Expert Knowledge in Long-Term Groundwater Monitoring Optimization Process: The Interactive Genetic Algorithm Perspective

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 most practical water resources optimization applications, a number of important subjective issues exist that cannot be represented in numerical optimization procedures. Considering these issues only in a post-optimization analysis of solutions by the expert (engineers, stakeholders, regulators, etc) does not ensure that the final set of optimal designs address all qualitative issues important to the problem. The Interactive Genetic Algorithm (IGA) promises to overcome these hurdles by involving the expert directly in the online search process to steer the genetic algorithm to a solution or set of solutions that address both quantitative and qualitative criteria.

This paper investigates the effect on the overall search process when a single user interacts with the IGA system. Some of the salient control parameters that affect performance of such a framework are algorithmic control parameters (i.e. the GA settings, visualization interfaces, etc.), human control parameters (i.e. the user's cognitive perception, user's degree of risk aversion, human fatigue, etc.), and external control parameters (i.e. environmental noise and uncertainty, etc.). This work begins a rigorous assessment of the effects of different control parameters on the IGA search process by simulating the human decision making process using fuzzy logic models of human preferences as 'pseudo humans'. Comparison of such a system with a conventional optimization framework (that lacks progressive user feedback) is made for a long-term groundwater monitoring optimization problem, and related ramifications are highlighted.

Introduction

Decision-making and optimization in the water resources management field is a difficult task that encompasses many engineering, social, and economic constraints and objectives. In groundwater monitoring applications, some of the important qualitative factors that play a role in deciding suitability of monitoring network designs include contaminant types, regulatory requirements, extent of plume spread in the aquifer, hydrogeological conditions of the site, proximity to potential exposed receptors (e.g. drinking water wells), surface water interactions, post remediation effects, social constraints, legal constraints, political constraints, etc. Representing all of these factors in numerical formulations for optimization can be a very non-trivial issue, and such formulations often fail to consider important qualitative and incomputable phenomena related to the management problem. These limitations can be partially overcome by involving experts in offline post-analyses of the solutions found by numerical optimization algorithms. However, this approach not only increases the likelihood that the algorithms will converge to sub-optimal results, but it also increases the possibility of making an unnecessarily large computational investment in the process.

In the past, decision making tools such as Fuzzy Logic (FL) have been used to handle uncertainties that have distinct and/or imprecise characteristics and to handle the ambiguity of a decision maker's linguistic reasoning logic [e.g. reservoir operation problems (Russel and Campbell, 1996; Shrestha et al., 1996; Fontane et al., 1997; Tilmant et al., 2001), contaminated groundwater risk assessment (Chen et al, 2003), solid waste management (Seo et al, 2003), water quality management (Chen et al, 1998, Sasikumar et al., 1998, Mujumdar et al 2002), air pollution management (Sommer et al, 1978)]. However, such approaches do not consider differences in perception for multiple experts or changing trends of preferences as an expert evaluates solutions and learns more about the system performance.

Interactive Genetic Algorithms (IGA) provide a framework for combining the advantages of the genetic algorithm's (GA) search techniques (Goldberg, 1989) for quantitative criteria with the advantages of a real expert's online involvement in decision making for qualitative criteria. In the IGA setup (Figure 1) the expert imparts information related to his or her qualitative knowledge of the problem by comparing various solutions and ranking them on a fitness scale based on their overall quality (which includes both numerical and non-numerical attributes of the design). The solutions are then evolved using the typical GA operations of selection, crossover, and mutation, based on the user's subjective information and/or numerical information. A number of researchers have applied this approach to applications such as geological model fitting, hearing aid fitting, lighting design, music generation, face image generation, etc (Takagi, 2001). Kamalian et al, 2004, have used an interactive evolutionary approach (IEC) in conjunction with an evolutionary multi-objective optimization approach, i.e. the Multi-Objective Genetic Algorithm (MOGA), to evolve microelectrical mechanical systems. They found that using designer participation in the IEC to further evolve designs generated by the non-interactive optimization can produce better results than those using numerical optimization alone.

Human interaction to include expression of expert knowledge in any system, including the IGA, is a two-way relay between the expert and the computational analysis tool. Various human factors (e.g. human's personal opinion of design requirements, human fatigue due to repetitive and tedious tasks of ranking designs, boredom because of analyzing 'uninteresting' designs, changing trends in preferences as visualization of new designs introduces new concepts of design quality, skillfulness of the expert, environmental hurdles, etc.) at the time of decision making determine what information is passed into the computational tool. The design of the analysis tool also determines what information is presented to the expert and in what form the feedback is translated into the computational tool. Some of the tool-related issues that can affect feedback information are quality of visualization interfaces, options for communicating user preferences (Takagi, 2001), diversity in solutions during interactive sessions (less diverse solutions provide few 'interesting' solution options to the user and increase chances of boredom and fatigue), work load (number of individuals required to be ranked by the user) also increases chances for human fatigue, boredom and errors in classification.

The main objective of this paper is to explore some of the initial questions that arise in a system like this: "Why do we need a real user's intervention in the search process and why is this involvement better than the alternative option of using human preference models alone." For investigating these questions rigorously we initially simulate the expert's decision making process using fuzzy logic tools and then make these 'pseudo humans' interact with the search process. The arguments for the first question have been explained in detail above and become more obvious in the subsequent discussions. To answer the second question we test the effects of different expert psychologies, that can vary from person to person and with time, on the optimization process. At the end of this investigation we point out the inability of machine learning models to adjust for these changes in human psychology by themselves, and how having real users involved in the search process is important to accommodate for such changes in the human's perception.

The rest of the paper describes the experimental setup of this work, followed by a discussion of results and conclusions.

Experimental Setup

Long-Term Groundwater Monitoring Design Application. The groundwater monitoring application used for this work involves a 1313 feet by 865 feet BP (formerly British Petroleum) site where the groundwater has been contaminated with benzene, toluene, ethylbenzene, and xylene (BTEX) over a period of 14 years. The groundwater flow gradient in this aquifer exists in the northwest direction. 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 problem is to remove any wells that are spatially redundant. A quantile kriging interpolation model (see Goovaerts et al, 1997, for details) was implemented to model benzene and BTEX concentrations throughout the plume. Figure 2 is a site map of the area with the current monitoring plan.

The monitoring decision variables for the problem are sampling flags ($x_i = 0/1$) for all 36 wells. Hence, if a flag is 1 then the well at that location is sampled. The numerical objectives for this problem were to minimize the number of wells sampled (Equation 1) and to minimize the maximum error between actual concentrations and those estimated with the benzene interpolation models using a subset of K wells (Equation 2). The error in Equation 2 is normalized by a user-specified allowable error limit to avoid any scaling issues with the genetic algorithm.

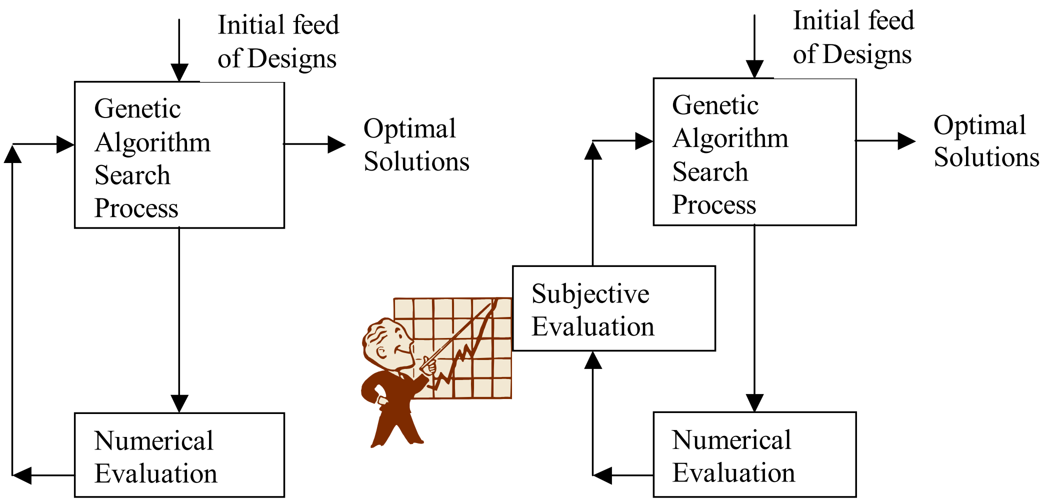

Figure 1. The traditional GA (left) and the interactive GA (right) frameworks.

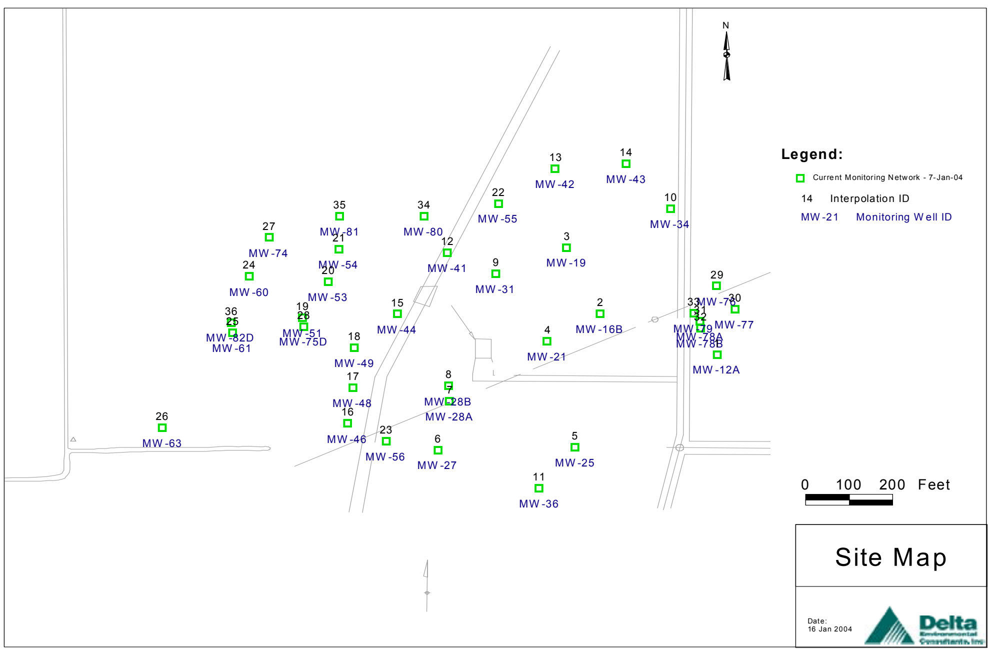

Figure 2. Current monitoring network plan at the BP site.

$$\text{Numerical Objective 1: } Minimize \sum_{i=1}^{n} x_i \text{, where } x_i = \begin{cases} 1 \text{ if well } i \text{ is sampled} \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

$$\text{Numerical Objective 2: } Minimize \left[\underset{K}{Max} \left\{ Error = \frac{\left| c_i^{\,actual} - c_i^{\,est}(K) \right|}{E_{allow,Benzene}} \right\} \right] \text{ for Benzene} \tag{2}$$

Initial results for this application indicated that benzene is the major constituent of concern in designing the monitoring system. To ensure that overall BTEX concentration errors still remain reasonable, a constraint was added to this problem to limit the maximum normalized BTEX error to a constant 1.

$$\text{Constraint: } \left[\underset{K}{Max} \left\{ Error = \frac{\left| c_i^{actual} - c_i^{est}(K) \right|}{E_{allow,BTEX}} \right\} \right] \leq 1 \quad \text{for BTEX} \tag{3}$$

In addition to these objectives, two qualitative criteria were considered in the IGA framework. The first criterion, which plays a vital role in determining a good monitoring design, is how similar the overall contaminant spatial distribution (plume map) appears before and after wells are removed from the sampling plan. The second criterion is how well defined the leading edge of the plume (the direction in which the groundwater is flowing) remains or how low the local interpolation errors at the leading north boundary are, 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. In the IGA framework, when a real user is involved, the human will make judgment based on their experience and visual perceptions of the plume maps. For the pseudo human involved in this work we simulate such analyses by including the local spatial errors at the leading edge boundaries (1100 feet by 100 feet north boundary region) into the fuzzy logic model and by setting decision rules for these local errors.

Interactive Genetic Algorithm. Most applications of IGAs to date have considered human preference in the form of a single qualitative objective. This human preference, usually expressed in the form of a ranking classification scheme (Takagi, 2001), is included in the optimization framework as an objective. Moreover, to control human fatigue from evaluating numerous potential solutions, the size of the populations used in IGAs has been kept low (e.g. 10 to 20, Takagi, 2001). For water resources problems where design possibilities are numerous, quantitative objectives can be important for prescreening alternatives that do not meet minimum design standards. In our work, we include both qualitative objectives of designs (in the form of rank evaluations varying from 1 (best design) to 5 (worst design) on five point ranking scale) and quantitative objectives within the Non-Dominated Genetic Algorithm II (NSGA II , see Deb, 2000, for more details) framework to rigorously solve conflicting multi-objective applications. The IGA setup for the experiments in this paper initializes the search process from a random population of designs, for small (e.g., 30) and large (e.g., 500 and 1000) population sizes. This IGA system is implemented within an automated learning system called 'Data to Knowledge' (D2K) (Welge et al. 2003).

Fuzzy Pseudo Human. Fuzzy logic provides a good framework for dealing with vague objects and approximate reasoning (see Zadeh, 1975, Zimmermann, 1987, for details). Fuzzy logic has been popularly adopted in applications that have insufficient data to characterize uncertainty using standard statistical measures. Also, when expert judgment based on experience and observed physical attributes is the best assessment method available for imprecision and contradictions, fuzzy logic allows for quantification of the expert's linguistic reasoning.

In this study, we implemented two fuzzy pseudo humans with different perceptions of the expert's linguistic reasoning. Expert A (or the 'conservative expert') included all 5 important attributes (i.e., benzene error, number of wells, BTEX error constraint violation, local benzene error in the northern boundary region, and local BTEX error in the same region) as fuzzy inputs to the model. Expert B (or the 'non-conservative expert') expressed his/her priorities for only 3 attributes (i.e., number of wells, local benzene error and local BTEX error). Hence, the first expert has a risk averse psychology through which he/she tries to satisfy and balance all features of the problem, while the second expert cares only for the overall cost of the monitoring plan and the local interpolation error predictions at the site boundary. Trapezoidal membership functions, based on expert judgment, were selected for these input attributes. The outputs for these fuzzy humans were simulated human ranks, which were also fuzzified by using trapezoidal functions for the five point ranking scale (1(best), 2(good), 3(average), 4(bad), and 5(worst)). Different rules were created for each fuzzy pseudo human to simulate the psychology of each expert.

Results and Discussion

Part 1 below explores the need for human feedback in the search process by involving no expert, expert A, or expert B (with different opinion from expert A) in the optimization process. Part 2 explores the effect of changing expert preferences on the search process.

Part 1. Three different types of experiments were performed for testing the effect of human preferences: (1) no human rank objective, (2) simulated expert A involved in online human ranking of solutions, and (3) less conservative simulated expert B involved in the online ranking. The results shown here demonstrate one of the experiments for a population size of 30, to highlight the main features of our findings. Figure 3 compares the normalized benzene error objective with the number of wells objective. The figure shows that not only is the benzene error vs. number of wells tradeoff different for GA searches with and without expert involvement, but the GA was able to find more solutions on the tradeoff curve when experts were involved. Note that when the number of wells is 25, the designs have somewhat different benzene errors for GA searches with and without expert involvement.

Figure 4 compares the human ranks of the final solution set with and without expert involvement. For the case in which no expert was involved, the final solutions were evaluated by expert A and expert B to obtain their perspective on the quality of the designs after the optimization was completed. When the human ranks of solutions found by the GA are compared, agreements and disagreements between the conservative and non-conservative decision maker become clear. In general, expert A tries to satisfy all 5 qualitative attributes of the design on average and ends up preferring solutions with larger numbers of wells because of their low global and local interpolation errors, in spite of high monitoring costs. Expert B on the other hand prefers the solutions with mid-range numbers of wells, where both monitoring costs and/or local boundary interpolation errors are relatively low. For most solutions found by the GA with no expert involvement, expert A and expert B disagree about the human ranks for the same solution (except when 22 wells are installed).

Note also that, even though both experts believe that the GA search without expert involvement found a ‘25 wells’ solution (Figure 4) with rank as good as their own designs, the magnitude of their ranks for this solution do not coincide with each other. In other words, Expert A is more satisfied with the quality of the solution found by the ‘no expert involved’ optimizer than Expert B is. This solution has low benzene error, but slightly higher local benzene and BTEX boundary errors. The magnitude of these errors is close to the errors of the ’25-well’ designs found by the two experts. However, Expert B heavily penalizes solutions that have high local errors, hence he/she ranks it harshly. Expert A, on the other hand, in considering performance of all attributes, ranks this solution favorably because of low monitoring cost and benzene error. Furthermore, for 28 wells, the solutions found by expert A and expert B have similar benzene errors (Figure 3), and both experts rank their own solutions to the same quantitative value (Figure 4). This solution has low benzene error and low local benzene and BTEX errors, hence both experts rank this solution favorably.

Figure 5 shows human ranks of solutions discovered by expert A, from both expert A and expert B’s perspective. Figure 6 shows a similar tradeoff of human ranks and number of wells for the solutions discovered by expert B, from both experts’ perspective. Note that expert B agrees with expert A’s rank for the ‘28-well’ solution found by expert A (Figure 5), but in expert A’s judgment (Figure 6) expert B’s solution should be ranked slightly worse than what expert B has ranked it. The ’28-well’ solution found by Expert A has low benzene error and local errors, and both experts in their analysis find this ‘mid-range’ (i.e. when number of wells are 25 to 28) equally fit. However, the ’28-well’ solution found by Expert B, despite having low local benzene and BTEX errors, is ranked slightly worse by Expert A because of slightly worse global benzene error. Expert B does not take into account global benzene error while evaluating solutions so he/she overlooks this attribute and ranks this solution favorably because of other attributes.

Part 2. This part of the experiments looks into the issue of nonstationarity of preferences in the human mind and its effect on the optimization. This nonstationarity occurs when human perception about decision rules changes during the optimization process, depending upon what the user learns from examining the evolving solutions. For example, the conservative expert A in the previous experiments might change his/her strategy and adopt the less-conservative strategy of expert B in the middle of his/her search process, if he/she is not satisfied with the designs that are being generated by the search.

Figures 7 and 8 provide examples of such circumstances for a small population size of 30. The IGA begins with expert A providing continuous feedback to the search process (similar to the experiment performed in Part 1). Suppose that, at generation 48, the user changes his/her perspective to Expert B’s less conservative ranking criteria. After 24 generations of adopting Expert B’s perspective, the solution set that is obtained (see icons for ‘Expert B perspective with initial 48 gen of Expert A perspective’ in Figure 7) looks very different from the final non-dominated set that the user would have obtained using either expert's perspective alone for the full 72 generations. Figure 7 also shows that when the user changes his/her perspective in the middle of the GA search, he/she is able to find solutions that have global benzene errors and local benzene and BTEX errors better than designs found by the expert when his/her perspective was that of Expert B throughout. Having an initial choice of Expert A's perspective helped the GA search for solutions that had lower global benzene errors, and when the expert changed his/her policy to that of Expert B the focus was changed to the improvement of local errors. The coverage of solutions on the tradeoff is also better for the mid range number of wells (i.e. 25, 26, 27 and 28 wells), that are not very costly, compared to the experiments when the user did not change his/her perspective at all.

Figure 8 shows that the tradeoffs between simulated human ranks and number of wells is also different for the 3 cases, with more solutions obtained between the range of 'Best' (i.e. rank 1) and 'Good' (i.e. rank 2) when Expert B perspective was adopted after an initial search based on Expert A's perspective, than the case when Expert B perspective was kept continuous throughout the search. The reasons for this phenomenon is similar to the one explained above. Such solutions that had low local and global errors automatically had a better human rank from Expert B's perspective. These new developments would not have happened if the user had not been interacting with the system and had not decided to intervene and modify his/her search strategy.

These results further emphasize the need for direct human interaction to capture changes in human preferences as designs are reviewed. A machine learning algorithm that simulates human preferences would not be able to detect a change in search strategies during the optimization process, unless a real user interacts and signals this modification. Hence, having real user participation becomes very critical for the overall performance of the system.

Conclusions and Future Work

This paper elucidates the need for an expert's interaction in optimization models that try to solve complex field scale applications. In part 1 of the results section, we saw how including an expert's preferences in the optimization process changes the kinds of solutions evolved by the genetic algorithm. Not only do the final non-dominated sets have different tradeoffs for numerical objectives in most cases, the quality of solutions (interpreted through human ranks) and the actual designs themselves change with expert involvement. Part 1 also revealed the effect of different kinds of expert opinions on the quality of solution sets evolved. Experts belonging to the same field of study can differ in their interpretation of what application features are important. These differences in perceptions can completely modify the structure of the final designs, even though the overall project objective of the experts might be the same. Part 2 shows the need for the user's explicit involvement to identify and respond to changes in preferences as the optimization progresses. This finding further highlights the disadvantage of using machine learning models of user preference alone in an optimization system, since such models cannot detect changes that may occur in the expert's mind as solution options are reviewed.

There are many more questions that this work puts forth for such an interactive system: Is it better to have an interactive system that accepts only the real expert's feedback or is it better to have a feedback system through which constant online interaction between the expert and the learning model improves solution evolution and performance criteria? What is the effect of population size on human fatigue? What kind of IGA setup is best for controlling human fatigue, encouraging diversity in solutions, and accommodating the effect of nonstationarity (when preferences change over time). The authors are currently exploring these questions.

Acknowledgements

We would like to thank 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 and office of Naval Research N00014-04-1-0437.

References

Babbar, M., B. S. Minsker, and H. Takagi, "Interactive Genetic Algorithm Framework for Long Term Groundwater Monitoring Design," American Society of Civil Engineers (ASCE) Environmental & Water Resources Institute (EWRI) World Water & Environmental Resources Congress 2004 & Related Symposia, Salt Lake City, UT, 2004

Chen, H. W., and Chang, N. (1998). "Water pollution control in the river basin by fuzzy genetic algorithm-based multiobjective programming modeling." *Water Sci. Technol.*, 37(3), 55–63.

Chen, Z., Huang, G.H., and Chakma, A. (2003). "Hybrid fuzzy-stochastic modeling approach for assessing environmental risks at contaminated groundwater systems." *Journal of Environmental Engineering*. 129(1), 79-88.

Deb, K., S. Agrawal, A. Pratap, and T. Meyarivan (2000) A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II, Kanpur Genetic Algorithm Laboratory (KanGAL) Report 200001, Indian Institute of Technology, Kanpur, India.

Fontane, D. G., Gates, T. K., and Moncada, E. (1997). ''Planning reservoir operations with imprecise objectives.'' *J. Water Resour. Plan. Manage.*, 123~3!, 154–163.

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

Goovaerts, P., Geostatistics for Natural Resources Evaluation, Oxford University Press, New York, 1997.

Kamalian, R., R., Takagi, H., and Agogino, A., M. (2004) ‘‘Optimized Design of MEMS by Evolutionary Multi-objective Optimization with Interactive Evolutionary Computation, ’’ Genetic and Evolutionary Computation (GECCO2004), Seattle, WA, 1030-1041.

Mujumdar, P. P., and Sasikumar, K. (2002). ‘‘A fuzzy risk approach for seasonal water quality management of a river system.’’ Water Resour. Res., 38~1!, 5-1–5-9.

Russel, S. O., and Campbell, P. F. (1996). “Reservoir operating rules with fuzzy programming.” *J. Water Resour. Plan. Manage.*, 122, 165–170.

Seo, S., Aramaki, T., Hwang, Y., and Hanaki, K. (2003). “Evaluation of solid waste management system using fuzzy composition”. Journal of Environmental Engineering, 129(6), 520-531

Shrestha, B. P., Duckstein, L., and Stakhiv, E. Z. (1996). ‘‘Fuzzy rule based modeling of reservoir operation.’’ *J. Water Resour. Plan. Manage.*,
122, 262–268.

Sasikumar, K., and Majumdar, P. P. (1998). “Fuzzy optimization model for water quality management of a river system.” *J. Water Resour. Plan. Manage.*, 124(2), 79-88.

Sommer, G. and Pollatschek (1978). “A fuzzy programming approach to an air pollution regulation problem.” In *Progress in Cybernetics and Systems Research*, Vol. III, Trappl, R. and Klir, G.J. (editors), 303-313.

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

Tilmant, A., Persoons, E., and Vanclooster, M. (2001). ‘‘Deriving efficient reservoir operating rules using flexible stochastic dynamic programming.’’ Proc., 1st Int. Conf. on Water Resources Management, WIT Press, U.K. 353–364.

Welge, M., L. Auvil, A. Shirk, C. Bushell, P. Bajcsy, D. Cai, T. Redman, D. Clutter, R. Aydt, and D. Tcheng. *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.

Zadeh, L. A. (1975). ‘‘The concept of a linguistic variable and its applications to approximate reasoning.’’ Inf. Science, 8, 199–249.

Zimmermann, H.-J. (1987). Fuzzy sets, decision making, and expert systems, Kluwer Academic, Boston.

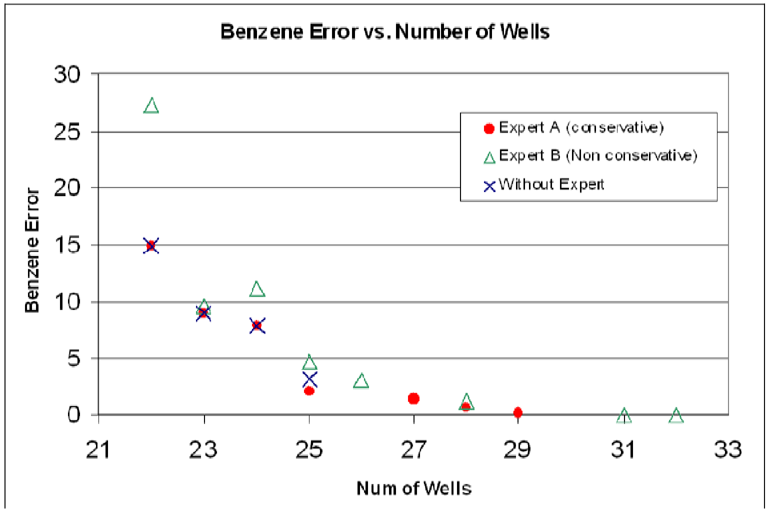

Figure 3. Effect of human feedback, Benzene error Vs. Number of wells.

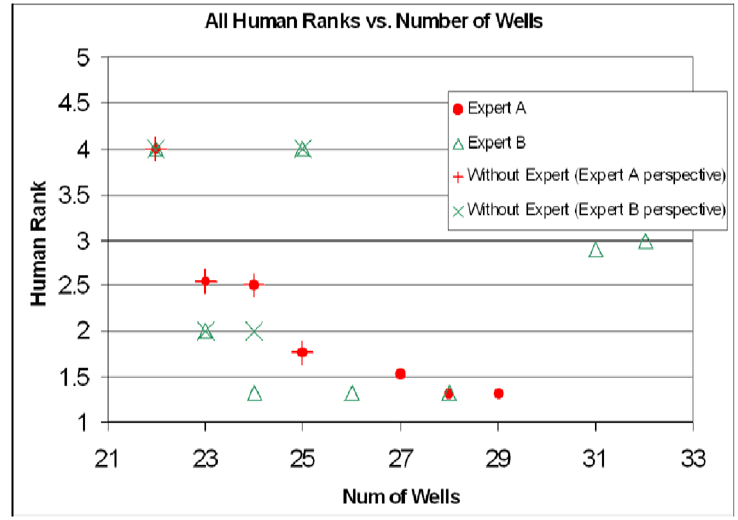

Figure 4. Effect of human feedback, Human ranks Vs. Number of wells.

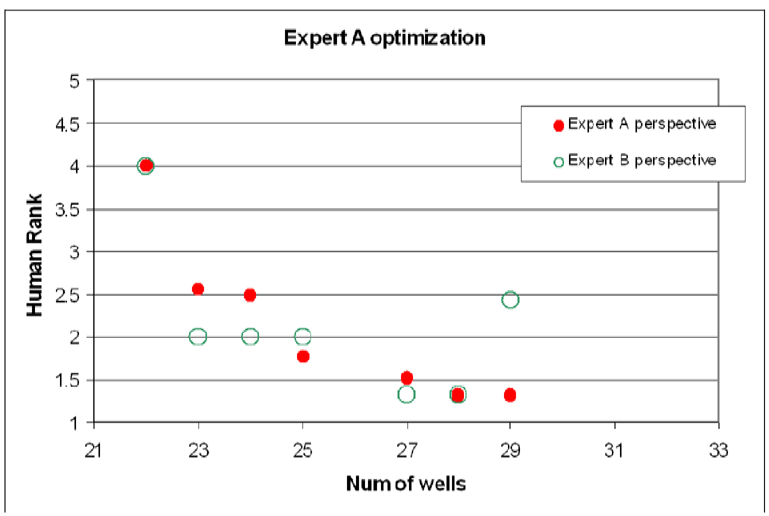

Figure 5. Human ranks for optimization involving Expert A.

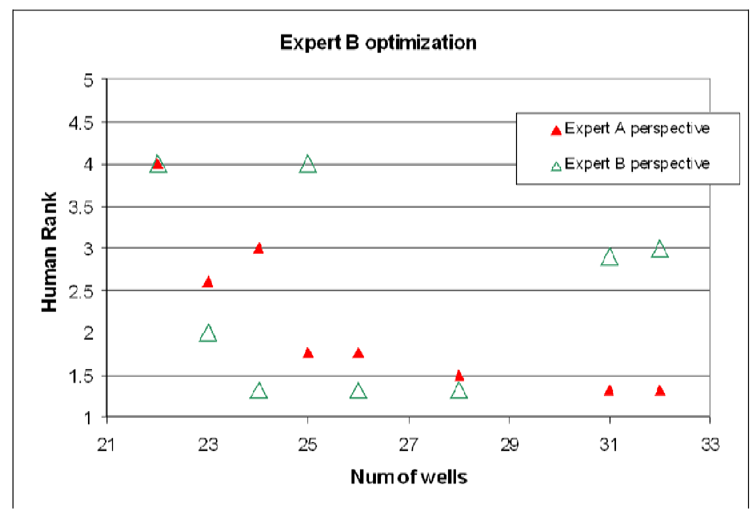

Figure 6. Human ranks for optimization involving Expert B.

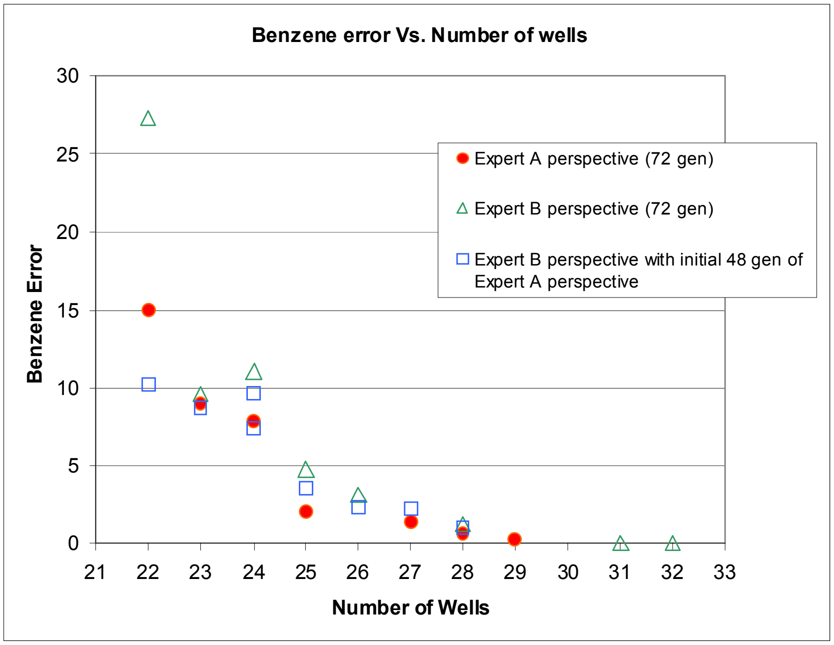

Figure 7. Effect of nonstationarity in human preference during optimization, Benzene error Vs. Number of wells, population size 30.

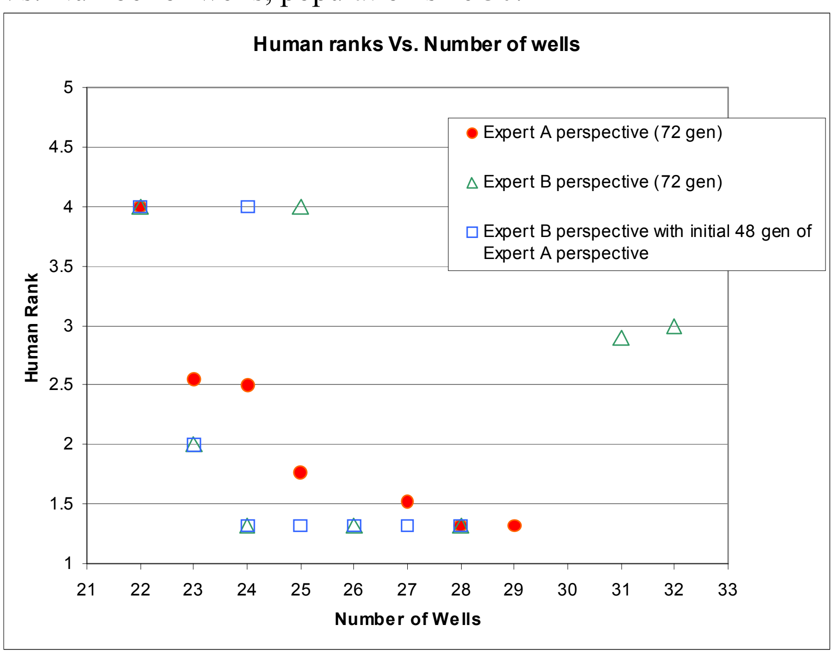

Figure 8. Effect of nonstationarity in human preference during optimization, Human ranks Vs. Number of wells, population size 30.