DOE Contract Number: EMSP 87023 Acct. 37-6000.511

Title: Final Report for “A New Framework For Adaptive Sampling And Analysis During Long-Term Monitoring And Remedial Action Management”

Time Period: October 2002 to September 2007

Published Date: January 23, 2008

Principal Investigator (PI) and Contact: Barbara Minsker, Minsker@uiuc.edu (3230d NCEL, MC-250, 205 N. Mathews Ave., Urbana, IL 61801, 217-265-5293)

Co-PI: Albert Valocchi, valocchi@uiuc.edu

Co-PI: Barbara Bailey, babailey@sciences.sdsu.edu

Abstract: DOE and other Federal agencies are making a significant investment in the development of field analytical techniques, nonintrusive technologies, and sensor technologies that will have a profound impact on the way environmental monitoring is conducted. Monitoring and performance evaluation networks will likely be based on suites of in situ sensors, with physical sampling playing a much more limited role. Designing and using these types of networks effectively will require development of a new paradigm for sampling and analysis of remedial actions, which is the overall goal of this project. Specifically, the objectives of this project were to create an adaptive framework that would: (1) enable effective interpretation of non-intrusive monitoring data, (2) improve predictions and assessment of remediation performance, (3) develop decision rules for on-site adaptive sampling and analysis, and (4) enable more informed decision making and risk analysis of long-term monitoring systems. These objectives were accomplished through development of a new framework for adaptive sampling and analysis, decision making, and risk assessment. The framework assumes that existing data have been used to develop some type of conceptual model for the site, be it an analytical, numerical, or statistical model. This project tested the effectiveness of new models that integrate multiple sources of process knowledge at the site, including geological data collected during site characterization, measurements of parameters collected during routine groundwater monitoring (including surrogate data and data at different spatial and temporal resolutions), and scientific understanding of biochemical transformation processes. An interactive model and parameter identification system was developed for creating the new models efficiently. The models were used to identify future sampling plans, using novel interactive and multi-objective optimization approaches. The approaches were tested using data from the 317/319 area at Argonne National Laboratory, the Waste Isolation Pilot Plant (WIPP) site, and a British Petroleum (BP) site.

Key Words: long-term monitoring, modeling, adaptive sampling, ground water, Bayesian statistics, genetic algorithms, decision support, optimization, uncertainty, phytoremediation, performance assessment

Methodology and Findings: Four PhD students were partially or fully supported by this project. Full details on the approaches and findings can be found in the following major references, which are summarized below each reference. A list of other references can be found at the end of this report.

Babbar, M. (2006). [Interactive Genetic Algorithms for Adaptive Decision Making in Groundwater Monitoring Design](#), PhD thesis, University of Illinois at Urbana-Champaign. In most real-world groundwater monitoring optimization applications, a number of important subjective issues exist. Most of these issues are difficult to rigorously represent in numerical

optimization procedures. Popular norms, such as objectives and constraints, implemented within current optimization methods make many simplifying assumptions about the true complexity of the problem. Hence, such norms fall short of characterizing all the relevant information related to the problem, which the expert (engineers, stakeholders, regulators, etc) might be aware of. Overcoming these limitations by merely performing a post-optimization analysis of solutions by the expert does not ensure that the final set of optimal designs will represent all qualitative issues important to the problem. Hence, there is a need for optimization and decision-aiding approaches that include subjective criteria within the search process for promising solutions. This research tries to fill this need by proposing and analyzing optimization methodologies, which include subjective criteria of a decision maker (DM) within the search process through continual online interaction with the DM. The design of the interactive systems are based on the Genetic Algorithm optimization technique, and the effect of various human factors, such as human fatigue, nonstationarity in preferences, and the cognitive learning process of the human decision maker, have also been addressed while constructing the proposed systems. The result of this research is a highly adaptive and enhanced interactive framework – Interactive Genetic Algorithm with Mixed Initiative Interaction (IGAMII) – that learns from the decision maker's feedback and explores multiple robust designs that meet her/his criteria. For example, application of IGAMII on BP's groundwater long-term monitoring case study in Michigan assisted the expert DM in finding 39 above-average designs from the expert's perspective. In comparison, Case Based Micro Interactive Genetic Algorithm (CBMIGA) and Standard Interactive Genetic Algorithm (SIGA) found only 18 and 6 above-average designs, respectively. Moreover, IGAMII used only 75% of the human effort required for CBMIGA and SIGA. IGAMII was also able to monitor the learning process of different human DMs (novices and experts) during the interaction process and create simulated DMs that mimicked the individual human DM's preferences. The human DM and simulated DM were then used together within the collaborative search process, which rigorously explored the decision space for solutions that satisfy the human DM's subjective criteria.

Singh, A. (2007). *An Interactive Multi-Objective Framework for Groundwater Inverse Modeling*, PhD thesis, University of Illinois at Urbana-Champaign.¹ This work presents a novel interactive multi-objective framework to solve the groundwater inverse problem - focusing on finding the optimal conductivity fields given measurements of aquifer response (such as hydraulic heads). The framework is based on an interactive multi-objective genetic algorithm (IMOGA) that considers model calibration as a multi-objective optimization process with user preference (expressing the plausibility of parameter fields) as an additional objective along with quantitative calibration measures such as prediction error and regularization. Given this information, the IMOGA converges to a set of Pareto optimal solutions representing the best trade-off among all (qualitative as well as quantitative) objectives. Results for the IMOGA show incorporating the site expert's subjective knowledge about the hydrogeology of the modeled aquifer leads to more plausible and reliable calibration results. Since the IMOGA is a population-based iterative search it requires the user to evaluate hundreds of solutions leading to the problem of 'user fatigue'. A two-step methodology is proposed to combat user fatigue. First the user is shown only a fraction of the total population in every generation by clustering potential solutions based on spatial similarity and only showing unique samples from distinct clusters. Next the unranked solutions are ranked using a surrogate model that 'learns' from the user preferences. Research from image processing is used to build clustering and learning algorithms based on the 2-D images of

hydraulic conductivity to closely mimic the human's visual evaluation of the parameter field. Such an approach is shown to reduce user fatigue by up to 50% without compromising the solution quality of the IMOGA. An important part of groundwater inversion is assessing parameter uncertainty and its effect upon model predictions. To assess the uncertainty in prediction it is necessary to generate multiple realizations and test the prediction for each realization. This work uses a multi-level sampling approach to incorporate uncertainty in both large-scale trends and the small-scale stochastic variability. The large-scale uncertainty is addressed using a model-averaging approach considering both calibration error and regularization. A geostatistical approach is adopted for the small-scale uncertainty, which is added to the large-scale conductivity to give the combined conductivity fields. The prediction model is then run using the simulated fields to get the distribution of predictions. These approaches are developed and tested on a hypothetical groundwater aquifer as well as a field-scale application based on the well known waste isolation pilot plant (WIPP) site.

Gopalakrishnan, G. (2008). *Nature's sensors: Using plants as an alternative monitoring approach for subsurface contamination*, PhD thesis, University of Illinois at Urbana-Champaign.¹ Soil and groundwater monitoring costs are a significant part of the remediation costs of a contaminated site, especially where the monitoring may need to be continued over many years. Traditional monitoring methods involve the use of soil borings and groundwater monitoring wells for soil and groundwater sampling respectively. These are expensive and often result in sampling schemes that focus on certain areas of the site and are not exhaustive. An alternative, exhaustive, less expensive method of monitoring contaminant fate and transport using plants is presented here. The objective of this work is to develop a method to predict chemical contamination in the subsurface as a function of the concentrations in the plant tissue. This has involved developing novel sampling and analytical techniques for plants; evaluating model parameters such as diffusion and sorption as a function of plant chemistry and developing optimal sampling schemes using both plants and soil borings at a real-world site. This strategy was tested at Argonne National Laboratory (Argonne), where phytoremediation is currently being used to remove tetrachloroethene (PCE), trichloroethene (TCE), and carbon tetrachloride (CCl₄) from soil and ground water. Branch samples were collected from 126 willows and 120 poplars. Contaminant concentrations from 31 soil borings and six monitoring wells were compared to those from branches of adjacent trees and correlation coefficients of at least 0.89 were obtained. Kriged profiles of the tree concentrations were compared to soil data for TCE and showed good agreement. An analytical model for subsurface contamination as a function of tree concentrations was developed. Model parameters included the diffusion and sorption coefficients and were calibrated with field data for a single tree species. This monitoring strategy was evaluated for a single tree species. The method was then extended to multiple tree species to make it as widely applicable as possible. A single contaminant, TCE, was evaluated. The first step in making the monitoring method more general involved developing an analytical technique that is robust, inexpensive and extracts almost all the contaminant mass present in tree tissue. Hot methanol was found to be the most effective analytical method for all tree species with mass recovery rates of 92-105% and sample variability of approximately 1%. In the second step, the diffusion and sorption coefficients used as parameters in the analytical model developed previously were evaluated as a function of plant chemistry. Results indicated that both sorption and diffusion coefficients were a function of the lipid content of the sample for the species evaluated. A relationship between the sorption coefficient, the lipid content and the octanol-

¹ Coming soon to <http://colab.ncsa.uiuc.edu/MinskerResearchPublication/documents/thesis-and-dissertations/>

water partitioning coefficient of the chemical was developed. Diffusion coefficients for the bark were 2-10 times less than the diffusion coefficients for wood, suggesting that bark acts as a barrier to mass transfer of contaminants in plant branches. Finally, methods to incorporate plant sampling with traditional monitoring techniques were explored at the Argonne phytoremediation site. The impacts on uncertainty estimation and optimal sampling schemes at the site from the inclusion of plant sampling was evaluated. Four methods were compared while evaluating uncertainty; entropy was found to be the most sensitive measure of uncertainty estimation at the site. Optimal sampling plans generated using a non-dominated sorted genetic algorithm found that uncertainty was reduced significantly with a marginal increase in cost when most of the samples collected were plant samples.

Demissie, Y., A. Valocchi, B. Minsker, and B. Bailey, “Integrating physically-based groundwater flow models with error-correcting data-driven models to improve predictions,” in preparation for *Journal of Hydrology*. Physically-based groundwater models such as MODFLOW contain numerous parameters which are usually estimated using statistically-based methods that assume that the underlying error is white noise. However, because of the practical difficulties of representing all the natural subsurface complexity, numerical simulations are often prone to large uncertainties that can result in both random and systematic model error. The systematic errors can be attributed to conceptual, parameter and measurement uncertainty, and most often it can be difficult to determine their underlying cause. In order to handle systematic error in physically-based groundwater flow model applications, we have developed a framework that uses error-correcting data-driven models in a complementary fashion. The data-driven models are separately developed and trained to fit the MODFLOW head calibration errors; these models are subsequently used to update temporal and spatial predictions of MODFLOW head. The framework is evaluated using a hypothetical case study based on a phytoremediation site at the Argonne National Laboratory (ANL). This case study includes structural, parameter, and measurement uncertainties. Results obtained indicate that the proposed complementary modeling framework provides substantially improved predictions (in terms of bias, prediction uncertainty and error correlation) than the original MODFLOW model, in both the calibration and the verification periods. For example, we achieved about 63% and 52% reduction in temporal prediction bias and uncertainty range, respectively. Even if a similar improvement is noted for spatial predictions, it is not as good as the temporal prediction, possibly due to the lack spatially relevant data for the data-driven model to learn the spatial pattern in the calibration error.

Research Implications

The research created novel approaches to long-term monitoring design and modeling that clearly demonstrate the value of including additional information sources into design and modeling efforts at DOE sites. The interactive monitoring design and model calibration approaches showed how combining expert judgment with more traditional quantitative objectives in a rigorous framework can substantially improve the plausibility and reliability of the results, as shown at the DOE WIPP site and a BP site. At DOE ANL site, alternative approaches to monitoring phytoremediation sites by sampling tree branches were shown to substantially increase knowledge about soil contamination at reduced cost over more traditional approaches. Lastly, a complementary modeling framework that combined information from neural networks

¹ Coming soon to <http://colab.ncsa.uiuc.edu/MinskerResearchPublication/documents/thesis-and-dissertations/>

with a MODFLOW model achieved substantial reduction in prediction bias and uncertainty bounds at the DOE ANL site.

Journal and Conference Publications

- Babbar, M., and B. S. Minsker, "Interactive Genetic Algorithm Framework for Long Term Groundwater Monitoring Design," American Society of Civil Engineers (ASCE) Environmental & Water Resources Institute (EWRI) World Water & Environmental Resources Congress 2004 & Related Symposia, Salt Lake City, UT, 2004.
- Babbar, M., B. S. Minsker, and H. Takagi, "Expert Knowledge in Long-Term Groundwater Monitoring Optimization Process: The Interactive Genetic Algorithm Perspective," American Society of Civil Engineers (ASCE) Environmental & Water Resources Institute (EWRI) World Water & Environmental Resources Congress, Anchorage, AK, 2005.
- Babbar, M., and B. S. Minsker, "A Collaborative Interactive Genetic Algorithm Framework for Mixed-Initiative Interaction with Real and Simulated Experts: A Case Study in Long-Term Groundwater Monitoring Design," American Society of Civil Engineers (ASCE) Environmental & Water Resources Institute (EWRI) World Water & Environmental Resources Congress, Omaha, NE, 2006.
- Babbar-Sebens, M., and B. S. Minsker, "Standard Interactive Genetic Algorithm - A Comprehensive Optimization Framework for Groundwater Monitoring Design," submitted to *J. of Water Resources Planning and Management*.
- Demissie, Y.K., Valocchi, A.J., Minsker, B.S. & Bailey, B.A. 2006. Data-driven modeling approach to enhance MODFLOW head predictions, pp. 51-55 in Proceedings of MODFLOW and More 2006: Managing Ground-Water Systems, v. 1, E. Poeter, M. Hill and C. Zheng (ed.), May 21-24, Colorado School of Mines, Golden, CO.
- Demissie, Y.K., A.J. Valocchi, B.S. Minsker, and B.A. Bailey, Bias Corrected Groundwater Model Prediction Uncertainty Analysis, International Conference ModelCARE 2007, Calibration and Reliability in Groundwater Modeling, Credibility of Modeling, Copenhagen, Denmark, Sept. 9-13, 2007.
- Demissie, Y., Valocchi, A., Minsker, B. & Bailey, B. (2008) Bias-Corrected Groundwater Model Prediction Uncertainty Analysis. In: K. Kovar (Editor), ModelCARE2007 Calibration and Reliability in Groundwater Modeling: Credibility in Modeling, Copenhagen, Denmark. IAHS Publ. 320
- Demissie, Y.K., Valocchi, A.J., Minsker, B.S. & Bailey, B.A. 2006. Data-driven modeling approach to enhance MODFLOW head predictions, pp. 51-55 in Proceedings of MODFLOW and More 2006: Managing Ground-Water Systems, v. 1, E. Poeter, M. Hill and C. Zheng (ed.), May 21-24, Colorado School of Mines, Golden, CO., 2006.
- Gopalakrishnan, G., M. C. Negri, B.S. Minsker, C.J. Werth, "Monitoring soil and groundwater plumes using trees in an existing phytoremediation system," Third International Phytotechnologies Conference, Atlanta, U.S.A., 2005
- Gopalakrishnan, G., M. C. Negri, B. S. Minsker, C. J. Werth, Monitoring subsurface contamination using tree branches, *Ground Water Monitoring and Remediation*, 27(1), 1-10, 2007.
- Gopalakrishnan, G., C.J. Werth, M.C. Negri, "Nature's sensors: Monitoring short-term temporal fluctuations in phytoremediation systems," Fourth International Phytotechnologies Conference, Denver, U.S.A., 2007
- Negri, M.C., G. Gopalakrishnan, S.F. Ahmad, J. Quinn, "Monitoring of a large-scale phytoremediation site over eight years," Fourth International Phytotechnologies Conference, Denver, U.S.A., 2007
- Singh, A., B. S. Minsker, and H. Takagi, "Interactive Genetic Algorithms for Groundwater Inverse Modeling," American Society of Civil Engineers (ASCE) Environmental & Water Resources Institute (EWRI) World Water & Environmental Resources Congress, Anchorage, AK, 2005.

- Singh, A., B. S. Minsker, and H. Takagi, "Interactive Multi-Objective Inverse Groundwater Modeling – Formulation and Addressing User Fatigue," American Society of Civil Engineers (ASCE) Environmental & Water Resources Institute (EWRI) World Water & Environmental Resources Congress, Omaha, NE, 2006.
- Singh, A, B.S. Minsker, A.J. Valocchi, and D.D.Walker, Interactive Multi-objective Inverse Groundwater Modeling for the WIPP Site, in: World Water & Environmental Resources Congress 2007 & Related Symposia, Tampa, FL. American Society of Civil Engineers (ASCE) Environmental & Water Resources Institute (EWRI), 2007.
- Singh, A., B.S. Minsker, and A.J. Valocchi, An Interactive Multi-Objective Framework for Expert-Based Hydrologic Model Calibration, in: Proceedings for The 32nd Congress of IAHR, The International Association of Hydraulic Engineering and Research: Harmonizing the Demands of Art and Nature in Hydraulics, July1-6 2007, Venice, Italy, 2007.
- Singh, A., B. S. Minsker, A. J. Valocchi, "An Interactive Multi-Objective Optimization Framework for Groundwater Inverse Modeling," *Advances in Water Resources*, in review.