal results was provided by the histograms (Figure 32) comparing patterns of optimal sampling intervals at individual wells. Despite differing user choices with respect to temporal bandwidths, confirmed outliers, and COCs, the comparative distributions of sampling intervals exhibit very similar quantiles at AFP44, and strong similarity at NOP. A Kolmogorov-Smirnov test of the hypothesis that both sets of optimal sampling intervals at AFP44 were drawn from a common distribution is clearly not significant, with approximate p-value ≈0.99. A similar test at NOP is also not significant, with approximate p-value ≈0.53. Thus, no clear statistical difference is evident at either site, even though the NOP analyst included two COCs (methylene chloride and TNT) in his analysis that were excluded by the ESTCP project team.

By contrast, the differing data sets used at Fernald by the ESTCP project team and independent analyst led to distinct distributions of optimal well-specific sampling intervals. The Fernald analyst found generally longer optimal intervals, and the Kolmogorov-Smirnov test of a common distribution was highly significant ($p<0.0001$), underscoring the different patterns that were computed.

Finally, it should be noted that iterative thinning was run on both versions of the AFP44 database by the ESTCP project team, though not discussed in Section 6.5.4. Given that the only difference in this case was the aquifer zone classification of certain wells—which does not impact iterative thinning—it is not surprising that the site-wide and aquifer zone-specific sampling interval recommendations from both runs were identical, only differing very occasionally at the individual well level. Based on these comparisons, this performance objective is met.

### 7.2.3 Reproducibility of Spatial Optimization

The expected performance metric is that GTS produces consistent, repeatable results during spatial optimization, such that different users analyzing the same data should generate substantially similar optimal sampling networks. The purpose of this objective is to determine whether the spatial optimization algorithms and features in GTS give valid results that can be replicated across multiple runs of the software or across multiple users. As detailed in Section 6.5.5, there was a close similarity at AFP44 in the percentages of redundant wells identified, whether the ESTCP project team used Version 1 of the database (24%), Version 2 of the database (26%), or whether the independent site analysts did the analysis (28% and 20%). At NOP, there was a much larger difference between the ESTCP project team (16%) and the site analyst (45%), largely attributable to the additional COCs optimized by the NOP analyst. When the independent analyst used the same COCs as the ESTCP project team, he arrived at a fairly similar redundancy percentage of 20%.

Additionally, analysis of the specific wells deemed redundant and the spatial pattern of redundant wells revealed substantial overlap and locational closeness at both AFP44 and NOP. Compared against Monte Carlo sampling of random, unprotected well subsets, the actual subsets of redundant wells in Versions 1 and 2 of the AFP44 database exhibited a highly statistically significant number of locations in common. This was also true of the comparison between the ESTCP project team results and that of the AFP44 site analyst, as well as the comparison of common locations at NOP between the ESTCP project team and the site analyst there. Monte Carlo testing further indicated that redundant wells at both sites were generally being selected from the same subareas, as indicated by highly statistically significant, low mean interwell distances between nearest neighbor location pairs (each pair formed from one well in each set of redundant locations).

The results for Fernald were exceptional, largely due to the differing data sets utilized by the ESTCP project team and independent analyst. When the Fernald analyst used the default GTS spatial bandwidths, he found less redundancy among a much smaller subset of wells and DPT locations than the ESTCP project team did using a much larger set of locations. When he re-did the analysis using the maximum spatial bandwidth for each map, the Fernald analyst found a higher level of redundancy than did the ESTCP project team.

While a detailed locational analysis could not be done on the Fernald analyst's base case, an analysis of the maximum bandwidth results found that though the number of redundant wells matched between the ESTCP project team and independent analyst was not significant, the relative closeness or spatial similarity was statistically significant ($p<0.025$) despite the differing data sets and choices of bandwidth parameters.

All in all, with the caveat that the choice of COCs can make a large difference in optimization results—especially if a user attempts to optimize COCs with very high non-detect rates and low optimization potential—the numeric similarity in spatial redundancy results indicates that this performance objective is met, to the degree it could be ascertained.

### 7.2.4 Predictability

The expected performance metric is that the Predict module in GTS will successfully project/extrapolate baseline trend and plume estimates to encompass at least 90% of near future measurements collected at the same site. The purpose of this performance objective is to determine whether GTS can accurately identify anomalous measurements, values that by definition are significantly different from previous trends and therefore should occur infrequently, especially if the future groundwater samples are collected close in time to the existing historical database. The Predict module in GTS v1.0 makes two kinds of extrapolations: (1) Baseline trends are extended linearly to the sampling dates of new measurements, based on the most recent slope and magnitude of each baseline trend. A prediction band is also estimated around the projected trend. (2) Base maps are projected by estimating a prediction envelope around the plume for each time slice. The plumes and their envelopes are then separately averaged across time slices to yield a joint prediction envelope around the predicted plume. New measurements falling outside the extrapolated prediction band are deemed trend anomalies. Likewise, those measurements falling outside the predicted plume envelope are denoted plume anomalies.

To evaluate this objective, the final and most recent year of sampling data was reserved at each demonstration site for testing of the trend flagging and plume flagging features of the Predict module, that is, all the previous years of historical data were utilized to construct baseline trends and base maps (as well as to perform the optimization studies), while the final year was treated in the demonstration as a set of new, future measurements. As detailed in Section 6.5.7, trend anomalies were detected in 11% of the reserved AFP44 data, 6% of the reserved NOP data, and 8% of the reserved Fernald data, for an overall rate of 8%. Plume anomalies were found respectively in 17%, 1%, and 2% of the same reserved data sets for an overall rate of 5%. Thus, while slightly less than 90% of the new measurements were correctly predicted at AFP44, the target was easily met at the other two sites, and for the project as a whole. So the stated objective appeared to be met.

Nevertheless, both the ESTCP project team and some of the independent analysts commented that too many anomalies were apparently flagged, a conclusion born out by further examination of the anomaly time series plots and plume prediction envelope limits. In Table 12, it was determined that perhaps only 30% of the trend anomalies and 65% of the plume anomalies were values deserving further investigation or verification. Improvements were also planned to the Predict module algorithms for a future version of GTS. So on this score, the performance objective is only partially met.

### 7.2.5 Optimization Effectiveness

The expected performance metric is that GTS is able to identify significant redundancy in larger groundwater monitoring networks and that it can generate optimized sampling programs. The purpose behind this objective is to ensure that GTS is "worth its salt" as an optimization tool, in that it can identify redundancies when they exist and generate relevant potential cost savings. The objective was assessed by computing the degrees of temporal and spatial redundancy identified at each demonstration site and translating these redundancies into estimated cost savings via the GTS cost comparison calculator. As discussed in Sections 6.5.4 and 6.5.5, each of the demonstration sites had a large groundwater monitoring network with significant annual monitoring expense. The number of wells analyzed at each site included 208 wells at AFP44, 250 wells at NOP, and a combination of 467 wells and DPT locations at Fernald. Optimized temporally by iterative thinning, GTS proposed a reduction in sampling frequency of approximately 75% at AFP44, 50% at NOP, and 67% at Fernald. Further, levels of spatial redundancy were estimated at 24 to 26% for AFP44, 16% for NOP, and 40% at Fernald. Each of these redundancies translates into a significant reduction in annual monitoring expense, particularly the decreases in minimum sampling frequency.

At each demonstration site, the iterative thinning results were translated by GTS into recommended optimal sampling intervals, not only on a site-wide basis, but also as recommendations for each aquifer zone, and, if so desired, as well-specific recommendations for each separate location. In a similar vein, spatial redundancies identified via the GTSmart algorithm were translated into optimal sampling networks, with a recommended list of essential wells at each site.

Finally, using the GTS cost comparison calculator (as discussed in Section 8.3), the optimized sampling programs computed using the software would translate into substantial annual cost savings compared to the current monitoring programs. At AFP44, the estimated savings would be 44% of an annual baseline program cost of $437,000 or approximately $191,000 per year. At NOP, the savings were estimated at 39% of an annual baseline program cost of $465,000 or approximately $181,000 per year. And at Fernald, savings were projected at 45% of an annual baseline program cost of $360,000 or approximately $162,000 per year. Clearly, this objective is met.

### 7.2.6 Accuracy

The expected performance metric is that there is good numerical and statistical agreement between the baseline trends and base maps GTS constructs and the original measurements from which they are estimated. In other words, the baseline trends and base maps accurately reflect or represent the underlying data. The purpose behind this objective is to ensure that GTS does not optimize a false or unrepresentative baseline. As noted in Section 6.2, GTS identifies redundancy based on its ability to accurately reconstruct concentration trends and maps. But if the starting point for optimization—either a baseline trend or base map—does not reflect actual site conditions, there is no reason to trust reconstructions of inaccurate trends or maps based on supposedly optimized sample sets. How, for instance, can a well location be considered redundant if a map to which it contributes is substantially off target?

To evaluate this objective, two key steps were taken: (1) extensive internal testing of the trend and map algorithms developed for the GTS v1.0 upgrade, including analysis of trend and map accuracy through minimization of weighted residuals and (2) building interface elements into GTS to allow users to check trend and map fits, and to override the GTS default temporal and spatial bandwidth selections. Since GTS uses local regression to estimate trends and maps, its trend-making and mapmaking tools are smoothers rather than interpolators. Regression is readily understood with respect to trends, but less common in geospatial mapmaking, where kriging is better known. As an interpolator, ordinary point kriging estimates always precisely match the observed data, so there are no residuals. Nevertheless, kriging-based concentration estimates between known data may or may not accurately reflect the overall spatial pattern or continuity in concentration values, nor are most measured groundwater concentration levels known with great precision. So interpolation via kriging can readily lead to inaccurate maps, despite the lack of residuals.

By contrast, local regression rarely matches the observed data, even as a linear regression trend may not precisely hit any of the observed data points. There are always residual differences (or error) between the regression fit and the measured concentrations. Nevertheless, it is designed to accurately capture the nature and direction of the trend, even as it attempts to minimize the residual error. GTS v1.0 employs this concept in both trend fitting and map estimation.

To ensure accurate trends, internal testing of the GTS algorithms was done using a variety of data sets, including data from the three demonstration sites. To minimize residual error between a given trend and its observed data, the GTS algorithm was designed to explore a series of possible bandwidths, with the default bandwidth value chosen to jointly best minimize (a) Mallows CP criterion (this is closely related to a scaled sum of squared residuals); (b) average bias in the residuals; (c) skewness in the residuals; (d) residual non-normality; and (e) correlation between the residuals and either the fitted concentrations or time of sampling. In the event of a tie between potential bandwidths, more weight was assigned to the Mallows CP and average bias diagnostic criteria.

This internal residual checking enables GTS to select the best-fitting local regression trend in terms of residual error. However, it does not always work to select the best-fitting trend. Occasionally, a trend may be close to its observed data and yet be radically inaccurate between certain sampling dates, as judged visually by the overall data pattern. To ensure accuracy in these cases—since they tend mostly to occur between more widely-spaced sampling events—GTS does both a sampling gap analysis, which attempts to eliminate data from trend fitting that occurs prior to a large gap between measurements, and allows the user to visually check and override the default bandwidth when necessary via the “check bandwidth” interface. Note in this regard that complex, nonlinear trend fitting is an inherently difficult statistical task. Two testers noted examples in their evaluations of wildly inaccurate default GTS trends (see Appendices D and E in the final report). This was seen as a drawback to GTS. In fact, in the very examples cited, the GTS interface offers alternate, much more accurate (and visually pleasing) trends that can be easily selected by the user.

To ensure accurate maps in GTS, similar internal testing was conducted to minimize the residual spatial error. In this case, as described in Section 6.2, the residuals were logged relative concentration errors, weighted by spatial density. The default bandwidth selection algorithm attempted to jointly best minimize: (a) the root mean squared error (RMSE); (b) the median absolute deviation in relative residuals; (c) the 90th percentile of the absolute relative residual distribution; and (d) the maximum absolute deviation. Ties in prospective bandwidths were broken by giving greatest weight to the RMSE and 90th percentile diagnostic criteria. An example diagram illustrating the minimization of these diagnostic criteria is shown in Figure 36.

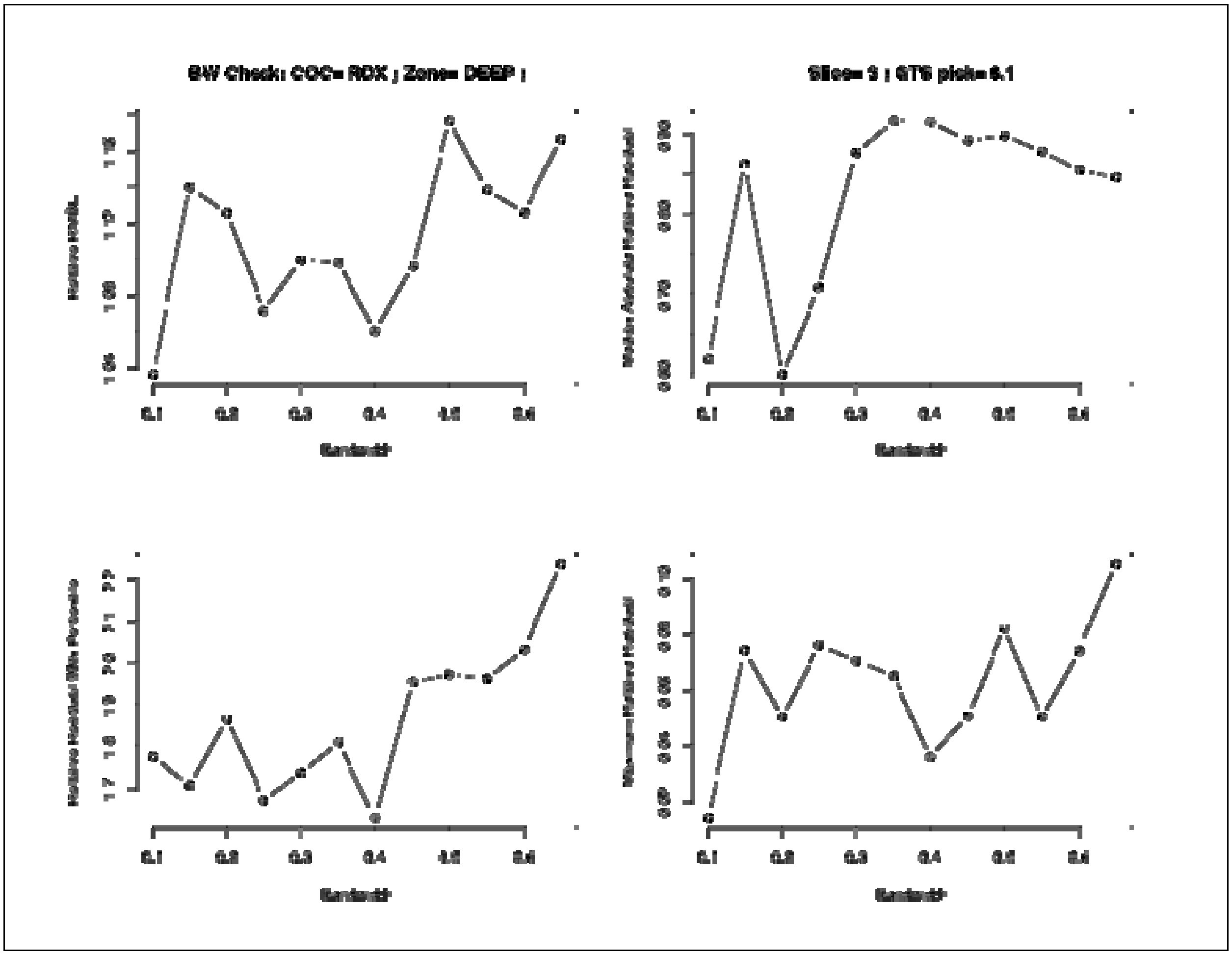



**Figure 36. Example of diagnostic spatial bandwidth selection.**

Like the trend fitting, maps with minimal residual error at observed wells may be inaccurate between sampling locations, where concentrations are unknown. In addition, as a three-dimensional object, it can be more difficult to judge the overall fit of a given map, especially when trying to assess residual error. It is also often true that high and low concentrations may be clustered together at nearby wells, perhaps due to lack of spatial continuity in concentration patterns, temporary spikes in concentration at one of the wells, differences due to variation in well screen depth, or low hydraulic connectivity. Such situations make it difficult to minimize residual error regardless of bandwidth and often necessitate user input to ensure a pleasing map. GTS has a built-in user interface for checking and, if necessary, overriding the default spatial bandwidths. Residuals are checked via color-coded post-plots of the relative errors.

Though these steps worked to ensure the general accuracy of GTS maps as measured by relative residual error, some testers either criticized the base maps as not well-matched to existing plume maps of their site or suggested improvements to the mapmaking features in GTS. At least three problems were evident:

- Given the need to create maps across an entire site area, there is no spatial gap analysis similar to the trend gap analysis. As such, inaccurate spatial trends can result between wells in sparsely sampled areas.
- Maps are currently extended to the site boundaries for all aquifer zones, even if one or more zones are only sampled within a smaller portion of the site. This can lead to inaccurate spatial extrapolation of the concentration estimates.
- The visible contours on GTS maps are selected from a fixed set of concentration levels, as opposed to being selected by the user based on site-specific criteria or regulatory limits. This can lead to GTS maps appearing rather different from traditional hydrogeologic maps, even if the underlying estimated concentration patterns are substantially the same.

Overall, while the trends and maps in GTS do minimize residual error as per the stated performance objective, several improvements to the mapmaking facility could be implemented. This objective is therefore rated as partially met.

### 7.2.7 Versatility

The expected performance metric is that the upgraded and revised GTS software is able to perform optimization studies at sites with more than 200 wells. The purpose for this objective is to ensure that GTS can be used at larger facilities, in addition to smaller ones. The previous beta version of the software, GTS v0.6, had a memory limitation due to its Fortran underpinnings that prevented its successful application to larger sites; in particular, it would fail at any site with more than 200 wells. So the new technologies in GTS—especially the R statistical computing environment—were specifically selected to ensure that GTS would no longer have this limitation. Each of the demonstration sites for this project was also selected with this aspect in mind; all of the sites have more than 200 wells, ranging from 208 at AFP44 to 467 at Fernald. In each case, optimization analyses were successfully run, as documented in previous sections, with no memory limitations or difficulties. Based on this success, the performance objective is met.

### 7.2.8 Return on Investment (ROI)

The expected performance metric is that the annual cost savings realized from implementing a GTS-recommended optimal sampling plan will more than offset the expense of utilizing GTS and performing an optimization study. In fact, the expectation is that an ROI will occur within 3 years of implementation at most sites and at each of the demonstration sites. The purpose behind this objective is to ensure that GTS provides a cost-effective and resource-saving optimization strategy. This was evaluated by importing the optimization results generated by the ESTCP project team into the GTS cost comparison calculator. The calculator is designed to compute ROI as one of its final outputs, as discussed in Sections 1.3, 1.4, and 8.3.

Calculation of ROI essentially weighs three components: (1) cost of performing the optimization study with GTS, including data retrieval, cleaning, and preparation, along with labor hours to run and interpret the software; (2) cost of installing and sampling any additional well locations proposed by GTS; and (3) yearly savings recaptured through reductions in sampling frequency and elimination of redundant wells from the monitoring network. As mentioned earlier and detailed in Section 8.3, none of the independent site analysts completed or submitted the GTS cost comparison calculator spreadsheet. Further, the analysts were not asked to keep a detailed log of hours they spent running the software (this would have been difficult in any case given the overlap between the GTS development and demonstration phases as discussed in Section 7.2.1). In addition, while each site was responsible for gathering and submitting electronic data for the project, the ESTCP project team was responsible for data cleaning and preparation. As a consequence, reasonable assumptions had to be made concerning labor hours and rates to perform the optimization study. The ESTCP project team further decided which new well locations suggested in the network adequacy analysis should be reasonably included in the cost benefit calculations.

Using these assumptions, the estimated ROI or payback easily met the performance objective. At AFP44, the total cost of new wells and doing the optimization amounted to $59,000, less than the expected annual savings of $191,000, leading to an ROI of less than 4 months. For NOP, the total cost of new wells and optimization was approximately $89,000, compared to an annual savings of $181,000, or an ROI of roughly 6 months. At Fernald, the additional expense was $49,000 versus an annual savings of $162,000, for an ROI of approximately 4 months. So this performance objective is clearly met, even if some of the assumptions made by the ESTCP project team as to optimization costs or numbers of new wells installed were different than what the site would choose in practice.

*This page left blank intentionally.*

# 8.0 COST ASSESSMENT

This section addresses the costs and benefits of implementing GTS for LTMO at typical DoD and government sites, including the potential cost savings that might result. Most of the expected savings will be derived from reductions in sampling frequency, and more generally from the temporal and spatial redundancies that GTS identifies. Additional costs will be associated with the installation, maintenance, and sampling of any new wells suggested by the network adequacy analysis, along with costs of performing the optimization study. The net cost-benefit balance for the three demonstration sites is discussed below.

## 8.1 COST MODEL

The GTS software is publicly funded, open-source freeware. As such, any user can download and use GTS at any site, public or private, without charge. The software is also designed to run on standard Windows-based desktop computing environments, so no capital purchases are required. Therefore, the cost of implementation is the estimated cost of applying the software at a typical site, with possibly some minor training costs for initial use.

The GTS cost comparison calculator was designed to quantify and automate a simple, but realistic cost model for implementing GTS. The key cost elements associated with performing an optimization study are listed in Table 16. These include start-up costs for downloading, installing, and learning the software; data retrieval and preparation, including formatting for GTS import, data importing, and removal of outliers and COC selection once within GTS; optimization, both temporal and spatial, along with analysis of any new wells suggested by the network adequacy analysis; populating site-specific cost factors into the GTS cost comparison calculator and importing the optimization results; and periodically conducting trend flagging and plume flagging on newly collected data. Note that the cost calculator spreadsheet itself does not break out these elements in the same way as Table 15. Rather, standard labor categories are listed, with options for the user to set site-specific labor rates and number of hours expended in each category.

**Table 16. Estimated costs to apply GTS at a typical site.**

| Cost Element | Estimated Level of Effort | Estimated Cost |
|---|---|---|
| Start-up | | |
| Software cost | Free | $0 |
| Software download/install | 1-2 hrs @ $100/hr | $200 |
| Training/learning | 16 hrs @ $100/hr | $1600 |
| | **Subtotal** | $1800 |
| Data preparation/import (per site) | | |
| Data retrieval/prep | 40 hrs @ $100/hr | $4000 |
| Data import | 2 hr @ $100/hr | $200 |
| Data exploration/massaging | 2-6 hrs @ $100/hr | $600 |
| | **Subtotal** | $4800 |
| Optimization (per site) | | |
| Temporal optimization | 4-10 hrs @ $100/hr | $1000 |
| Spatial optimization | 6-24 hrs @ $100/hr | $2400 |
| Network adequacy | 2 hr @ $100/hr | $200 |
| Interpret results/write-up | 20 hrs @ $100/hr | $2000 |
| | **Subtotal** | $5600 |
| Cost-benefit analysis | | |
| Populate cost calculator | 1-2 hrs @ $100/hr | $200 |
| Import/format optimization results | 1-2 hrs @ $100/hr | $200 |
| Write-up results | 1 hr @ $100/hr | $100 |
| | **Subtotal** | $500 |
| Trend/plume flagging (periodic) | | |
| Create GTS-ready file for new data | 8 hrs @ $100/hr | $800 |
| Import data and run trend/plume flagging | 1-2 hrs @ $100/hr | $200 |
| Export reports, write-up results | 5 hrs @ $100/hr | $500 |
| | **Subtotal** | $1500 |
| **Optimization Study Total** | **110-142 hrs @ $100/hr** | **$14,200** |

## 8.2 COST DRIVERS

The cost estimates provided in Table 16 are rough upper limit estimates based on the testing performed at the three demonstration sites as part of this project. Costs of applying GTS at typical DoD and government sites may vary but should significantly exceed the estimates in Table 15 only at very complex and very large facilities (e.g., thousands of wells, hundreds of potential COCs, more than five aquifer layers, etc.). Cost drivers that would potentially impact the cost of applying GTS would include:

- *Labor Mix and Computing Costs* — Table 16 assumes that much of the effort in a typical optimization study will be conducted by mid-level and junior-level analysts, thus the assumption that labor rates will average $100 per hour across the project. Further, it is assumed that physical computational time will be billed in labor hours and that multiple variations in optimization formulation and strategy may be attempted. Should the labor mix include a higher proportion of senior-level time, the cost structure may be higher. On the other hand, should optimization runs be conducted overnight with no labor charge attached to physical computing time, costs could be significantly less than those estimated above.

- *Quality and Format of Site Data* — Data preparation cost is highly dependent on the quality and existing format of the available historical data. During data preparation, site data are converted into ASCII text files that can be imported into GTS. This includes an analytical data input file, and a water level file if those measurements exist. Obviously, the level of effort will depend on the format of the site data and the extent to which site data have previously been screened for data quality. At many sites, historical analytical sampling data are already available electronically, and reformatting those data into the proper format for input into GTS is a straightforward exercise using software such as Microsoft Excel or a robust text editor.

  Nevertheless, since GTS also requires fields and field names consistent with the ERPMS data structure, some sites may need to reformat their data to fit ERPMS conventions. Further, if some site data are not in digital format, then those data may need to be converted into electronic format, which could substantially increase the data preparation cost. The estimate provided in Table 16 of $4000 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 or missing well names or well coordinates; inconsistent aquifer designations; missing detection status [PARVQ]). If more substantial problems with data quality are found, data preparation costs could be higher.

- *Number of Distinct Sites and Aquifer Zones* — The three demonstration sites were analyzed as single, discrete areas (as encompassed by a single site boundary). AFP44 has essentially four aquifer layers, though one layer is too sparsely sampled to be reasonably analyzed by itself. NOP has three layers, and Fernald has one (based on initial data supplied to the ESTCP project team). Run times for GTS optimization were thus based on these site configurations. Since each additional aquifer layer or discrete site area increases run times linearly, costs will be higher at installations with greater numbers of site-aquifer layer pairs.

- *Number of COCs* — Each COC optimized adds linearly to GTS run times. Since the maximum number of COCs that can be simultaneously analyzed is currently capped at four, and AFP44 was analyzed with this configuration, Table 16 should accurately reflect the upper cost limit as it pertains to number of COCs. However, should a site choose to make multiple runs on more than four COCs, costs would be higher.

- *Number of Wells, Amount of Historical Data* — The number of wells in a data set adds greater than linear complexity to GTS optimization run times. At the demonstration sites, the maximum number of wells analyzed was 467 (376 unprotected and eligible for optimization). Sites with larger numbers of wells will incur more run time and hence higher cost. The length of the historical data record at each well impacts temporal optimization run times using iterative thinning. Sites with extensive histories will incur the longest run times. Since there were numerous wells at the demonstration sites with 15-20 year histories, run times may not be much longer than Table 16 for the majority of prospective facilities.

## 8.3 COST ANALYSIS

A cost-benefit analysis for applying GTS as an LTMO tool must account for the costs of doing an optimization study, the costs of any new wells added as a result of the study, and cost savings likely to be realized from identifying and eliminating redundancy. The estimated costs of performing an optimization study are presented in Table 16. The GTS cost comparison calculator is designed to balance these costs against the other two components: (1) cost of new wells and (2) cost savings from eliminating redundancy in sampling and analysis.

Actual costs and savings are subject to many site-specific factors such as the number of aquifers, numbers of wells and contaminants, cost of sampling and laboratory analysis, labor rates, and several other factors. Since these factors vary from site to site, a definitive cost analysis cannot be provided. However, it is possible to describe the factors and assumptions incorporated into the GTS cost comparison calculator and illustrate the cost analysis derived for each of the three demonstration sites.

An annual cost summary using the GTS cost comparison calculator is built from the following elements and assumptions:

- *Input of the GTS optimized network status report.* This text file includes all of the distinct baseline wells used in the analysis, their baseline and optimized sampling frequencies, and which wells were deemed redundant.
- *Analytes or analyte groups and their relative frequency of sampling.* Users are asked to input each analyte or group of analytes being monitored (e.g., metals by analytical method), as well as the laboratory analysis cost per sample for each one. Users can also input a relative frequency factor between 0 and 1 for each analyte (default=1) to indicate those contaminants or groups that are sampled either less often than the analyte sampled most frequently (e.g., metals sampled quarterly, VOCs sampled semi-annually), or that are sampled in only a portion of the site (e.g., wells in lower southwest quadrant).
- *Optimal sampling frequencies.* Although the cost comparison calculator automatically inputs optimized sampling frequencies from the optimized network status report file, users can choose to employ either a site-wide frequency, aquifer zone-specific frequencies, or well-specific frequencies, depending on which type best fits the operational profile and configuration of the site. Well-specific frequencies delineate an optimal sampling frequency for each and every well, but also then require well-specific sampling schedules. Often, operational constraints dictate a single sampling frequency for the site as a whole (site-wide), or perhaps for each aquifer (aquifer zone-specific).
- *Suggested new wells and their proposed sampling frequencies.* Users are asked to input a text file listing the number and coordinates of all new well locations. This file is exported from the GTS application as the new well location report. Each new location can be assigned its own sampling frequency, generally either the optimal site-wide frequency or an aquifer zone-specific value.

- *Costs to install new wells.* Common industry default unit costs are provided for mobilization and demobilization, monitoring well installation per foot of depth, dedicated pump, well survey, and well development. Users can override any of these defaults, including the average depth of drilling, in order to build a realistic, site-specific cost structure.
- *Quality control samples.* A default rate of 20% is used to compute the number of field QC samples to be collected each year for each analyte or analyte group. The user can override with a site-specific rate if desired. The QC samples are added to the number of samples per year collected from both essential wells and new well locations to derive a total number of samples per year per analyte and their associated analytical cost.
- *Labor rates.* Default hourly rates are provided for senior level, mid level, junior level, and technician. Users can override these rates with site-specific values.
- *Field sampling costs.* Default values are provided for the number of hours typically spent annually per well to do field sampling for each labor category (e.g., 0.1 hour for senior level, 3 hours for technician). Total field sampling costs are built up from the labor rates per hour and the number of wells sampled per year.
- *Other labor costs.* Default values are given for number of hours by labor category spent on chemistry data management (users can override). Similar input slots are also provided for typical hours spent on reports and meetings, as well as project management, administration, and QA. GTS assumes that reports, meetings, and project management costs are essentially constant regardless of whether an optimized sampling program is adopted.
- *Non-labor costs.* Default values are provided for sample shipping costs and sampling equipment and materials on a unit basis. Users can override defaults for samples per cooler and shipping cost per cooler, as well as those for materials and equipment per well.
- *Optimization study costs.* Users can input hours by labor category necessary to run a GTS optimization study. They can also input others costs, such as site visits, photocopies, etc.
- *Cost Summary.* All unit costs are escalated to compute both a baseline (i.e., current) cost summary (including all analytical and sampling costs) using the current well network and sampling frequencies, and an optimized cost summary using both the essential wells and the newly proposed well locations coupled with the GTS-optimized sampling frequencies. The overall annual net balance is derived by adding the costs of the baseline monitoring program to the costs of the optimization study, then subtracting the costs of the proposed optimized monitoring program.

The GTS cost comparison calculator was applied to each of the three demonstration sites for this project, based on the optimization analyses conducted by the ESTCP project team. Because detailed information on all the cost elements could not be obtained from every site, default values and assumptions were utilized to fill in the gaps. Thus, the cost summaries presented below should be regarded as hopefully reasonable estimates but not actual dollar amounts. It should also be noted that contractors working at AFP44 did review the GTS cost comparison calculator and provided some site-specific cost data for that installation. They noted that the defaults utilized in the calculator were quite similar to their own cost structure.

**AFP44 Estimated Cost Analysis**

Use of the GTS cost comparison calculator at AFP44 (Figure 37) involved the following site configuration and assumptions:

- In the baseline monitoring program, 208 wells were analyzed, two of which were designated as protected based on recommendation of site representatives. Within this network, a suite of VOCs was regularly and extensively sampled, including two contaminant drivers—TCE and 1,1-DCE. Two other COCs, total chromium and 1,4-dioxane, were sampled either less often or only across a portion of the network. These last two contaminants were given fractional relative sampling rates for purposes of the cost analysis (chromium=0.5, 1,4-dioxane=0.25). All four of the COCs—TCE, 1,1-DCE, chromium, and 1,4-dioxane—were optimized using GTS. Analytical costs per sample were estimated by SAIC and then confirmed by AFP44 site representatives, amounting to $25 per chromium sample, $150 per 1,4-dioxane sample, $90 for TCE and 1,1-DCE, and $115 for other VOCs. A rate of 20% for field QC sampling was also assumed.
- Three semi-distinct aquifer zones were optimized, representing a deeper layer (LZ-UZLU), an upper layer (UZUU), and a topmost layer present over a portion of the site (SGZ). Optimal sampling frequencies were computed with iterative thinning. By aquifer zone, the optimized number of annual samples per well was computed equal to one for wells in the LZ-UZLU and SGZ layers, and 0.8 for wells in the UZUU layer.
- Based on version 2 of the AFP44 database, 155 wells were deemed essential and thus part of the optimal sampling network. For purposes of costing the optimal program, aquifer zone-specific optimal sampling frequencies were selected.
- Six of 20 new well locations were retained from the network adequacy analysis. Those eliminated were either very close to existing wells or located in areas where the SGZ aquifer zone did not extend. The same aquifer zone-specific sampling frequencies were applied to these proposed wells. Default values were assumed for new well installation costs, amounting to $9000 per well.
- Labor rates by category were supplied by AFP44 representatives, along with unit labor costs for field sampling, chemistry data management, and administrative hours. Reports, meetings, and project management hours were assumed to be constant regardless of optimization.

**Summary**

<< Back | Home

| AFP44-2 | | |
|---|---|---|
| | Baseline Program | Optimized Program |
| Wells Monitored Per Year | 208 | 161 |
| Average Sampling Frequency (per well, per year) | 3.0 | 1.0 |
| Annual Costs | | |
| Sample Analysis | $ 189,080 | $ 48,125 |
| Field Sampling Labor | $ 104,146 | $ 80,613 |
| Sample Shipping | $ 12,750 | $ 3,250 |
| Sampling Materials and Equipment | $ 11,440 | $ 8,525 |
| Chemistry Data Management | $ 18,422 | $ 4,680 |
| Reports and Meetings | $ 80,400 | $ 80,400 |
| Project Management, Administration, and QA | $ 17,664 | $ 17,664 |
| **Total Annual Program Cost** | **$ 433,902** | **$ 243,257** |
| **Potential Annual Cost Savings** | | **$ 190,645** |
| **Percentage Reduction from Baseline** | | **43.94%** |
| | | |
| **Return on Investment (Payback) Analysis** | | |
| Cost of New Well Installation | | $ 45,000 |
| Cost of Optimization Analysis | | $ 13,835 |
| **Total Investment** | | **$ 58,835** |
| **Optimization will pay for itself in less than:** | | **4 months** |

**Figure 37. AFP44 cost analysis summary.**

The cost analysis at AFP44 suggests that almost 44% of the baseline monitoring program cost might be eliminated by adopting the GTS optimized sampling plan, or an approximate total of $191,000 per year. Less savings would be realized in any year in which an optimization study was conducted or new wells were installed. Assuming this study was conducted at the start of the first year of a multiyear monitoring horizon, the net savings for the first year would amount to roughly $132,000, after installing six new wells and paying for the study. Still, the estimated ROI is less than 4 months.

**NOP Estimated Cost Analysis**

Use of the GTS cost comparison calculator at NOP (Figure 38) involved the following site configuration and assumptions:

- In the baseline monitoring program, 250 wells were analyzed, 77 of which were designated as protected by directive of site representatives. Within this network, a suite of VOCs is regularly and extensively sampled, including one contaminant driver, TCE. Another suite of explosives, including COC RDX, is also regularly sampled. The two COCs—TCE and RDX—were optimized as part of the demonstration. Analytical costs per sample were initially estimated by SAIC but then slightly revised by NOP site representatives. These amounted to $100 per VOC sample and $250 per explosives sample. A rate of 20% for field QC sampling was assumed.
- Three distinct aquifers were optimized, representing SHALLOW, MEDIUM, and DEEP layers. Optimal sampling frequencies were computed with iterative

thinning. By aquifer zone, the optimized number of annual samples per well was computed as one for wells in the MEDIUM and DEEP layers, and 1.33 for wells in the SHALLOW layer.

- Including the 77 protected locations, 222 wells were deemed essential and thus part of the optimal sampling network. For purposes of costing the optimal program, aquifer zone-specific optimal sampling frequencies were selected.
- Ten of 14 new well locations were retained from the network adequacy analysis. Those eliminated were very close to existing wells. The same aquifer zone-specific sampling frequencies were applied to these proposed wells. Default values were assumed for new well installation costs, amounting to $7500 per well.
- Default labor rates by category were utilized, along with default unit labor costs for field sampling, chemistry data management, and administrative hours. Reports, meetings, and project management hours were assumed to be constant regardless of optimization.

**Summary**

<< Back | Home

| NOP Cost Summary | | |
|---|---|---|
| | **Baseline Program** | **Optimized Program** |
| **Wells Monitored Per Year** | **250** | **232** |
| **Average Sampling Frequency (per well, per year)** | **2.6** | **1.1** |
| **Annual Costs** | | |
| Sample Analysis | $ 272,650 | $ 111,300 |
| Field Sampling Labor | $ 80,000 | $ 74,240 |
| Sample Shipping | $ 9,750 | $ 4,000 |
| Sampling Materials and Equipment | $ 13,750 | $ 12,210 |
| Chemistry Data Management | $ 10,322 | $ 4,214 |
| Reports and Meetings | $ 64,300 | $ 64,300 |
| Project Management, Administration, and QA | $ 14,640 | $ 14,640 |
| **Total Annual Program Cost** | **$ 465,412** | **$ 284,904** |
| **Potential Annual Cost Savings** | | **$ 180,508** |
| **Percentage Reduction from Baseline** | | **38.78%** |

| Return on Investment (Payback) Analysis | |
|---|---|
| Cost of New Well Installation | $ 75,000 |
| Cost of Optimization Analysis | $ 13,500 |
| **Total Investment** | **$ 88,500** |
| **Optimization will pay for itself in less than:** | **6 months** |

**Figure 38. NOP estimated cost summary.**

The cost analysis at NOP suggests that almost 39% of the baseline monitoring program cost might be eliminated by adopting the GTS optimized sampling plan, or an approximate total of $180,000 per year. Most of the savings is realized through reduction in sampling frequencies. Less savings would be realized in any year in which an optimization study was conducted or new wells were installed. Assuming this study was conducted at the start of the first year of a multi-year monitoring horizon, the net savings for the first year would amount to roughly $92,000, after installing 10 new wells and paying for the study. The estimated ROI is less than 6 months.

**Fernald Estimated Cost Analysis**

Use of the GTS cost comparison calculator at Fernald (Figure 39) involved the following site configuration and assumptions:

- At least some historical data existed for 467 wells and DPT locations in the baseline monitoring program. Of these, 91 were designated as protected because they had recently been abandoned but were still part of the database. To ensure that these abandoned locations were not included as part of either the current baseline or optimized sampling programs, all 91 were manually removed from the GTS optimized network status report prior to importing into the GTS cost comparison calculator. This left 376 active locations as part of the baseline monitoring program. Within the current network, the single contaminant driver and COC was uranium. Analytical costs per sample were estimated by SAIC at $75 per sample. A rate of 20% for field QC sampling was assumed.
- Although uranium was the only COC at Fernald and the only contaminant assessed in the cost analysis, the historical database contained a few other contaminants sampled sporadically at a much more limited subset of well locations. Including these contaminants in the cost analysis would tend to increase the overall cost savings but has not been estimated in Figure 39.
- Based on the data that was initially provided to the ESTCP project team, all locations at Fernald were analyzed as if part of a single aquifer (2D analysis). Optimal sampling frequencies were computed with iterative thinning. The optimized number of annual samples per well was computed as 1.33.
- The number of active wells and DPT locations deemed essential and thus part of the optimal sampling network was 231. For purposes of costing the optimal program, a site-wide optimal sampling frequency was selected.
- Four new well locations were retained from the network adequacy analysis. The same site-wide sampling frequency was applied to these proposed wells. Default values were assumed for new well installation costs, amounting to almost $9000 per well.
- Default labor rates by category were utilized, along with default unit labor costs for field sampling, chemistry data management, and administrative hours. Reports, meetings, and project management hours were assumed to be constant regardless of optimization.

Summary

<< Back Home

| Fernald | | |
|---|---|---|
| | Baseline Program | Optimized Program |
| Wells Monitored Per Year | 376 | 231 |
| Average Sampling Frequency (per well, per year) | 3.5 | 1.3 |
| Annual Costs | | |
| Sample Analysis | $ 119,475 | $ 27,750 |
| Field Sampling Labor | $ 120,320 | $ 73,920 |
| Sample Shipping | $ 10,000 | $ 2,350 |
| Sampling Materials and Equipment | $ 20,680 | $ 12,485 |
| Chemistry Data Management | $ 10,554 | $ 2,451 |
| Reports and Meetings | $ 64,300 | $ 64,300 |
| Project Management, Administration, and QA | $ 14,640 | $ 14,640 |
| **Total Annual Program Cost** | **$ 359,969** | **$ 197,896** |
| **Potential Annual Cost Savings** | | **$ 162,072** |
| **Percentage Reduction from Baseline** | | **45.02%** |

| Return on Investment (Payback) Analysis | |
|---|---|
| Cost of New Well Installation | $ 35,000 |
| Cost of Optimization Analysis | $ 13,568 |
| **Total Investment** | **$ 48,568** |
| **Optimization will pay for itself in less than:** | **4 months** |

**Figure 39. Fernald estimated cost analysis.**

The cost analysis at Fernald suggests that 45% of the baseline monitoring program cost might be eliminated by adopting the GTS optimized sampling plan, or an approximate total of $162,000 per year. Savings are realized both through reduction in sampling frequencies and elimination of redundant wells. Less savings would be realized in any year in which an optimization study was conducted or new wells were installed. Assuming this study was conducted at the start of the first year of a multiyear monitoring horizon, the net savings for the first year would amount to roughly $113,000, after installing four new wells and paying for the study. The estimated ROI is less than 4 months.

# 9.0 IMPLEMENTATION ISSUES

This section discusses issues related to future implementation of the GTS software technology at prospective sites. Relevant issues discussed below include:

- Software availability and documentation
- Ease of use
- Limitations of GTS v1.0
- Proposed and recommended changes to the software
- Regulatory issues.

**Software Availability and Documentation**

The anticipated end users of GTS include both government personnel and support contractors managing groundwater monitoring programs, whether at public or private facilities. A copy of the software executable, GTS cost comparison calculator spreadsheet, and users guide is available on the AFCEE website. Sample input data files—preformatted according to GTS specifications—are also available at the website.

Anyone with legal access to the AFCEE website can download and install GTS for free onto their desktop computer. As publicly funded, open source freeware, there are no restrictions on GTS usage, nor does a license need to be secured or purchased. The software and users guide were previously submitted as a separate deliverable under this ESTCP project.

Although the software and its usage are free, there is no technical support or training available for GTS at this time. Such support and/or training can be purchased separately from MacStat Consulting, Ltd.

**Ease of Use**

Overall, the GTS software was found to be easy to use by the testers and mid-level site analysts. None of these users was formally trained on the software; questions regarding usage (and other project matters, including software bugs and development) were fielded in weekly conference calls sponsored by the ESTCP project team. Experience with other LTMO software varied among the testers; most had some previous experience running MAROS. Users commented that:

> This tester rates the general usability of GTS as very good considering it is in beta form. Its modular structure is logical and relatively easy for the minimally experienced geostatistical practitioner to use.
>
> The five major modules coupled with Windows menu and dialog boxes allow an environmental professional with limited statistical training and expertise to navigate successfully through the many spatial and temporal elements of GTS. The GUI appears to be highly functional and user friendly.

The software is quite user-friendly. The screens are easy to navigate and read. The screen sequence is logical and appears to be structured to prevent a novice user from bypassing necessary steps. On the other hand, the ability to jump to other steps that have either already been conducted or that can be conducted based on the steps already completed make the program easy to navigate.

Apart from bugs encountered during the Fernald application, GTS was easily used. The interface made sense and was clear.

The overall ease of use is good, as familiarity with the 5 main modules and their underlying windows comes fairly quickly.

### Limitations of GTS v1.0

GTS v1.0 has certain limitations that will impact its use at prospective sites. Many of these concerns and limitations have been mentioned earlier in this report but are listed here for completeness:

- *Data Importing* — GTS requests input of a large number of data fields, though users have not always been clear on which fields are required versus optional. In addition, the data fields must be named and formatted according to ERPMS-consistent conventions. Some users suggested that the importing process could be simplified and better explained.
- *Exporting Graphics* — GTS is predicated on significant graphical analysis of data and generates a large number of statistical graphics when applied at medium to larger sites. Yet there is no current feature allowing for easy export of batches of related plots and maps. Instead users must capture screen shots of individual graphics they would like to save and import into other documents or software. In addition, GTS maps are static images and not configured for import into GIS software.
- *Map Displays* — Users commented that "maps created by GTS do not always have consistent spacing along the easting and northing axes, leading to distorted views." Users also mentioned lack of control over colors, symbols, fill patterns, and contours. Inability to contour areas of regulatory exceedance was cited as a reason for GTS base maps looking different and inferior to traditional plume maps, along with the coarse default grid over which GTS computes map estimates.
- *Compatibility* — GTS was designed to be fully compatible with desktop systems running Windows XP. However, the architecture was finalized prior to the adoption of either Windows Vista or Windows 7. Some users expressed difficulties in getting GTS properly loaded and running on Vista or Windows 7 systems, especially with 64-bit machines, although others seemed to have little difficulty.
- *Optimization Runtimes* — At large sites (>200 wells), optimization runs—using iterative thinning or especially GTSmart—may take several hours to complete (perhaps necessitating overnight runs). This is a limitation for users needing to

complete a project on a tight deadline or for those who want to test out several variations of parameter choices or data configurations.

- *Technical Guidance* — Multiple users commented that since the current users guide does not include any technical appendices, they were sometimes unsure of what GTS was doing or computing at particular steps, or that they were unsure how to interpret GTS results (e.g., why were certain wells flagged as redundant but not others; how to select and interpret temporal and spatial bandwidths, etc.).
- *Minimum Data Requirements* — Effective spatial optimization in GTS requires a minimum of 20-25 wells and at least two sampling events per well; temporal optimization requires at least one well and four to eight distinct sampling events per location.
- *Radiochemical Data* — GTS does not offer sophisticated handling of radiochemical data, particularly measurements recorded with non-positive values (i.e., zeros or negatives). These data must first be converted to positive values, unless they represent non-detects with a known, positive detection or reporting limit.
- *Temporal Optimization* — Optimized sampling intervals from temporal variograms in GTS often do not match the optimized sampling intervals from iterative thinning using the same data. Further improvements to the temporal variogram algorithm may be needed, especially to account for sites with spatial trends that are actively changing over time.
- *Cost-Accuracy Trade-offs* — Cost-accuracy trade-off curves in GTS are not interactive. Although the bias limits can be adjusted by the user, the spatial optimization must be completely re-run each time those limits are changed, in order to see the impact of the revised limits and to generate a new optimal network.
- *Plume Mass* — GTS v1.0 does not track changes in contaminant or plume mass, nor does it allow users to specify contaminant mass as an optimization criterion.

**Proposed and Recommended Changes to the Software**

Based on the limitations of the current v1.0 release of GTS, along with additional user feedback, several changes are proposed for a future v2.0 release in order to increase its ease of use, flexibility, and adaptability to real-life environments and messy data sets. These include the following items:

- **System-wide Upgrades**
  - Make GTS fully compatible with Windows 7.
  - GUI — Add menus to provide direct access to GTS features and components. Users will be allowed to set preferences and options.
  - Add context-sensitive user help throughout GTS.

- Restructure the GUI to more easily allow users to perform only a temporal optimization if a site has less than 20 wells, or only a spatial optimization if there are fewer than four to six separate sampling events.
- Improve sorting and display of SQLite database tables (these house data imported into GTS).
- Improve sorting and display of GTS analysis reports.
- Improve user navigation and searching through batches of GTS plots. A typical GTS analysis generates a large volume of plots that the user may desire to electronically save and/or print for use outside the application. Add ability to save and print graphics from GTS output, including automated batches of graphics when desired.
- Graphics — Add more user control over graph options and appearance; improve display of maps and shapefile map overlays; expand interactivity between paired graphs and tables (e.g., if user clicks on a well in a post-plot, highlight that well in the associated table).
- Users Guide — Expand to include technical appendices and additional material on how to judge and interpret GTS optimization results.

- **Module A (Prepare) Upgrades**
  - Expand checks for inconsistent or missing data, such as dilution outliers, unusual lab qualifiers, inconsistent elevations and depths, duplicate records, etc.
  - Improve computation and display of GTS time slices (i.e., time snapshots used to subset data for analysis); allow users to manually adjust time slice ranges, in order to account for site-specific changes to the monitoring program (e.g., installation of new treatment system).
  - Improve display and documentation of data import capability. Streamline and improve user interface for data import, making it easier for users to navigate the import process.
  - Improve display of well post-plots, including addition of separate plots by vertical zone.
  - Restrict spatial mapping and display to expanded convex hull around existing well locations.
  - Outliers — Combine current temporal and spatial outlier searches into one; simplify GTS interface for identifying and confirming suspected outliers; perform outlier searches separately by vertical zone for each COC.
- **Module B (Explore) Upgrades**
  - Improve GTS interface for displaying data summary statistics.
  - Display post-plots of concentration levels and MCL exceedances by vertical zone.

- Improve vertical horizon analysis; check for consistency of vertical zone designations; improve display of current box plots.

- **Module C (Baseline) Upgrades**
  - Sampling gaps — Improve ease of use by eliminating current "sampling gaps" diagnostic interface. Revise trend-fitting algorithms to better account for large sampling gaps.
  - Improve usability of table of trend types and "Check Bandwidths" interface.
  - Improve display of baseline trends; link each trend with a displayed numeric table of trend results; hot-link locations on each trend map with their associated baseline trends.
  - Spatial Bandwidth interface — Improve user ability to select appropriate bandwidth parameters by adding new diagnostic plots and improving existing display of map residuals.
  - Improve display of base maps and existing color bar legends; expand viewing options to improve handling of highly skewed data.
  - Test and deploy water-flow aided spatial mapping, GTS does not require numerical flow and transport models, yet will provide improved spatial mapping by combining information about the potentiometric surface along with observed patterns of contaminant levels. Install as an additional user option for data sets that include water level measurements.

- **Module D (Optimize) Upgrades**
  - Temporal variograms — Improve computation and accuracy by a) enabling option to compute variograms on transformed data (e.g., log, square root); test option of computing variograms on de-trended data, using baseline trend to de-trend each COC-well pair.
  - Improve display of iterative thinning optimization results by adding graphic that overlays baseline trend, optimized trend, and confidence band utilized in the thinning algorithm.
  - Temporal optimization — Revise iterative thinning algorithm to allow optimization of both Theil-Sen and LWQR trends; as part of this change, perform exhaustive thinning on small data sets ($n \leq 10$) to expand flexibility and improve accuracy of iterative thinning technique.
  - Spatial optimization — Current GTSmart optimization strategy is a quasi-genetic algorithm. Improve by developing and deploying a full genetic algorithm that retains the computational benefits of GTSmart. This will improve the accuracy and defensibility of GTS spatial optimization results.
  - Add option for user to separately optimize water level data if available. This will allow for more efficient potentiometric surface mapping.

- Increase flexibility by adding option for user to pick alternative critical index threshold by which GTS delineates critical versus redundant well locations.
- Trade-off curves — develop and test option of combining current trade-off curves into single, weighted curve for use in determining points of optimality; link points on trade-off curve to specific sampling plans; this will allow user to compare different possible optimal plans without having to re-run entire optimization routine.
- Improve display of spatial optimization results by adding a graphical and tabular summary of the numbers of essential or redundant wells by vertical zone.
- Cost Comparison Calculator — Integrate current cost calculator Excel spreadsheet into GTS interface. This will allow seamless computation of optimization benefits from within the GTS application, instead of user having to export results and then import into a separate spreadsheet in Excel.

- **Module E (Predict) Upgrades**
  - Trend anomalies — Improve current prediction band used to flag potential anomalies by revising code to add a flat linear extension. This will cover cases where the apparent trend has recently flattened out instead of continuing a past rise or descent.
  - Improve display of trend anomalies by hot-linking the time series plots which currently display prediction bands to locations graphed on the trend anomalies post-plot (i.e., if a user clicks on a particular location, the hot-linked time series plot would then display).
  - Improve display and usefulness of uncertainty envelopes by expanding viewing options to include either log-scale or concentration-scale displays.
  - Hot-link well-specific time series plots also to locations displayed on plume anomalies post-plot. This will allow user to gain longitudinal perspective on potential plume anomalies.

**Regulatory Issues**

Regulatory approval of a GTS-optimized sampling plan typically boils down to three concerns: (1) Is there an existing general consensus among stakeholders that sampling redundancy might be present and a regulatory willingness to consider alternate approaches? (2) Will removing wells and sampling events from regular monitoring preclude obtaining data needed for remedial decision-making or site characterization? (3) How can GTS plume/site maps be trusted if they don't look like traditional hydrogeologic maps?

Interaction with regulators regarding implementing the GTS results at the three demonstration sites was not a specific part of this ESTCP project. However, each site was interested in evaluating the optimization results to determine whether changes would be justified in its sampling program. Preliminary findings of the optimization study were also presented to joint meetings of regulators and site personnel at AFP44. Both in that presentation and in talks given to other (non-ESTCP) sites, site personnel have generally been very receptive to GTS as an LTMO tool and have desired to use GTS results as a line of evidence in regulatory discussions/negotiations.

Obtaining regulatory acceptance of GTS will probably require two major steps: (1) increasing awareness of LTMO in general, and awareness of GTS v1.0 in particular, within the regulatory community and (2) individual sites agreeing to petition regulators for modifying their LTM program based on a GTS-optimized sampling plan. As discussed in the section on current limitations above, there may also be a need to improve the mapping tools within GTS so that users can set site-specific contours for visualizing areas of regulatory exceedance and so that hot spots are mapped more accurately.

To achieve the first step, AFCEE is actively promoting and advertising GTS as an available software tool. Efforts are also underway to develop an Interstate Technology & Regulatory Council (ITRC) project that will spotlight GTS under the larger umbrella of analyzing groundwater monitoring data and meeting groundwater regulatory requirements.

With respect to the second step, each of the demonstration sites indicated they would be reviewing the GTS results to determine applicability and usability of the recommendations. AFP44 contractors 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. This was because they wanted to include site-specific factors not available to the ESTCP project team. Also, given the 3-year schedule of this ESTCP project and the fact that the most recent year's worth of data at each site was reserved for validation and testing of the trend and plume flagging features, the demonstration sites would be advised to repeat the optimization analysis using up-to-date data before incorporating the results into a revised LTM sampling plan proposal.

Improving the mapping capabilities in GTS will require an upgrade to the existing version. Efforts are underway to secure funding for such improvements.

*This page left blank intentionally.*

## 10.0 REFERENCES


1. RPO Inventory and Performance Software [RIPS], Air Force RPO Outreach Office.

2. DoD. 2006. *Defense Environmental Programs Fiscal Year 2005 Annual Report to Congress*. Office of the Deputy Under Secretary of Defense, Washington, D.C., 2006.

3. Cameron, K. 2004. Better optimization of LTM networks. *Bioremediation Journal*, 8 (03-04), 89-108.

4. Loader, C. 1999. *Local Regression and Likelihood*. New York: Springer-Verlag, 1999.

5. Goovaerts, P. *Geostatistics for Natural Resources Evaluation.* Oxford University Press, New York, 1997

6. Cameron, K., P. Hunter. 2002. Using spatial models and kriging techniques to optimize long-term ground-water monitoring networks: a case study. *Environmetrics*, 13, 629-656, 2002.

7. Cameron, K. 1999. Optimization of long-term monitoring networks: a statistical approach. Conference on Subsurface Remediation: Improving Long-Term Monitoring & Remedial Systems Performance, St. Louis, MO, June 1999.

8. Cameron, K., P. Hunter. 2000. Optimization of LTM networks: statistical approaches to spatial and temporal redundancy. Spring National Meeting of American Institute of Chemical Engineers, Atlanta, GA, March 2000.

9. Cameron, K. 2002. Geostatistical optimization at AFP06 using GTS. 2002 AFCEE Technology Transfer Workshop, San Antonio, TX, March 2002.

10. Cameron, K. Optimization at AFP06 using improved GTS. 2003 AFCEE Technology Transfer Workshop, San Antonio, TX, February 2003.

11. Cameron, K.M., P. Hunter. 2003. Optimization of LTM networks at AF Plant 6 using GTS. In V.S. Magar & M.E. Kelley (Eds.), *In Situ and On-Site Bioremediation — 2003.* Proceedings of the Seventh International In Situ and On-Site Bioremediation Symposium (Orlando, FL; June 2003). ISBN 1-57477-139-6, Columbus, OH: Battelle Press, 2003.

12. Cameron, K., P. Hunter. 2004. Optimizing LTM networks with GTS: three new case studies. Conference on Accelerating Site Closeout, Improving Performance, and Reducing Costs Through Optimization, Dallas, TX, June 2004.

13. Cameron, K. 2006. Optimizing LTM networks: applying GTS at Tinker AFB. World Environmental & Water Resources Congress, Omaha, NE, May 2006.

14. USEPA. 2005. *Roadmap to long-term monitoring optimization*, Office of Superfund Remediation and Technology Innovation (5102G), USEPA 542-R-05-003, 2005.

15. Minsker, B. 2003. *Long-Term Groundwater Monitoring: The State of the Art,* ASCE/EWRI Task Committee Report. American Society of Civil Engineers: Reston, VA, 2003.

16. Yamamoto, J.K. 2000. An alternative measure of the reliability of ordinary kriging systems. *Mathematical Geology,* 32(4): 489-509, 2000.

17. Stewart R., T. Doley, and P. Hunter. 2004. Estimating cost savings from the optimization of long term monitoring programs at U.S. Air Force bases. Conference on Accelerating Site Closeout, Improving Performance, and Reducing Costs Through Optimization, Dallas, TX.

18. Herrera, G.S., and G.F. Pinder. 1998. Cost-effective groundwater quality sampling network design. *Proc. XII International Conference on Computational Methods in Water Resources, Vol. I,* Crete, June 1998.

19. Rizzo, D.M., D.E. Dougherty, and M. Yu. 2000. An adaptive monitoring and operations system (aLTMOs) for environmental management. In *Building Partnerships, Proc. 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), 2000.

20. Deschaine, L.M. 2003. Simulation and optimization of large scale subsurface environmental impacts; investigations, remedial design and long term monitoring. *Mathematical Machines and Systems*, National Academy of Sciences of Ukraine, No. 3,4: 201-218, 2003.

# APPENDIX A

# POINTS OF CONTACT

| Point of Contact | Organization | Phone<br>Fax<br>E-Mail | Role In Project |
|---|---|---|---|
| Philip Hunter, P.G. | HQ AFCEE/TDV<br>3300 Sidney Brooks Road<br>Brooks City-Base TX 78235 | Phone : 210-395-8441<br>E-mail : philip.hunter@us.af.mil | Principal Investigator |
| Robert B. Stewart | SAIC<br>8301 Greensboro Drive<br>Mclean, VA 22102 | Phone : 703-676-6965<br>Fax : 703-736-0815<br>E-mail : robert.b.stewart@saic.com | SAIC Project Manager |
| Michael Kenny | SAIC<br>1901 S 1st Street, Suite D-1<br>Champaign, Il 61820 | Phone : 217-337-9520<br>E-mail : michael.j.kenny@saic.com | Lead Computer Scientist |
| Kirk Cameron, Ph.D. | MacStat Consulting<br>10330 Mill Creek Court<br>Colorado Springs, CO 80908 | Phone : 719-532-0453<br>Fax : 719-532-0453<br>E-mail : kcmacstat@qwest.net | Lead Statistical Scientist; R Programmer |
| Dave Becker, P.G. | US ACE HTRW<br>12565 W Center Road<br>Omaha, NE 68144 | Phone : 402-697-2655<br>E-mail : dave.j.becker@nwd02.usace.army.mil | Government Partner; GTS Tester |
| Andrea Leeson, Ph.D. | ESTCP Office<br>901 North Stuart Street<br>Suite 303<br>Arlington, VA 22203 | Phone : 703-696-2118<br>Fax : 703-696-2114<br>E-mail : Andrea.Leeson@osd.mil | Environmental Restoration Program Manager |

ESTCP Office
901 North Stuart Street
Suite 303
Arlington, Virginia 22203
(703) 696-2117 (Phone)
(703) 696-2114 (Fax)
E-mail: estcp@estcp.org
www.serdp-estcp.org