HydroImage, a User-Friendly Hydrogeophysical Characterization Software

Prepared by:
Chin Man Mok[1] (Principal Investigator), Susan Hubbard[2], Jin-song Chen[2], Raghu Suribhatla[1], and Dawn Kaback[1]

A Small Business Technology Transfer Research Project
Grant Number DE-FG02-05ER86244

AMEC Environment & Infrastructure, Inc.
2000 South Colorado Boulevard, Suite 2-1000
Denver, CO 80222
Tel: (303) 935-6505

January 2014

[1] AMEC Environment and Infrastructure
[2] Lawrence Berkeley National Laboratory

EXECUTIVE SUMMARY

HydroImage, user-friendly software that utilizes high-resolution geophysical data for estimating hydrogeological parameters in subsurface strata, was developed under this grant. HydroImage runs on the personal computer platform to promote broad use by hydrogeologists to further understanding of subsurface processes that govern contaminant fate, transport, and remediation. The unique software provides estimates of hydrogeological properties over continuous volumes of the subsurface, whereas previous approaches only allow estimation of point locations. Thus, this unique tool can be used to significantly enhance site conceptual models and improve design and operation of remediation systems.

The HydroImage technical approach uses statistical models to integrate geophysical data with borehole geological data and hydrological measurements to produce hydrogeological parameter estimates as 2-D or 3-D images. During the project, the HydroImage software was developed and then successfully tested using real field-site data. The software is now available to be used by scientists throughout the world to provide great benefits to the public interested in protecting human health and the environment. Results of the project have been presented at international conferences including the Battelle Chlorinated Solvents Conference and the American Geophysical Union Annual Conference.

1.0 INTRODUCTION

HydroImage is user-friendly software, developed during this project, to systematically utilize high-resolution geophysical data for estimating hydrogeological parameters in subsurface strata. HydroImage runs on the personal computer platform and was designed for use by hydrogeologists. This final report describes the work performed during Phase II of the project.

1.1 Grant/Contract Information

This project was performed under grant/contract no. DE-FG02-05ER86244, which was initially awarded to Geomatrix Consultants, Inc. (Geomatrix) in collaboration with the Lawrence Berkeley National Laboratory (LBNL) in June 2005. After AMEC Earth and Environmental acquired Geomatrix in June 2008, Geomatrix became AMEC Geomatrix, Inc. (AMEC Geomatrix), and the staff of AMEC Geomatrix, Inc., continued to work on this project. The project was completed in August 2010. In 2012, AMEC Geomatrix merged into AMEC Environment and Infrastructure (AMEC).

1.2 Problem Statement and Background

The shallow subsurface of the Earth serves as the repository for a large percentage of our water supplies. Unfortunately, this zone also houses wastes disposed intentionally as well as contaminants that have migrated from waste disposal units or been unintentionally introduced into the subsurface. The protection of water resources and the clean-up and long-term stewardship of contamination require effective characterization and monitoring of subsurface properties and processes. It is widely recognized that natural heterogeneity and spatial variability of hydraulic parameters in the subsurface control 1) infiltration within the vadose zone, 2) groundwater flow, 3) the spread of contaminants, and 4) remediation efficacy. Subsurface heterogeneity plays a significant role in long-term processes, such as monitored natural attenuation. Conventional sampling techniques for characterizing or monitoring the shallow subsurface typically involve collecting core samples and acquiring hydrological measurements and/or geophysical log data from boreholes. When the size of the contaminated site is large relative to the scale of the hydrological or geological heterogeneity, data obtained at point locations are unlikely to capture key information about field-scale heterogeneity, thus challenging our ability to adequately design effective remediation systems or monitor processes. As such, there is a continuing need for cost-effective subsurface characterization and monitoring techniques that can provide information about properties and processes with a reasonable resolution and over reasonable field scales.

Numerical models of fluid flow and contaminant transport are also used as critical tools for designing effective remediation programs. However, input data for these models often rely upon point measurements as described above or upon predicted values, which both introduce significant uncertainties into the models. The inability to adequately parameterize the numerical models in a way that represents natural heterogeneity is perhaps one of the largest contributors to the frequent failure of field-scale transport predictions. The limitations (both technical and economical) of the ability to satisfactorily characterize and monitor subsurface variability indirectly impact the cost of remediation design, remediation rehabilitation, and long-term stewardship at thousands of contaminated sites across the United States.

Over the last two decades, geophysical methods have successfully been used for shallow subsurface characterization at many contaminated sites. In most cases, these geophysical characterization activities produce images of a geophysical attribute, such as seismic velocity or electrical conductivity. More recently, the field of hydrogeophysics, which strives to more rigorously integrate geophysical data with direct hydrogeological or biogeochemical datasets within an estimation framework, has developed. Through such procedures, the dense geophysical data can be used to estimate properties or to monitor processes that are important for flow and transport studies, such as lithology, water content, permeability, or sediment geochemistry. A recent book called Hydrogeophysics (Rubin and Hubbard, 2005) describes the state-of-the-art of this new research discipline.

Although hydrogeophysical advances have been developed and seismic, radar, and electrical methods are all now commercially available, the routine use of these methods is hindered by difficulties in transferring the state-of-the-discipline into practice. A key obstacle in this transfer is the lack of a user-friendly software tool that can be broadly utilized across the environmental industry. We focus herein on advancing HydroImage, which is a software package that enables the estimation of hydrogeological properties given geophysical measurements and limited direct measurements, such as from boreholes. Development of HydroImage should facilitate the transfer of the research advances into practice across DOE, DOD, and commercial sites.

To address this problem, we have developed a user-friendly software package, which integrates continuous geophysical data with limited borehole data to estimate hydrogeological parameters of interest in the subsurface. The HydroImage software package can be used to significantly enhance site conceptual models and thus improve design and operation of remediation systems.

2.0 TECHNICAL APPROACH

Our technical approach focuses on integration of spatially extensive geophysical data with direct (geological, hydrological, biogeochemical, and geophysical) borehole measurements to improve characterization and monitoring of the subsurface over a variety of resolutions and spatial scales. Our technical approach uses statistical models to integrate geophysical data with borehole geological and hydrological measurements to produce hydrogeological parameter estimates as 2-D or 3-D images (Figure 1). Statistical models are able to provide good estimated values as well as their associated uncertainty. In contrast to conventional site characterization approaches that rely solely on borehole measurements, our approach can combine all available data.

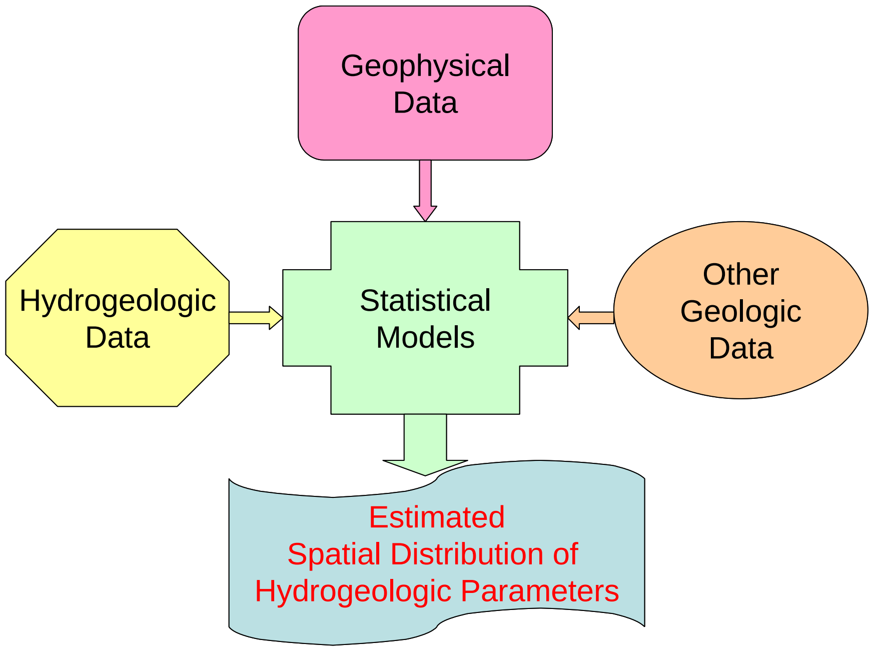

Figure 1: Schematic of the technical approach

Because the utility of geophysical data for estimating hydrogeological parameters varies on a case-by-case basis, HydroImage includes quality control steps that will allow users to assess the utility of a certain geophysical dataset for a particular application and the quality of the geophysical data inversion approach. These quality control steps in HydroImage avoid the use of the software in a 'black box' fashion, while informing the user about the validity of a particular dataset to estimate a particular hydrogeological parameter.

2.1 Overview of the Bayesian Framework

Bayesian methods provide a systematic, consistent, and general framework for data integration or assimilation (Bernardo and Smith, 2002). There are many successful applications in hydrology (e.g., Lortzer and Berkhout, 1992; Copty et al., 1992; McLaughlin and Townley, 1996; Chen et al., 2001). As shown in Figure 2, this framework is conceptually simple.

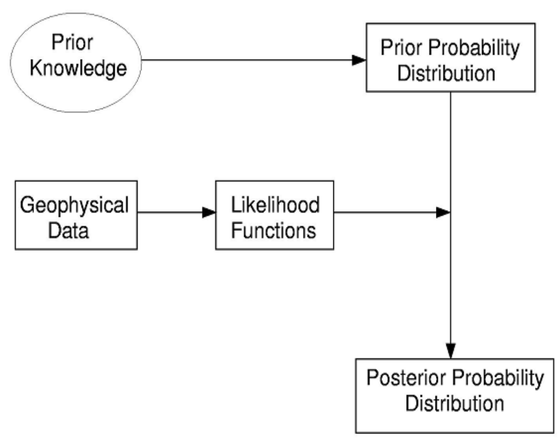

Figure 2: General framework of Bayesian models

Suppose, we have an unknown vector $\boldsymbol{\theta}$, which may include many components, we need to estimate the parameter vector given data set

$\mathbf{d}$ (e.g., geophysical data). Before we look over or collect the data, we may have some knowledge about the unknown. We can quantify the prior knowledge using a probability distribution $f(\boldsymbol{\theta})$, referred to as 'prior'. According to Bayes' Theorem (Bernardo and Smith, 2002), we can get the conditional probability of vector $\boldsymbol{\theta}$ given the data set $\mathbf{d}$ as follows:

$$f(\boldsymbol{\theta}\,|\,\mathbf{d}) = \frac{f(\mathbf{d}\,|\,\boldsymbol{\theta})\,f(\boldsymbol{\theta})}{f(\mathbf{d})} = Cf(\mathbf{d}\,|\,\boldsymbol{\theta})\,f(\boldsymbol{\theta}) \ . \qquad (1)$$

Equation 1 defines a joint conditional probability distribution of the unknown, which is referred to 'posterior.' The first term on the right side is a normalizing constant, and the second term is the likelihood function of data $\mathbf{d}$, which link the data to the unknown. The last term on the right side is the prior probability distributions, derived from other sources of information. Because the data integration results are probability distribution, we have exhaustive information about the unknown parameter, for example, means, modes, predictive intervals, etc.

2.2 Geophysical Inversion

To use geophysical attributes for hydrogeological parameter estimation, geophysical measurements obtained through analysis of tomographic data (e.g., radar or seismic crosshole travel time) must first be inverted to obtain estimates of the geophysical attributes (i.e., radar or seismic velocity) along the wellbore traverse. Many different types of approaches exist for inverting different types of geophysical data. We have chosen to include the Algebraic Reconstruction Technique (ART, Peterson, 1985) inversion approach within HydroImage to enable users to invert radar or seismic tomographic attributes. In the following, we provide a short description of the method.

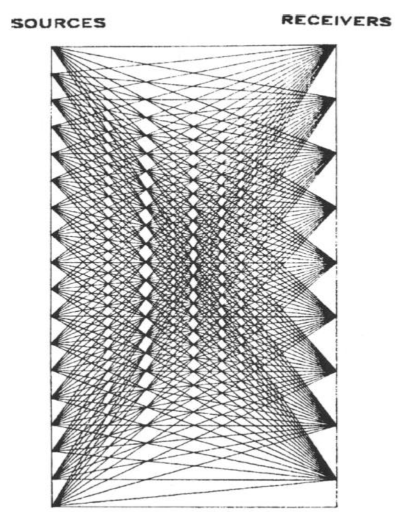

Figure 3: Configuration of 2D crosswell survey (from Peterson et al.,

As shown in Figure 3, there are two boreholes, with sources being deployed in one borehole and multiple receivers being deployed in the other. Radar or seismic signals are sent out from each source and various types of responses can be recorded at receivers. For ease of description, we consider only traveltime from the source to the receivers as data in the tomography inversion and strive to estimate radar or seismic velocity from the traveltime data. The crosswell region is first divided into many rectangle pixels, for which radar or seismic velocity is to be estimated. If the velocity contrast within the region is small, which is valid for many environmental applications, we can assume the signals pass the region as a straight-ray. Therefore, we can write the traveltime of a ray-path as the summation of the time that the ray passes through. In general, let y_k, $k = 1, 2, \cdots, N$ presents the measured traveltime for N paths and let x_i represent the slowness (i.e., reciprocal of velocity) at pixel i that a given ray-path passes through. Thus, we have

$$y_k = \sum_{i=1}^{I} \Delta\alpha_{ki} x_i \,. \tag{2}$$

In equation 2, $\Delta\alpha_{ki}$ is the length of the ray k that penetrates pixel i, and I is the total number of pixels intersected by ray k. Equation 2 forms a set of linear equations and through common inversion techniques, the equation may be solved in principle for x_i, $i = 1, 2, \cdots, m$.

Because the linear equations involve a large, sparse matrix, it is usually impractical. Instead, we use the Algebraic Reconstruction Technique (ART), which is well-studied in linear algebra. These techniques are iterative, i.e., where one equation (one ray-path) is analyzed at a time. Starting from an initial value $x_i^{(0)}$, the key is to update x_i iteratively as follows:

$$x_i^{(n+1)} = x_i^{(n)} + \Delta x_{ki}^{(n+1)} \,. \tag{3}$$

In equation 3, $\Delta x_{ki}^{(n+1)}$ is the correction from path k in iteration $n+1$. There are many different ways to updating each pixel. One of them is given below:

$$x_i^{(n+1)} = x_i^{(n)} + \frac{\Delta y_k^{(n)}}{\sum_{i=1}^{I} \Delta\alpha_{ki}^2} \Delta\alpha_{ki} \,. \tag{4}$$

The more detailed methodology is given by Peterson et al. (1985).

2.3 Petrophysical Model

Petrophysical relationships that link geophysical properties (such as radar velocity, seismic velocity, and electrical resistivity) and hydrological parameters (such as lithology, hydraulic conductivity, and water content) play a central role in the use of geophysical data for subsurface characterization. Although a few relationships that are commonly used to translate geophysical parameters into hydraulic properties exist, such as the Topp (Topp, 1980), and Archie (Archie, 1942) relationships, development of these relationships for low pressure, low temperature, un- to semi-consolidated materials is still in an early stage of development. The most common approach for hydrogeophysical studies is to develop site-specific empirical relationships between the geophysical measurements and the parameters of interest, using co-located field data, such as coincident geophysical attributes from inverted tomograms at the wellbore location and wellbore measurements.

Many different approaches can be utilized to develop site-specific petrophysical relationships given coincident hydrogeological and geophysical datasets. These approaches include regression-based, fuzzy, cluster, discriminate analysis, and principle components analysis. As a first step, we developed capabilities to use a regression model with a stepwise deletion method to find the most suitable petrophysical relationship between continuous geophysical and hydrological measurements. This allows users to develop site-specific relationships from given data sets.

The petrophysical model in HydroImage is mainly based on one response variable (y) and one exploratory variable (x). This method certainly can be extended to the case that has one response variable and multiple exploratory variables. Suppose we can fit variable y with a polynomial of variable x in the following format:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p . \tag{5}$$

We use the stepwise deletion model selection methods to determine the highest power of variable x that fit the data best. The detailed methods are given below:

Step-1: Start from a given number of power N and obtain a design matrix $\mathbf{x}$:

$$\mathbf{x} = \begin{Bmatrix} 1 & x_1 & \cdots & x_1^N \\ 1 & x_2 & \cdots & x_2^N \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_n & \cdots & x_n^N \end{Bmatrix} . \tag{6}$$

Step-2: Use the sweep method to find coefficients and p-values for each coefficient: Let vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_N)^T$ and let vector $\mathbf{y} = (y_1, y_2, \cdots, y_N)^T$. Thus, the linear regression model is given by $\mathbf{y} = \mathbf{x}\boldsymbol{\beta}$. By sweeping the first N columns of the augmented new matrix given as below, we can obtain the least square estimates of coefficients and the residual sum of square RSS .

$$\begin{pmatrix} \mathbf{x}^T\mathbf{x} & \mathbf{x}^T\mathbf{y} \\ \mathbf{y}^T\mathbf{x} & \mathbf{y}^T\mathbf{y} \end{pmatrix} \xrightarrow{sweep} \begin{pmatrix} V(\hat{\boldsymbol{\beta}})/\hat{\sigma}^2 & \hat{\boldsymbol{\beta}} \\ -\hat{\boldsymbol{\beta}}^{\mathbf{T}} & RSS \end{pmatrix} \tag{7}$$

Step-3: Obtain p-value for each coefficient using the sweep method and compare the p-value corresponding to the item with the highest power to the cutoff value 0.05. If the p-value is larger than 0.05, delete the item and repeat steps 1 and 2; otherwise, stop and use the highest power as the best fitted model.

2.4 Bayesian Integration

The Bayesian model developed by Chen et al. (2001) was used to combine hydrogeological and geophysical data. Compared to geostatistical integration, Bayesian methods are more effective for integrating diverse datasets, especially when the problem under consideration is complex. However, these methods have not been used as widely as geostatistical methods, primarily because there is a gap between the state of the research and the state of the practice. For example, many hydrogeophysical studies still work within deterministic frameworks, even though the utility of quantifying uncertainty is well recognized. In the following, we briefly describe the Bayesian method used in HydroImage, and more details can be found in Chen et al. (2001).

Suppose we want to estimate logarithmic hydraulic conductivity from borehole direct hydrological measurements (e.g., flowmeter test data) and indirect crosshole geophysical data, such as ground-penetrating radar (GPR) velocity and seismic velocity. Let the random variable Y denote the log-conductivity at a pixel $\mathbf{x}$ along a cross-section between two boreholes. Let V_g and V_s denote GPR and seismic velocity, respectively. All data are mean-removed and normalized by their corresponding standard deviations. The log-conductivity estimate at the given location $\mathbf{x}$, in terms of probability density function (pdf), is obtained using the Bayes theorem as follows (Bernardo and Smith, 2002):

$$f(y(\mathbf{x}) \mid v_g(\mathbf{x}), v_s(\mathbf{x})) = CL(y(\mathbf{x}) \mid v_g(\mathbf{x}), v_s(\mathbf{x}))\, f(y(\mathbf{x})) . \tag{8}$$

where $y(\mathbf{x})$ is an unknown value of Y being estimated at the location $\mathbf{x}$, and $v_g(\mathbf{x})$ and $v_s(\mathbf{x})$ are the geophysical data at the same location. Letter C is a normalizing coefficient, and $L(y(\mathbf{x}) \mid v_g(\mathbf{x}), v_s(\mathbf{x}))$ is the likelihood function given the co-located geophysical data. The functions $f(y(\mathbf{x}) \mid v_g(\mathbf{x}), v_s(\mathbf{x}))$ and $f(y(\mathbf{x}))$ are the posterior and prior pdfs of Y at the location $\mathbf{x}$. Note that only co-located geophysical data have been used to update the prior pdf, because they are most informative compared to the measurements at adjacent locations.

The Bayesian method has been used for many years in the water resources field. One of the earliest applications in groundwater hydrology was provided by Kitanidis (1986) for analyzing parameter uncertainty in estimation of spatial functions. In that work, the mean and covariance matrix of the posterior distribution were derived analytically by choosing a prior distribution that is conjugate to the likelihood function in the sense that the posterior has the same form as the prior. In this study, we have developed a new approach, which allows for large flexibility in the form of the likelihood function and posterior pdf, to get numerical rather than analytical posterior mean and variance.

The prior pdf $f(y(\mathbf{x}))$ was estimated based on the hydraulic conductivity data at boreholes using kriging (Journel, 1989). The prior distribution is Gaussian if Y has a multivariate Gaussian distribution (Deutsch and Journel, 1998). The likelihood function $L(y(\mathbf{x})\,|\,v_g(\mathbf{x}), v_s(\mathbf{x}))$ plays a central role in the Bayesian method and was inferred from the hydrological and co-located geophysical data. It can be expressed as follows by chain rules (Bernardo and Smith, 2002) if geophysical data available are ground penetrating radar (GPR) and seismic velocity:

$$L(y(\mathbf{x})\,|\,v_g(\mathbf{x}), v_s(\mathbf{x})) = f(v_g(\mathbf{x})\,|\,y(\mathbf{x}))\, f(v_s(\mathbf{x})\,|\,y(\mathbf{x}), v_g(\mathbf{x})). \tag{9}$$

If uncertainty in $v_g(\mathbf{x})$ and $v_s(\mathbf{x})$ are conditionally independent of each other given the co-located $y(\mathbf{x})$, the inference of the likelihood function becomes simple because each conditional pdf involves only two variables. In this case, we have the following formula:

$$L(y(\mathbf{x})\,|\,v_g(\mathbf{x}), v_s(\mathbf{x})) = f(v_g(\mathbf{x})\,|\,y(\mathbf{x}))\, f(v_s(\mathbf{x})\,|\,y(\mathbf{x})). \tag{10}$$

In general cases, we need to develop multivariate relationship to link multisource of geophysical data with hydrological parameters, we need to update the prior pdf based on all the co-located geophysical data.

3.0 HYDROIMAGE DEVELOPMENT

The HydroImage software was developed based on a modular approach to make future enhancements easy. All relevant data are stored in a central database with utility tools for importing and exporting data. Various modules extract needed data from the database, perform the intended task, and put the results into the database for use by other modules.

3.1 HydroImage Framework

As depicted in Figure 4, the HydroImage framework involves inverting the geophysical data using geophysical inversion methods, so that a geophysical estimate is obtained at each pixel on the geophysical traverse. The geophysical estimate is subsequently used as input to a Bayesian integration routine with the direct hydrogeological measurements at boreholes to estimate hydrogeological parameters along the tomographic transects. HydroImage then outputs estimates of the hydrogeological parameters and the associated uncertainty.

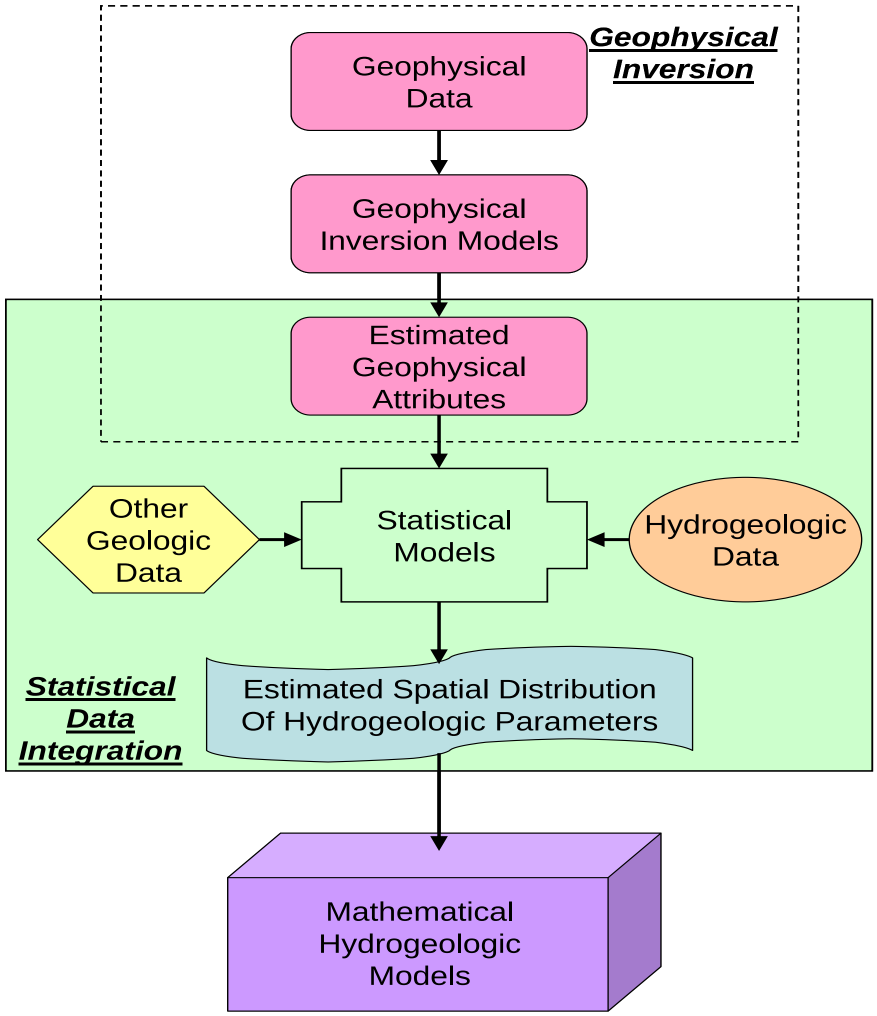

Figure 4: HydroImage framework

3.2 HydroImage Structure

Figure 5 illustrates the current HydroImage structure reflecting the framework. The main menu includes five submenus, each of which performs independent tasks for integration of hydrogeological and geophysical data.

- Under the hydrogeological data submenu, users can perform basic analysis of hydrogeological data, such as inputting, editing, and resampling borehole logs.
- Under the geophysical data submenu, users can perform basic checks on crosshole geophysical data following Peterson (2001), correct or edit on recorded data if needed, and perform tomographic inversion following the already developed LBNL algorithm (Peterson, 1985).

- Under the petrophysical model submenu, users can develop relationships between the hydrogeological and geophysical data. The key submenu is data integration, including both Bayesian and geostatistical integration approaches. For geostatistical integration, users are allowed to perform variogram and co-variogram analysis. For Bayesian integration, users are allowed to build Bayesian models and to select unknown variables and prior distributions. Our experience suggests that Bayesian approaches are more flexible and robust for subsurface hydrogeological characterization.

  The hydrogeological parameter estimates are displayed as static graphics.

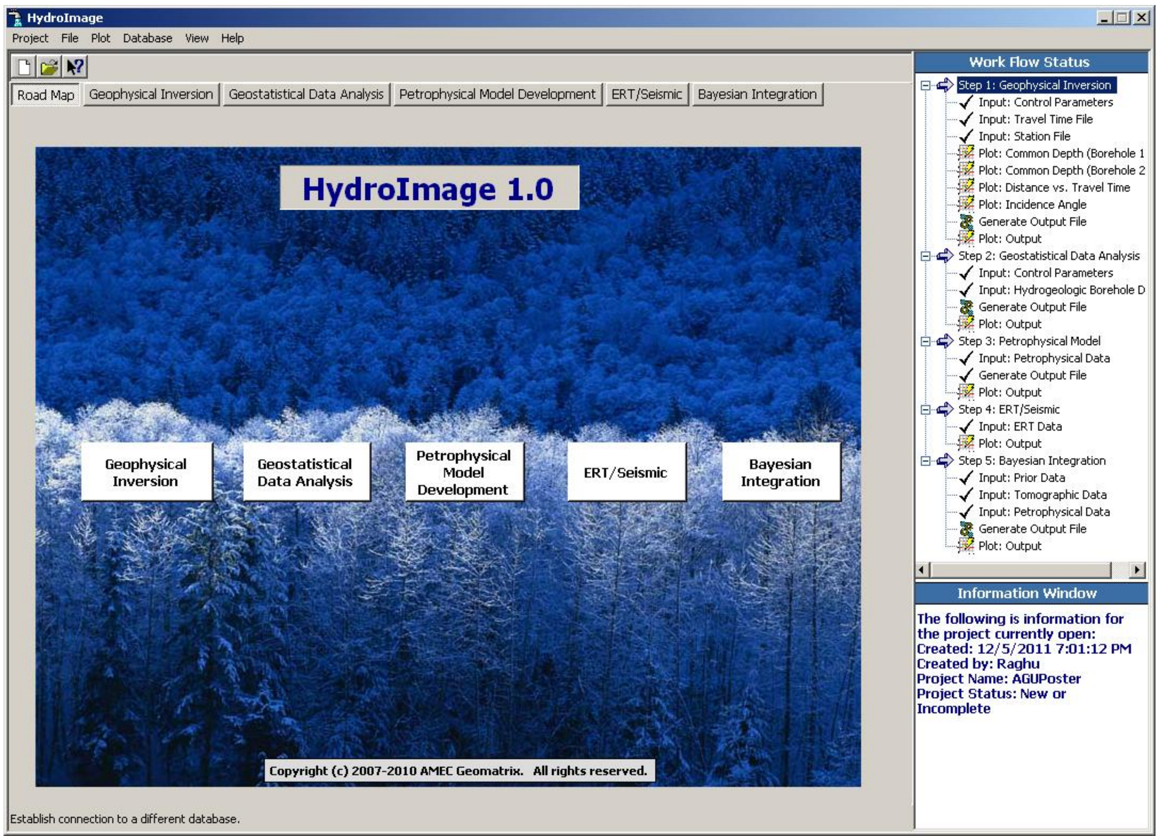

Figure 5: HydroImage structure

3.3 Database Structure

The basic structure of HydroImage was constructed in an expandable and modular way. The code is developed around a centralized database. Instead of passing data through multiple data files in a wide range of formats, all data can be imported into a core database. An advantage of using a core database is data bookkeeping. Data from various sources can be stored coherently and logged systematically. In addition, the data in the database can be accessed outside HydroImage using Geographical Information System (GIS) tools and other CAD software.

Figure 6 shows the conceptual data and control structure of HydroImage.

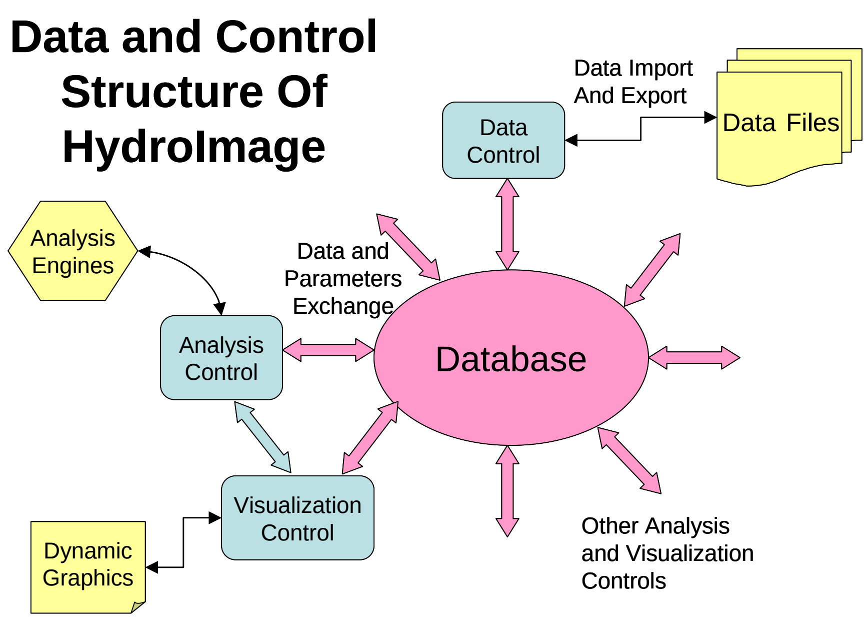

Figure 6: Conceptual structure of HydroImage

The codes were developed to allow users to import data into a core database from data files in various common data format (such as Excel, fixed-width text, Access database, Geo-eas, Surfer grid, and Arcview shapefile). All analyses are performed in a modular way. Each analysis module invokes a routine to extract needed data and parameters from the database and to pass the extracted data to the analysis engine. After the analysis is completed, it transfers the results to the database for storage. There are additional routines to export data from the database to files in various common formats, such as Excel, fixed-width text, Access database, Geo-eas, Surfer grid, and Arcview shapefile. Visualization is performed in a similar way. Data are first extracted from the database, and then are passed to the visualization module to generate graphics.

Figure 7 shows a schematic layout of the database structure. Specific standard tables have been created for each set and type of available data. Specific standard tables have been generated to store the output from various analysis engines. We have developed codes to allow users to import data into the core database from data files in various common data format. We have constructed the relationship between the data fields in tables developed in the database. In addition, standard queries and report templates have been built.

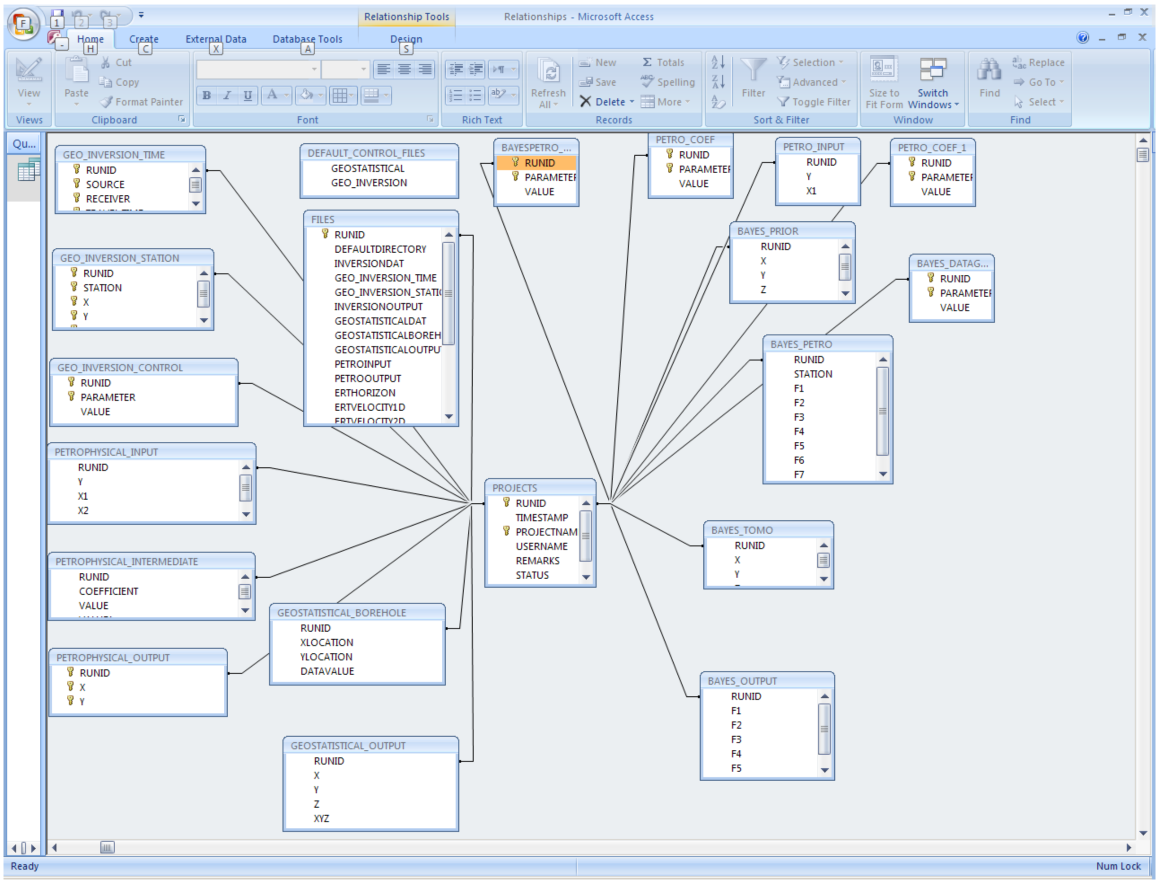

Figure 7: Schematic layout of database structure

3.4 Graphical User Interface (GUI) Implementation

We have developed GUIs that allow users to (1) load borehole hydrogeological measurements and crosshole geophysical data into the system, (2) easily develop the relationships to link geophysical and hydrological parameters, and (3) estimate hydrological parameters. The interface of the main program controls the other modules of HydroImage, including file input and output, data manipulation, and computation. Many HydroImage modules are packaged as dynamic link libraries (DLL) that are available for access as needed. Each of these modules can be updated independently from each other in the future. This flexible structure allows HydroImage to be easily expanded. Figure 8 shows an example of the GUI for crosshole geophysical tomographic inversion.

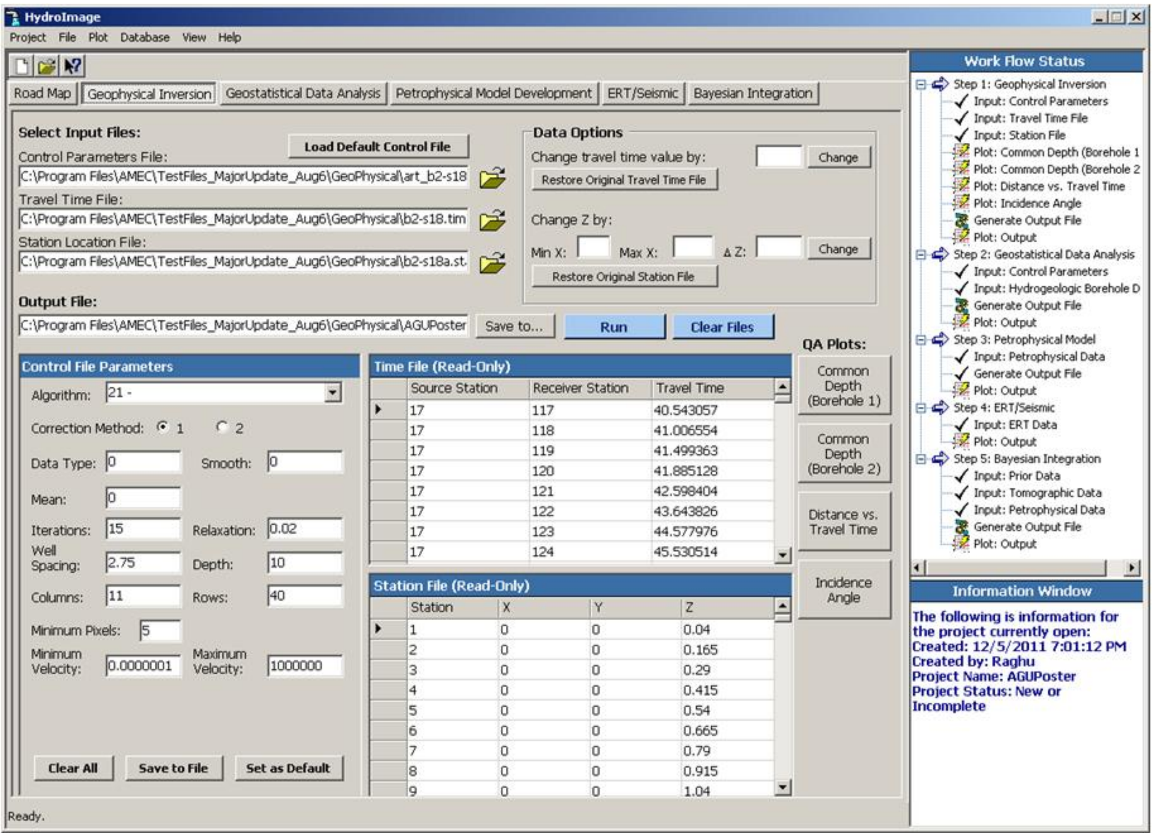

Figure 8: GUI developed for geophysical inversion

The main module of HydroImage includes the codes that accept the information from the database interface, arrange them as arguments/parameters for the DLLs, and send the results to the database interface. HydroImage allows the users to export the results into data files that can be used by other programs for other purposes. The GUIs can be coded using Microsoft Visual Studio .NET, which is a powerful software development environment that supports multiple languages, such as Visual C++, BASIC, and FORTRAN. The GUI accepts the user's selected parameters and triggers various analysis, visualization, and data management control modules. Default parameter values are provided to the user and stored in the database. If the user selects to change these parameters, the updated parameter values will be stored in the database. A warning message will be displayed if the values selected by the users are not within the normal range common adopted. In addition, if the selected values are not within the acceptable range, an error message will be displayed and the user needs to re-enter a new value for the parameter.

3.5 Modular Linkage Protocol

The protocol of data transfer between the GUI is written in Visual Basic .NET and various analysis and visualization modules written in Visual Fortran and Visual C++. Additional codes are needed to link all the modules according to the software framework described. When the user instructs the software to perform an action through the GUI, it will invoke the appropriate control module in response to the action. The control module will extract data from the database and pass the data and user's selected parameters to the various engines (such as the analysis engines, visualization engines, and data management engines).

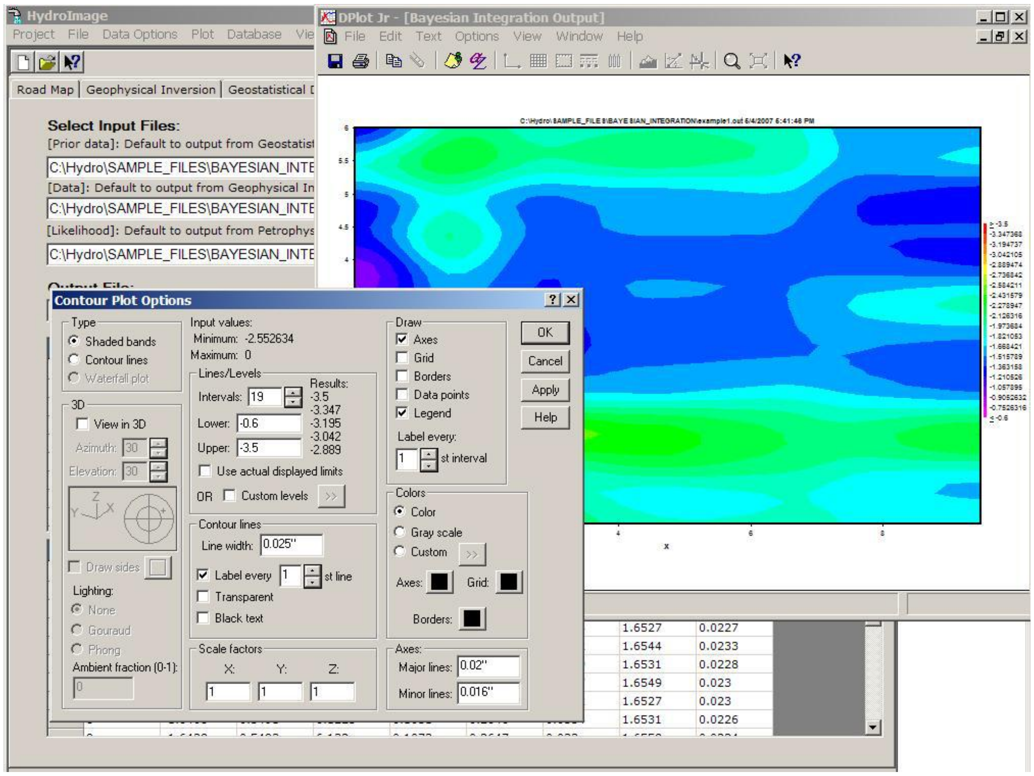

Figure 9: Visualization of results after completion of analysis

Some of the codes developed were originally written in C for Unix/Linux systems. These codes were compiled using the Visual C++ .NET compiler into Dynamic Linked Library (DLL) modules in a format that is commonly referred by software engineers as 'unmanaged DLL.' Unmanaged DLLs are efficient and functional, but they lack compatibility standards. However, in contrast to 'managed DLL,' they did not take advantages of the .NET technology and functionalities. In contrary, 'managed' codes are more standardized, mutually accessible and transparent, but they are less efficient. Linking unmanaged DLLs with other modules developed in languages that only produce managed codes (such as Visual Basic .NET and Visual C#) often requires a wrapper. We modified some of these C codes to generate managed DLLs, if doing such was more advantageous. Figure 9 shows an example of the graphical visualization after completion of data analysis.

3.6 Geophysical Inversion Module

HydroImage includes the Algebraic Reconstruction Technique (ART) inversion method developed at LBNL (Peterson, 1985), based on the assumption of straight raypaths, which is commonly assumed when inverting crosshole radar and seismic data. Figure 10 shows an example of the GUI developed for crosshole geophysical tomographic inversion. We incorporated steps that can be used to assess the quality of the geophysical inversion following Peterson (2001), such as errors associated with incorrect information about borehole locations. The procedures developed allow easy input/output portals to the codes and build initial QC steps into the estimation procedures. The geophysical inversion procedure yields a set of geophysical attribute values that minimizes the difference between the observed measurements and the simulated measurements given that distribution of attributes in a least squares sense. As with all geophysical inversion procedures, inversion parameters are chosen (such as the level of smoothing, the discretization, etc.), and the choice of parameters can lead to different geophysical data inversion results. Quantification of geophysical inversion error and translation of that error into the hydrogeological parameter estimate are addressed.

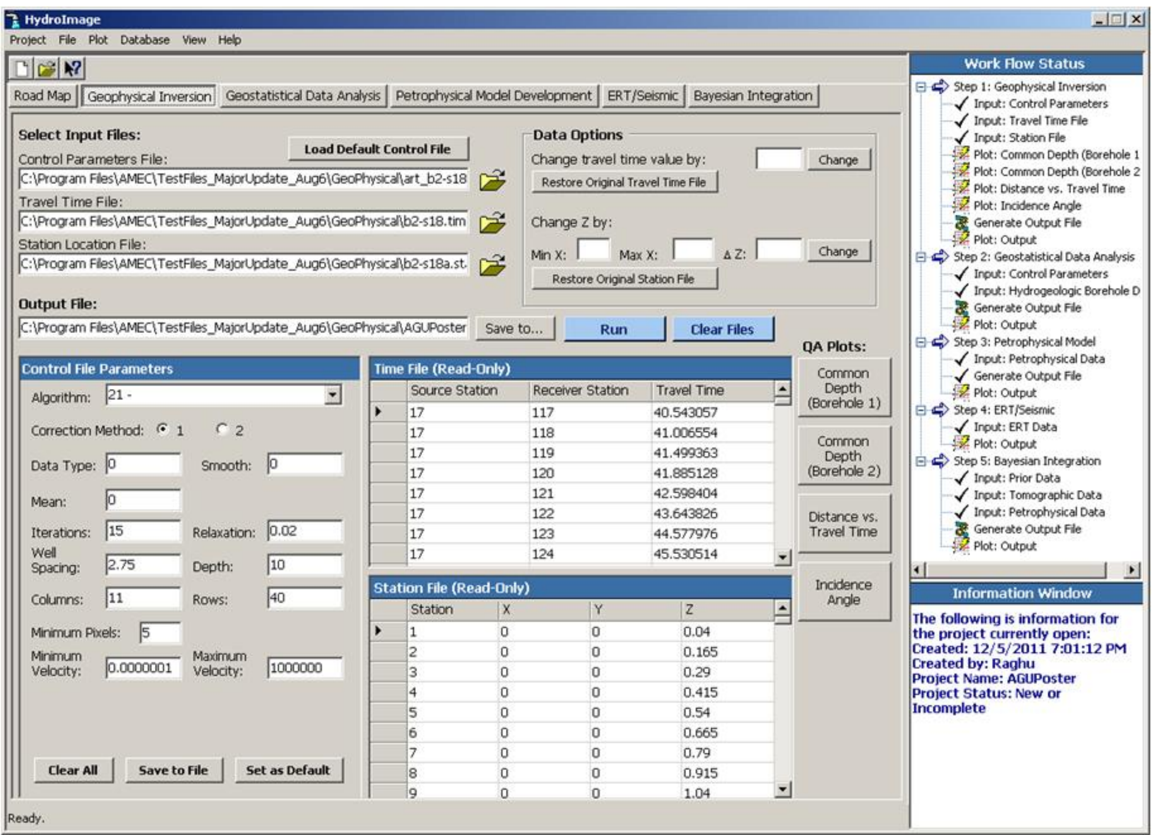

Figure 10: Geophysical inversion module

3.7 Geostatistical Analysis Module

Geostatistical methods are commonly used to interpolate between measured datapoints based on an assumed or estimated spatial correlation. We included kriging routines to allow users to (1) work with a routine that is commonly employed; (2) develop prior functions for use within the Bayesian routine, and (3) compare results obtained using kriging of borehole data with the results obtained using hydrological and geophysical data fused through Bayesian approaches. In variogram analysis routine of HydroImage, users can identify the spatial structure of given data sets. In the kriging subtask, hydrogeological parameters can be estimated at each location. In the cokriging subtask, a simple tool is provided to integrate geophysical tomographic data with hydrogeological measurements. The codes were developed using FORTRAN 90 program language, and are based on the codes developed in GSLIB. Figure 11 shows an example of the GUI for geostatistical data analysis.

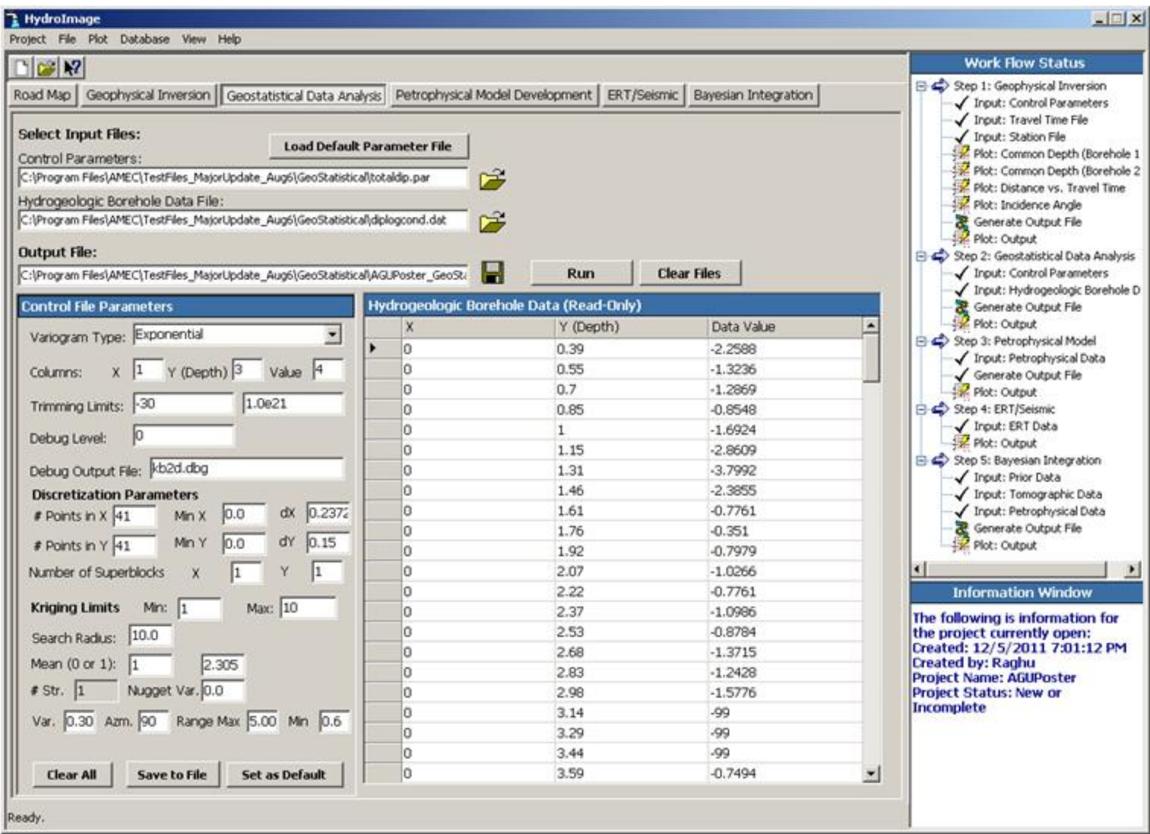

Figure 11: Geostatistical Data Analysis Module

3.8 Petrophysical Modeling Module

As described in Section 2.3, Petrophysical Model, petrophysical relationships that link geophysical properties and hydrological parametersplay a central role in the use of geophysical data for subsurface characterization. Development of relationships to translate geophysical parameters into hydraulic properties for low pressure, low temperature, un- to semi-consolidated materials, which are common to hydrology investigations, is still in an infancy stage. The most common approach for hydrogeophysical studies is to develop site-specific empirical relationships between the geophysical measurements and the parameters of interest, using co-located field data (such as coincident tomographic and borehole measurements) or by developing relationships using site materials at the laboratory scale. As many studies have illustrated that geophysical measurements can be associated with more than one hydrogeological property (e.g., Marion, 1992; Copty et al., 1993; Knoll, 1996; Hubbard et al., 1997; Prasad, 2002), methods for handling non-unique petrophysical relationships will be implemented in HydroImage.

We initially developed capabilities to use a regression model with a stepwise deletion method to find the most suitable petrophysical relationship between continuous geophysical and hydrological measurements. During Phase II, we enhanced the capabilities of the petrophysical toolbox by permitting more flexibility in the analysis and design of the petrophysical model and through implementation of scale-matching routines. A built-in module was developed to allow users to explore different model options, such as to develop relationships between continuous geophysical and hydrological parameters, or to consider unknowns as indicators and using discrete probability distribution functions to characterize petrophysical models. Prior to comparing different types of datasets, it is important to scale data to a similar discretization. We developed tools that allow users to use various re-sampling or up-scaling techniques to transform data, for example, moving average methods, 1-D or 2-D interpolation methods, and filtering methods.

The length scale between hydrogeological measurements, crosshole geophysical data, and borehole geophysical data are often different. To develop a relationship between measured hydrological data and co-located inverted geophysical-attribute values near the boreholes, upscaling or downscaling must be performed. This system includes options for simple scale matching, such as filtering, resampling, or smoothing, before developing petrophysical relationships. We developed visualization techniques to illustrate cross-correlations between geophysical and hydrological data.

We have developed computer codes to input surface electrical resistivity data to develop site-specific petrophysical relationships. Electrical resistivity methods have been widely used for environmental purposes, because these surveys are relatively easy to carry out and can be set up for autonomous monitoring, instrumentation is relatively inexpensive, data processing tools are widely available, and the relationships between resistivity and hydrological properties, such as porosity and moisture content, are reasonably well established. In particular, electrical resistance tomography (ERT) methods have been commonly utilized for subsurface characterization. Figure 12 shows an example of the GUI for estimating petrophysical model parameters.

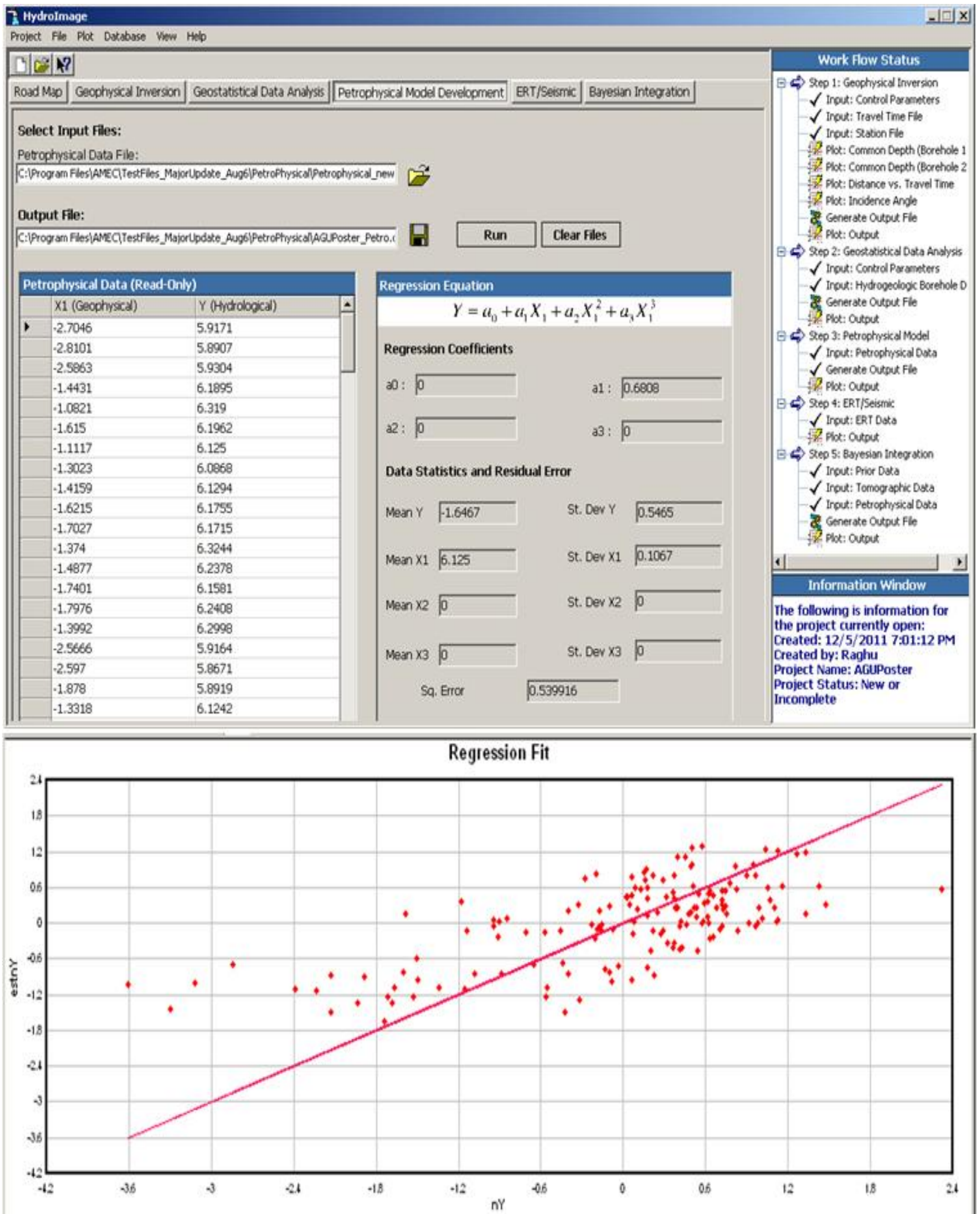

Figure 12: Petrophysical model module

3.9 Bayesian Integration Module

This component is the novel workhorse of HydroImage. We have developed several Bayesian models for combining crosshole tomographic geophysical data and borehole hydrogeological measurements. Compared to geostatistical integration, Bayesian methods are more effective for integrating diverse datasets, especially when the problem under consideration is complex.

The main idea of Bayesian methods is simple and straightforward, as depicted in Figure 13. Based on the prior knowledge or information from the sources other than the data, users can define prior probability distribution. For example, a user can borrow information from a neighbor analogue site. Alternatively, kriging could be used to interpolate borehole data to obtain a prior probability distribution of the hydrogeological parameter of interest at each point in space.

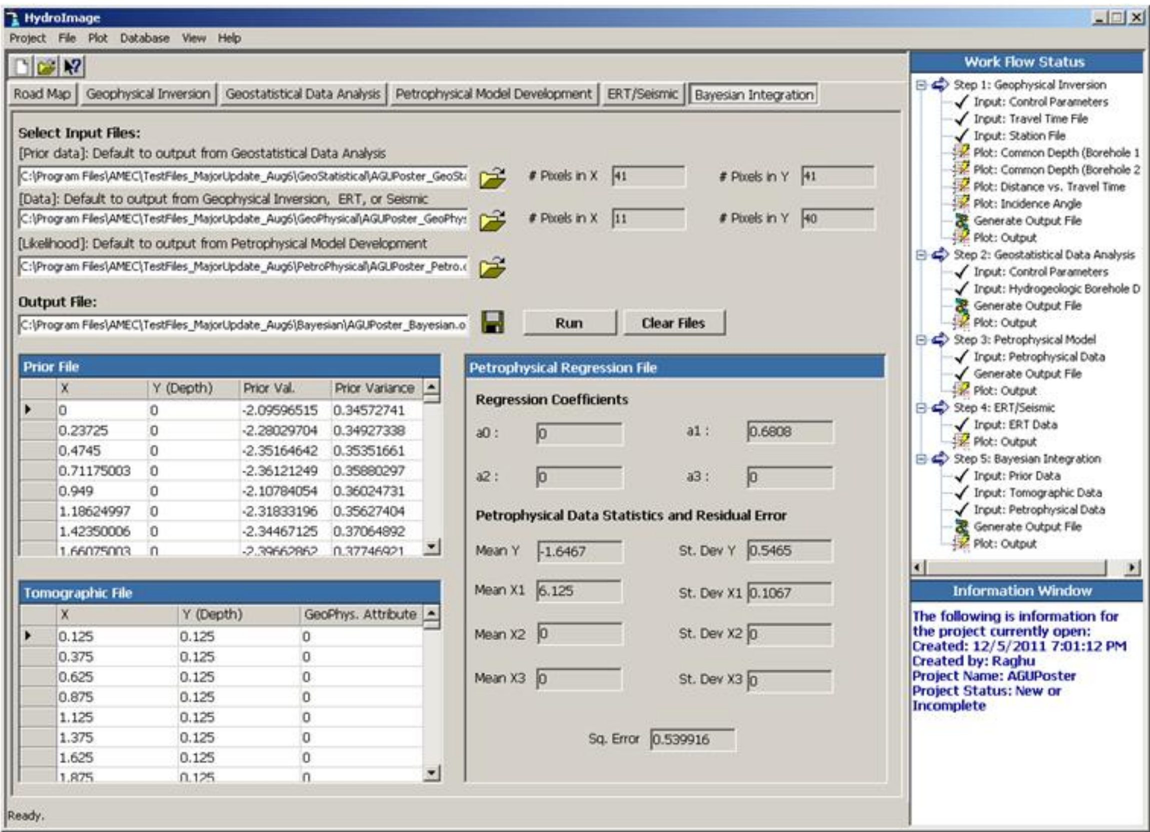

Figure 13: Bayesian integration module

With the use of data at hand, the user can update the prior probability distribution through likelihood functions to get posterior probability distribution. The user can incorporate the 2-D estimates of geophysical attributes, together with the relationship between the geophysical attributes and the hydrogeological parameters of interest, into the Bayesian routine.

Bayesian methods estimate probability distribution for each unknown, rather than a single value for each unknown parameter. At each point in space, the user can obtain an entire probability distribution of the unknown parameter of interest, such as hydraulic conductivity. To get a single value for each unknown variable, the mean, mode, or median from its probability distribution can be calculated. Therefore, Bayesian methods yield not only the estimate of unknowns, but also the uncertainty associated with the estimation.

3.10 Quality Control Tools

HydroImage includes quality control steps that allow users to assess the utility of a certain geophysical dataset for a particular application and the quality of the geophysical data inversion approach. These quality control steps in HydroImage avoid use of the software in a 'black box' fashion, while informing the user about the validity of a particular dataset to estimate a particular hydrogeological parameter.

Similar to hydrological data, geophysical data involve a large degree of uncertainty. The uncertainties may come from geophysical data acquisition and pre-processing, inversion, and petrophysical models. HydroImage includes some tools for users to detect possible errors in geophysical data collection and to help users to understand or interpolate uncertainty associated with geophysical data inversion. Specifically, we incorporated methods for quality control of cross-hole geophysical data. The need and techniques of controlling the errors and uncertainty in cross-hole geophysical surveys have been documented in Peterson (2001). We automated other quality control steps associated with correction of time-zero drift, correction of wrong station coordinates, and correction of high angle ray-path effects**.**

To produce 'reliable' tomographic geophysical estimates, which are in turn used to estimate hydrogeological properties within HydroImage, it is first important to assess the quality of the 'raw' geophysical measurements. We have developed several algorithms that permit the assessment of seismic and radar tomographic data quality following Peterson (2001). Given the input tomographic travel time picks and station file, HydroImage can be prompted to display plots, such as propagation angle versus velocity, travel time versus travel distance, and station location versus travel time. An example of these plots is shown on Figure 14. Analysis of these plots can indicate systematic 'errors' associated with incorrect station geometry or wellbore location/deviation, zero time definition, and high angle effects. The module includes options for 'cleaning' the datasets, such as adding a bulk travel time to all records or deleting specific records.

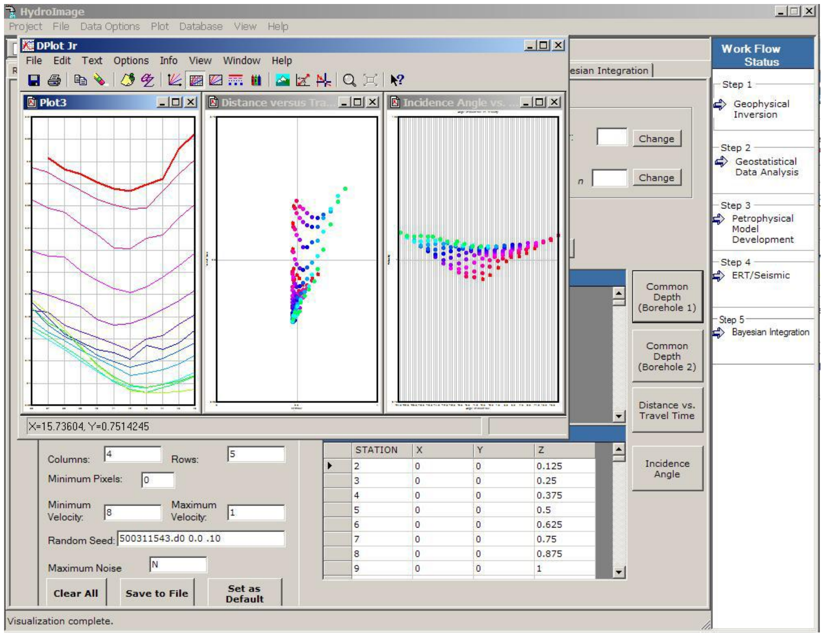

Figure 14: Graphical display for geophysical quality control

4.0 SOFTWARE TESTING

4.1 Verification

HydroImage verification was performed by: (1) doing a fresh installation on Microsoft Windows machines (2) testing with field-derived and previously used datasets for the Geophysical Inversion, Geostatistical Data Analysis, Petrophysical Model Development, and Bayesian Integration modules, and (3) validating the results of HydroImage output against those generated using the existing codes on which the development of HydroImage is based. The software was installed on a machine that was not used for the HydroImage code development and HydroImage had not been installed. The testing was performed by team members at the Lawrence Berkeley National Laboratory. HydroImage installation files include a Windows Installer named ‘HydroImageSetup.msi’ and an executable file named ‘setup.exe’. These files were successfully installed on a Windows 7 of 64-bit machines. All the required dynamically-linked-library files (DLLs) and the main HydroImage database were installed successfully to appropriate folders. Following installation, permission for the program to access the default database file named ‘HydroImage.mdb’ was set to be unrestricted. A new project was created in the database and all the modules were tested sequentially. Results of individual module testing are described below.

4.1.1 Geophysical Inversion Module

The Geophysical Inversion module was tested by reading an input parameter file that is prepared before inversion, containing iteration and convergence criterion, a travel time data file, and a station file showing seismic source and receiver positions. The travel time and station location files were parsed correctly into the database, and loaded and displayed correctly on their respective data tables. The parameters from the control file were loaded correctly on their respective textboxes and the inversion algorithm combo-box. After the data were loaded, the four Quality Assurance (QA) graphs generated by HydroImage were verified by clicking on each of the buttons sequentially. After a new output file name was specified, the inversion computation was performed by clicking the Run button. The resulting output file, which contains GPR Slowness values, is consistent with the results from the existing FORTRAN code, ART, using the same set of input files. A maximum difference of 2.2% was occasionally noted between those results. This is acceptable considering uncertainty in the least-squared based inversion. The graphs showing the final output result was generated successfully.

4.1.2 Geostatistical Data Analysis Module

The Geostatistical Data Analysis module was tested by importing a GSLIB input control parameter file, which is same as the one used in GSLIB, and a borehole data file consisting of log-conductivity values. The borehole data file was loaded into the database, and the log-conductivity values were displayed correctly in the borehole data table on the form. The parameters required by the KB2D kriging routines were also loaded correctly into their respective textboxes, and the combo-box for variogram type correctly converted the type specified in the control file to text. A new output file name was specified and the kriging routine was tested by clicking the Run button. Again, we compared the kriged values and their corresponding variances in the specified output file with the results obtained from using the original KB2D routine in the GSLIB package. The maximum absolute difference between the two outputs was 9.4E-07 for log-conductivity and 5.5E-07 for variance, and this was considered acceptable. The final kriged log-conductivity output was plotted successfully by HydroImage. The original KB2D control file specified an Exponential variogram; other variogram types were tested successfully by selecting each from the combo-box for variogram type.

4.1.3 ERT/Seismic Module

The ERT module was tested by importing an ERT data file that was obtained by using other software of inversion methods from direct electrical measurements. The file was parsed into the database and was displayed correctly in the data table. The ERT data were plotted successfully by HydroImage.

4.1.4 Petrophysical Model Development Module

The Petrophysical Model Development module was tested by importing a new input data file consisting of co-located geophysical and hydrological data from the South Oyster site in Virginia (Chen et al., 2001). A new output file was specified and the Run button was clicked to generate the output file. This output file was parsed into the database and the petrophysical regression coefficients were displayed correctly on the form along with the regression statistics. The regression coefficients were same as those obtained from an existing C++ code. The regression fit was plotted successfully by HydroImage.

4.1.5 Bayesian Integration Module

The Bayesian Integration module first verifies data files for priors and Likelihood from the outputs generated from Geostatistical Data Analysis, Geophysical Inversion and Petrophysical Model Development modules correctly. All the data files were retrieved from the database correctly and displayed on their respective data tables and textboxes. A new output filename was specified and the Run-button was clicked to produce the output successfully and subsequently results are plotted by using HydroImage.

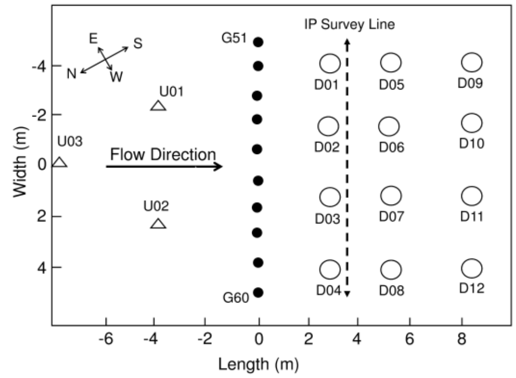

Figure 15: Borehole Locations and the Geophysical Survey Profile (Chen et al., 2013).

4.2 Field Data Testing

4.2.1 Rifle Integrated Field Research Challenge (IFRC) Site

The Rifle Integrated Field Research Challenge (IFRC) site is located in northwestern Colorado at the site of a former uranium and vanadium mill, which operated between 1924 and 1958. Although mill tailings and other contaminated surface materials were removed from the site by 1996, residual contamination remains, as evidenced by locally persistent levels of dissolved U(VI) in groundwater (Williams et al., 2011). While natural flushing led to modest decreases in dissolved U(VI) over time, numerous laboratory-based studies (e.g., Anderson et al. 2003; Holmes et al. 2005; Vrionis et al. 2005) have demonstrated that stimulated bioremediation, coupled to the enrichment of members of the *Geobacteraceae*, is able to rapidly remove U(VI) from groundwater following injection of acetate.

To further investigate the potential for bioremediation of uranium in groundwater at the site, several field-scale bioremediation experiments have been conducted at the DOE Rifle IFRC site near Rifle, Colorado (USA) from 2002 to 2009. The shallow subsurface consists of an unconfined alluvial aquifer that includes sandy-gravelly unconsolidated sediments with variable silt and clay content. Underlying the aquifer is a relatively impermeable aquitard (i.e., silt and mudstones of the Eocene Wasatch Formation) located at spatially variable depths of 5.9-7.0 m below ground surface [*Williams et al.*, 2011; *Chen et al.*, 2012]. During the field experiments, acetate as an electron donor and organic carbon source was injected into the groundwater through a series of injection wells, along with the conservative tracer bromide.

The field test data used for the HydroImage testing focused on the biostimulation field experiments that were conducted between August of 2007 and December of 2009; the timeline of different amendment injections and details on the acquisition of geophysical measurements are given in Flores-Orozco et al. [2011]. Specifically, we focused on the geochemical and geophysical data collected from 19 July 2008 to 8 December 2009. Figure 15 illustrates the well field used to conduct the bioremediation experiments, where the ten solid circles (i.e., G51-G60) are acetate injection boreholes and the twelve open circles (i.e., D01-D12) are down-gradient monitoring wells. Acetate was injected into the unconfined aquifer over the saturated interval of 3.5-6.0 m below ground surface. The three open triangles (i.e., U01-U03) are up-gradient monitoring wells. Time-lapse surface spectral induced polarization data were collected along the dashed line, which is located 2.7 m down gradient from the injection wells. Geochemical sampling and geophysical data collection (both described below in detail) occurred before, during, and after the period of acetate injection.

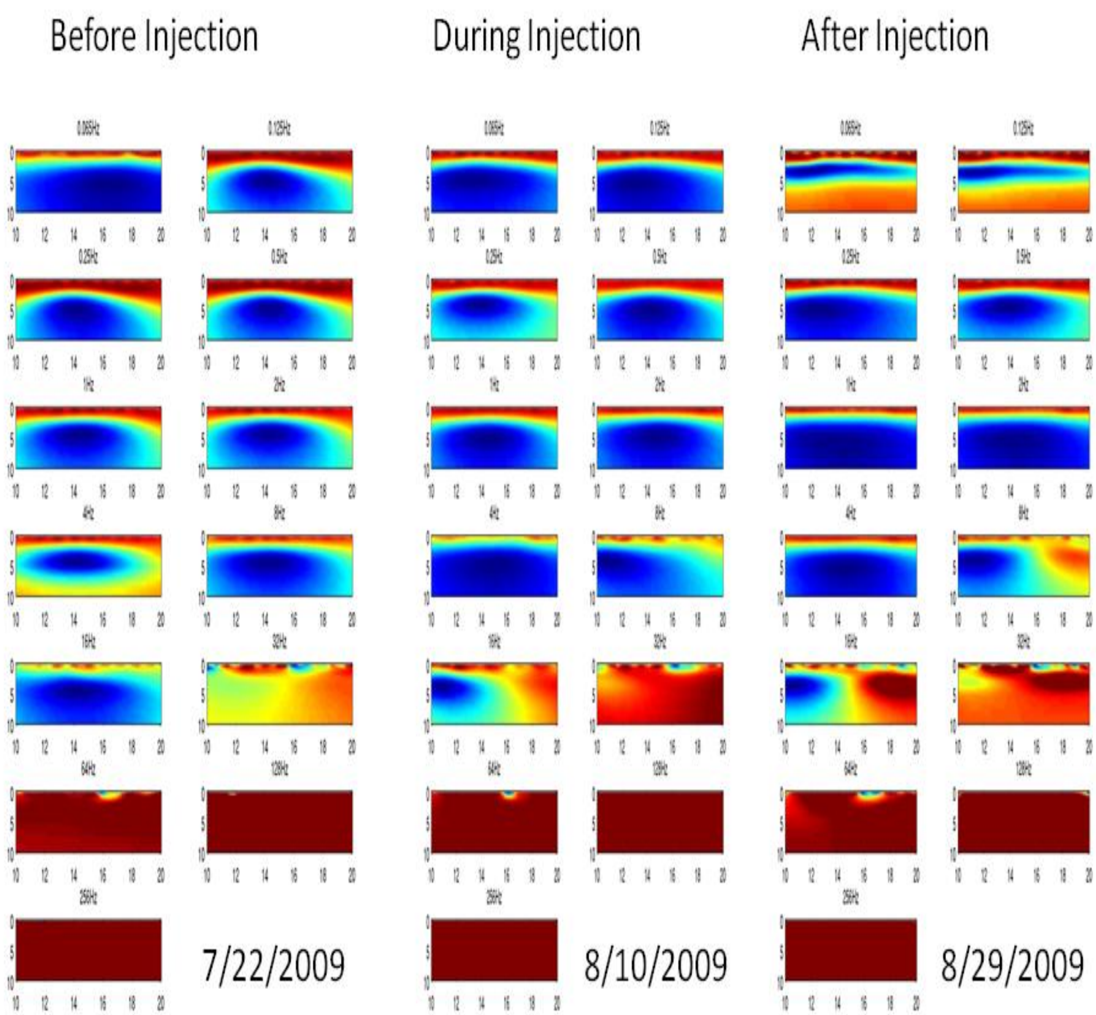

Figure 16: Spectral induced polarization data at thirteen different frequencies along the survey profile.

4.2.2 Borehole Aqueous Geochemical and Surface Spectral Induced Polarization Data

Surface spectral induced polarization data were collected along the profile showing as a dashed line in Figure 15 using thirteen different frequencies, ranging from 0.065 Hz to 256 Hz on three days, July 22, 2009 (before injection), August 10, 2009 (during injection), and August 29, 2009 (after injection). Figure 16 shows the phases in milliradians (mrad) as a function of frequencies before, during, and after the acetate injection along the cross-section from depth z=0 m to z=10 m and horizontal distance from x=10 m to x=20 m, with grid sizes of dx=dz=0.5 m. The domain that we focus on in this study nearly traverses the sampling boreholes D1, D2, D3, and D4 (see Figure 15).

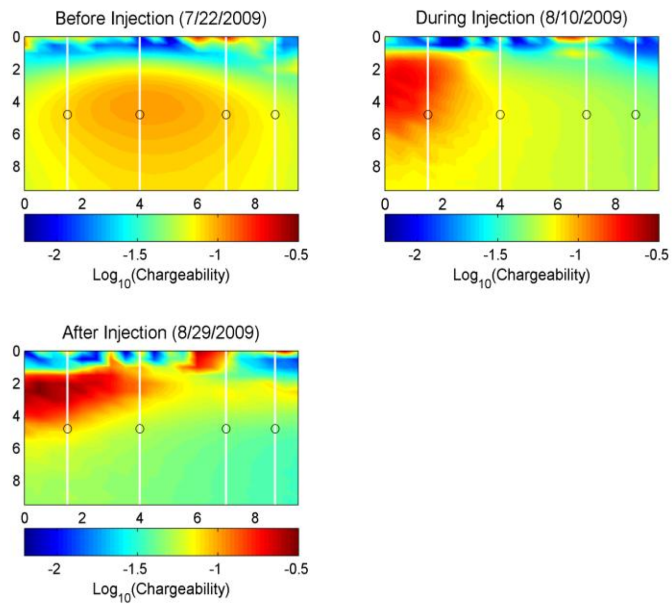

Figure 17: Inverted Cole-Cole parameters using the stochastic inversion algorithm developed by Chen et al. (2008). These are the data sets used for current demonstration.

The amplitude and phase values were inverted using the stochastic method developed by Chen et al. (2008) to get Cole-Cole parameters, such as zero-frequency resistivity, chargeability, time constant, and dependence factor. Figure 17 shows the estimated zero-frequency resistivity, chargeability, normalized chargeability, and time constant along the survey profile before, during, and after acetate injection. The white vertical line segments show the locations of boreholes D1-D4 and the red circles show the groundwater sampling locations within the boreholes. We consider the inverted chargeability along the 2D profile as data for this demonstration.

Groundwater samples were collected for chemical analysis, including Fe(II), sulfate, sulfide, acetate, uranium, chloride, and bromide concentrations, at ~5 m depth within the four boreholes, starting from July 10, 2009 and ending on December 8, 2009. Figure 18 shows the time-series of logarithmic concentration of various chemical concentrations collected from boreholes D1, D2, D3, and D4, respectively, where Day 0 corresponds to July 10, 2009 and the initiation of the biostimulation is on Day 12.

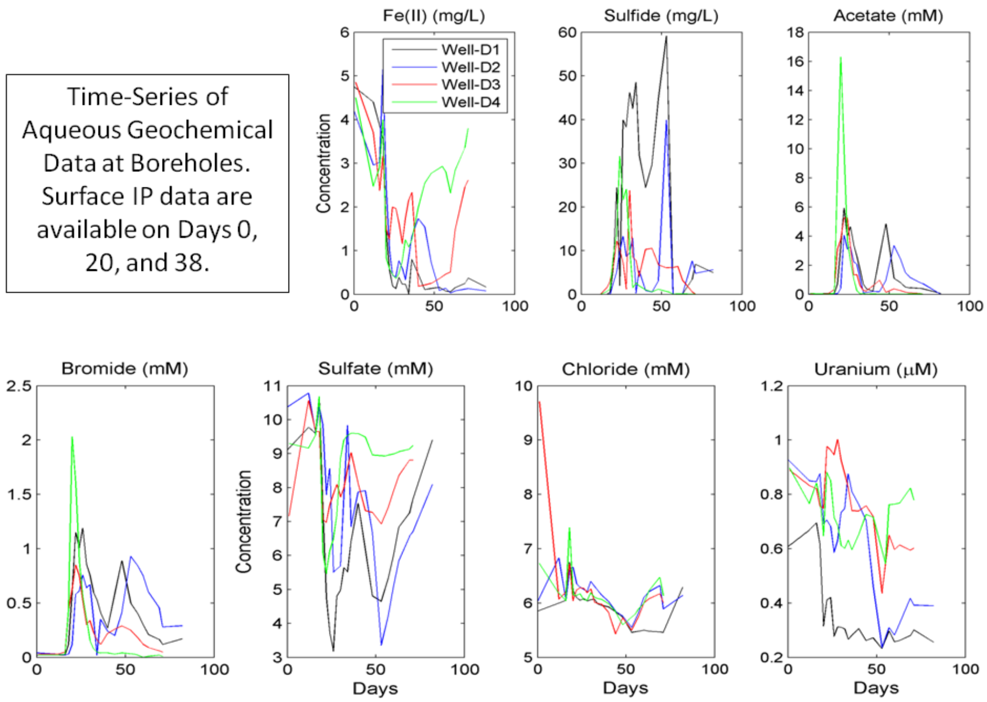

Figure 18. Borehole chemical concentrations at four wells as functions of elapsed time.

4.2.3 Petrophysical Relationship Derived from Co-located Geophysical and Borehole Geochemistry Measurements Using HydroImage

The petrophysical analysis submenu in HydroImage was used to develop the linkage between surface geophysical and borehole geochemical data. Three steps were followed: (1) Extract chargeability data from the pixels near the four boreholes; (2) Extract geochemical data at the time where geophysical surveys were carried out; (3) Load both extracted geophysical and geochemical data into HydroImage; (4) Fit a polynomial function by using the step-wise deletion method. Figure 19 shows the relationship derived between Fe(II) concentrations and chargeability from surface induced polarization data using HydroImage. Figure 20 shows the submenu for the petrophysical analysis. Because we only had twelve data points, the derived relationship is subject to a large degree of uncertainty. This relationship was used to integrate 2D geophysical data with borehole geochemical concentrations to obtain spatial distribution of geochemical parameters along the 2D profile.

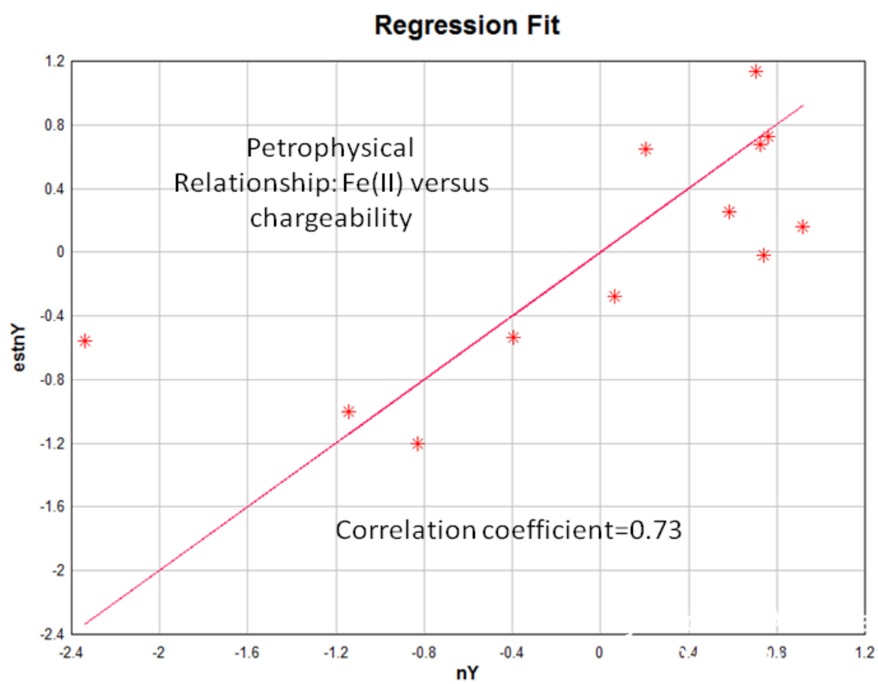

Figure 19: Cross-correlation derived from co-located geochemical and geophysical data using HydroImage.

4.2.4 Geostatistical Analysis Using HydroImage

HydroImage provides a set of tools for carrying out geostatistical analysis, such as kriging, by using given variogram models and observation data at some locations. Here, the kriging tool is used to spatially interpolate borehole geochemical measurements at the four boreholes to other locations along the 2D profile. This provides a baseline for Bayesian integration. Figure 21 shows the result and plots from the geostatistical submenu. Because there are not large data sets to develop a reliable variogram for the estimation, the exponential variogram is picked with correlation lengths of 2 m for the horizontal direction and 0.5 m for the vertical direction based on our experience from other sites.

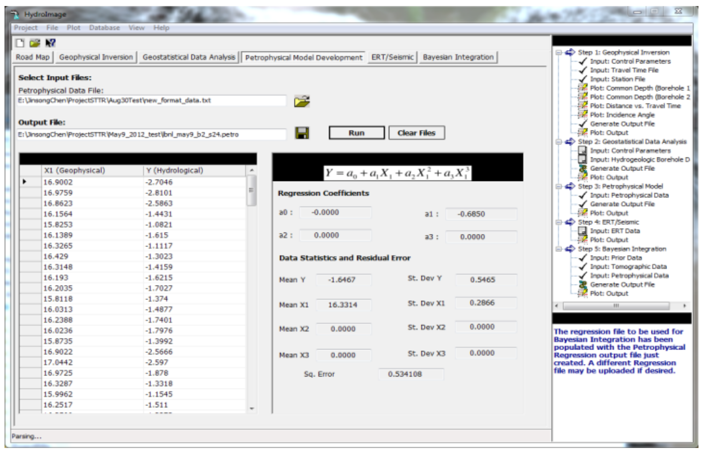

Figure 20: Petrophysical Analysis Submenu for deriving the linkage between geophysical and geochemical data.

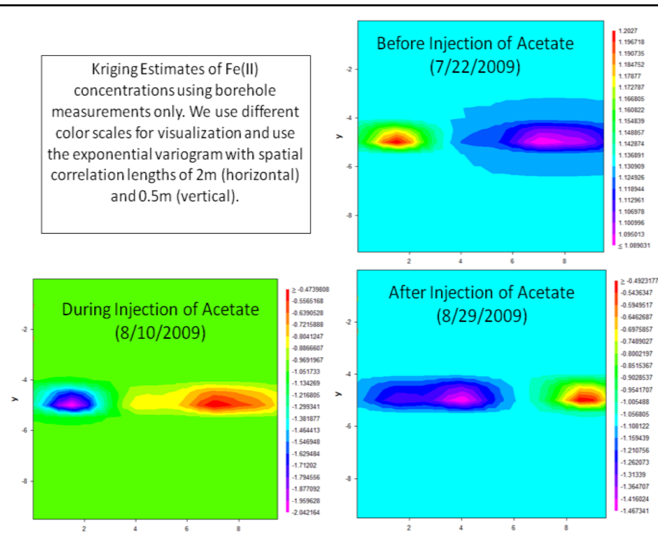

Figure 21: Estimated Fe(II) Concentrations using the Geostatistical Submenu.

4.2.5 Bayesian Integration of Surface Geophysical and Borehole Geochemical Data Using HydroImage

Although borehole geochemical data are dense in time (see Figure 18), they are sparse in space. As shown in Figure 21, they only provide information near the four wells. To have information away from the wells, we need to use the surface induced polarization data along the 2D profile. We used HydroImage to combine 2D geophysical data with 1D time-series of geochemical data by following four steps below: (1) generate prior images using the geostatistical submenu; (2) develop petrophysical relationship using the petrophysical analysis submenu (see Figure 20); (3) obtain 2D geophysical data sets either from a given file prepared by other methods or directly invert crosswell seismic data provided by the geophysical inversion toolbox; and (4) combine geophysical and geochemical data using Bayesian integration toolbox. Figure 22 shows the integration results.

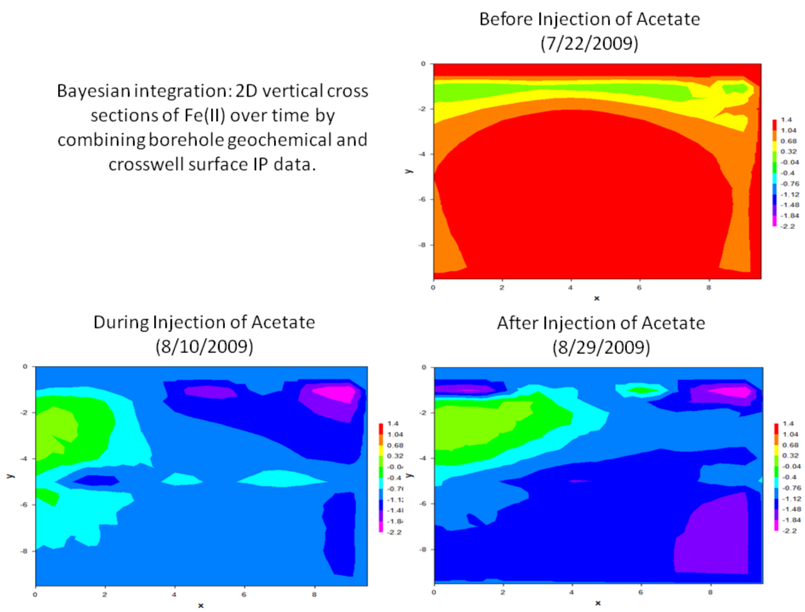

Figure 22: Estimated Fe(II) concentrations before, during, and after acetate injection.

Compared to the three priors shown in Figure 21, we can see that the posterior estimates have much more detail, especially at locations away from boreholes. It is clear that with the use of HydroImage, we are able to combine geophysical data with borehole geochemical data effectively and the results are significantly improved.

5.0 References

Anderson, R. T., H. A. Vrionis, I. Ortiz-Bernad, C. T. Resch, P. E. Long, R. Dayvault, K. Karp, S. Marutzky, D. R. Metzler, A. Peacock, D. C. White, M. Lowe, and D. R., Lovley (2003), Stimulating the in-situ activity of *Geobacter* species to remove uranium from the groundwater of a uranium-contaminated aquifer, *Applied and Environmental Microbiology*, 69 (10), P. 5884-5891.

Bernardo, J. M., and F. M. Smith (2002), Bayesian Theory, John Wiley and Sons Ltd, New York.

Chen, J., S. S. Hubbard, and K.Williams (2013), Data-driven approach to identify field-scale biogeochemical transitions using geochemical and geophysical data and hidden Markov models: Development and application at a uranium contaminated aquifer, Water Resources Research, Vol. 49, 1-13

Chen, J., S. Hubbard, K. Williams, A. Flores-Orozco, and A. Kemna (2012), Estimating the spatiotemporal distribution of geochemical parameters associated with biostimulation using spectral induced polarization data and hierarchical Bayesian models, Water Resources Research, Vol. 48, W05555

Chen, J., A. Kemna, and S. S. Hubbard (2008), A comparison between Gauss-Newton and Markov Chain Monte Carlo based methods for inverting spectral induced polarization data for Cole-Cole parameters, *Geophysics*, 73(6).

Chen, J., S. Hubbard, and Y. Rubin (2001), Estimating the hydraulic conductivity at the South Oyster Site from geophysical tomographic data using Bayesian techniques based on the normal linear regression model, Water Resources Research, 37(6)

Copty, N., Rubin, Y., and Mavko, G. (1993), Geophysical-hydrological identification of field permeabilities through Bayesian updating, Water Resources Research, 29(8), 2813--2825.

Deutsch, C. and Journel, A. (1998), GSLIB: Geostatistical Software Library and User's Guide, Oxford University Press, New York.

Lovley, D. R., F. H. Chapelle, and J. C. Woodward (1994), Use of dissolved H2 concentrations to determine distribution of microbially catalyzed redox reactions in anoxic groundwater, *Environ. Sci. Technol.*, 28, P. 1205– 1210.

Peterson, J. E. (2001), Pre-inversion processing and analysis of tomographic radar data, Journal of Environmental and Engineering Geophysics, 6(1), 1-18

Peterson, J. E., Paulsson, B. N., and T. V. McEvilly (1985), Applications of algebraic reconstruction techniques to crosshole seismic data, Geophysics, 50, 1566--1580.

Vrionis, H. A., R. T. Anderson, I. Ortiz-Bernad, K. R. O'Neill, C. T. Resch, A. D. Peacock, R. Dayvault, D. C. White, P. E. Long, and D. R. Lovley (2005), Microbiological and Geochemical Heterogeneity in an In Situ Uranium Bioremediation Field Site, *Applied and Environmental Microbiology*, 71(10), P. 6308-6318.

Williams, K. H., P. E. Long, J. A. Davis, C. I. Steefel, M. J. Wilkins, A. L. N'Guessan, L. Yang, D. Newcomer, F. A. Spane, L. J. Kerkhof, L. McGuinness, R. Dayvault, and D. R. Lovely (2011), Acetate availability and its influence on sustainable bioremediation of uranium-contaminated groundwater, *Geomicrobiology Journal,* in press.