Interactive Genetic Algorithms for Inverse Groundwater Modeling: Issues with Human Fatigue and Prediction Models

Abhishek Singh[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: asingh8@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

This paper presents an interactive multi-objective evolutionary optimization based approach to solve the inverse problem of estimating heterogeneous aquifer parameters (in this case - hydraulic conductivity) for a groundwater flow model. A hypothetical aquifer, for which the 'true' parameter values are known, is used as a test case to demonstrate the usefulness of this method. It is shown that using automated calibration techniques without using expert interaction leads to parameter values that are not consistent with site knowledge. In such cases, it is desirable to incorporate expert knowledge in the estimation process to generate more reasonable estimates. An interactive approach is proposed within a multi-objective framework that allows the user to evaluate trade-offs between the expert knowledge and other measures of numerical errors. For the hypothetical aquifer, this type of expert interaction is shown to produce more plausible estimates. A major issue with interactive approaches is 'human fatigue'. One way of dealing with human fatigue is to use machine learning to model user preferences. This work presents some initial results that show that machine learning models can be used to augment user interaction, allowing the IGA to find good solutions with much less user effort.

Introduction

Inverse problems are encountered whenever physical phenomena are mathematically modeled to predict the response of such a system. Typically the model has parameters corresponding to certain characteristics of the real system and these are calibrated based on existing observed data. Groundwater models have many spatially distributed and heterogeneous parameters such as hydraulic conductivity, recharge, evapotranspiration, storage, aquifer thickness, as well as boundary and initial conditions. Unfortunately it is usually impossible to have enough measurements of these parameters to characterize the spatial variability and heterogeneity reliably. Of these the hydraulic conductivities is arguably the parameter with the most uncertainty, the maximum variability and a significant effect on the prediction of groundwater models.

In this study we investigate the inverse problem related to the estimation of hydraulic conductivity given head measurements in an aquifer, although the methodology discussed in this paper can be easily extended to the estimation of other parameters and models. Many studies such as *Yakowitz and Duckstein* (1980), *Carrera and Neuman*

(1986), *Sun* (1995), and *Zimmerman et al.* (1998) have demonstrated that the inverse problem for hydraulic conductivity is ill posed, i.e. it has non-unique solutions, which are unstable for small variations in the observation data. This is primarily due to the small-scale heterogeneity and the large-scale variability of conductivity values that give this parameter a potentially infinite domain.

To address this problem, it is necessary to reduce the dimensionality of the problem (*McLaughlin and Townley*, 1996). The simplest way of doing this is the block or regionalized approach (*Carrera and Neuman*, 1986), where the site is divided into blocks of uniform conductivities. Spatial interpolation can be used to come up with continuous descriptions of the conductivity field (*Yeh*, 1986, *Hill et al.*, 1998, *Tsai et al.,* 2003). The pilot point approach discussed by *Doherty* (2003) assigns conductivity values to a finite set of dummy points distributed throughout the model domain and then interpolates these values using geostatistical methods. Other geostatistical methods used to characterize parameter heterogeneity include *de Marsily* (1984)*, Kitanidis* (1997), and *Zimmerman et al.* (1998) among many others.

The basic idea in all of these approaches is to try to describe the spatial variability structure of the parameter using a functional form with fewer parameters than the available measurements. However, the type of functional form to be used is often left to the 'expert judgment' of the modelers and site hydrogeologists. For example when using kriging for spatial interpolation, it is unclear what kind of kriging (simple, ordinary, or universal), variograms, or search methods to use for a particular problem. In their extensive review of geostatistical inverse approaches *Zimmerman et al*. (1998) conclude 1) *"the greatest attention should be given to the semivariogram to be used in the inversion*", and 2) "*identifying the proper parameterization (topological/geometric structure of the T field) can be more important than estimating parameter values. It has been shown that the calibration of the model (head matching) can be very good even with a T field that is not very representative of reality*." Even in non-geostatistical approaches like zonation or universal parameterization it is left to the expert judgment of the modeler to come up with appropriate zones or interpolation functions. Unfortunately there is no formal structure to address the contributions of the expert in inverse modeling, due to the fact that usually this judgment or intuition is 'fuzzy' and difficult to quantify as objectives.

The basic motivation of our ongoing research is to explore interactive approaches that combine mathematical optimization with expert knowledge in a systematic methodology. Previous research has 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, *Singh et al.*, 2003, *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). This makes GAs ideal to be used in the inverse problem, which can depend on complex simulations, and can have multiple data sources and objectives to be satisfied. Interactive evolutionary algorithms (*Takagi*, 2001) use the evolutionary optimization paradigm but also incorporate user preference in the optimization process. Interactive evolutionary algorithms have been applied to the fields of geology, hearing aid fitting, machine design, lighting design, music generation, face image generation, etc. *Wijns et al.* (2001) have applied interactive inverse modeling to a geodynamic model of fault structure. This paper presents the on-going development of an interactive multi-objective genetic algorithm system for solving the groundwater inverse problem, as well as some promising results that we obtained on a synthetic aquifer model.

Synthetic Groundwater Aquifer Case Study

The case study for this project is based on the aquifer discussed by *Freyberg* (1988) and shown in Figure 1. This test case was chosen because previous studies (*Freyberg,* 1988) have shown that the inversion of conductivity based on head observations is an ill-posed problem with little or no correlation either between calibration error and root mean square error (RMSE) of hydraulic conductivity, or accuracy of post-calibration predictions. Since numerical calibration led to unsatisfactory results this test case may demonstrate the importance of expert interaction in improving calibration process. The aquifer is considered to be the ground truth for the parameterization, and all observations are taken from this aquifer. Moreover it is desirable that the conductivity scenario looks similar (captures the large scale trends) of the original aquifer.

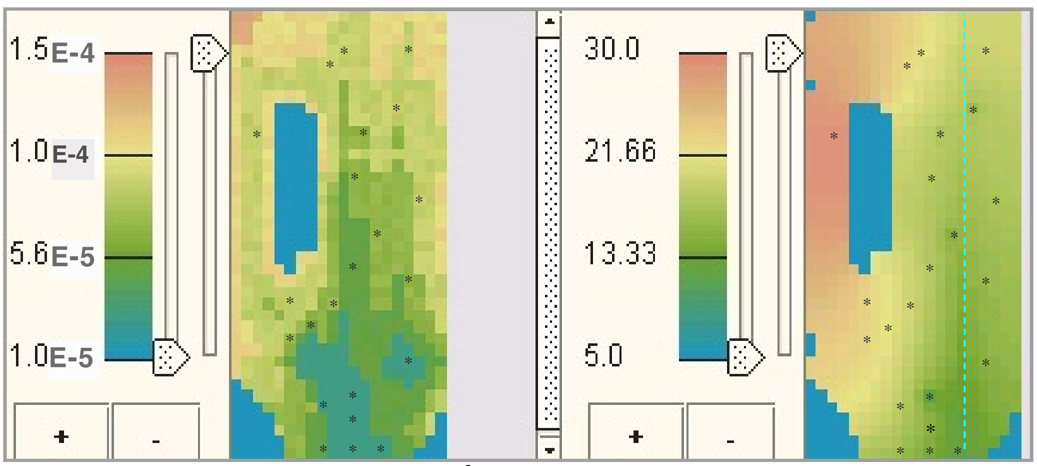


Figure 1. (Left) Conductivities in m^2/sec and (Right) steady state heads in m in the original aquifer. The stars represent the observation points, the line on the right is the river, and the areas in blue are the boundaries and outcrop.

The hypothetical aquifer is a shallow, 2-D, phreatic aquifer 10 km in length and 5 km across underlain by an impervious stratum with no leakage. The computational grid is square with a grid spacing of 250 meters giving a 20*40 gridded computational domain. The aquifer is assumed to be isotropic and vertically homogenous with steady state hydraulic heads. There are no flow boundaries along the north, east, and west boundaries, and a constant head boundary along the south. In addition there is an outcrop in the middle, which introduces an internal no-flow region. Uniform and constant recharge is assumed. There is a river with a steady, non-uniform stage that runs north to south across the aquifer. The hydraulic conductivities as well as the bottom elevations are highest in the northwest, with a decreasing trend to the southeast. There are 6 wells pumping at constant rates, and 16 observation wells spread across the aquifer (giving a total of 22 observation points). Since only the broad ranges of the hydraulic conductivities and bottom elevations were known, some random noise was added to make the conductivities and elevations more 'realistic'. The hydraulic heads in this hypothetical aquifer were then solved using MODFLOW (*McDonald and Harbaugh*, 1988).

Once the hydraulic heads have been calculated for the 'true case', the values at the observation points can be used as calibration targets for the inverse method. For this project only the hydraulic heads were considered unknown, while all other parameters, boundary conditions, and observations were assumed to be perfectly known. The hydraulic heads and conductivity measurements at the observation points were taken as the calibration data for the inverse model.

To make the inverse problem tractable it is necessary to represent the continuous conductivity field as a function of finite parameters. In this study we use a geostatistical description of the conductivity field to come up with kriging-based linear estimates of log conductivity. KT3D (*Deutsch and Journel,* 1998) was used to krige the log-conductivity values based on the 22 locations. The kriged estimates are finally back-transformed (the original data was log-conductivity) to give the conductivity field to be used in the groundwater flow model (MODFLOW, *McDonald and Harbaugh*, 1988). The following parameters for the kriging are decision variables within the GA – type of kriging (simple kriging, ordinary kriging, universal kriging with linear trend), kriging search parameters (maximum/minimum radii and direction of search ellipse), maximum/minimum data within search space to be used for kriging, types, scaling coefficients, ranges and direction of anisotropy for the nested variogram models. This leads to a total of 24 decision variables.

Interactive Multi-Objective Genetic Algorithms

Babbar et al., 2004, have explored the applicability and utility of Interactive Multi-objectives GAs (IMOGA) for long term monitoring design. The IGA system is being built into an automated learning system called 'Data to Knowledge' (or D2K) (*Welge et al*, 2003). Details about the IMOGA can be obtained from *Babbar et al.*, 2004.

The objective function for the inverse problem (*McLaughlin and Townley*, 1996) consists of a measure of prediction error (estimation objective) and some measure of deviation of parameter values (regularization objective) from observed data. *Madsen and Kristensen* (2002) have demonstrated that using a multi-objective framework in calibration can lead to significant improvement of the estimate and better understanding of the site and the data.

In this study we consider three separate objectives: head error, cross-validation error, and variogram error. The head error is the RMSE of head predictions with respect to head observations. Since kriging is an exact estimator, the estimation error for the conductivity field at the measurement location is zero. The cross-validation error is calculated for the kriged estimates of log conductivities by removing one point at a time and using all other points to predict the missing point. The total mean square error between the missing and the predicted points is the cross-validation error. Finally, the variogram error is the fitting error between the modeled and empirical variograms for all lag distances.

Figure 2 shows a flow chart of the IMOGA solution process. 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. The IMOGA uses the elitist non-dominant sorting genetic algorithm (NSGA-II) (*Deb et al, 2000*), which has been shown to be a reliable and efficient algorithm for solving multi-objective problems. For this application, there are three numerical objectives and one human objective. For the human objective, multiple panels consisting of the image of the (back-transformed) kriged conductivity field and the hydraulic heads for the given conductivity are shown to the user. Each panel corresponds to a particular variogram model and kriging parameters. The user then ranks each solution from 1 to 5 (where 1 is the best solution and 5 is the worst).

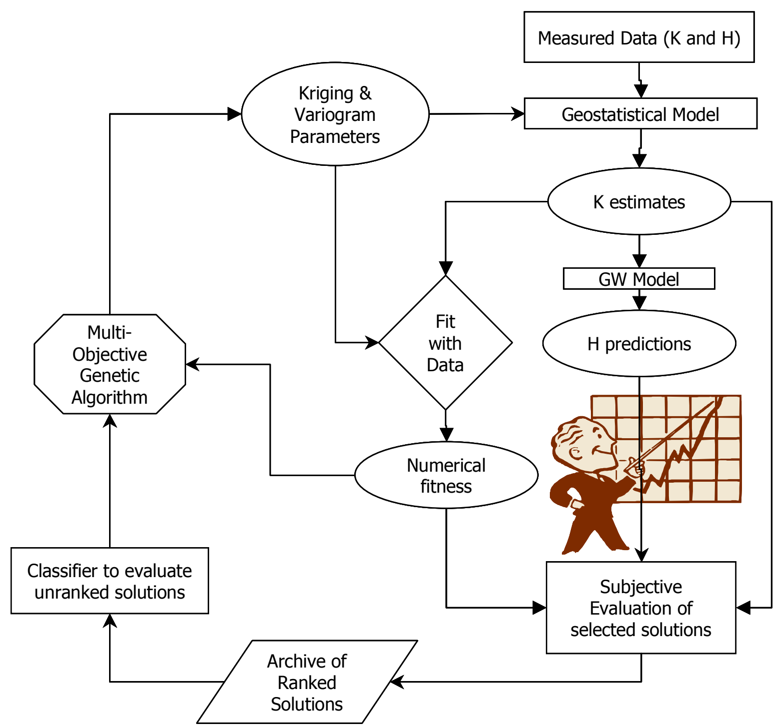


Figure 2: Framework for interactive multi-objective genetic algorithm for the inverse problem.

The solutions ranked by the user are then stored in the archive. If the population size is large, only a portion of the population will be ranked and a portion will remain unranked. To reflect user preferences, these individuals need to be ranked based on the existing ranked solutions. Since the user will typically rank only a few individuals (typically 10 to 20) in every generation, only simple rule-based classification algorithms can be used.

This study uses a decision tree (*Quinlan*, 1986) to model user preference. Decision trees recursively select the most predictive feature based on information gain theory and split the training sets into subsets. Splitting continues until the information in the inputs is exhausted and the terminal nodes are the classification of the final instances.

In our case the final instances that need to be predicted correspond to the ranks (from 1 to 5), the predictive features consist of all the decision variables and the numerical objectives used in the GA, and the training is performed based on the already ranked individuals stored in the archive. To accommodate changes in user preferences as more solutions are viewed, the decision tree is trained only on the most recently ranked individuals.

Results

Some preliminary results were obtained from the IMOGA system for the Freyberg test case. These runs consisted of running the IMOGA without human interaction (in which case it is simply a multi-objective inverse solver) and with continuous human interaction of all individuals in all generations. Human ranking was based on how qualitatively similar a particular conductivity and head field appeared relative to the original fields. No direct comparisons were made, but the user ranking the images had modeling experience with the test case. To reduce the burden of human ranking, a small population size (40) and number of generations (20) were used, which still required almost 800 images to be ranked. Because of this significant burden, additional runs were completed with less human interaction (user ranking only half of the population) to investigate whether the machine learning model could be used as a surrogate for some user interaction. For both these cases only 4 sessions of user rankings were stored in the archive to reflect the most current 'state of mind' of the expert. This is important because expert preference is relative and changes as better solutions are found is later generations of the GA. For this paper only the runs with user ranking half the population are shown. More results will be shown at the conference.

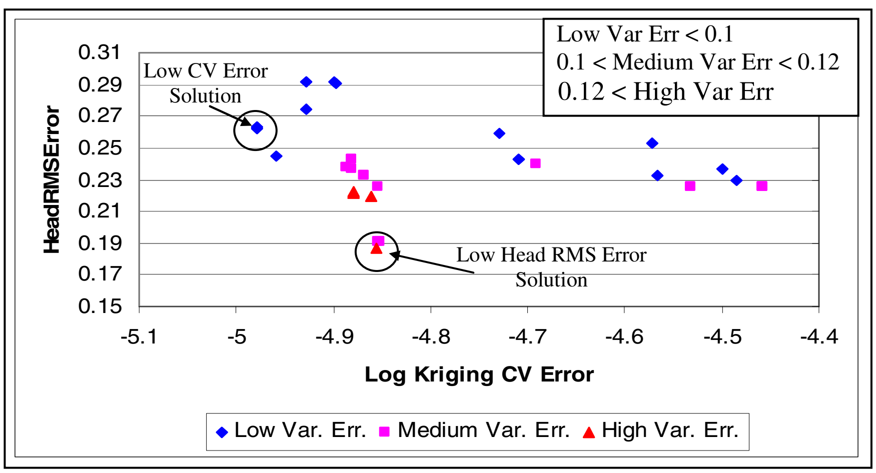
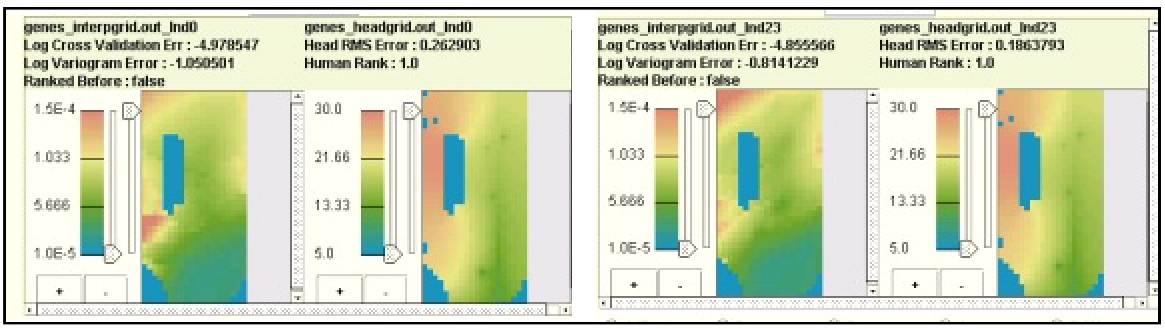


Figure 3: Final Pareto front for MOGA with no user interaction.

Figure 3 shows the trade-off curve (Pareto front) for the final generation from the simple MOGA with no human interaction. The figure shows the kriging cross-validation error and the head root mean square error as the primary objectives for visualization. The third objective of variogram fitting error (the fit between the model variogram used for the solution and the empirical variogram from the data) is shown by the color of the solutions (with blue for the low range and red for the high range values). It is interesting to note that there are tradeoffs among all three objectives. For example, there are solutions that have high head root mean square error but low cross-validation error (i.e. even though the kriged field is accurately modeled the head predictions are not very accurate). Also interesting is the fact that there are solutions with high variogram fitting errors that have low cross-validation and head prediction errors. Usually modelers use a regression tool to fit the parameters of the variogram to the data. This would correspond with the blue solutions that have the smallest variogram fitting error. Thus this standard practice actually constrains the accuracy of the kriged field and head prediction because allowing for some error in the variogram fitting could lead to better solutions for those two objectives.

This is an important demonstration of the usefulness of the multi-objective approach in finding tradeoffs between different objectives and how they affect the accuracy of the model. For comparison with subsequent results with human interaction, two solutions, corresponding to the lowest head RMS error and the lowest kriging cross-validation error, are visualized in Figure 4.



(a) **(b)**

Figure 4: Kriged K field and head predictions for a) low kriging cross-validation error and b) low head RMS error.

Comparing Figure 4 with the original heads in Figure 1 reveals some interesting features. First it is clear that the head predictions are relatively insensitive to the different K spatial patterns, confirming that this inverse problem *is* ill-posed. Secondly, the solution with lowest kriging CV error (Figure 4a) has an unreasonably high conductivity in the western (left) portion of the site. This would be the conductivity field we would get if we simply kriged the data and optimized it based on the cross-validation results (as many geostatistical approaches tend to do). The second solution corresponding to the low head RMSE (Figure 4b) does a better job at predicting the western portions of the site but has unreasonably high conductivity in the northwestern (top-left) portion of the site. This type of 'boundary effect' is very commonly seen in estimation and inverse problems where there are not enough data to characterize the estimated parameters near the boundary of the site. Often it is these boundaries that play an important role in determining containment of contaminants within the site.

The next step was to introduce expert interaction into the optimization. The criterion for ranking the solutions was how qualitatively close they were to the original kriged and hydraulic head fields. Figure 5 shows the Pareto front for the case where the user ranked every individual in every generation, along with the previous simple MOGA trade-off curve. To make a better comparison, the same user ranked the final population of the MOGA off-line, using the same criteria that were used for the online interaction. This graph indicates that the final Pareto front from the IMOGA is significantly different from the final front from the MOGA. Thus user interaction has led to the discovery of different solutions than non-interactive optimization. It is also important to note that the MOGA did find some solutions that attained a human rank of 1 (the best ranking). Most of the new solutions discovered by the IMOGA are in the objective space marked by the ellipse on the graph. This ‘zone of discovery’ corresponds to previously sub-optimal solutions that were preferred by the user. On further inspection it is revealed that these solutions generally have a high variogram fitting error. Thus by choosing these solutions, the human expert allows solutions with different variogram structures to be explored that may not necessarily fit well with the empirical variogram. Thus the usual practice of fitting the variogram before kriging the surface and predicting heads leads to sub-optimal solutions from the expert’s point of view.

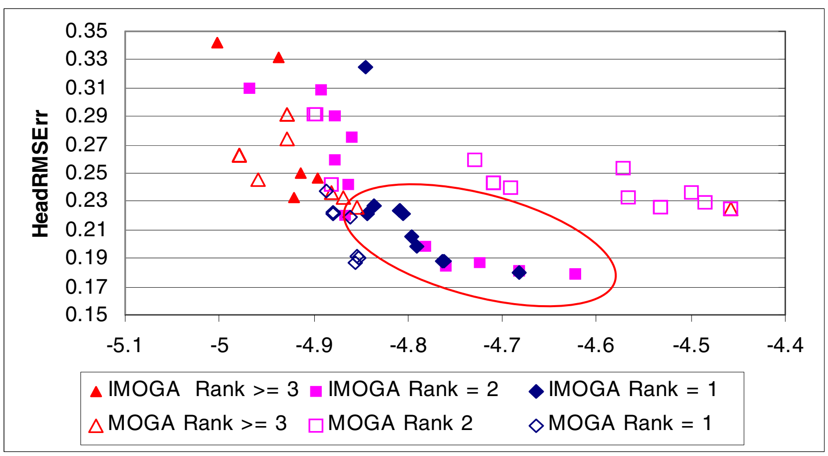
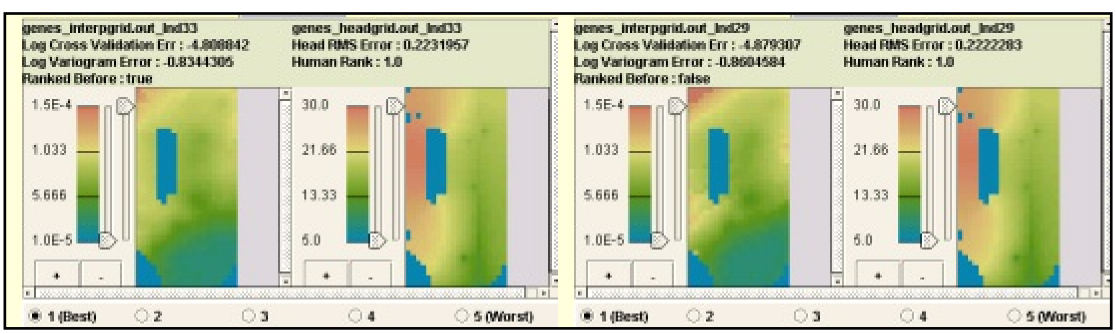


Figure 5: Pareto front for the IMOGA (solid diamonds, boxes, and triangles) compared with the MOGA (open diamonds, boxes, and triangles).



(a) **(b)**

Figure 6: Comparison of kriged field conductivity and head predictions for a) solution found best by the user from IMOGA and b) from MOGA.

It is also important to note that the IMOGA has more rank-1 solutions (14 in all) than the MOGA (7 in all). However the human rank is a subjective measure, thus two rank-1 solutions in one population could differ in their actual conductivity structure. To make a fair comparison the two best (according to the expert) solutions are compared in Figure 6. It is clear that the best solution for the IMOGA is closer to the true field (see Figure 1) than that for the MOGA. Specifically, the two areas in the north and northwest have more reasonable conductivity values. Also the solution from the MOGA has a pronounced anisotropy in the northeasterly direction that is not present in the IMOGA results. In ranking each solution the user considers multiple factors that MOGA cannot consider based solely on quantitative criteria. It is a well-known fact in the IGA community that human subjects are excellent at image processing and can easily pick up spatial patterns. Information about the spatial pattern is lost in the non-interactive approach that can only look at the values of the kriged surface and the predicted heads at a limited number of observation points.

Finally, in Figure 7 the results with complete user ranking are compared with the results from a run with user ranking of only half the solutions per generation. Recall that in this case, unranked solutions are given rankings from the decision tree model fit to the ranked solutions. Figure 7 shows that the Pareto front for the case with only half the user interaction is very similar to the original Pareto front with full user interaction. However some of the rank-1 solutions are missing, which could be due to classification errors in the decision tree.

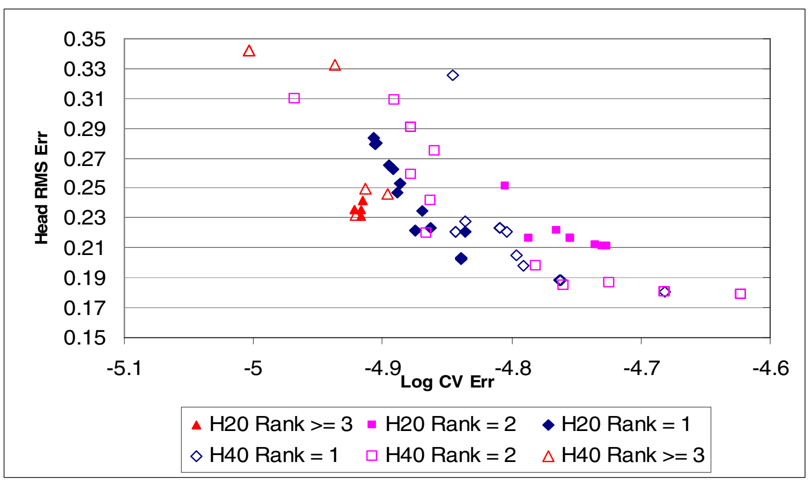
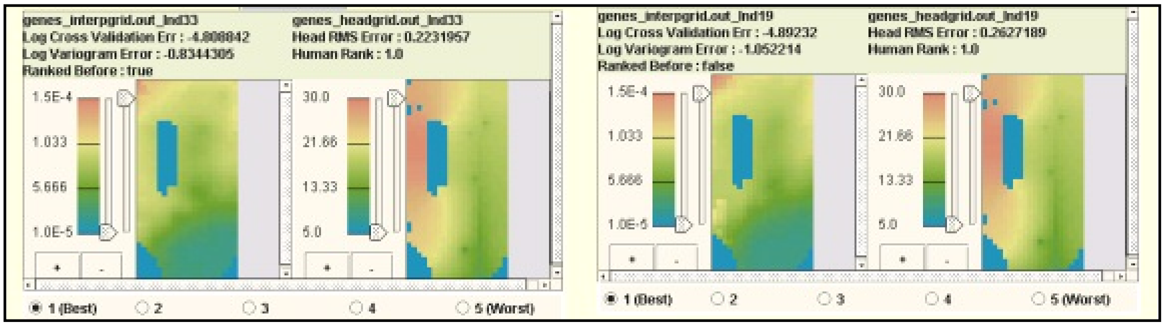


Figure 7: Comparison of the Pareto front for the final generation with 50% user interaction (solid diamonds, boxes, and triangles) and full user interaction (open diamonds, boxes, and triangles).

The best solutions with full and partial user interaction are compared in figure 8. The results show that the best solution found from the IMOGA with only half the user interaction is almost identical to the solution found with continuous user interaction. While less user interaction can lead to a difference in the final diversity of the trade-off curve (Figure 7), it is clear that the classification algorithm is still able to identify solutions with similar solution quality.



(a) **(b)**

Figure 8: Comparison of kriged field conductivity and head predictions for the best solution from a) IMOGA with full user interaction and b) IMOGA with half interaction.

Conclusions

This paper introduces an innovative interactive approach that aims at striking a balance between automated parameterization of models, which gives numerically optimal results, and the traditional trial and error approach that allows experts with site knowledge to incorporate their understanding of the site hydrogeology. This is done within a multi-objective framework that allows a range of Pareto dominant solutions to be found. Specifically for this example, adding variogram fit as a third objective allowed the discovery of solutions that did not fit the empirical variogram well but performed well in terms of estimation and prediction. Such a discovery would not have been possible if the variogram was 'fit' to the data beforehand in a typical geostatistical model fitting approach, which would preclude solutions with high variogram fit errors from the search.

This study also demonstrates that using the proposed interactive inverse modeling approach could lead to significant improvements in the solutions found. Automated parameterization results in parameter values that only honor the data at the observation points. The spatial structure of the parameters, which is of utmost concern in modeling groundwater flow and transport processes, can usually not be explicitly included in the solution search. Allowing the user to systematically interact with the optimization process allows 'user knowledge' to drive the search towards more 'realistic' solutions. However for this approach to be feasible, user interaction needs to be kept at a minimum. Using machine learning classifiers such as decision trees to learn and augment the user interaction is one way to deal with the issue of human fatigue. Initial results indicate that there is a trade-off between the amount of user-interaction and the final solution diversity. However, the solution quality in terms of the best solutions found with only half the user interaction was almost identical to that with continuous user interaction. This is a promising result and indicates that further research needs to be conducted in this direction. Future research will investigate the performance of different types of classifiers and how model error affects solution quality. The effect of the size and memory of the archive used to train the classifiers on the performance of the IMOGA is an important issue that also needs further analysis.

Acknowledgements

We wish to thank Dr. Albert Valocchi for his help in formulating the inverse problem, and Meghna Babbar for her work on the D2K IGA setup.

References

Babbar, M., B. S. Minsker, and H. Takagi, (2004) 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 2003 & Related Symposia, Salt Lake City, Utah.

Carrera, J., and S. P. Neuman (1986) Estimation of aquifer parameters under transient and steady state conditions, Maximum likelihood method incorporating prior information, *Water Resources Research*, 22(2), 199–210.

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

Doherty, J. (2003) Groundwater model calibration using pilot points and regularization, *Ground Water*; 41(2): 170-177.

Deutsch, C. V. and A. G. Journel (1998) GSLIB: Geostatistical Software Library and User Guide; Oxford University Press: New York.

Freyberg, D. L. (1988) An exercise in Ground-Water Model Calibration and Prediction, *Ground Water*, V. 26, No. 3, May-June.

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

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

Hill, M. C., R. L. Cooley, and D. W. Pollock (1998) A controlled experiment in groundwater flow model calibration, *Ground Water*, 36(3), 520–535.

Huijbregts, C. J. and G. Matheron (1971) Universal kriging—An optimal approach to trend surface analysis, in *Canadian Institute of Mining and Metallurgy, decision making in the mineral industry*, Spec. Vol. 12, p. 159–169.

Kitanidis, P.K. (1997) Introduction to geostatistics: applications to hydrogeology. Cambridge University Press, Cambridge ; New York, xx, 249 pp.

McDonald, M.G., and A. W. Harbaugh (1988) *A modular three-dimensional finite-difference ground-water flow model*. Techniques of Water Resources Investigations 06 A1, United States Geological Survey.

de Marsily, G. (1984) Spatial variability of properties in porous media: A stochastic approach, in *Fundamentals of Transport in Porous Media*, edited by J. Bear and M. Y. Corapcioglu, pp. 719–769, Martinus Nijhoff, Boston.

McLaughlin, D., and L. R. Townley, (1996) A reassessment of the groundwater inverse problem, Water Resources Research, 32(5), 1131–1161.

Quinlan, J. R. (1986), Induction of decision trees. *Machine Learning*, 1:81-106.

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

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. (2003) 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.

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

Sun, N.-Z. (1995) "*Inverse Problems in Groundwater Modeling*" Kluwer Academic Publishers, the Netherlands.

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

Tsai, F.T-C., N-Z. Sun, and W.W-G. Yeh (2003). Parameter structure identification in groundwater modeling: A universal parameterization method, EWRI World Water & Environmental Resources Congress, Philadelphia, Pennsylvania, June 23-26.

Yakowitz, S., and L. Duckstein (1980) Instability in aquifer identification: Theory and case study, *Water Resources Research*, 16(6), 1045–1064.

Yeh, W. W-G. (1986) Review of parameter identification procedures in groundwater hydrology: The inverse problem, *Water Resources Research*, 22(1), 95–108.

Welge, M. et al. (2003) 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).

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

Zimmerman, D. A., et al. (1998) A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resources Research*, 34(6), 1373–1413.