Interactive Genetic Algorithm Framework for Long Term Groundwater Monitoring Design

Meghna Babbar[1], Barbara Minsker[2], and Hideyuki Takagi[3]

[1] Department of Civil and Environmental Engineering, University of Illinois at Urbana Champaign, Urbana, USA, PH (217) 333-6979, email: mbabbar@uiuc.edu
[2] Department of Civil and Environmental Engineering, University of Illinois at Urbana Champaign, Urbana, USA, PH (217) 333-9017, email: minsker@uiuc.edu
[3] Kyushu University, Fukuoka, Japan, email: takagi@design.kyushu-u.ac.jp

Abstract

In standard optimization approaches for water resources management problems, the designer is responsible for correctly formulating mathematical equations to describe the system objectives and constraints. The search for optimal or near-optimal solutions is made under the assumption that these formulated objectives and constraints completely describe the system. However, in real systems that is often not true. Many qualitative criteria can be integral parts of the design analysis that numerically based algorithms cannot capture. For such problems, designer interaction with the search algorithm can help the search be more creative and inclusive. Genetic algorithms are ideally suited for incorporating such interaction in their usual search process, and can successfully evolve solutions that are optimal with respect to both qualitative and quantitative objectives.

Under an interactive approach, the genetic algorithm performs the usual operations of selection, crossover, and mutation, but the user evaluates the suitability ('fitness') of candidate solutions, enabling objectives that cannot be quantified to be included in the search process. In multi-objective problems, where quantitative objectives can be as important as qualitative fitness of designs, analysis of designs is done based on tradeoff fronts made from both quantitative and qualitative information. In this paper, we demonstrate the use of interactive genetic algorithms for long term groundwater monitoring problems, which have multiple numerical and subjective objectives. We also analyze the effects on the optimal monitoring designs of using an interactive optimization approach instead of more traditional numerical optimization approaches.

Introduction

Environmental decision support systems are expert systems that are used in strategy and goal planning for active pollution management, site monitoring and restoration. For most water resources applications that use such expert systems, knowledge representation of the problems can be a very non-trivial issue. Numerical and analytical models are extensively used to describe the data, processes, and responses of an environmental system. The artificial intelligence tools used in the expert systems (e.g., genetic algorithms) depend on the ability of such models to reliably explain the physics of the complex phenomena, the missing information, and the uncertainties associated with the process. However, these systems often fail to consider important qualitative and incomputable phenomena related to the management problem, and are also not very supportive in involving site owners, regulators, and other stakeholders in the evaluation and optimization process. Such limitations can impair the creditability and acceptance of these systems. The work described in this paper examines a novel approach that aims at bridging the gaps in problem analyses between important qualitative and quantitative aspects. A new expert system is proposed that takes information from the 'calculable' models to obtain the quantitative information and provides a framework for users to interact with the system and impart qualitative knowledge to the optimization process. The optimization model searches for management plans that are optimal in both quantitative and qualitative objectives using expertise gathered from the users.

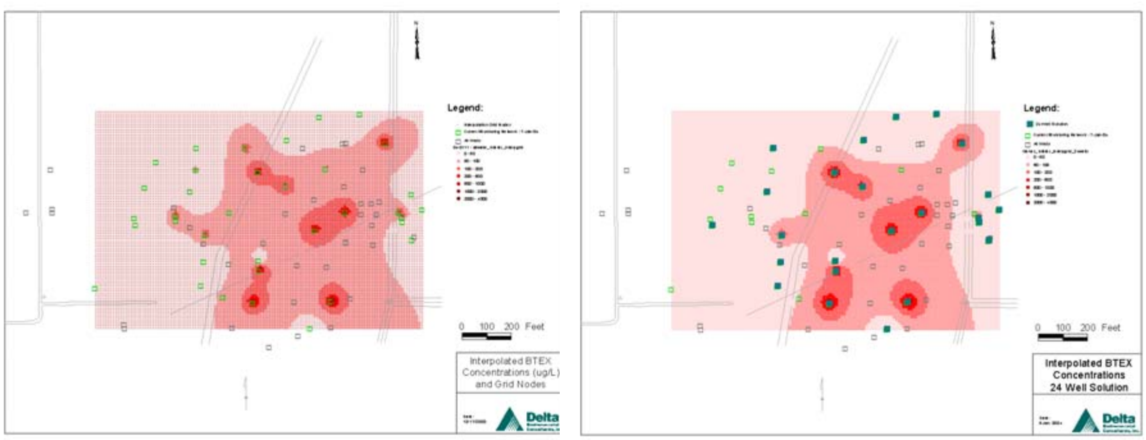
From previous research, it has been shown that Genetic Algorithms (GAs) are flexible and powerful optimization tools for complex water resources problems that have single and multiple objectives (e.g., Reed et al, 2001, Hilton et al, 2000, Ritzel et al., 1994). GAs can also consider uncertainty in identifying optimal solutions (Smalley et al., 2000, Gopalakrishnan et al., 2003, Hilton et al., 2000, Singh et al., 2003). Our new expert system combines these advantages of GAs with other artificial intelligence algorithms (also called machine learning algorithms) to interpret and absorb the qualitative and quantitative information into the system. Since direct interaction with the end user is an important component of this process, these types of GAs are called Interactive Genetic Algorithms (IGAs). This paper presents an IGA approach to solving a long-term groundwater monitoring case study. The rest of the paper discusses important implementation issues related to this type of system.

Groundwater Monitoring Case Study

The case study for this paper is a British Petroleum site where the groundwater has been contaminated with benzene, toluene, ethylbenzene, and xylene (BTEX) over a period of 14 years. Active remediation has been completed in recent years and the site has reached a stage where there is a need for long term monitoring. Data are currently available from about 36 wells in the region and the objective of the project is to remove any wells that are spatially redundant. A deterministic interpolation model was implemented to interpolate BTEX concentrations using inverse distance weighting.

The numerical objectives for this problem are to minimize the number of wells sampled and to minimize root mean square errors for the BTEX interpolation models (similar to Reed et al, 2001), where the errors are measured relative to sampling all 36 existing wells. The decision variables for the problem are installation flags (0/1) for all 36 wells. Hence, if a flag is 1 then the well at that location is sampled. One criterion for assessing the fitness of a monitoring design are based on how similar the overall contaminant spatial distribution looks when compared to the case when all wells are sampled. Another criterion for assessing a particular design is how well defined the leading edge of the plume (the direction in which the groundwater is flowing) remains after removing particular wells. Accurate prediction of concentrations on this boundary is important to ensure that the plume does not cross the site boundary onto public property. However, these subjective criteria are difficult to include within a traditional multiobjective GA framework that can incorporate only quantifiable objectives.

When optimization is done only on the basis of the two numerical objectives, we got a tradeoff of monitoring network designs that are optimal in terms of both BTEX overall interpolation error and cost of wells (or number of wells installed). Figure 1 compares the interpolated concentrations from the optimal 24-well solution with those interpolated from all 36 wells. The 24-well solution is less costly and has a reasonably low overall interpolation error (root mean square error). However, we can notice that even though the interpolation was more or less similar to the case when all 36 wells are installed, the front end of the plume is not as well defined (top left of the plume, which is also the direction of the groundwater flow). This can be an important criterion for decision making, since it will dictate how close the plume is to the boundary. Such issues cannot easily be accommodated in the traditional implementation of optimization algorithms, and are usually handled post-optimization. However, in the IGA system this kind of a subjective analysis is handled within the ongoing search process, which helps to ensure that the GA evolves solutions that are also favorable with respect to more subjective criteria.



<u>36 Well Interpolation</u> <u>24 Well Interpolation</u>

Figure 1. All wells (36 wells) interpolation Vs. 24 wells solution interpolation for BTEX.

Interactive Genetic Algorithm

Under an interactive approach, the GA performs the usual operations of selection, crossover, and mutation (Goldberg, 1989), but the user evaluates the suitability ('fitness') of candidate solutions, enabling objectives that cannot be quantified to be included in the search process. Takagi et al have applied this approach to the fields of geology, hearing aid fitting, lighting design, music generation, face image generation, etc (Takagi, 2001). However, most applications of IGAs to date have considered only qualitative objectives. Moreover, to control human fatigue from evaluating numerous potential solutions, the size of the populations used in IGAs has been kept to low numbers that the user can manage (e.g. 20 to 25). For water resources problems where quantitative objectives are as important as the qualitative criteria, large population sizes for the GA are important to efficiently search the complex numerical objective space. In this situation, while the quantitative fitness is evaluated for all individuals, only a selected few individuals are evaluated by the user for qualitative knowledge, so that user fatigue does not impair user feedback. Once a user evaluation is obtained, the algorithm uses these labeled individuals to predict the 'best possible' user evaluation for those individuals not explicitly ranked or classified by the human. Figure 2 shows the conceptual difference between the traditional GA and the IGA.

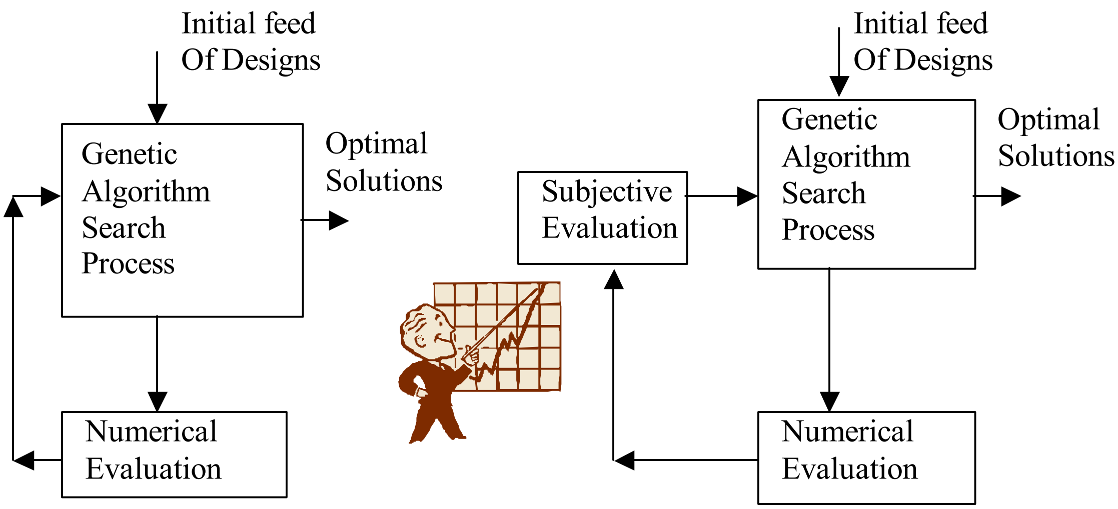


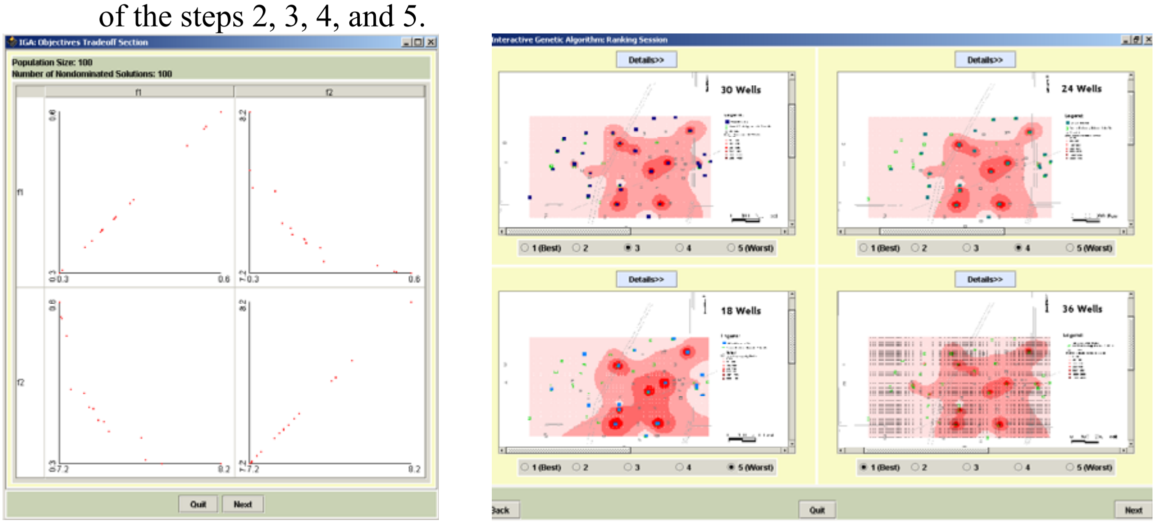
Figure 2. The traditional GA (left) and the IGA (right) frameworks.

Interactive Genetic Algorithm Experimental Framework. An IGA system is being built into the National Center of Supercomputing Applications (NCSA, Champaign, IL) automated learning system called 'Data to Knowledge' (or D2K) (Welge et al. 2003). The IGA system is a multiobjective approach, where the addition of human preferences adds another objective to the quantitative objectives. The algorithm can be outlined as follows:

1. The GA first performs standard operations (selection, crossover, and mutation) considering only the quantitative objectives for a few generations to find a set of initial solutions fit for ranking. These solutions might not be optimal, but are fair when judged based on their numerical fitness values.
2. The user interaction session then calls for user attention. The user initially views the entire population in objective space. For multiple numerical objectives, there will be multiple 2-D tradeoff frontiers (also called pareto fronts), as shown in Figure 3a. From this visualization, the user makes an initial selection of individuals that he/she might want to view for further analysis. This is done to decrease the number of solutions that the user must consider, so that human fatigue is minimized during the ranking session. Figure 3a shows an example of this interface.
3. The next visualization (ranking session) consists of viewing more detailed information (such as sampling locations or a visualization of spatial interpolation

errors) related to a small pool of selected individuals in detail and then ranking them based on different qualitative criteria. Figure 3b shows an example of this interface.

4. Once the ranking session is over, a classification model is trained on the ranked ('labeled') samples. The GA then uses the model to predict 'surrogate ranks' for other 'unlabeled' individuals in the population.
5. The GA then asks the users if they are satisfied with the current level of discovery or if they want to further analyze the solutions. If they are satisfied and wish to stop, the GA produces all recently found solutions in its output. Otherwise, the GA again starts its evolution loop. This time it uses both numerical objectives and qualitative objectives (by predicting labels from the earlier explicitly labeled individuals) to find new solutions. Since the number of new solutions encountered increases substantially as the optimization progresses, the set of unlabeled data will increase as well. Hence, the reliability of the classification of new data based on an initial small set of labeled data will decrease over time. At that point, based on some stopping criteria (e.g. when testing error of the supervised learning algorithm increases beyond a threshold value), the GA can initiate another interactive session to collect more labeled data sets. This would lead to repetition of the steps 2, 3, 4, and 5.



(a) **(b)**

Figure 3. (a) Selection Session and (b) Ranking Session.

The modification of the GA algorithm to incorporate interactivity is conceptually simple, but a number of important issues arise in designing and using such systems that must be considered. These issues, discussed in more detail in the following sections, are related to the visualization and user interfaces for interactive sessions, which act as bridges between the user and the GA; evaluation data (or 'user data') collected from the user through the interactive session; and user modeling that aims at understanding the user preferences and subjective analyses.

User Data Collection. While there can be well behaved decision rules for differentiating designs based on their numerical fitness, analysis of qualitative fitness can be much more challenging. Figure 4 explains this phenomenon by showing how design A could be better than design B in the numerical objective space based on the design parameters P1 and P2, but could be worse than design B in the psychological space based on some cognitive factors 'factor 1' and 'factor 2'. Also, every user can have his/her own cognitive perception about a subjective issue, which can affect his/her preference about how good a design is in the qualitative objective space. For example, a regulator may have very different perceptions from a site owner about the suitability of candidate sampling plans. This perception is also prone to noise due to environmental factors and the condition of the user at the time of ranking (discussed in more detail below). The number of users interacting with the system can also bias the search process. Hence user data collected will be a function of how many users contributed to it and under what circumstances the users made their decisions. It is also known that " … humans are poor judges of absolutes but very good judges of relative differences" (Stone et al., 1993). This unreliability when evaluating individual designs can be worsened when the user has gone through multiple designs already. At that point, the user's perception about the quality of a design will be influenced by the history of designs he/she has already evaluated. This adds an extra dimension of labor if the user has to keep referring to his/her past evaluations while examining new solutions.

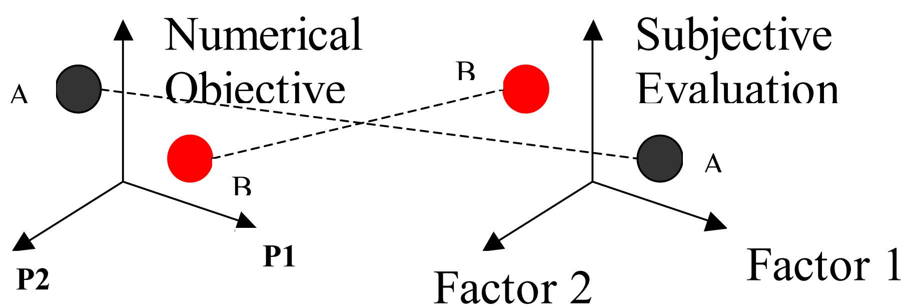


Figure 4. Dynamic of Qualitative and Quantitative Evaluation.

Communicating user preferences to the GA is usually done by letting the user rank the solutions based on whatever cognitive criteria he/she might have (Tagaki et al., 2001, Corney, 2002). Ranking allows solutions to be labeled and classified into groups according to user preference. This becomes an indirect way of understanding what 'kinds' of solutions are desirable to the user's cognitive perceptions. Ranking can be as coarse as labeling solutions merely 'good' or 'bad', or they can be more fine; e.g. a 5 rank classification would have a scale of solution quality with five levels progressing from worst to best (Takagi, 2001). Ranking can also be done to assess overall performance of a solution, or it could be done for multiple criteria that are relevant to the problem. These multiple criteria act as multiple qualitative objectives that the user is trying to consider in selecting the best solutions.

What rank a user gives to a solution, besides his/her personal proclivity, also depends on what kind of information about the solution has been presented to the user. Incomplete or unclear representation can adversely affect the user's choices. Visualization of solutions thus becomes an important factor, especially for water resources problems that deal with hydrologic and geologic attributes. The kind of information that the user might need to make a decision can be the numerical objective values, constraint violations, actual decision variable values, model results, geologic conditions (e.g., bore hole records), or any other information that affects the choice of decision variables. The user might also need a collection of comparative information for certain parameters to see how the solution performs with respect to other solutions created by the GA. It is usually better to first display the most important information about all of the solutions to the user, with a link to a 'details' page on the interactive session window (as shown in Figure 4b). This avoids too much cluttering of information on the same visual window. Also, the number of solutions shown to the user at one time on the screen should be not too large. Continually displaying the highest- (and perhaps lowest-) ranked solution in the session can help promote consistency in user rankings when more than one screen is required to rank all of the solutions.

User Data Analysis. Hand's (Hand D.J., 1998) definition of data analysis refers to obtaining information from the data so that certain questions about the problem can be answered successfully. Summarizing data or predictions of data in the presence or absence of an underlying theoretical explanation can do this. Hence correct interpretation of data collected through user evaluation is critical for the performance of the entire IGA system to answer the questions that the users have for their subjective criteria. The ability to predict by creating models of user preference, without knowing the underlying causes, can assist in avoiding the task of determining the relationship between solution ranks and various cognitive, physical, chemical, or numerical, etc. properties of the problem. Also, as discussed before, for larger population sizes the human explicitly evaluates only small sets of individuals. Hence, for the remaining individuals it becomes imperative to use surrogate models that can predict the 'most probable rank' for the unranked/unlabeled individuals.

User data sets (that get 'labeled' for qualitative criteria) are usually sparse, noisy and small compared to all the individuals evaluated for numerical objectives. Because of this there can be a tendency to create overly complex models than what the data set can support. Thus feature selection (Corney, 2002) can reduce the number of features used in the models to predict user preference, and avoid over-fitting. This is an issue that should be examined while creating supervised prediction models (discussed below) based on the information obtained from the small data sets.

Prediction of unlabeled individuals (i.e., those that were not evaluated by the user) based on the ranks of the labeled individuals can be considered a classification problem. The deterministic ranks assigned to labeled individuals represent their subjective quality, which otherwise cannot be numerically calculated. Hence, other unlabeled individuals that have features similar to these labeled individuals will have a high probability of being in the same deterministic rank class as that of the labeled individual. Standard supervised classification methods (e.g., weighted K nearest neighbor method, decision trees, etc), that use the information from the labeled data sets to predict the labels for the unlabeled data, can fall short on accuracy because of the small labeled data sets. It has been shown that semi-supervised algorithms work best for such data sets (Corney, 2002, Hand et al., 1998), where both labeled and unlabeled data are exploited for classification.

This way the data set size used for classification becomes large and the problems of feature selection become less of an issue. Other methods used for classifying small and noisy data sets are Bayesian methods. Bayesian belief networks have been successfully applied in the fields of robot guidance (Berler et al., 1997), software reliability assessment (Neil et al., 1996), data compression (Frey, 1998), fraud detection (Ezawa et al., 1995), and food product design (Corney, 2002). Corney used them to build a model of user preference and food product characteristics to come up with new ideas for product invention. Our initial investigation, which will be presented at the conference, will examine some of the simple supervised learning algorithms to build models of user preferences. Semi-supervised methods and Bayesian methods will be left as a task for future research.

Another issue that arises while handling data related to user preferences is that of outlier detection. User feedback can be inconsistent because of various environmental factors like misunderstanding the task assigned to the participant, biases, human fatigue, etc. Also, the participants may not be very skilled at differentiating between distinct designs and might falter while trying to provide a well-founded preference rank (Corney, 2002). Adding to this complexity is the absence of some kind of noise model that can describe the distribution of these outlier scores. Some methods to detect such extreme unrepresentative data will be tested in future research related to this topic.

Classification Models. Classification models are important for the IGA system to predict user preferences for those individuals that are not labeled explicitly, as discussed before. The features used for classifying designs are based on the actual decision variables and the numerical objectives. Hence unlabeled individuals that are similar to labeled individuals in both design space and the numerical objective space are assumed to be similar to the labeled individual in their classes. For this initial work we will use the following two simple supervised classification methods: weighted K-nearest neighbor and decision trees. The weighted K-nearest neighbor method finds the nearest K (where K is an integer) labeled individuals and assigns a weighted class to the unlabeled individual (Mitchell, 1997). Decision trees are tree-structured plans of the features that can be tested to predict outputs (classes). Information theory is usually used to decide which feature should be tested in what order to find the most probable class of that design, while keeping the depth of the tree small (Mitchell, 1997). The initial set of labeled data is usually small, since very few individuals are ranked at every session, so there is a need to keep rebuilding the tree as more and more labeled data are collected through subsequent sessions. Hence, the training and testing data sets will be smaller for the initial trees, which can adversely affect the initial prediction capability of the trees. This issue will be investigated in an effort to design learning models that are robust and stable for the monitoring problem.

This is an ongoing work and more results will be shown at the conference.

Conclusions

IGAs are valuable tools that should be added to existing management systems when there is a call for analyzing problems from both quantitative and qualitative perspectives. The lack of ability to represent complete knowledge about a site or problem domain effectively is a very realistic one. Because of these gaps, the designers usually make simplifying assumptions about a problem before they try to model or optimize it. These assumptions can lead to misleading conclusions and can affect the reliability of such tools. Through this research, we aim to analyze these issues related to groundwater monitoring problems and create a hybrid expert system that allows freedom to adopt 'subjective' information in identifying optimal monitoring plans. This will create a new paradigm for optimization that approaches the task of problem solving from both computational and philosophical perspectives.

Acknowledgements

We would like to thank Abhishek Singh and the Automated Learning Group staff at NCSA, who helped us in creating the infrastructure for the experiments. This research was supported by the Department of Energy under grant number DE-FG07-02ER635302.

References

Berler, A, and Shimony, S.E., Bayes networks for sonar sensor fusion, in Geiger, D, Shenoy, P (eds) Thirteenth Conference on uncertainty in artificial intelligence, Morgan Kaufmann 1997.

Chan Hilton, A.B., and Culver, T.B., Optimizing groundwater remediation design under uncertainty. Proceedings of the Joint Conference on Water Resource Engineering and Water Resources Planning and Management, Minneapolis, MN, 2000.

Corney, D.P.A., Intelligent analysis of small data sets for food design, PhD thesis, University College, London, 2002.

Ezawa, K.J., and Schuermann, T., Fraud/Uncollectible debt detection using a Bayesian network based learning system: A rare binary outcome with mixed data structures, Besnard, P., Hanks, S. (eds) Eleventh Conference on uncertainty in artificial intelligence, Morgan Kaufmann, 1995.

Frey, B. J., *Graphical models for machine learning and digital communication*, MIT Press, 1998.

Goldberg, D. E., *Genetic Algorithms in Search, Optimization & Machine Learning*, Addison-Wesley, Reading, MA, 1989.

Gopalakrishnan G., Minsker, B.S., and Goldberg, D.E., Optimal sampling in noisy genetic algorithm for risk based remediation design, Journal of Hydroinformatics, 2003.

Hand, D.J., Intelligent data analysis: Issues and opportunities. Intelligent Data Analysis 2:2,1998.

Mitchell, T.M., *Machine Learning*, McGraw-Hill, 1997.

Neil, M., Littlewood, B., and Fenton, N., Applying Bayesian belief networks to systems dependability assessment, Safety Critical Systems Club Symposium, Leeds, Springer-Verlag, 1996.

Reed, P. M., Minsker, B. S. and Goldberg, D. E., A multiobjective approach to cost effective long-term groundwater monitoring using an elitist nondominated sorted genetic algorithm with historical data. Invited paper, Journal of Hydroinformatics, 3, 71-89, 2001.

Ritzel, B. J., Eheart, J. W. & Ranjithan, S. (1994) Using Genetic Algorithms to Solve a Multiple Objective Groundwater Remediation Problem, Water Resource Research, 30(5), 1589-1603.

Singh, A., Minsker, B. S., and Goldberg, D. E., Combining Reliability and Pareto Optimality - An Approach Using Stochastic Multi-Objective Genetic Algorithms, American Society of Civil Engineers (ASCE) Environmental & Water Resources Institute (EWRI) World Water & Environmental Resources Congress 2003 & Related Symposia, Philadelphia, PA, 2003.

Smalley J. B., Minsker B. S., and Goldberg D. E., Risk-based In Situ bioremediation design using a noisy genetic algorithm, Water Resources Research, 36(20), 3043-3052, 2000.

Stone, H., and Sidel, J. L. *Sensory Evaluation Practices*, 2nd ed., Academic Press, 1993.

Takagi, H., Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation, Proceedings of the IEEE, 89(9), 1275-1296, 2001.

Welge, M., Auvil, L., Shirk, A., Bushell, C., Bajcsy, P., Cai, D., Redman, T., Clutter, D., Aydt, R., and Tcheng, D, *Data to Knowledge (D2K)* An automated learning group report, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign. (available at http://alg.ncsa.uiuc.edu), 2003.

Wijns, C., Moresi, L., Boschetti, F, and Takagi, H., Inversion in geology by interactive evolutionary computation, IEEE International Conference on Systems, Man, and Cybernetics, 2, 1053-1057, 2001.