

Reducing Spatial Sampling in Long-Term Groundwater Monitoring Networks Using Ant Colony Optimization

Yuanhai Li¹ and Amy B. Chan Hilton²

¹Department of Civil and Environmental Engineering, FAMU-FSU College of Engineering, Florida State University, Tallahassee, FL 32310 USA
yuanhai@eng.fsu.edu

²Department of Civil and Environmental Engineering, FAMU-FSU College of Engineering, Florida State University, Tallahassee, FL 32310 USA
abchan@eng.fsu.edu

Abstract: Groundwater long-term monitoring (LTM) is required to assess human health and environmental risk of residual contaminants after active groundwater remediation activities are completed. However, LTM can be costly because of the large number of sampling locations that exist at a site from previous site characterization and remediation activities. The cost of LTM may be reduced by identifying redundant sampling locations. However, care must be taken so that the elimination of specific individual wells from the monitoring network does not result in unacceptable levels of data loss and errors. An ant colony optimization (ACO) algorithm is proposed to identify optimal sampling networks that minimize the number of monitoring locations while maintaining the overall data loss below a given threshold. ACO is inspired by the ability of an ant colony to identify the shortest route between its nest and a food source through indirect communication and positive feedback. Metrics for quantifying well redundancy and overall data loss after optimization are quantified and used in the ACO heuristics. To demonstrate its effectiveness, the ACO developed for LTM optimization is applied to a case study with 30 existing monitoring wells. The LTM optimization problem was solved using different data loss thresholds to identify solutions with 27 to 21 wells remaining in the LTM network. Contour mapping of the contaminant plume using the remaining wells show that the ACO solutions are effective and practical. These results demonstrated that ACO is a promising method for solving LTM optimization problems.

Keywords: Ant colony optimization, Computational intelligence applications, Groundwater monitoring.

I. Introduction

Groundwater long-term monitoring (LTM) has become increasingly important and prevalent in recent years, especially as active remediation concludes and the use of

monitored natural attenuation increases. LTM is required to assess human health and environmental risk of residual contaminants after active groundwater remediation activities are completed. LTM can be costly given the large number of sampling locations (from dozens to hundreds), frequency of sampling (as often as quarterly), and number of constituents monitored at a given site. Biotic, chemical, and physical data are collected during monitoring, in which these data should be relevant to site-specific objectives [24]. The U.S. Department of Energy (DOE) estimated that the total costs for monitoring at their sites where long term stewardship has been mandated may be up to \$100 million per year [20]. The U.S. Navy estimated the costs of LTM and remedial active operation in Navy's contaminated sites are from \$46 million for 1999 to \$77 million for 2003 [22]. Since LTM is required for scores of years at many sites, the cumulative costs can be significant. Thus a need exists to reduce LTM costs through optimization.

The overall goal of LTM optimization is to reduce monitoring costs while still capturing sufficient information about contaminant levels and plume movement. Since the majority of monitoring costs is from sampling, the focus is to reduce number and frequency of monitoring points and to decide which contaminants and constituents require sampling. An existing monitoring network typically has more than necessary sampling locations for the purpose of LTM. Thus LTM costs may be reduced by identifying redundant sampling locations, which is the focus of this work.

Two common approaches to LTM optimization are mathematical optimization and statistical analysis. In mathematical optimization, optimal LTM sampling is identified by applying search algorithms that maximize or

minimize a given objective function subject to constraints. These optimization algorithms use plume predictions based on numerical simulation models of contaminant transport and groundwater flow and/or geostatistical interpolation to guide the search. For example, [6] optimized groundwater monitoring networks using a genetic algorithm stochastic search method combined with Monte Carlo simulation of plume movement. Reference [16] developed an approach using simulated annealing for stochastic global optimal search and statistical methods to reduce temporal redundancy and increase spatial accuracy of monitoring networks. Reference [17] optimized sampling networks using geostatistical interpolation through inverse distance weighting (IDW), ordinary kriging combined with a fate-and-transport model, and genetic algorithms. Reference [25] introduced new spatial moment constraints based on [17] to obtain robust long-term monitoring optimization designs. Reference [12] designed an improved detection network for a landfill using a two-dimensional simulation model alone without mathematical optimization methods. The drawback of using simulation models to predict contaminant plume movement is that limited site data or complex hydrogeological conditions will lead to uncertainty in the model and input parameters, causing errors in the predications and possibly unreliable optimal monitoring networks.

Statistical methods have been applied to improve existing monitoring networks by analyzing current or historical data. Reference [2] developed the monitoring and remediation optimization system (MAROS), which is a decision support system that uses a ranking rule-based approach combined with Delaunay triangulation to reduce sampling locations and statistical methods to reduce sampling frequency. Reference [5] developed the Geostatistical Temporal/Spatial (GTS) optimization algorithm, which is a site specific technique based on kriging and is applicable to sites with a large number of monitoring wells. Reference [19] developed the cost-effective sampling (CES) method to reduce sampling frequency through a decision support system based on trend analysis and simple statistical measures. However, MAROS, GTS, and CES are not mathematical optimization methods; they are decision support tools in which manual iterative steps rather than automated processes are used to improve existing LTM networks in a sequential ranking procedure. As such, these decision support tools do not focus on identifying global optimum solutions and do not have the benefits of mathematical optimization methods, such as the ability to evaluate multiple options while considering the interactions between decision variables, objectives, and constraints simultaneously. More importantly, ranking methods that eliminate the most redundant well(s) each iteration are similar to greedy search procedures, which typically result in suboptimal and local optima solutions.

This work presents the development of a groundwater LTM network optimization methodology based on a recent and

promising global stochastic search algorithm. The goal is to present a proof of concept that the ant colony optimization (ACO) paradigm can be effectively created and applied to solve the LTM spatial optimization problem of maximizing cost-effectiveness through reducing sampling locations while maintaining sufficient data to limit errors. To the authors' knowledge, this is the first development of ACO for groundwater remediation and monitoring problems. In the following sections, an overview of ACO is presented, and then the mathematical formulation of the LTM optimization problem is described and the metrics for data redundancy and overall data loss are introduced. The results of the application of the developed ACO for LTM network optimization to a case study are presented and discussed.

II. Ant Colony Optimization (ACO)

ACO is an evolutionary computation optimization method based on ants' collective problem-solving ability. This global stochastic search method is inspired by the ability of a colony of ants to identify the shortest route between the nest and a food source. Individual ants contribute their own knowledge to other ants in the colony by depositing pheromones, which act as chemical markers, along the paths they traverse. Through indirect communication with other ants via foraging behavior, a colony of ants can establish the shortest path between the nest and the food source over time with a positive feedback loop known as stigmergy. As individual ants traverse a path, pheromones are deposited along the trail, altering the overall pheromone density. More trips can be made along shorter paths and the resulting increase in pheromone density attracts other ants to this paths. Furthermore, shorter paths will tend to have higher pheromone densities than longer paths since pheromone density decreases over time due to evaporation [3]. This shortest path represents the global optimal solution and all the possible paths represent the feasible region of the problem. This stigmergy behavior was observed by [9], in which the movement of live ants was observed.

ACO algorithms have been successfully adapted to solve difficult combinatorial optimization problems with discrete search space. The first ant colony simulation algorithm was developed by [8] to solve the classic NP-hard traveling salesman problem (TSP). In the TSP, the goal is to obtain a shortest path that connects all the cities while visiting each city only once. The distance between two cities is used as a heuristic for determining ant movement and pheromone density. The pheromone is updated based on evaporation of the existing pheromone along the segment, the quality of the overall solution as determined by the ant's overall path, and any pheromone deposited by the best ant. The iterative procedure of ants traversing different paths, based on pheromone density and local solution quality, and pheromone updating guides the stochastic search to the optimal solution.

Reference [4] compared ACO with other stochastic search

algorithms, including the genetic algorithm, evolutionary programming, and simulated annealing, by solving the TSP with 50, 75, and 100 cities. Results showed that ACO identified the best solution for each TSP case. ACO algorithms also have been developed for other classical optimization problems, including the quadratic assignment problem, job-shop scheduling problem, vehicle routing problem, and graph-coloring problem [4]. Moreover, [11] showed that under certain conditions, solutions from ant-based optimization converge to the global optimum with a probability close to 1.

More recently, ACO algorithms have been applied to solve a wide range of engineering and science problems. To date, a limited number of works have been published in which ACO or swarm intelligence has been used to solve water resources and hydrology problems; nevertheless, these do not focus on groundwater management and remediation design optimization problems. In water resources, [15] used ACO to optimize water distribution systems designs and [23], which used particle swarm optimization to determine pump speeds to minimize the total costs in water distribution systems. Reference [1] used ACO to solve an inverse modeling problem of identifying unsaturated soil parameters.

III. LTM Spatial Optimization Problem

This work focuses on LTM spatial optimization through reducing the number of sampling locations in an existing LTM network during a given sampling event. As in other LTM optimization works [2], this study assumes that the contaminant plume is stable because of previous contaminant transformation or remediation activities and that the groundwater monitoring data used for optimization has undergone review and standard quality assurance/quality control practices. Thus LTM optimization is considered during the later stages of remediation projects. While redundant data are valuable during early periods of remediation to assess performance, they often are unnecessary during later periods when LTM is the focus. By identifying redundant data at individual sampling locations, the overall LTM network may be optimized to reduce the overall LTM costs. To do so, data redundancy and overall data loss of the reduced network must be quantified, which both are based on concentration interpolation.

A. Concentration Interpolation

When monitoring wells are eliminated from the LTM network, concentration of contaminants or other constituents at unsampled locations must be estimated based on the concentration values from nearby measured locations. The challenge in interpolation is determining the weights of the measured data used to estimate values at unsampled locations. One commonly used interpolation method is inverse distance weighting (IDW), which is based on the concept that closer points tend to be more alike than those that are farther apart.

One advantage of IDW is that it is a computationally simple deterministic interpolation method. IDW assumes that each measured point has a local influence that diminishes with distance between unmeasured and measured locations. Points closer to the interpolation location are weighted more than those farther away. To estimate a value at any unmeasured location, IDW uses the measured values from surrounding locations and weighs them according to the distance between the measured points and the interpolation location (1), (2). For example, the concentration at location (x_0, y_0) may be estimated using IDW interpolation by:

$$C_0 = \frac{\sum_{i=1}^n \frac{C_i}{d_{i0}^p}}{\sum_{i=1}^n \frac{1}{d_{i0}^p}} \quad (1)$$

$$\text{where } d_{i0} = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2} \quad (2)$$

C_0 is the estimated concentration at location (x_0, y_0) ; C_i is the measured concentration at location (x_i, y_i) ; d_{i0} is the distance between points (x_0, y_0) and (x_i, y_i) ; p is a parameter (typically $p = 2$); and n is the number of neighbors around location (x_0, y_0) included in the interpolation.

Another common interpolation method is kriging, which uses statistical data analysis to determine spatial correlations and variances between sampling locations. In kriging the weights of measured data are identified by the distance between measured points and estimation location and variogram functions. Kriging is an empirical geostatistical method that may be computationally intensive for larger problems. Reference [14] compared three commonly used interpolations techniques, including kriging, IDW, and natural neighbor interpolation, on five monitoring data sets using cross-validation. Reference [18] compared ordinary, multigaussian, intrinsic, and quantile kriging with nonlinear least squares inverse distance weighting and IDW by applying the interpolation approaches to small, medium, and large cases of LTM optimization problems. Because IDW provides reasonable interpolation results, is applicable for sites with small data sets that may be insufficient for geostatistical analysis, and is easy to implement, IDW is the interpolation method used in this work which focuses on the ACO development rather than interpolation methods. Nevertheless, other estimation methods, such as kriging and simulation modeling, may be substituted in place of IDW in the methodology presented here.

B. Data Redundancy

Individual sampling locations are evaluated with respect to their redundancy in the LTM network in order to identify candidate monitoring wells to eliminate. The Relative Estimation Error (*REE*) is developed here to quantify the spatial redundancy of measured data. The *REE* is the

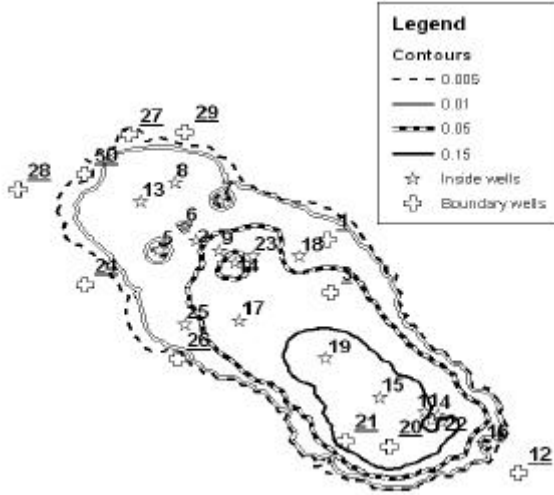


Figure 1. The original 30-well LTM network and contaminant concentration plume (with contours in mg/L) based on monitoring data.

normalized difference between the estimated and measured concentrations and is defined as

$$REE = \frac{|C_{est,i} - C_i|}{\min(C_{est,i}, C_i)} \quad (3)$$

where C_i is the measured concentration of the well i and $C_{est,i}$ is the estimated concentration of the well i . Note that the estimated concentration is compared with the measured value, which is assumed to be the “true” concentration during a given sampling period. Since the difference between $C_{est,i}$ and C_i is normalized by the minimum of these two values, the REE is very sensitive to the residual. REE values may range from 0 to more than 1,000. Individual monitoring wells with low REE values are potential redundant sampling locations. Acceptable REE values may vary among individual monitoring wells depending on their location. For example, for boundary wells to be considered potential redundant wells, the acceptable REE values may be lower than the values for interior wells since monitoring wells along the boundary of a contaminant plume help to define the extent of the plume.

C. Overall Data Loss

In addition to evaluating the importance of individual monitoring wells, the overall data loss of the reduced LTM network due to interpolating concentrations at removed redundant wells needs to be quantified. The overall data loss resulting from optimizing the number of sampling locations is quantified using the interpolation root mean square error ($RMSE$), which is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^m \left(\frac{C_{est,i} - C_i}{\min(C_{est,i}, C_i)} \right)^2}{m}} \quad (4)$$

where m is the number of removed wells, C_i is the measured

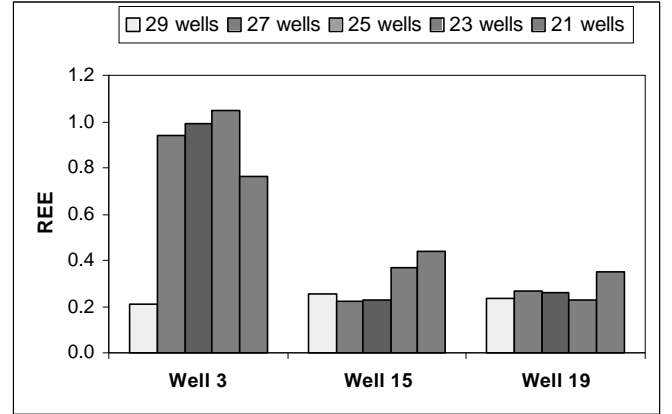


Figure 2. Comparison of REE values from solutions using a ranking method to eliminate wells.

concentration of well i , and $C_{est,i}$ is the estimated concentration of well i estimated using data from the remaining wells. It is expected that acceptable optimized LTM networks have relatively low $RMSE$ values. The $RMSE$ increases due to increased data loss from eliminating too many monitoring wells and/or wells that are not redundant with respect to the overall monitoring network.

D. Optimization Formulation for a LTM Example Problem

A field site in the Upper Aquifer at the Fort Lewis Logistics Center in Pierce County, Washington [21] is used as an example application and case study in this work. This same site was used in a study comparing other LTM network improvement approaches [21]. The existing LTM network is optimized using the ACO developed in this work. The contaminant of concern is trichloroethylene (TCE), which was used as a degreasing agent at the site until 1970’s and has a maximum contaminant level (MCL) of 0.005 mg/L. Regular monitoring was conducted during the period between November 1995 and October 2001. Data from the September 2000 monitoring period is used in this work to optimize the existing LTM network of 30 monitoring wells (Figure 1). By inspection of the existing data, some wells may be categorized as redundant or important by calculating the REE values based on estimating the concentration at each well using the other 29 data points. For example, wells 3, 15, and 19 each have low REE values that are less than 0.3. However, well 3 is defined as boundary well (Figure 1) and may not be presumed to be a redundant well. On the other hand, some wells have REE values over 1,000, such as well 12 and 16, and may be considered to be very important to the whole monitoring network since their values cannot be estimated reliably based on neighboring data. More importantly, using a ranking of each well’s REE values with concentrations interpolated with data from the other 29 wells can significantly underestimate the actual REE for individual wells (Figure 2). This is because the impacts of eliminating wells on individual wells are not taken into account. This demonstrates that mathematical optimization is necessary to

simultaneously analyze the effects of eliminating multiple wells and results in identifying solutions with better performance than those found through a ranking approach.

The existing LTM network is optimized by solving the following combinatorial optimization problem. The objective of this LTM optimization problem is to minimize monitoring cost by way of minimizing the number of sampling locations while maintaining the overall data loss (*RMSE* of concentration estimations) below a specified threshold. Mathematically, this optimization formulation is described as

$$\text{Min } Z = n \quad (5)$$

$$\text{s.t. } \sqrt{\frac{\sum_{i=1}^m \left(\frac{C_i - C_{est,i}}{\min(C_i, C_{est,i})} \right)^2}{m}} \leq T \quad (6)$$

where n is the number of remaining monitoring wells in the reduced LTM network, m is the number of removed wells, C_i is the measured concentration of removed well i , $C_{est,i}$ is the estimated concentration of removed well i interpolated using data from the remaining wells, and T is the threshold *RMSE*. This optimization formulation (5) and (6) is solved with different *RMSE* thresholds to identify optimal reduced LTM networks with wells different number of remaining wells. As the *RMSE* threshold decreases, it is expected that more wells remain in the LTM network. In this work, we focus on developing an ACO algorithm to solve this single-objective formulation. However, it may be possible to reformulate this problem into a multi-objective problem and adapt multi-objective ACO algorithms developed for other applications (for example, [7], [10], [13]) to this problem.

IV. ACO Development for LTM Optimization

An ACO algorithm for solving the groundwater LTM spatial optimization problem is developed in this work. In order to adapt ACO to the LTM optimal sampling problem, heuristics for the ACO paradigm need to be made. Unlike the TSP, the path distance itself is not important in the LTM optimization problem. The developed ACO paradigm developed here is based loosely on [1]. The developed method is a mathematical optimization approach that uses IDW as the interpolation method. The overall goal of LTM optimization is to minimize the number of monitoring wells used for sampling by minimizing spatial redundancy while data loss is minimized (5) and (6).

The general representation of the LTM problem in the ACO framework is shown in Figure 3. Ants visit a series of locations that represent actions at the monitoring wells. When at well i ($i=1, 2, \dots, M$) in the monitoring network, an ant has two choices: either eliminate the well ($j=0$) or include the well ($j=1$) in the updated monitoring network (Figure 3). As the ant travels through the monitoring network, it decides to visit

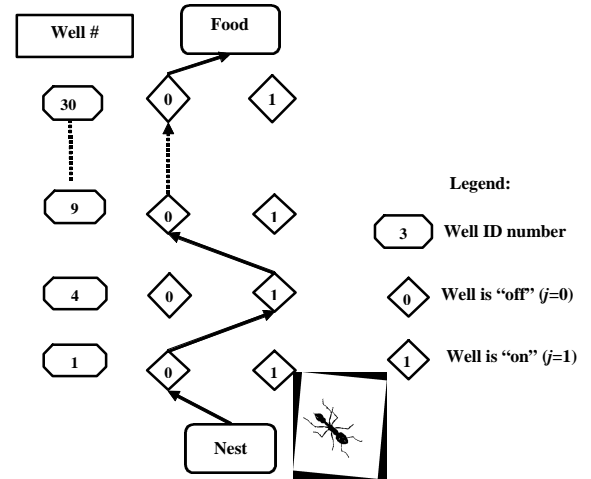


Figure 3. An ant’s trail is comprised of visits to nodes that represent decision j at each well i and identifies the monitoring wells retained or eliminated in the reduced network.

location $j=0$ or $j=1$ for each well i , therefore including or excluding each well from the reduced LTM network. The number of decision variables is the total number of wells considered for optimization, with each variable having a binary option, resulting 2^M combinations of monitoring well networks. A challenge in applying ACO for solving optimization problems is the development of a heuristic for the foraging and stigmergy behavior of ants. In this work, the *REE* at individual wells and overall *RMSE* of the monitoring network are used to regulate pheromone density and ant movement. The overall ACO algorithm is summarized in the following steps:

1. The order in which each ant visits the wells is randomized for each individual ant to ensure that the colony explores different regions of search domain, thus each ant may have a different visiting order from other ants. Each ant must visit all wells once and only once before arriving to the food source from the nest (i.e. starting node), considering all wells in the original monitoring network. The initial pheromone at the beginning of the first iteration at each node is equal. In this work, the number of ants in one colony is equal to the number of monitoring wells in the original LTM network.
2. Data redundancy is quantified using the relative estimation error (*REE*) (3). If the current well is classified as a boundary well, then the *REE* is calculated first; if and only if the *REE* is less than a specified boundary threshold, T_{edge} (in this work $T_{edge} = 0.1$), then the decision of whether to keep or eliminate this boundary well will be determined by ants by continuing onto step. Otherwise, the boundary well is retained in this ant’s solution. Since wells on the edge of the LTM network are important for plume delineation, the probability of eliminating these wells from the monitoring network is lower than that of interior wells.

Evaluation of the *REE* is part of the heuristic developed for the LTM problem to allow ants to make decisions about individual wells. An individual well's contribution to or impact on its neighboring locations is evaluated via a cross-validation procedure. Ant k identifies well x , which is the closest well to its current location well i . The *REE* at well x is calculated twice: $REE(x,0)$ is determined using the concentration at well x interpolated without the data from the current well i , and $REE(x,1)$ is determined by including the concentration data from current well i . A local assessment of current well i 's contribution to the local area is made by comparing the $REE(x,0)$ and $REE(x,1)$ values. A low $REE(x,0)$ value indicates that well i provides little additional information (low contribution) when used to estimate the concentration at well x , whereas a high $REE(x,0)$ value indicates that well i may be important to well x . Moreover, a large difference between $REE(x,0)$ and $REE(x,1)$ indicates that well i contributes to interpolating the concentration at well x .

- a) 3. An ant decides its next step based on a weighted combination of the current pheromone density along a segment and the redundancy of a well. It is this balance between these two heuristics that differentiates ACO from other search methods or a sequential ranking method, which takes a greedy approach to identifying redundant wells. Ant k decides to visit option $j=0$ or $j=1$ for well i , indicating that well i is excluded or included the overall monitoring network, respectively, based on a probability function defined by:

$$P(i,0) = \frac{t(i,0)^a (REE(x,0))^b}{t(i,0)^a (REE(x,0))^b + t(i,1)^a (REE(x,1))^b} \quad (7)$$

where $t(i,0)$ and $t(i,1)$ are the pheromone levels during the current iteration for the cases that represent eliminating well i ($j=0$) or retaining well i ($j=1$), respectively, and \hat{a} and \hat{b} are parameters that weight the pheromone and *REE* components. The values $P(i,0)$ and $P(i,1)$ represent the probability well i will be eliminated or retained in k 's monitoring network. Since $P(i,0) + P(i,1) = 1$, only $P(i,0)$ is directly calculated. Note that if well i is eliminated, then well i will no longer be considered as a neighbor or existing well by ant k for additional well assessments in the current iteration.

4. Steps 2-4 are repeated for each well in the LTM network for ant k , in the order randomly assigned in step 1.
5. After ant k has visited all wells in the existing LTM network, the overall data loss, which quantifies the overall quality of the reduced set of monitoring wells, is evaluated by calculating the interpolation *RMSE*. The concentrations at the removed wells are estimated using data from the remaining wells. The *RMSE* is affected by the number and spatial distribution of removed wells.

Table 1. Parameters used in the ACO for this example problem

| Parameter | Value |
|---|-------|
| Number of neighbors for IDW, n | 8 |
| Exponent parameter in IDW, p | 2 |
| \hat{a} parameter | 0.5 |
| \hat{b} parameter | -1 |
| Initial pheromone | 0.1 |
| Total pheromone, Q | 500 |
| Elitism parameter, e | 5 |
| Pheromone evaporation rate, \tilde{n} | 0.5 |

6. Steps 2-5 are repeated for all other ants in the colony, with the path of all ants affected by the pheromone deposited along trails during the previous iteration.
7. After all ants have completed their tours in the iteration, the pheromone trails are updated. The pheromone for decision j at well i during iteration $t+1$ is updated using:

$$t_{(ij)}(t+1) = (1 - \tilde{n})t_{(ij)}(t) + I\Delta t_{(ij)} + e\Delta t_{(ij)}^e \quad (8)$$

where \tilde{n} is the pheromone decay coefficient ($0 \leq \tilde{n} \leq 1$); $\tau_{(ij)}(t)$ is pheromone during iteration t for well i and case j ($j=0$ indicates well i is eliminated; $j=1$ indicates well i is included). The *RMSE* is calculated using newly interpolated concentrations using the data from the remaining wells. The pheromone density deposited on the path taken by ant k depends on how well the *RMSE*

the *RMSE* constraint (6). For a solution that violates the *RMSE* constraint ($RMSE > T$), the value of \tilde{e} is less than one, decreasing the pheromone density of this ant's trail. The term \tilde{e} is the penalty for *RMSE* constraint violation (6) and is defined as $I = \left(\frac{T}{RMSE}\right)^2$. Similarly, an ant's solution is rewarded if it is feasible. If $RMSE \leq T$, then

In the second term of (8), $\Delta t_{(ij)}$ is the total change in pheromone associated with decision j at well i , which is defined as

$$\Delta t_{(ij)} = \sum_{k=1}^K \Delta t_{(ij)}^k \quad (9)$$

where $\Delta t_{(ij)}^k$ is the change in pheromone due to ant k selecting decision j at well i ; and K is the total number of ants in one colony. The quantity $\Delta t_{(ij)}^k = Q/n_k^2$ denotes the case when ant k uses well i with decision j during the current iteration, Q is a constant related ant pheromone density, and n_k is the number of remaining wells in ant k 's monitoring network. The third term in (8) allows for elitism, in which the pheromone of the ant with the best solution found so far is included in the pheromone update. In this term, e is an integer parameter and $\Delta t_{(ij)}^e = Q/n_e^2$ is the pheromone of the best path (from this elitist ant) found so far, where n_e is the number

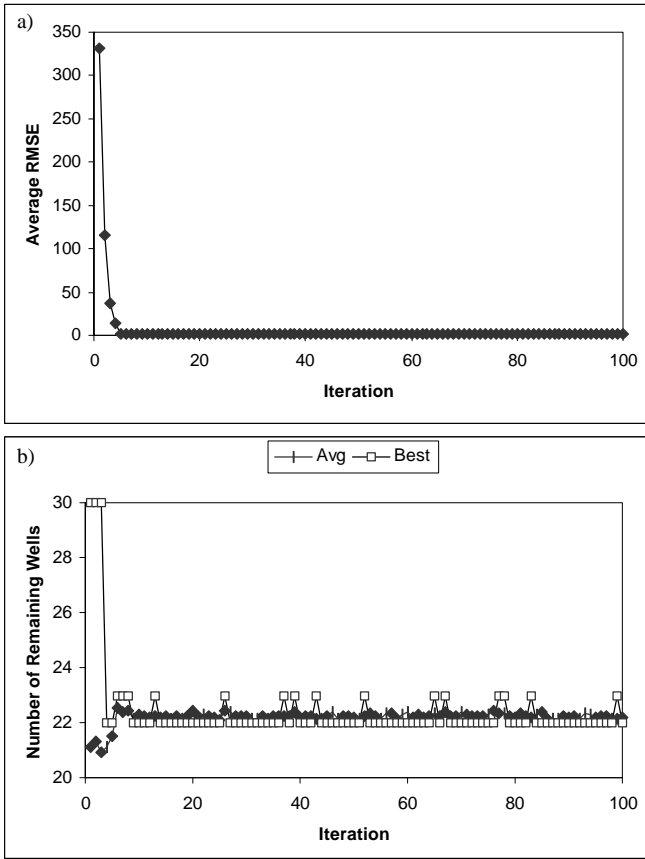


Figure 4. Examples of the convergence of the a) average $RMSE$; b) objective function (number of remaining wells).

of remaining wells in the elitist ant's solution. This elitism term helps direct future iteration ants toward a good solution with higher probability.

8. Continue to the next iteration $t+1$ using the updated pheromone trails and return to step 1. This entire process continues for a preset number of iterations.

This iterative procedure guides the stochastic search to the optimal solution. In this work, the set of ACO parameters were identified by prior evaluations and based on the guidelines presented in the literature [4]. Parameter values used in this work are summarized in Table 1.

V. Results and Discussion

The ACO algorithm developed for solving the optimization formulation defined by (5) and (6) was applied to the Fort Lewis case study [21]. Results indicate that the developed algorithm successfully identified feasible solutions that satisfy the specified overall data loss ($RMSE$) constraint (6). Figures 4a and 4b show a typical convergence of the average $RMSE$ and best feasible solution of a colony during each iteration of the ACO search for the case when the $RMSE$ threshold (T) is 1.5. For this example, an ant colony consists of 30 individual ants each iteration. In other words, the activity of each ant colony is equivalent to a search with a

Table 2. Summary of optimal LTM networks identified by ACO algorithm

| Remaining Wells | Reduction in Wells | Wells Eliminated | $RMSE$ | Average (Std Dev) REE |
|-----------------|--------------------|-----------------------------------|--------|-------------------------|
| 27 | 10 % | 2, 15, 19 | 0.383 | 0.361 (0.157) |
| 26 | 13 % | 2, 10, 11, 25 | 0.595 | 0.578 (0.166) |
| 25 | 17 % | 10, 11, 15, 19, 25 | 0.559 | 0.523 (0.219) |
| 24 | 20 % | 2, 10, 11, 15, 19, 25 | 0.545 | 0.515 (0.197) |
| 23 | 23 % | 2, 3, 10, 11, 15, 18, 25 | 0.637 | 0.562 (0.323) |
| 22 | 27 % | 2, 10, 11, 13, 15, 18, 19, 25 | 0.725 | 0.649 (0.345) |
| 21 | 30 % | 2, 10, 11, 13, 15, 18, 19, 23, 25 | 1.165 | 0.862 (0.832) |

colony containing a single ant over 30 iterations, with each ant following a different random order of visiting all wells and updating the pheromone density every 30 iterations or ant tours. Thus while Figure 4 indicates that the solution converged after 5 iterations for this case, this is approximately equivalent to a search using 150 iterations of an ACO using a single-ant colony. In the early iterations, the average $RMSE$ of the colony was high (Figure 4a), with values over 100, and no feasible solutions identified. After a few additional iterations the average $RMSE$ of the colony rapidly decreased to a range between 1.26 and 1.37, while the best feasible ant solution (i.e., minimum number of remaining wells) in the colony decreased to a range between 22 and 23 wells (Figure 4b). During these later iterations, variations of 8% of the average $RMSE$ and 4% of the minimum number of wells occurred for this example case. The average number of wells remaining in the solutions found during the early iterations is lower (Figure 4b) since these included infeasible solutions in which a high $RMSE$ resulted from eliminating too many wells (Figure 4a).

Optimal solutions for several cases of number of remaining wells were identified by solving the LTM optimization formulation with different values in the overall data loss threshold (T) constraint (7). As expected, results indicate that

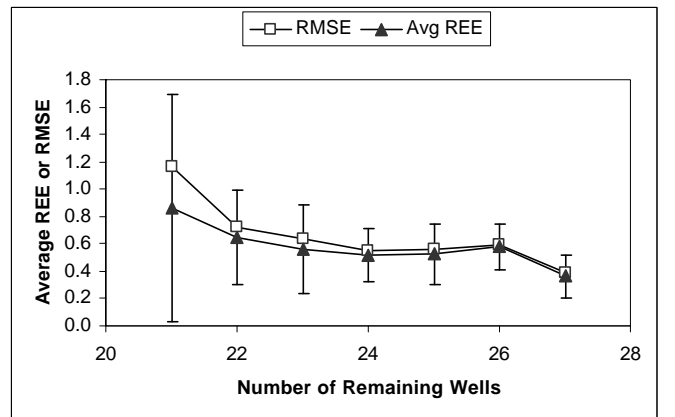


Figure 5. Variation of overall data loss as quantified by the $RMSE$ with the number of remaining wells in the optimized LTM network. The average REE of the interpolated concentrations for the eliminated wells, along with error bars represented one standard deviation, is shown for comparison.

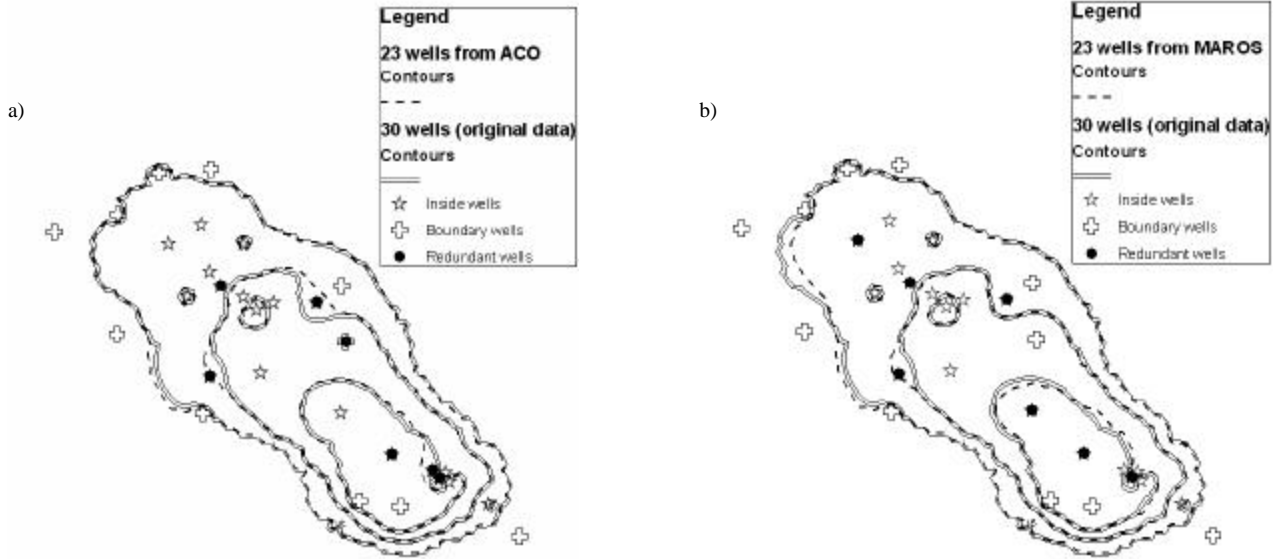


Figure 6. Comparison of concentration contours based on data from the original 30 wells with the 23 wells and interpolated values at the 7 redundant wells from the a) ACO solution and b) MAROS-based solution.

as the number of remaining wells decreases, the average *REE* for the eliminated wells and overall *RMSE* increase non-linearly (Figure 5 and Table 2). Since concentrations at wells eliminated from the monitoring network are estimated using the remaining wells, fewer remaining wells may increase the *REE* of each deleted well, and consequently, the *RMSE* increases. The *RMSE* values here are the lowest for a given number of remaining wells based on a minimum of 50 replicate runs of the ACO for each case. For the cases with 25 and 26 remaining wells, the reduced LTM networks identified by the ACO are near-optimal in terms of *RMSE* compared to the solution for the 24-well case. However, the solutions for the 24-, 25-, and 26-well cases are all similar; the average *REE* for the 25- and 26-well solutions fall within one standard deviation of the average *REE* for the 24-well solution (Table 2). Figure 5 indicates that after reducing the number of wells in the LTM network to a certain point, the trade-off between eliminating additional wells and maintaining an acceptable level of error changes such that the resulting increase in overall *RMSE* becomes too large. For example in this problem, eliminating more than 8 wells out of the original 30 results in a sharp increase in the *RMSE* as well as the average and standard deviation of the *REE* of the interpolated concentrations (Figure 5 and Table 2).

The reduced LTM networks identified by the ACO also were evaluated using concentration contours (Figure 6a). The contours were developed using interpolated concentrations calculated using the same IDW method employed in the ACO algorithm at the eliminated wells and the measured data at the remaining monitoring wells. Results show that the eliminated wells are distributed in the areas inside the plume boundary and among clustered wells. Differences between the contours resulting from the reduced networks and the original contours were minor for most cases (Figure 6a). In the extreme case with 21 remaining wells, which represents a 30% reduction in

monitoring wells, there is some contour area loss at the 0.05 mg/L concentration isopleth. This result verifies that the trade-off between well reduction and data loss through interpolation error may not be beneficial or desirable when too many wells are eliminated, as also indicated in Figure 5 and Table 2. Decision makers need to assess what their error tolerance is and whether this higher data loss, as indicated by the contour map (Figure 6a), is a reasonable trade-off for cost savings through eliminating wells. Thus the *RMSE* value alone may not be sufficient to evaluate a solution; additional analysis such as through contour mapping, which shows the *RMSE* and individual *REE* qualitatively, may be necessary.

The same LTM optimization example problem also was solved following a procedure based on the MAROS procedure for sampling optimization outlined in [2]. In particular, a ranking of slope factor (*SF*) values was used here to sequentially eliminate wells using. The *SF* is defined as the normalized difference between the measured and estimated log-concentrations and is used to quantify an individual well's contribution to the entire monitoring network [2]. Note that for comparison to the ACO algorithm, the estimated concentrations were interpolated using IDW here. In MAROS a well may be considered redundant if the value of *SF* is smaller than a specified value. However, the *SF* is not very sensitive to differences between $C_{est,i}$ and C_i . The case with 23 wells remaining in the monitoring network was analyzed (Figure 6b). A limitation of MAROS is that the impact of removing additional wells on previously removed wells is not explicitly considered. Thus there is a possibility that the *SF* of previously removed wells may increase and even might violate *SF* thresholds, such that a well should not be considered a redundant well. For example, well 15 was eliminated during the fourth iteration with a *SF* value of 0.120, but its *SF* value increased to 0.169 after 3 more wells were deleted. Therefore the estimated concentration and

error (SF) of previously removed wells needs be recalculated each iteration. Eliminating the well with the lowest SF value each iteration is analogous to a greedy search procedure. There is potential that if a well other than the well with the lowest SF is selected for elimination then additional wells may be removed without violating overall constraints and thresholds. In MAROS [2], after a well is eliminated based on SF the overall information loss metrics Average Concentration Ratio (CR) and Area Ratio (AR) are calculated. However, these are averaged metrics, and individual estimated errors of wells removed in previous iterations are not explicitly assessed.

When comparing the results from the ACO and the aforementioned procedure based on MAROS for the case with 23 remaining wells (Figure 6), it is seen that the ACO results are very similar to and slightly better than the MAROS-based results. The resulting overall $RMSE$ is 0.637 for the ACO solution and 0.721 for the MAROS-based solution. The difference is due to the location of two removed wells: wells 13 and 19 are chosen by MAROS-based procedure while wells 3 and 11 are selected by the ACO. Concentration contours based on the data from sampled wells and interpolated values at the deleted wells for the 23-well networks identified by the ACO and MAROS-based procedure are shown in Figure 6. These contours are overlain on the contours created using data from the original 30-well network. The contours from the ACO solution slightly overestimate contaminant contour regions, in particular at the 0.01 mg/L isopleth, in comparison to the contours developed using data from original 30-well LTM network. The ACO results are slightly more conservative compared to the minor underestimation of the 0.005 mg/L isopleth of the MAROS-based solution. Overall, the two sets of resultant contours match the original contours very well.

VI. Conclusions

A new algorithm for LTM spatial optimization using ACO is presented in this work as a proof of concept. ACO is a global search procedure inspired by ant colony foraging activity and stigmergy. The goal of the LTM optimization problem is to minimize the number of sampling wells, as a surrogate to minimizing sampling costs, while satisfying constraints on overall data loss. The developed ACO algorithm is a stochastic search procedure used to optimize LTM networks by identifying redundant sampling locations. Metrics used to quantify individual and overall data loss due to interpolating concentrations at unsampled locations are presented and used to develop heuristics that guide the ACO search for optimal reduced LTM networks.

This work demonstrates that the ACO is a promising method for optimizing monitoring networks. This is significant given the predicted costs of LTM in the coming decades. The ACO was applied to a case study with 30 wells

in the original LTM network, and the results indicate that the developed algorithm can identify optimal and near-optimal monitoring networks. The resulting reduced sampling locations are feasible solutions that also are intuitively logical solutions in which the eliminated wells typically are located in areas inside the contaminant plume boundary and among clustered wells. The ACO solution performs slightly better than the solution identified by a ranking method based on the MAROS decision support system [2]. By solving the LTM optimization problem with the ACO using different data loss thresholds, trade-offs between number of wells eliminated from the monitoring network and overall data loss, as quantified by the interpolation $RMSE$, can be evaluated. Both the $RMSE$ and comparisons of plume contours based on the original LTM network and the reduced network help convey to the decision maker the performance of the optimized monitoring networks. It can be seen that a limit to the number of wells eliminated exists such that the trade-off between cost reductions via reduced sampling and data loss is no longer acceptable.

The ACO heuristic may be extended to apply additional formulations of the LTM optimization problem that address additional environmental monitoring real-world issues. For example, a new ACO paradigm for LTM optimization based on an analogy to the TSP may be adapted based on the ACO developed in this work. In this formulation, an ant may choose to visit any well not previously visited based on similar redundancy metrics and pheromone density. In this way, the LTM problem is not limited to a binary decision variable formulation in which the order in which wells are visited are determined randomly for each ant each iteration. Moreover, an ACO algorithm may be developed to solve both the spatial and temporal LTM optimization problem in which the sampling locations and frequency are optimized.

References

- [1] Abbaspour, K.C., Schulin R., and van Genuchten M.Th., "Estimating unsaturated soil hydraulic parameters using ant colony optimization," *Advances in Water Resources*, 24, 827-841, 2001.
- [2] Aziz, J., Ling, M, and Rifai, H., "MAROS: A Decision Support System for Optimizing Monitoring Plans," *Ground Water*, 41(3), 355-367, 2003.
- [3] Bonabeau, E., Dorigo, M., and Theraulaz, G., "Inspiration for optimization from social insect *Nature*, 406, 39-42, 6 July 2000.
- [4] Bonabeau, E., Dorigo, M., and Theraulaz, G., *Swarm Intelligence, from Natural to Artificial Systems*. New York: Oxford University Press, 1999.
- [5] Cameron, K., and Hunter, P., "Optimization of LTM networks using GTS: Statistical approaches to spatial and temporal redundancy," Air Force Center for Environmental Excellence, Brooks AFB, TX, 2000.

- [6] Cieniawski, S.E., Eheart, J.W., and Ranjithan, S., "Using genetic algorithms to solve a multiobjective groundwater monitoring problem," *Water Resources Research*, 31(2), 399-409, 1995.
- [7] Doerner, K.F., Gutjahr, W.J., Hartl, R.F., Strauss, C. and Stummer, C., "Pareto ant colony optimization with ILP preprocessing in multiobjective project portfolio selection," *European Journal of Operational Research*, to be published. doi:10.1016/j.ejor.2004.09.009
- [8] Dorigo, M., "Optimization, Learning and Natural Algorithms" (in Italian), Ph.D. thesis, DEI, Politecnico di Milano, Italy. 140 pp., 1992.
- [9] Gutjahr, W.J., "A graph-based ant system and its convergence," *Future Generation Computer Systems*, 16, 873-888, 2000.
- [10] Goss, S., Aron, S., Deneubourg, J.-L., and Pasteels, J.M., "Self-organized shortcuts in the Argentine ant," *Naturwissenschaften*, 76, 579-581, 1989.
- [11] Gravel, M., Price, W.L., and Gagné, C., "Scheduling continuous casting of aluminum using a multiple objective ant colony optimization metaheuristic," *European Journal of Operational Research*, 143(1), 218-229, 2002.
- [12] Hudak, P.F., "Effective contaminant detection networks in uncertain groundwater flow fields," *Waste Management*, 21, 309-312, 2001.
- [13] Ippolito, M.G., Sanseverino, E.R., and Vuinovich, F., "Multiobjective ant colony search algorithm optimal electrical distribution system planning," in *Proceedings of the Congress on Evolutionary Computation*, CEC2004, IEEE, Portland, OR, 2004, Vol. 2, 1924-1931.
- [14] Jones, N.L., Davis, R.J., and Sabbah, W., "Comparison of Three-Dimensional Interpolation Techniques for Plume Characterization," *Ground Water*, 41(4), 411-419, 2002.
- [15] Maier, H.R., Simpson, A.R., Zecchin, A.C., Foong, W.K., Phang, K.Y., Seah, H.Y., and Tan, C.L., "Ant Colony Optimization for Design of Water Distribution Systems," *Journal of Water Resources Planning and Management*, 129 (3), 200-209, 2003.
- [16] Nunes, L.M., Cunha, M.C., and Ribeiro, L., "Groundwater monitoring network optimization with redundancy reduction," *Journal of Water Resource Planning and Management*, 130(1), 33-43, 2004.
- [17] Reed, P., Minsker, B., and Valocchi, A., "Cost-effective long-term groundwater monitoring design using a genetic algorithm & global mass interpolation," *Water Resources Research*, 36(12), 3731-3741, 2000.
- [18] Reed, P., Ellsworth, T., and Minsker, B., "Spatial interpolation methods for nonstationary plume data," *Ground Water*, 42(2), 190-202, 2004.
- [19] Ridley, M. and MacQueen, D., "Sampling Plan Optimization: A Data Review and Sampling Frequency Evaluation Process," *Groundwater Monitoring and Remediation*, 1, 74-80, 2004.
- [20] U.S. Department of Energy (DOE), "A Report to Congress on Long-term Stewardship: Volume I," Office of Environmental Management, Office of Long-Term Stewardship, United States Department of Energy, Washington, D.C., DOE/EM-0563, 2001. Available: <http://www.fernaldcab.org/FRReport/keydocuments/OtherDOE1.pdf>
- [21] U.S. Environmental Protection Agency, "Demonstration of Two Long-Term Groundwater Monitoring Optimization Approaches," Office of Solid Waste and Emergency Response, EPA 542-R-04-001a, 2004. Available: <http://www.epa.gov/tio/download/char/542-r-04-001b.pdf>
- [22] U.S. Navy, "Guide to Optimal Groundwater Monitoring," 2000. Available: http://enviro.nfesc.navy.mil/erb/erb_a/support/wrk_grp/raoltm/case_studies/Int_Final_Guide.pdf
- [23] Wegley, C., Eusuff, M., and Lansey, K., "Determining pump operations using particle swarm optimization," in *Proceedings of the Joint Conference on Water Resources Engineering and Water Resources Planning and Management*, ASCE, Minneapolis, MN, 2000, CD-rom.
- [24] Williams, G., "Chapter 2: Introduction," in *Long-Term Groundwater Monitoring, the State of the Art*, B. Minsker *et al.*, Eds., Reston, VA: American Society of Civil Engineers, 2000, pp. 1-4.
- [25] Wu, J., Zheng, C, and Chien, C.C., "Cost-effective sampling network design for contaminant plume monitoring under general hydrogeological conditions," *Journal of Contaminant Hydrology*, 77, 41-65, 2005.

Author Biographies

Yuanhai Li is a Doctoral candidate in the Department of Civil and Environmental Engineering at Florida State University in Tallahassee, Florida, USA. He earned his M.S. degree from Shantou University in Shantou, China, and B.S. degree from Tsinghua University in Beijing, China, both in environmental engineering.

Amy B. Chan Hilton is an Assistant Professor of Environmental Engineering at Florida State University in Tallahassee, Florida, USA. She earned her Ph.D. degree from the University of Virginia in Charlottesville, Virginia, USA, and M.S. degree and two B.S. degrees from the Massachusetts Institute of Technology (MIT) in Cambridge, MA, USA, all in civil and environmental engineering. Dr. Chan Hilton has published numerous papers on the development and application of computational intelligence, with a focus on evolutionary computation, for environmental engineering and groundwater management problems.