Report for 2005PA43B: A Bayesian Framework for Cost Effective Groundwater Monitoring Design

Publications

- Conference Proceedings:
    - Kollat, J. B. and P. M., Reed, 2005, The Value of Online Adaptive Search: A comparison of NSGA-II, å-NSGAII, and åMOEA, Third International Conference on Evolutionary Multi-Criterion Optimization (EMO 2005), Guanajuato, Mexico, March 2005, Editors Coello Coello, C. A., Hernandez, A., Zitzler, E., Lecture Notes in Computer Science, Springer-Verlag, 386-398.
- Articles in Refereed Scientific Journals:
    - Kollat, J. B., and P., Reed, 2006, Comparison of Multi-Objective Evolutionary Algorithms for Long-Term Monitoring Design, Advances in Water Resources, 29(6), 792-807.
    - Reed, P. J. B., Kollat, and V. Devireddy, In-Press, Using Interactive Archives in Evolutionary Multiobjective Optimization: A Case Study for Long-Term Groundwater Monitoring Design, Environmental Modeling & Software.
    - Kollat, J. B., and P. M., Reed, In-Review, A Computational Scaling Analysis of Multiobjective Evolutionary Algorithms in Long-Term Groundwater Monitoring Applications, Advances in Water Resources.
    - Tang, Yong, P. M., Reed, and J. B. Kollat, In-Review, Parallelization Strategies for Rapid and Robust Evolutionary Multiobjective Optimization in Water Resources Applications, Advances in Water Resources.
- Dissertations:
    - Kollat, J. B., 2005, The Epsilon Non-Dominated Sorted Genetic Algorithm II: A Highly Effective Multi-Objective Evolutionary Algorithm for Water Resources Applications, "MS Dissertation," Department of Civil and Environmental Engineering, College of Engineering, The Pennsylvania State University, University Park, PA, 71 pages.

Report Follows

ABSTRACT.

The goal of this proposed research is to develop a decision support framework for designing cost-effective monitoring (LTM) networks. The monitoring framework allows hydrologic scientists' to balance multiple design objectives while characterizing complex hydrologic systems and adapt their objectives and system design to account for advances in real-time sensing. The objectives of this research are to:

1. Initiate the Real-Time Hydrologic Monitoring Network in the Shale Hills experimental watershed.
2. Develop, assess, and extend decision support tools for long-term monitoring network design.

The monitoring framework couples multiobjective optimization with the geostatistical visualization and uncertainty modeling to inform cost-benefit analysis for sensor investments. This research has yielded robust monitoring design tools that have been shown to be superior to existing tools. Additionally, computational scaling studies have shown that the framework can be used to reduce growth of computational demands from being quadratic with increasing network design problem sizes to linear, making it feasible to solve large-scale spatiotemporal monitoring problems.

Ongoing extensions of this work will utilize the multiobjective design framework to adaptively design a real-time groundwater monitoring campaign within the Shale Hills experimental watershed. The Shale Hills 19.8 acre experimental watershed was established in the 1970's by the Forest Hydrology group at Penn State to experimentally determine the physical mechanisms of streamflow generation for an upland forested catchment and to evaluate the effects of antecedent soil moisture on storm flow volume and timing. Historical and current research at the site will provide the opportunity to demonstrate our methodology for the two scenarios discussed below:

<u>Scenario 1:</u> An extensive monitoring network exists and the framework will be used to analyze the value of existing and proposed monitoring stations in reducing site uncertainties.

<u>Scenario 2:</u> Given fixed funds and a new experimental site, the monitoring framework will be used to identify the locations, sampling rates, and observation technologies for new monitoring stations. Note this will be part of a long-term research effort that extends beyond the time frame of this proposal.

The multiobjective decision support tools will provide stakeholders and policy makers with a better understanding of the value of adding monitoring points into existing networks and allow them to exploit a broader array of information sources.

STATEMENT OF CRITICAL NEED.

Groundwater is a vital resource for the Commonwealth of Pennsylvania (PA), supplying nearly 50-percent of the residential drinking water demand (PADEP 1997). Long-term monitoring of the quantity, quality, and susceptibility of groundwater has been and continues to be an issue of paramount importance within the state. Figure (1) illustrates this point showing Pennsylvania's state-wide effort to develop long term, real-time groundwater monitoring stations.

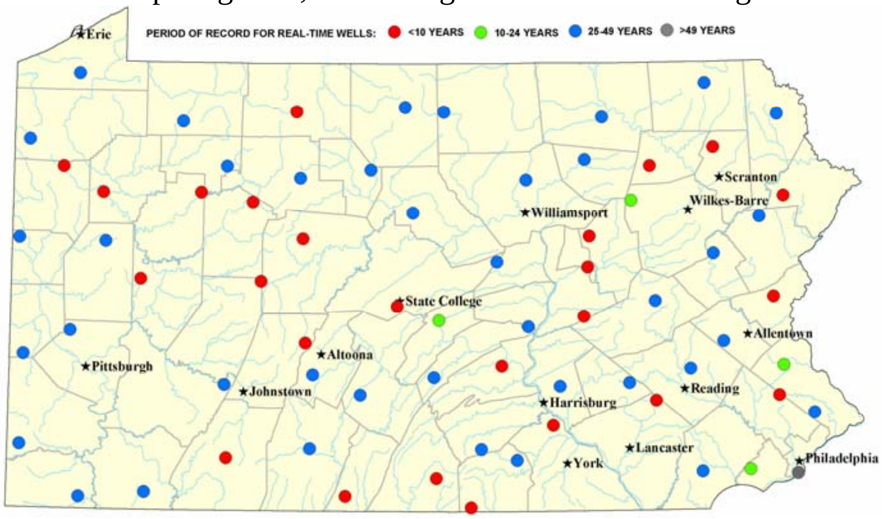

Figure 1: PA long-term, real-time groundwater monitoring stations (Courtesy of Patricia Lietman, PA District Chief USGS).

Broad initiatives such as Act 220 the Water Resources Planning Act and the Pennsylvania Wellhead Protection Program (WHPP) mandated by the Federal Safe Drinking Water Act will continue to increase the need for long-term groundwater monitoring records at both regional and local scales. Signed in 2002, Act 220 will require the Department of Environmental Protection to develop a state water plan by 2007, which will establish water budgets within critical planning areas throughout Pennsylvania. Given that groundwater is the largest store of fresh water within the state, increased investment in long-term groundwater monitoring will be required to better understand recharge and baseflow contributions to streams in critical watersheds. In the context of the WHPP, management of the wellhead protection areas will require rigorous, long-term groundwater quality assessments. The success of Act 220 and the WHPP as well as future groundwater resources management initiatives will require new tools for optimally balancing monitoring costs and data uncertainty. This is particularly important when monitoring groundwater for pollutants such as volatile organic chemicals, pesticides, and other expensive analytes over very long time periods. Nationally, ASCE (Task Committee on Long-Term Groundwater Monitoring Design 2003) highlights that projected federal expenditures on long-term groundwater monitoring for the decade beginning in the year 2000 will be more than $5 billion (Task Committee on Long-Term Groundwater Monitoring Design 2003).

Notice in Figure (1) that there has been a phased or "adaptive" addition of monitoring stations over the past fifty years as societal and scientific needs placed a greater emphasis on monitoring PA's groundwater resources. The figure also illustrates a shifting emphasis toward spatially distributed, real-time monitoring technologies. Two fundamental questions had to be considered with the addition of each new monitoring station in the network shown in Figure (1):

1. Is the cost of the proposed station justified by value of information it will contribute?
2. Where should the proposed monitoring station be located and at what frequency should sampling occur to maximize the information gained?

The goal of this proposed research is to aid monitoring stakeholders in answering these questions by developing a decision support framework for designing cost-effective long-term groundwater monitoring (LTM) networks. The long-term goal of this research is to provide the first *adaptive observation system* design paradigm that will enhance hydrologic scientists' abilities to (1) balance multiple design objectives while characterizing complex groundwater systems across space-and-time, (2) merge physical model predictions with a broad range of data sources (e.g., indicator contaminant samples, expert knowledge, real-time data series), (3) consider a much broader range of model and data uncertainties, and (4) adapt their objectives and system design to account for advances in real-time sensing.

STATEMENT OF RESULTS OR BENEFITS.

The expected outcome of this proposed research will be a new monitoring design framework that will enhance our ability to cost-effectively characterize the quantity, quality, and susceptibility of groundwater resources within Pennsylvania and throughout the United States. The multiobjective decision support tools will provide stakeholders and policy makers with a better understanding of the value of adding monitoring points into existing networks and allow them to exploit a broader array of information sources. This project is expected to generate the following deliverables: (1) Monitoring Design Software Library, (2) Journal Publications, and (3) Conference Publications. The software tools have a general applicability across scientific disciplines, ensuring that they can be used to optimize large-scale investments into a broad array of environmental observation systems (e.g., the CLEANER, CUAHSI, or NEON initiatives).

NATURE, SCOPE, AND OBJECTIVES.

The goal of this proposed research is to develop a decision support framework for designing cost-effective long-term groundwater monitoring (LTM) networks. The monitoring framework allows hydrologic scientists' to balance multiple design objectives while characterizing complex groundwater systems and adapt their objectives and system design to account for advances in real-time sensing. The objectives of this research are to:

- Initiate the Real-Time Hydrologic Monitoring Network in the Shale Hills experimental watershed.
- Develop, assess, and extend decision support tools for long-term groundwater monitoring network design.

The monitoring framework couples multiobjective optimization with the geostatistical visualization and uncertainty modeling to inform cost-benefit analysis for sensor investments. This research has yielded robust monitoring design tools that have been shown to be superior to existing tools. Additionally, computational scaling studies have shown that the framework can be used to reduce growth of computational demands from being quadratic with increasing network design problem sizes to linear, making it feasible to solve large-scale spatiotemporal monitoring problems.

Ongoing extensions of this work will utilize the multiobjective design framework to adaptively design a real-time groundwater monitoring campaign within the Shale Hills experimental watershed. The Shale Hills 19.8 acre experimental watershed was established in the 1970's by the Forest Hydrology group at Penn State to experimentally determine the physical mechanisms of streamflow generation for an upland forested catchment and to evaluate the effects of antecedent soil moisture on storm flow volume and timing. Historical and current research at the site will provide the opportunity to demonstrate our methodology for the two scenarios discussed below:

Scenario 1: An extensive monitoring network exists and the framework will be used to analyze the value of existing and proposed monitoring stations in reducing site uncertainties.

Scenario 2: Given fixed funds and a new experimental site, the framework will be used to identify the locations, sampling rates, and observation technologies for new monitoring stations. Note this will be part of a long-term research effort that extends beyond the time frame of this proposal.

METHODS, PROCEDURES, AND FACILITIES.

The objectives of this proposed research will be addressed using the two research tasks:

Task 1: Develop the monitoring framework.
Task 2: Demonstrate the monitoring framework at the Shale Hills experimental watershed located in north central PA within the Valley and Ridge Province of the Susquehanna River Basin (task ongoing beyond the time allocated for this project)

Detailed descriptions of these tasks and the methods that will be employed in this research are provided below.

Task 1: Develop the monitoring framework.

The purpose of this task is to couple the optimization capabilities of the NSGA-II with the advanced geostatistical mapping and uncertainty modeling. The NSGA-II and the mapping library, in combination, will help hydrologic scientists to (1) balance multiple design objectives while characterizing complex groundwater systems across space-and-time, (2) merge physical model predictions with a broad range of data sources (e.g., indicator contaminant samples, expert knowledge, real-time data series), (3) consider a much broader range of model and data uncertainties, and (4) adapt their objectives and system design to account for advances in real-time sensing. A brief background for evolutionary multiobjective optimization methods are given below to justify our proposed methodology. The NSGA-II will potentially evaluate thousands of sampling strategies using the enhanced geostatistical estimation routines to determine the optimal tradeoffs for an application; therefore this as part of this task we will ensure that our software estimation routines are highly robust.

Background on Evolutionary Multiobjective Optimization

Within the monitoring framework, multiobjective genetic algorithms will be used to search for optimal LTM tradeoffs similar to the tradeoff in Figure (2). Genetic algorithms' population-based search enables them to efficiently evolve entire design tradeoffs for nonlinear, discrete problems with huge decision spaces (Back et al. 2000; Goldberg 1989; Salomon 1998). This is particularly relevant for LTM design since the problem is discrete and the total number of possible monitoring designs (i.e., the decision space) grows according to the function 2^n where n is the number of potential samples in space-and-time. For a simple spatial monitoring network design, choosing one new sampling point from 30 potential monitoring sites will yield more than 1,000,000,000 (or 2^{30}) possible designs. Reed and Minsker (Reed and Minsker 2004) have specifically shown these solution methods have significant promise for LTM design. Schaffer (Schaffer 1984) developed one of the first evolutionary multiobjective optimization (EMO) algorithms termed the vector evaluated genetic algorithm (VEGA), which was designed to search decision spaces for the optimal tradeoffs among a vector of objectives. Subsequent innovations in EMO have resulted in a rapidly growing field with a variety of solution methods that have been used successfully in a wide range of applications (as reviewed by (Coello Coello et al. 2002; Deb 2001; Fonseca and Fleming 1995; Van Veldhuizen 1999)). These solution methods have garnered increased attention over the past decade and have been applied successfully in a variety of water resources and environmental applications (Cieniawski et al. 1995; Erickson et al. 2002; Halhal et al. 1997; Horn and Nafpliotis 1993; Loughlin et al. 2000; Reed and Minsker 2004; Reed et al. 2001; Ritzel et al. 1994).

Task 2: Demonstrate the monitoring framework at the Shale Hills experimental watershed located in north central PA within the Valley and Ridge Province of the Susquehanna River Basin.

The monitoring framework's design capabilities will be tested at the Shale Hills experimental watershed. The test case will be used to carefully validate the monitoring framework's effectiveness relative to current design methods and justify broad dissemination of its software tools. The Shale Hills 19.8 acre experimental watershed was established in the 1970's by the Forest Hydrology group at Penn State to experimentally determine the physical mechanisms of streamflow generation for an upland forested catchment and to evaluate the effects of antecedent soil moisture on storm flow volume and timing. The experiment consisted of a comprehensive network of piezometers, neutron access tubes, and 4 weirs. A spray irrigation network was installed to apply a controllable amount of rainfall over all or part of the entire watershed. The data collected have been used for many years for teaching and research.

Shale Hills will provide an excellent test bed for the monitoring framework because of the historical research infrastructure illustrated in Figure (2) and current instrumentation initiatives. Historical and current research at the site will provide the opportunity to demonstrate our methodology for the two scenarios discussed below:

Scenario 1: An extensive monitoring network exists and the framework will be used to analyze the value of existing and proposed monitoring stations in reducing site uncertainties.

Scenario 2: Given fixed funds and a new experimental site, the framework will be used to identify the locations, sampling rates, and observation technologies for new monitoring stations. Note this will be part of a long-term research effort that extends beyond the time frame of this proposal.

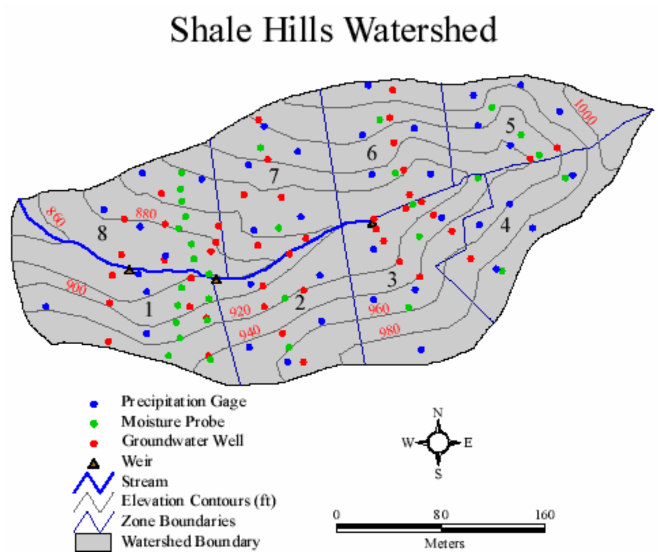

Figure 2: Illustration of the historical research infrastructure used to monitor the Shale Hills Experimental Watershed (Courtesy of Chris Duffy, Penn State University).

The scenarios discussed above are representative of the types of problems practitioners will face when seeking to improve PA's long-term groundwater monitoring systems. For Scenario 1, we will evolve Cost—Uncertainty tradeoffs using historical data from the 40 groundwater monitoring wells [see Figure (2)] used to initially characterize the Shale Hills site. As part of this research, we will analyze the impact of data uncertainty, the value of physical models, and the value of alternative sampling technologies (e.g., real-time versus synoptic sampling). For Scenario 2, the framework will synergistically support current efforts (i.e., Reed, Duffy, and Helly NSF/EAR-0418798, 2005-2007) to develop long-term monitoring infrastructure in several sites within the Shavers Creek watershed, which includes Shale Hills. This research will test the framework's ability to find solutions that maximize the value of fixed annual investments in new groundwater monitoring stations. In Scenario 2, Cost—Uncertainty tradeoffs will arise when selecting a subset of proposed observation sites while considering cost constraints and alternative monitoring technologies (i.e., manual sampling, pressure transducers connected to data loggers, or wireless sampling with internet-based data access). This research task will exploit the NSF-supported research of Duffy and Reed (Duffy and Reed 2004), in which a Monte Carlo physics-based groundwater simulation is currently being developed for the Shale Hills site. This task is part of a long-term monitoring effort that will continue well beyond the scope of this funded project.

PRINCIPAL FINDINGS AND SIGNIFICANCE.

The goal of this proposed research is to develop a decision support framework for designing cost-effective long-term monitoring (LTM) networks. The monitoring framework allows hydrologic scientists' to balance multiple design objectives while characterizing complex water resources systems and adapt their objectives and system design to account for advances in real-time sensing. The objectives of this research are to:

1. Initiate the Real-Time Hydrologic Monitoring Network (RTH_Net) in the Shale Hills experimental watershed.
2. Develop, assess, and extend decision support tools for long-term monitoring network design.

The monitoring framework couples multiobjective optimization with the geostatistical visualization and uncertainty modeling to inform cost-benefit analysis for sensor investments. This research has yielded robust multiobjective monitoring design tools that have been shown to be superior to existing tools. Computational scaling studies have shown that the framework can be used to reduce growth of computational demands from being quadratic with increasing network design problem sizes to linear. Additionally, the multiobjective network design tools can be used effectively on parallel computing platforms. Advances in this project have substantially enhanced the computational tractability of solving large-scale spatiotemporal monitoring problems. The multiobjective decision support tools developed in this research have also been used to calibrate an integrated hydrologic model of the Shale Hills experimental watershed.

The hydrologic modeling tools developed for the Shale Hills have been used to initiate long-term adaptive monitoring within the RTH_Net experiment (http://www.engr.psu.edu/rth_net/). The goal of the RTH_NET adaptive monitoring experiment is to promote multidisciplinary environmental research into the optimal design of real-time sensor systems that are capable of characterizing coupled processes (e.g. atmosphere, soil, groundwater, and streams) and improving predictions of hydrologic response. Support from the Pennsylvania Water Resources Center has aided in garnering additional support from the National Science Foundation to extend the experimental and theoretical scope of the RTH_Net experiment.

The monitoring design concepts developed in this project and extended in the long-term RTH_Net experiment will be used to develop river-basin scale observatory network design tools, with a specific emphasis on the Susquehanna River Basin and Chesapeake Bay (http://www.srbhos.psu.edu/default.asp). Ongoing work will develop statistical simulation-optimization network design frameworks formulated for multiples scales and processes that will be capable of conditioning model-based predictions on real-time observations.

STUDENTS SUPPORTED.

Joshua Kollat, Civil and Environmental Engineering, Master Degree.

PRESENTATIONS AND OTHER INFORMATION TRANSFER ACTIVITIES.

Reed, P. M. and Kollat, J. B., "Comparison of State-of-the-Art Multi-Objective Evolutionary Algorithms for Long-Term Monitoring Design." Fall 2005 Meeting of the American Geophysical Union, San Francisco, California, December 2005.

Kollat, J. B., Tang, Yong, and Reed, P. M., "Tools for Comparing Evolutionary Multiobjective Optimization Algorithms." In Proceedings of the ASCE World Water and Environmental Resources Congress, Anchorage, Alaska, May 2005.

Kollat, J. B., and Reed, P. M., "Comparison of Multi-Objective Evolutionary Algorithms for Long-Term Monitoring Design." In Proceedings of the ASCE World Water and Environmental Resources Congress, Anchorage, Alaska, May 2005.