Optimal Groundwater Sampling Network Design through Ant Colony Optimization

Amy B. Chan Hilton, Ph.D.
Florida A&M University – Florida State University
College of Engineering
Civil and Environmental Engineering
2525 Pottdsdamer St., Tallahassee, FL 32310 USA
850.410.6121
abchan@eng.fsu.edu

Yuanhai Li
Florida A&M University – Florida State University
College of Engineering
Civil and Environmental Engineering
2525 Pottdsdamer St., Tallahassee, FL 32310 USA
850.410.6204
yuanhai@eng.fsu.edu

ABSTRACT
Groundwater long-term monitoring (LTM) is required to assess the performance of groundwater remediation and human being health risk at post-closure sites where groundwater contaminants are still present. The large number of sampling locations, number of constituents to be monitored, and the frequency of the sampling make the LTM costly, especially since LTM may be required over several decades. An optimization algorithm based on the ant colony optimization (ACO) paradigm for solving the traveling salesman problem (TSP) is proposed to reduce the number of monitoring wells while minimizing the overall data loss due to fewer sampling locations. The ACO method is inspired by the ability of ant colony to identify the shortest route between their nest and a food source. Ants depositing pheromones along their paths act as a form of indirect communication. The developed ACO-LTM algorithm is applied to a field site with an existing 30-well LTM network. Optimal LTM networks with 27 to 21 wells, which represent a 10% to 30% reduction in sampling locations, resulted in overall data losses ranging from 0.383 to 1.74. Results from developed ACO-LTM algorithm provide a proof-of-concept for the application of the general ACO analogy to the groundwater LTM sampling location optimization problem.

Categories and Subject Descriptors
J.2 [**Computer Applications**]: Physical Sciences and Engineering
– *Earth and atmospheric sciences, Engineering.*

General Terms
Algorithms, Management, Design, Economics.

Keywords
Optimization, Ant Colony Optimization, Sampling Network, Groundwater Monitoring.

1. INTRODUCTION
Long-term monitoring (LTM) of contaminated groundwater sites is an increasingly important issue in environmental remediation. LTM has become more important in recent years as active remediation concludes and the use of monitored natural attenuation has increased. 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 given the large number of sampling locations, and number of constituents monitored at a given site. For example, the U.S. Department of Energy (DOE) estimates that for sites where DOE has been mandated to conduct long term stewardship, total costs may be up to $100 million per year [17]. Thus the optimization of LTM networks may provide significant cost savings.

The objective of this work is to demonstrate a proof-of-concept of the application of the ant colony optimization (ACO) paradigm to solving the long-term groundwater monitoring sampling network optimization problem. Background information on LTM optimization and ACO are briefly described. Then the specific LTM problem addressed in this work is presented. An ACO algorithm developed to solve this LTM optimization problem is described. Results of the application of the developed ACO-LTM algorithm to a field site are presented and discussed.

2. BACKGROUND
2.1 Long-term Groundwater Monitoring
The overall goal of LTM optimization is to reduce the monitoring costs while still capturing sufficient information about the contaminant plume. 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.

In previous works, approaches for LTM network optimization include heuristic decision support tools and mathematical optimization. These approaches are combined with numerical groundwater flow and contaminant transport simulation models, estimation methods, and/or statistical analysis to predict or interpolate groundwater contaminant concentrations. Through these estimations, monitoring wells may be eliminated from the LTM network.

Heuristic decision support tools identify improvements to an existing LTM network but do not use mathematical optimization techniques. For example, Azia et al. [2] developed the monitoring

and remediation optimization system (MAROS), which reduces the number of spatial sampling locations using a set of heuristics with the Delaunay interpolation method. Cameron et al. [5] proposed the Geostatistical Temporal/ Spatial (GTS) Optimization algorithm, which is a site specific statistical method for reducing large monitoring networks. GTS uses kriging, which is a geostatistical interpolation method. However, since both MAROS and GTS are decision support tools, they use manual iterative procedures rather than automatic optimization, and therefore no global optimal search and sensitivity analysis under different constraints are included.

In general, LTM optimization is a non-linear combinatorial problem and therefore well-suited for heuristic optimization methods. For example, stochastic search methods including genetic algorithms (GAs) and simulated annealing have been used to solve LTM problems. Cieniawski et al. [7] optimized ground-water monitoring networks using GAs combined with Monte Carlo simulation. Nunes et al. [14] used simulated annealing with statistical methods to reduce temporal redundancy and increase spatial accuracy of LTM networks. Some simulation-optimization works also included geostatistical interpolation methods. Reed et al. [15] optimized sampling networks using inverse distance weighting (IDW) and ordinary kriging with GAs and simulation models. Wu et al. [20] improved the work of [15] by introducing new constraints with spatial moment to increase the accuracy of interpolation estimates. The problem with using numerical simulation models is that it often is difficult and time-consuming to calibrate model parameters. Since the evaluation of the objective function and constraints are dependent the predictions of the simulation models, uncertainty in simulation model parameters is propagated to the optimal solution, leading potential reliability issues.

2.2 Ant Colony Optimization

Ant colony optimization (ACO) is an evolutionary optimization method based on ants' collective problem-solving ability. This global search method is inspired by the ability of an ant colony to identify the shortest route between their nest and a food source [3]. A single ant randomly chooses one path to visit from all the possible routes from the nest to food source. Individual ants contribute information to the colony by dropping chemical markers, or pheromones as they traverse a path. In addition, the pheromone decreases over time at a given evaporation rate. Thus a shorter path means higher pheromone density, and may make it more likely to be chosen by other ants [3]. Through indirect communication to other ants via pheromones and foraging behavior, a colony of ants can establish the shortest path between the nest and the food source over time. This shortest path represents the global optimal solution, and all the possible paths represent feasible region.

The first ant colony simulation algorithm was developed by Dorigo [8] to solve the classic traveling salesman problem (TSP), which is an NP-hard problem. In the TSP, the goal is to obtain a shortest path that connects all the cities while visiting each city only once. Bonabeau et al. [4] compared ACO with other stochastic searching algorithms, including the GAs, 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. Ant colony simulation 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].

Gutjahr [9] proved that under certain conditions, solutions from ant-based optimization converge with a probability. More recently, ACO algorithms have been applied to solve a wide range of engineering and science problems such as random number generators [10], autonomous decentralized shop floor routing [6], bandwidth minimization problem in a large scale power transmissions system [12], redundancy apportionment problem in electrical and mechanical systems [21], and capacitated minimum spanning tree problems, which is applied to telecommunication networks [16]. To date, ACO has not been applied to groundwater management and design optimization problems, with the exception of Li et al. [11], which developed an ACO for LTM based on the approached used in [1]. The only related works include Maier et al. [13], which used ACO to optimize water distribution systems designs, Abbaspour et al. [1], which used ACO to solve an inverse modeling problem, unsaturated soil parameters, and Wegley et al. [19], which used particle swarm optimization to determine pump speeds to minimize the total costs in water distribution systems.

3. LTM OPTIMIZATION

An optimization model is formulated to describe the LTM problem studied in this work. The goal is to identify the optimal reduced set of monitoring wells from an existing monitoring network that retains sufficient measured data such that the removed wells may be estimated with minimal error. The objective of the LTM optimization problem is to minimize the overall data loss of the reduced network (Equation 1) given a fixed number of monitoring wells (Equation 2).

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

$$\text{s.t.} \qquad s = S_{max} \qquad (2)$$

where m = the number of removed monitoring wells (sampling locations), C_i = the measured concentration of removed well i, $C_{est,i}$ = the estimated concentration of removed well i based on the remaining wells, s = the number of remaining wells, and S_{max} = desired number of remaining wells, which may be predetermined according to available budget.

In this work, concentrations are estimated using IDW interpolation. To estimate a value at any unsampled location, IDW uses the measured values from surrounding locations (Equations 3-4). IDW assumes that each measured point has a local influence that diminishes with distance between unsampled and measured locations. Points closer to the estimated location have more influence (*i.e.*, weight) to the estimation location than those farther away.

$$C_0 = \frac{\sum_{i=1}^{n} \dfrac{C_i}{d_{i0}^p}}{\sum_{i=1}^{n} \dfrac{1}{d_{i0}^p}} \qquad (3)$$

where $\qquad d_{io} = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2} \; ; \qquad (4)$

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

In particular, this optimization formulation is applied to a field site in the Upper Aquifer at the Fort Lewis Logistics Center in Pierce County, Washington, USA [18]. The existing LTM network consists of 30 monitoring wells (Figure 1). The contaminant of concern is trichloroethylene (TCE), which was used as a degreasing agent at the site until the 1970s. 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 (Figure 1).
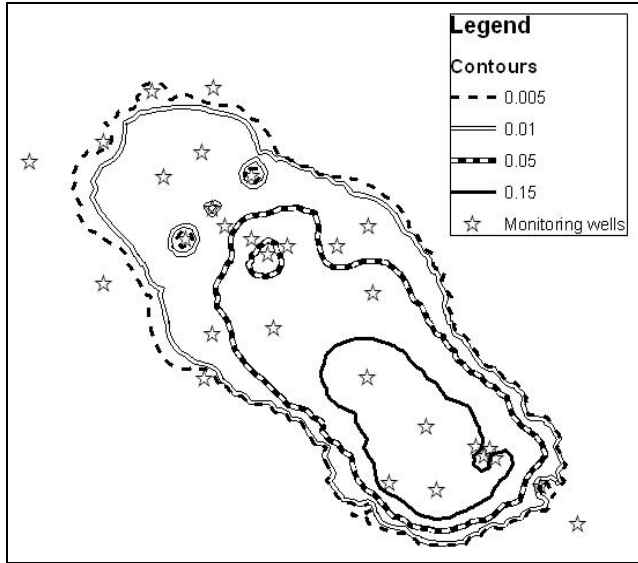

Figure 1. Concentration contours of TCE (in mg/L) based on data from the original LTM network of 30 monitoring wells.

4. ACO DEVELOPMENT
An ACO algorithm for solving the groundwater LTM spatial optimization problem is presented in this work. The developed ACO-LTM algorithm is analogous to the ACO paradigm for the classic traveling salesman problem (TSP) developed in [4]. In the ACO-TSP paradigm, each ant at city i is an agent who places pheromone on a visited path, and then chooses to visit the next city j with a probability that is a function of the distance between cities i and j (dij) and the pheromone density on this path. In the TSP, the distance between cities and the order in which they are visited are significant and affect the solution quality (*i.e.*, total

path length). However in the ACO optimization problem, the distance between wells and the order in which they are visited are not explicitly relevant to the objective function and constraints. Nevertheless, distance between wells is significant in concentration estimation (Equations 3-4). In the ACO-LTM paradigm, ants select wells to include in the reduced LTM network based on the relative importance of a monitoring well to its neighboring wells. An ant at well i may choose from multiple wells not already selected to visit at each step along its path (Figure 2) based on a local error resulting from removing next well j (η_{ij}) and the pheromone density along path ij. A summary of the developed ACO-LTM algorithm is described below.

1. An ant's starting point (*i.e.*, first selected well) is randomly chosen. The order in which an ant visits the wells is stochastically determined, depending on the pheromone density and relative error value. Each individual ant will visit only the specified number of wells (S_{max}), which become the remaining wells of the reduced LTM network.

2. The set of neighboring wells L around an ant's current location are identified. These neighbors are candidate wells for an ant to visit next. Only wells not yet visited are included in the candidate list (Figure 2). The number of neighboring wells considered increases over the iterations to prevent an ant from being limited in path choices.
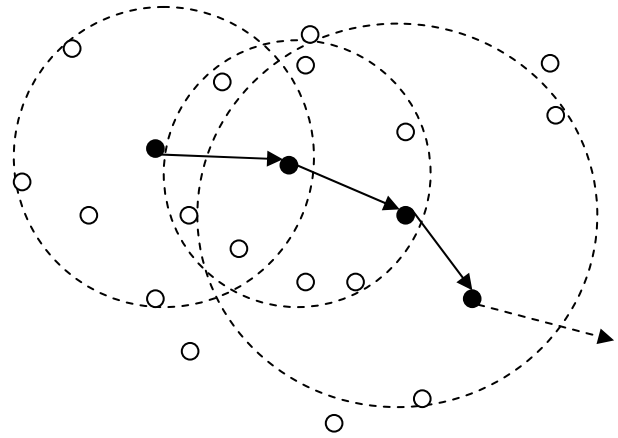

Figure 2. Representation of the ACO paradigm for the LTM network optimization problem. An ant's path of identifies the selected wells (solid circles) from among a subset of 8 neighboring monitoring wells.

3. In this step, the ant decides which well will be selected among the candidates. A set of relative errors η_{ij} is calculated for current well i and its neighbors. Calculate the error one by one as follows. Suppose well j is one well in the candidate list and well i is the center well where the ant is currently located (Figure 3). The concentration at well i is estimated through IDW interpolation several times; each time, one of the candidate wells is excluded from the interpolation. For example, $C_{est,ij}$ is the estimated concentration of center well i when well j is excluded. The relative error from eliminating candidate well j when at well i (η_{ij}) is characterized by

$$\eta_{ij} = \frac{\left|C_{est,ij} - C_i\right|}{\min(C_{est,ij}, C_i)} \quad , \qquad (5)$$

where C_i = measured concentration of well i. A high η_{ij} value indicates that the estimated error without well j is very high, which implies that candidate well j is important to the well i.
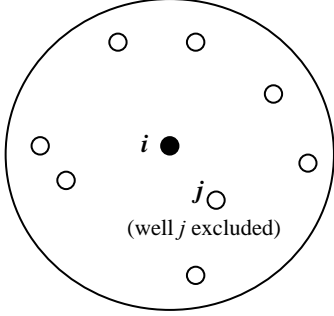

Figure 3. Neighboring wells around an ant's current location at well i. Relative error η_{ij} at well i due to excluding information from well j is calculated to quantify the relative importance of well j to well i.

4. The next monitoring well j in the reduced LTM network is selected from among the candidate list stochastically based on the above relative error (η_{ij}) and pheromone deposited in the individual path ij (τ_{ij}). The probability well j is chosen when an ant currently is at well i (p_{ij}) is defined by

$$p_{ij} = \frac{\left(\tau_{ij}\right)^{\alpha} \left(\eta_{ij}\right)^{\beta}}{\sum\limits_{l \in L} \left(\tau_{il}\right)^{\alpha} \left(\eta_{il}\right)^{\beta}} \qquad (6)$$

where α and b are parameters ($\alpha = 1$ and $\beta = 1$).

5. After a well among the candidates list is selected, the well is marked as a remaining well. Then this well becomes the current well for this ant.

6. Repeat steps 2 through 5, until the number of visited wells is equal to the predetermined number of remaining wells (Equation 2).

7. After an ant has visited the prespecified number of wells (S_{max}), the overall data loss of the reduced LTM network is determined. The data loss is due to estimating the concentrations at the eliminated wells based on information from the remaining wells. The overall goal of the LTM optimization is to reduce this overall loss, which is quantified by the root mean square error (*RMSE*) of the estimated concentrations of removed wells. The *RMSE* is given by

$$RMSE = \sqrt{\frac{\sum\limits_{i=1}^{m}\left(\dfrac{C_{est,i} - C_i}{\min(C_{est,i}, C_i)}\right)^2}{m}} \qquad (7)$$

where m = number of removed wells, C_i = measured concentration of well i, and $C_{est,i}$ = estimated concentration of well i based on data from the remaining wells.

8. The pheromone for each segment τ_{ij} of an ant's path is updated for iteration $t+1$ by the following rule:

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \Delta\tau_{ij} + e\Delta\tau^{e}_{ij} \qquad (8)$$

where ρ = pheromone evaporation rate; $\tau_{ij}(t)$ = the pheromone density for path ij during current iteration t; and $\Delta\tau = Q/RMSE$, in which Q is a parameter ($Q = 100$), and the goal is to minimize the *RMSE* value. The idea here is that after an ant finishes a tour, if the total data loss *RMSE* value is low, then the pheromone density of the path of the ant will be high. Lower total data loss means higher density of pheromone, which will attract more ants to follow the same path. The term $e\Delta\tau^{e}$ includes elitism to the ACO search, where e is a parameter that is the number of elitist ants in an iteration; and $\Delta\tau^{e}$ is the pheromone density of the recent elitist ant. The elitist ant is the ant which has obtained the best solution so far. An old elitist ant's pheromone will be replaced by an ant with a lower *RMSE* value.

9. Return to step 2 to implement the next iteration. The procedure terminates after a specified number of iterations.

In this work, each ant colony is comprised of 30 ants, and the ACO search is continued over 100 iterations. By solving the LTM optimization problem with varying desired number of remaining wells (S_{max}) using this ACO-LTM algorithm, different optimal reduced LTM networks are identified.

5. RESULTS

The results of solving the LTM optimization problem described by Equations 1-2 with different desired number of remaining wells (S_{max}) using the developed ACO-LTM algorithm are shown in Figures 4 through 6. In particular, optimal LTM networks with 27, 23, and 21 remaining wells are presented, which represents 10%, 23%, and 30% reductions in monitoring wells, respectively. Since the majority of LTM costs results from sampling the reduction in the number of monitoring wells reflects a similar reduction in monitoring costs. The results presented here are from replicate runs of the ACO-LTM. As expected the resulting *RMSE* increases as the number of remaining wells in the LTM network decreases. The resulting overall *RMSE* for 27, 23, and 21 remaining wells are 0.383, 0.976, and 1.745, respectively. Although the locations of the removed wells for the cases with 21 and 23 remaining wells are similar, the resulting overall *RMSE* values are quite different. The *RMSE* increased from 0.976 to 1.745 (79% increase) when the number of remaining wells decreased from 23 to 21 (8.7% decrease). This indicates that a non-linear relationship between number of remaining wells and *RMSE* values exists, with a greater rate of *RMSE* increase when at lower numbers of remaining wells (Figures 5 and 6).

Additionally, the solutions identified by the ACO are evaluated by comparing the resulting concentration contours (Figures 4 through 6) with the original contours from the existing LTM network of 30 wells (Figure 1) at several isopleths (for example 0.05 and 0.15 mg/L). The contours for the 27-well LTM network (Figure 4) are almost identical to the contours based on the data of

30 wells (Figure 1), with the exception of the small region near the leftmost well selected for removal. The most significant difference is for the 0.15 mg/L isopleths based on the 21-well and 30-well LTM network (Figures 1 and 6).
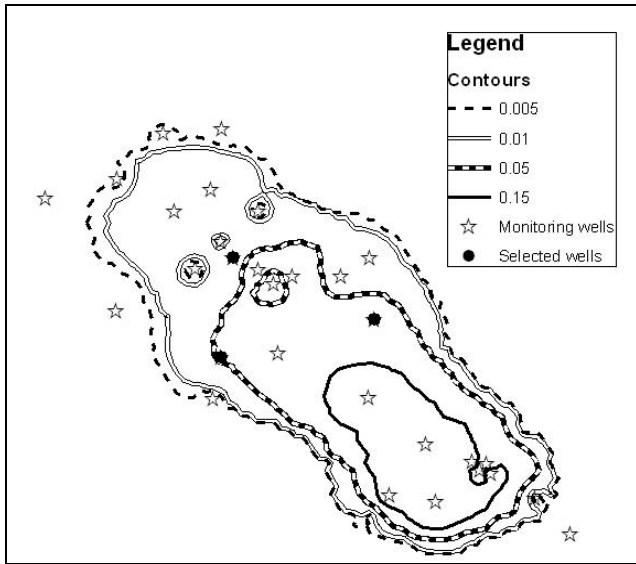

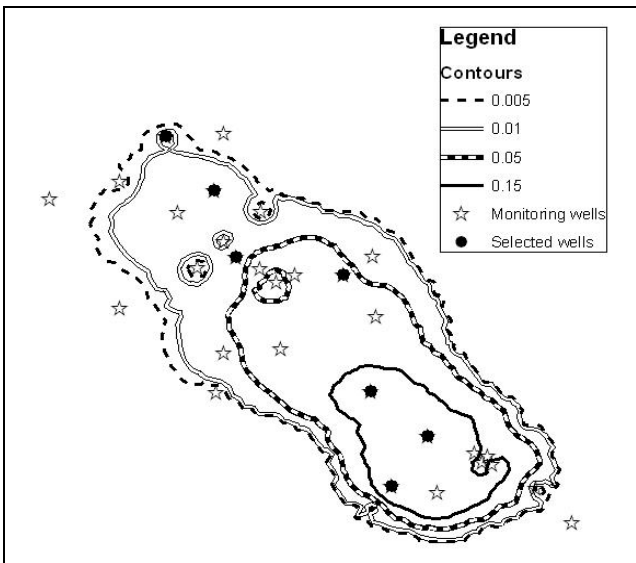**Figure 4. Concentration contours (mg/L) based on data from the optimal LTM network of 27 wells (*RMSE* = 0.383).**

Figure 5. Concentration contours (mg/L) based on data from the optimal LTM network of 23 wells (*RMSE* = 0.976).

6. CONCLUSIONS

This paper presents an ACO-LTM algorithm developed based on the ACO paradigm for solving the TSP. Based on similarities between the TSP and LTM problems, analogies are made. An ant acts as an agent identifying the most significant sampling locations as it travels around the existing LTM network. An ant's path is based on the local importance of a well on estimating the contaminant concentration at a neighboring well. The developed

ACO-LTM algorithm is applied to a field site with an existing 30-well LTM network. Optimal LTM networks with 27 to 21 wells, which represent a 10% to 30% reduction in sampling locations, resulted in overall data losses ranging from 0.383 to 1.74. Results from developed ACO-LTM algorithm provide a proof-of-concept for the application of the general ACO analogy to the groundwater LTM sampling location optimization problem. Future work includes the implementation of additional features to the ACO search algorithm to improve the search efficiency and solution quality. Additionally, the ACO-LTM algorithm will be applied to additional field sites and expanded to include temporal optimization of LTM.
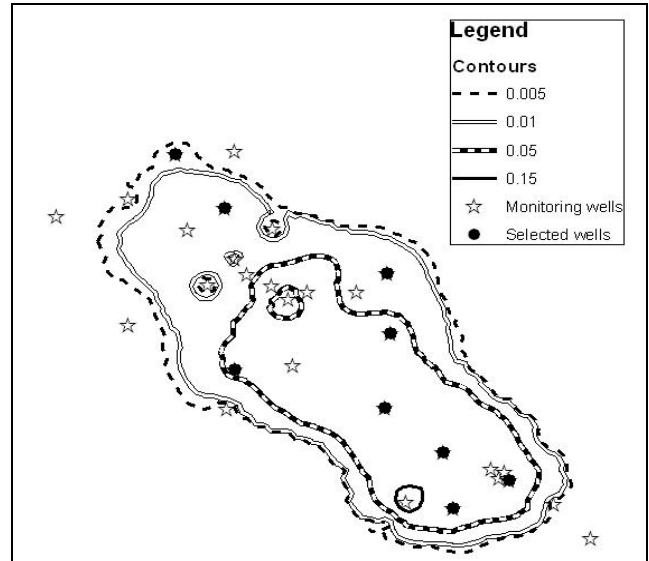

Figure 6. Concentration contours (mg/L) based on data from the optimal LTM network of 21 wells (*RMSE* = 1.745).

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Abbaspour, K.C., Schulin R., and van Genuchten M.Th. (2001). Estimating unsaturated soil hydraulic parameters using ant colony optimization. *Advances in Water Resources*, 24, 827-841.

[2] Aziz, J., Ling, M, and Rifai, H.S. (2002). MAROS: A decision support system for optimizing monitoring plans. *Ground Water*, 41(3), 355-367.

[3] Bonabeau, E., Dorigo, M., and Theraulaz, G., (2000). Inspiration for optimization from social insect behavior. *Nature*, 406, 6 July 2000, 39-42.

[4] Bonabeau, E., Dorigo, M., and Theraulaz, G. (1999). *Swarm Intelligence, from Natural to Artificial Systems*. Oxford University Press, New York.

[5] Cameron, K., and Hunter, P. (2000). Optimization of LTM networks using GTS: Statistical approaches to spatial and

temporal redundancy. Air Force Center for Environmental Excellence, Brooks AFB, TX.

[6] Cicirello, V.A. and Smith, S.F. (2001) Ant colony control for autonomous decentralized shop floor routing. In *ISADS-2001: International Symposium on Autonomous Decentralized Systems* (March 2001), 383-390. IEEE Computer Society Press. Available at http://www.cs.drexel.edu/~cicirello/cicirello-isads2001.pdf

[7] Cieniawski, S.E., Eheart, J.W., and Ranjithan, S. (1995). Using genetic algorithms to solve a multiobjective groundwater monitoring problem. *Water Resources Research*, 31(2), 399-409.

[8] Dorigo, M. (1992). *Optimization, Learning and Natural Algorithms* (in Italian), Ph.D. thesis, DEI, Politecnico di Milano, Italy. 140 pp.

[9] Gutjahr, W.J. (2000) A Graph-based ant system and its convergence. *Future Generation Computer Systems*, 16, 873-888.

[10] Isaacs, J.C., Watkins, R. and Foo, S.Y. (2002). Evolving ant colony systems in hardware for random number generation. In *Proceedings of the 2002 IEEE World Congress on Computational Intelligence* (WCCI 2002) (Honolulu, Hawaii, May 12-17, 2002), IEEE.

[11] Li, Y., Chan Hilton, A.B., and Tong, L. (2004). Development of ant colony optimization for long-term groundwater monitoring. In *Proceedings of the ASCE EWRI 2004 World Water & Environmental Resources Congress*, (Salt Lake City, Utah, June 27-July 1, 2004). ASCE, Reston, VA, CD-rom, 10 pp.

[12] Lim, A., and Lin, J., Rodrigues, B. and Xiao, F. (2005) Ant colony optimization with hill climbing for the bandwidth minimization problem. *Applied Soft Computing*, in press, doi:10.1016/j.asoc.2005.01.001

[13] Maier, H.R. et al. (2003) Ant colony optimization for design of water distribution systems. *Journal of Water Resources Planning and Management*, 129 (3), 200-209.

[14] Nunes, L.M., Cunha, M.C., and Ribeiro, L. (2004). Groundwater monitoring network optimization with redundancy reduction. *Journal of Water Resource Planning and Management*, 130(1), 33-43.

[15] Reed, P., Minsker, B., and Valocchi., A.J. (2000). Cost-effective long-term groundwater monitoring design using a genetic algorithm and global mass interpolation. *Water Resources Research*, 36(12), 3731-3741.

[16] Reimann, M., and Laumanns, M. (2005) Savings based ant colony optimization for the capacitated minimum spanning tree problem. *Computers & Operations Research*, in press, doi:10.1016/j.cor.2004.11.019

[17] U.S. Department of Energy (2001). *A report to Congress on long-term stewardship: Volume I.* Office of Environmental Management, Office of Long-Term Stewardship, United Sates Department of Energy, Washington, D.C., DOE/EM-0563. Available at http://www.fernaldcab.org/FRReport/keydocuments/OtherDOE1.pdf

[18] U.S. Environmental Protection Agency (2004). *Demonstration of Two Long-Term Groundwater Monitoring Optimization Approaches.* Available at http://www.epa.gov/tio/download/char/542-r-04-001b.pdf

[19] Wegley, C., Eusuff, M., and Lansey, K. (2000). Determining pump operations using particle swarm optimization. In *Proceedings of the Joint Conference on Water Resources Engineering and Water Resources Planning and Management*, (Minneapolis, Minnesota, July 30-Aug 2, 2000). ASCE, Reston, VA.

[20] Wu, J., Zheng, C, and Chien, C.C. (2005). Cost-effective sampling network design for contaminant plume monitoring under general hydrogeological conditions. *Journal of Contaminant Hydrology*, 77, 41-65.

[21] Zhao, J., Liu, Z., and Dao, M. (2005). Reliability optimization using multiobjective ant colony system approaches. *Reliability Engineering and System Safety*, in press, doi:10.1016/j.ress.2004.11.011