Advances in Water Resources 31 (2008) 1269–1283

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

An interactive multi-objective optimization framework for groundwater inverse modeling

Abhishek Singh *, Barbara S. Minsker, Albert J. Valocchi

Department of Civil and Environmental Engineering, University of Illinois, Urbana-Champaign, IL 61801, United States

ARTICLE INFO

Article history:
Received 7 April 2007
Received in revised form 28 April 2008
Accepted 7 May 2008
Available online 3 June 2008

Keywords:
Groundwater calibration
Inverse modeling
Interactive optimization
Multi-objective
Genetic algorithms
Pilot points
Regularization

ABSTRACT

The groundwater inverse problem of estimating heterogeneous groundwater model parameters (hydraulic conductivity in this case) given measurements of aquifer response (such as hydraulic heads) is known to be an ill-posed problem, with multiple parameter values giving similar fits to the aquifer response measurements. This problem is further exacerbated due to the lack of extensive data, typical of most real-world problems. In such cases, it is desirable to incorporate expert knowledge in the estimation process to generate more reasonable estimates. This work presents a novel interactive framework, called the 'Interactive Multi-Objective Genetic Algorithm' (IMOGA), to solve the groundwater inverse problem considering different sources of quantitative data as well as qualitative expert knowledge about the site. The IMOGA is unique in that it looks at groundwater model calibration as a multi-objective problem consisting of quantitative objectives – calibration error and regularization – and a 'qualitative' objective based on the preference of the geological expert for different spatial characteristics of the conductivity field. All these objectives are then included within a multi-objective genetic algorithm to find multiple solutions that represent the best combination of all quantitative and qualitative objectives. A hypothetical aquifer case-study (based on the test case presented by Freyberg [Freyberg DL. An exercise in ground-water model calibration and prediction. Ground Water 1988;26(3)], for which the 'true' parameter values are known, is used as a test case to demonstrate the applicability 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. Adding expert interaction is shown to not only improve the plausibility of the estimated conductivity fields but also the predictive accuracy of the calibrated model.

1. Introduction

Many groundwater modeling applications require the use of complex flow and transport models that have scale-dependent and spatially distributed parameters such as hydraulic conductivity/transmissivity, recharge, evapo-transpiration, storativity, aquifer dimensions, boundary/initial conditions, etc. Among these, hydraulic conductivity is one of the most important parameters for flow and transport models, controlling not only hydraulic head responses of the modeled aquifer but also transport characteristics of sub-surface contaminants. Characterization of this parameter is a critical and challenging part of subsurface modeling. However, there are almost never enough measurements to adequately (with respect to the predictions of the model) characterize the aquifer's conductivity field. This difficulty occurs because: (1) conductivity is known to be highly heterogeneous – often varying over many orders of magnitude within a small region, (2) direct (and even indirect) measurements of sub-surface properties are typically sparse and may be contaminated with noise in most real-world applications, and (3) even when such measurements *are* available they may be at scales different from those required for prediction (for example, calibrating a transport model based on spatially integrated pump tests can lead to non-unique parameter sets and erroneous prediction). In such cases most modelers resort to 'inverse modeling' – finding one or more conductivity fields such that the output of the modeled aquifer is as close as possible to field measurements of aquifer response (such as hydraulic heads, flux, or tracer concentrations) that are in general easier to obtain. Inverse modeling has been the area of extensive research and the interested reader is referred to Yeh [48], Sun [39], Mclaughlin and Townley [26], Zimmerman et al. [49], Marsily et al. [22], Carrera et al. [5] for details on the theory and practice of inverse groundwater modeling. The groundwater inverse problem is well-known to be an 'ill-posed' problem [4,39,47,49]. This means that the estimated conductivity field is often non-unique and highly sensitive to small changes in head measurements. In such cases minimizing the difference between the predicted aquifer response and the measured aquifer response is not sufficient to

* Corresponding author. Tel.: +1 2177210301.
E-mail address: asingh8@gmail.com (A. Singh).

0309-1708/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2008.05.005

guarantee a unique and stable conductivity estimate. Due to this non-uniqueness there is no guarantee that the estimated conductivity field estimate is close to the 'true' conductivity, leading to unreliable and potentially erroneous model predictions. In other words, reducing the calibration error in itself is not sufficient and we need an additional objective that ensures the conductivity estimate is close to reality. Neuman [29] was the first to pose the inverse problem as a multi-objective optimization problem where the calibration error is minimized along with a 'physical plausibility criterion'. Neuman [29] defines this second objective as a measure of how close the solution is to reality or an 'objective or subjective estimate of the expected (in a statistical sense) parameter values as obtained from field test and other sources'. Studies such as McLaughlin and Townley [26] and Marsily et al. [22] have pointed out that this is consistent with a Bayesian formulation, which conditions the posterior estimate of the conductivity on a prior conductivity field. Of course, since the true parameter field is unknown, the challenge is to define an appropriate plausibility heuristic that measures how far the parameter is from reality. It is in this context that 'prior information' becomes of utmost importance to inverse modeling. Studies, including Woodbury and Ulrych [45], Marsily et al. [22], Kowalsky et al. [17], Alcolea et al. [1], have discussed and demonstrated the importance of prior information for inverse modeling.

Prior information itself can be of three types. The first type consists of direct measurements of the parameters to be estimated. Such 'hard' data provides a direct measure of the true conductivity field and should be incorporated in inverse modeling. The second type of prior information – known as 'soft' or 'indirect' data – consists of indirect indicators such as topology, land-use, and geophysical images that can be used to give some information about the parameter of interest. In this study our primary concern is a third type of prior information – expert knowledge about the hydrogeology of the site. Due to sparse and noisy field measurement data, expert knowledge is often critical in model calibration and selection. Experts may have knowledge about the geological and sub-surface characteristics of the site (such as predominant rock or soil types, geological history of the site, existence of high conductivity flow channels, etc.) that may not be reflected in the field data. Direct and indirect prior information has been incorporated into inverse modeling using Bayesian approaches [17,26,45] or by imposing preferred value 'regularization' constraints and objectives [1,28]. Though both these approaches provide frameworks to include expert knowledge, this requires the expert to express essentially qualitative knowledge in purely quantitative terms. Bayesian techniques require the expert to represent their knowledge of the site in terms of a 'prior' distribution of conductivity fields – something that the experts may not feel comfortable or confident doing. Regularization and calibration objectives require the expert to express their knowledge in terms of low-order metrics (such as sum of squared residuals from measurement data), objectives, and constraints. Such representation will always suffer from what is known as the 'blindness of norms' – the loss of information when multi-dimensional information is represented by a low-dimensional statistic or norm; using such metrics to assess the fit of the model with multiple calibration data points may further compound the non-uniqueness problem. In fact, Brill et al. [3] have demonstrated that optimizing a multi-dimensional problem in a lower dimensional space leads to loss of solutions that would be optimal in the higher-dimensional space. They considered a hypothetical three-dimensional problem and solved it successively using one and two objectives. Most of the optimal solutions for the three-objective problem projected into the inferior region for the two-dimensional objective space. This was the motivation for their proposal of a 'joint cognitive' approach where computers were used to generate multiple optimal (albeit in the low-dimensional objective space) alternative solutions that human experts could search to find solutions that are optimal in the higher dimensional domain.

Finally, expert knowledge is dynamic and adaptive and can evolve as the expert is allowed to interact and learn from the model. Thus, there is a need for a qualitative and adaptive environment that allows site experts to give maximum information during inverse modeling, while simultaneously evaluating and updating their own understanding of the model.

In this study we propose a novel interactive framework to adaptively incorporate qualitative expert knowledge in the inversion process. The proposed methodology is based on the concept of 'interactive optimization' [41], where responses from users are used to drive the optimization search. Such an interactive framework is essentially similar to the 'joint cognitive systems (JCS)' proposed by Woods et al. [46] and Brill et al. [3], combining the cognitive abilities of humans with the computational powers of the computer to solve complex real-world problems. In the case of the groundwater inverse problem expert knowledge is used to assess plausibility of conductivity fields, while the computational powers of the computer are used to optimize the quantitative calibration objectives and generate the best alternatives for the expert to evaluate. Interactive optimization has been successfully applied to many real-world applications [41] including geological modeling [44]. This work, however, represents the first attempt to investigate its application to groundwater inverse modeling. Moreover, our approach introduces a very important distinction from previous interactive optimization approaches: instead of using an over-riding 'human objective' (as used by [41,44]), we treat quantitative and qualitative criteria as multiple objectives within a 'Pareto optimization' framework. This means that both quantitative and qualitative criteria are treated as equally important and the goal of the optimization algorithm is to find a diverse set of solutions that represents the best trade-off among all of the objectives (for further information on multi-objective Pareto optimization see reviews by Deb [7,8,30]). Such an approach is especially appropriate for the groundwater inverse problem where quantitative objectives, such as calibration error, and agreement with measured prior data are as important as qualitative expert preferences. In addition, we believe an important advantage of such a multi-objective approach is that it lays the framework for the modeler to explicitly evaluate the trade-off between his or her conceptual understanding of the site and different sources of field information that feed into the modeling process. With the proposed methodology, which we call the interactive multi-objective genetic algorithm (IMOGA) it is possible to generate multiple alternatives for the parameter field, each representing an optimal trade-off between different quantitative and qualitative criteria. The goal of this paper is to present, implement and test such an algorithm for a hypothetical case involving estimation of hydraulic conductivity, and demonstrate that with such an approach it is indeed possible to learn qualitative preferences from the user and include this information in inverse modeling. For this study we focus on the inverse problem involving estimation of hydraulic conductivity given measurements of hydraulic heads.

The rest of the paper is organized as follows. First, we discuss a synthetic case study from the literature that we use to demonstrate the applicability of interactive optimization for inversion. Next, the parameterization and formulation of this inverse problem is presented along with the methodology for inverse modeling with IMOGA. Finally, the results for the synthetic case study are presented, followed by conclusions and future research directions.

2. Methodology

The first step in solving the groundwater inverse problem involves parameterization – defining the decision variables for the inverse problem – and formulation – defining the objectives and constraints for the optimization problem. The process of parameterization defines the search space for inverse modeling (see McLaughlin and Townley [26] for a comparison of different parameterization schemes and Tsai et al. [40] for a generalized parameterization scheme for the groundwater inverse problem). In our case the parameterization chosen is based on the concept of 'pilot points'. Details of the pilot-point implementation are given in Section 2.1.

Once the parameterization has been selected, the next step is to formulate the objectives for the inverse problem. For this study, the quantitative objectives are given by the calibration error – a measure of the difference between the model predictions and measurements of aquifer response – and a physical plausibility (regularization) objective – a measure of how far the conductivity field deviates from direct field measurements of conductivity. Finally, the user's subjective judgment of the plausibility of conductivity fields is treated as a third *qualitative* objective. Details of the calibration error, the regularization objective, and the subjective qualitative objective are given in Section 2.2.

Given the decision variables and the quantitative and qualitative objectives, a two-stage optimization strategy is used to find the best conductivity fields. In the first stage, the inverse problem is solved using only the quantitative objectives (calibration error and regularization) using an efficient optimization strategy, ensuring that these objectives are adequately optimized. The interactive phase then starts when the non-interactive stage ends and allows the expert to further optimize the solutions by incorporating his or her subjective knowledge, while still ensuring good performance on the quantitative objectives. Details about the interactive optimization framework are given in Section 2.3.

2.1. Parameterization

The parameterization used for this study is based on pilot points [10,31]. Since its introduction by Marsily [21], pilot-point inversion has become increasingly popular because of its flexibility in representing the heterogeneity of the field without too many restrictive assumptions. Pilot points have been applied to a vast domain of problems [15,42] and have been shown to work for complex and highly heterogeneous conductivity fields [10].

Pilot-point inversion starts with a given set of control (or pilot) points over the spatial domain. The values at the control locations are interpolated (using an interpolation algorithm such as kriging) over the entire spatial domain, giving a parameter field. The values at these control locations (and consequently the parameter field) are then optimized by minimizing the calibration error of the resulting field.

Two approaches exist when locating pilot points. The first locates a pilot point based on the sensitivity of the calibration target for those locations [19,20,31]. An alternative approach (proposed by Gomez-Hernandez et al. [13] and Doherty [10]) specifies a large number of pilot points over the entire domain, making it possible to capture complex spatial features (guidelines for choosing pilot points are given by Doherty [10]). However, with too many pilot points there is the risk of over-parameterization and instability and additional 'regularization' criteria need to be used to make the problem stable [1,10,28]. The regularization process begins with a large number of parameters (as many as the modeler deems appropriate for representing the heterogeneity of the underlying aquifer) and then constrains the variability in many of these parameters by 'regularizing' them either with respect to prior field information or through the principle of parsimony (see [27] for a detailed discussion of regularization for pilot points).

For this study we use the method proposed by Doherty [10] with a large number of pilot points. We assume the conductivity field is log-normal and log-transform the conductivity measurement (even though the methodology is presented for log-conductivity estimates it applies without any loss of generality to other transformations). Ordinary kriging [9] is used to interpolate the pilot-point values *and the conductivity field measurements* to get the final conductivity field, ensuring that the resulting field honors the conductivity measurements exactly, deviating from the kriged field only at the pilot-point locations (for example, Fig. 1 shows two fields kriged from the same K data with different pilot-point values – thus honoring the field measurements perfectly but differing at the pilot-point locations). Details of the kriging methodology are given in Appendix A.

The conductivity field obtained from kriging the data and the pilot points can then be used to predict hydraulic heads using groundwater flow model (MODFLOW [23] in this case). These predictions are then used to calculate calibration errors for inverse modeling. Different values for pilot points give different conductivity fields all of which perfectly match the field measurements but give different calibration errors and deviate from the 'ground truth-data' conductivity field to different degrees. The goal of the optimization can be defined as finding the conductivity field(s) with best calibration error and minimum deviation from the prior field.

2.2. Calibration formulation

We follow the multi-objective formulation proposed by Neuman [29] and optimize the conductivity field using *both* the quantitative objectives of calibration error and plausibility. Given direct measurements of the conductivity these can be used to formulate a 'quantitative' plausibility criterion (the traditional regularization term in inverse frameworks such as Doherty [10]). In addition, the subjective information that the field expert has provided represents an *additional* plausibility objective that reflects his or her qualitative preferences for certain geospatial characteristics of the aquifer.

The calibration objective consists of minimizing calibration error with respect to field measurements of hydraulic heads. This is given by the weighted root mean square error between measured field data and the prediction of the groundwater flow model for a given conductivity field (Eq. (1)):

$$\underset{\mathrm{PP}\in R^p}{\mathrm{Min}}\, H_{\mathrm{err}} = \left(\frac{[H\langle K_{\mathrm{PP}}\rangle - H^{\mathrm{obs}}]^T C_{\mathrm{err}}^{-1} [H\langle K_{\mathrm{PP}}\rangle - H^{\mathrm{obs}}]}{n_{\mathrm{obs}}} \right)^{1/2} \tag{1}$$

where PP is a vector of p pilot points, H^{obs} is the head measurement vector, C_{err} is the head measurement error-covariance matrix, and $H\langle K_{\mathrm{PP}}\rangle$ is a vector with groundwater head predictions (from a model such as MODFLOW) using the conductivity field, K_{PP}, interpolated from the pilot points (PP). It is typical to assume that C_{err} is dependent on the measurement accuracy of the head, and these errors are uncorrelated and unbiased – making the C_{err} matrix diagonal, with the diagonal terms representing the measurement error variance. If information is available on measurement error correlation, the off-diagonal terms can easily be added to the matrix.

As discussed earlier, when using a large number of pilot points it is necessary to impose some form of 'regularization' [1,28] to reduce unwarranted oscillations in the pilot-point values. Moore and Doherty [28] and Alcolea et al. [1] have proposed a 'preference-based' regularization scheme that aims at minimizing the deviations of the pilot-point values from some prior conductivity field

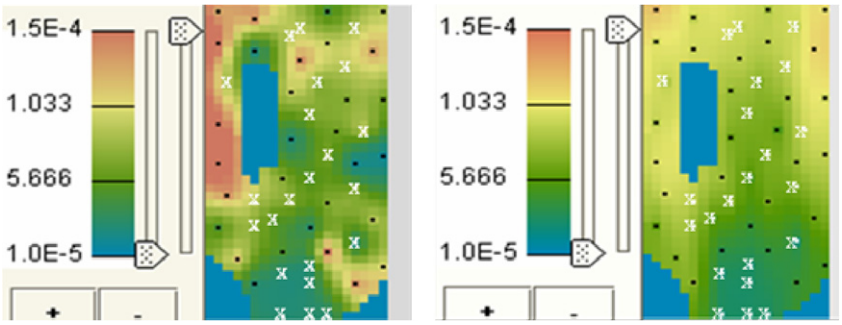

Fig. 1. Two conductivity fields that fit the measured values (white crosses) exactly, but have different values for the pilot points (black dots) resulting in different spatial fields.

(that can either be specified by the expert or be obtained from existing field data). This preference-based regularization is essentially a quantitative measure of the plausibility of the pilot-point values with respect to prior information. For the preference based regularization objective it is first necessary to come up with a prior conductivity field as given by the measurement data. To do this, the conductivity measurements are kriged using the ordinary kriging Eqs. (1) and (2) (except that this time only n data points are used). This gives an estimate of the log conductivity $\mathrm{Ln}(K'_{pp})$ at each of the pilot points. The regularization objective K_{err} is given by the weighted root mean square of the difference between the pilot-point values and the kriging estimate at those locations (Eq. (2)):

$$\underset{PP\in R^p}{Min}\, K_{err} = \left(\frac{[\mathrm{Ln}(K_{pp}) - \mathrm{Ln}(K'_{pp})]^T C_{pp}^{-1} [\mathrm{Ln}(K_{pp}) - \mathrm{Ln}(K'_{pp})]}{n_{pp}} \right)^{1/2} \quad (2)$$

where $\mathrm{Ln}(K_{pp})$ is the value at the pilot points (this is the decision variable for the optimization algorithm), $\mathrm{Ln}(K'_{pp})$ is the prior estimate at the pilot-point locations based on field data (in this case, based on kriging of the measurements of K), n_{pp} is the number of pilot points, and C_{pp} is the estimation covariance for the pilot-point locations. While using the field measurements to estimate K'_{pp} (kriging estimate at the pilot points) care should be taken to ensure that the field data are at the appropriate scale – i.e. if the pilot points represent spatial averages for the conductivities at the model grid scale then the kriging estimate at the pilot points should also be at the same scale (note that this may require up-scaling or down-scaling the field data used for kriging). An important distinction between this study and previous work (such as [10]) is that instead of the typical diagonal matrix (with the diagonal elements specifying the kriging estimation variance) we consider C_{pp} to be a full matrix specifying the *estimation covariance* – thus ensuring that the pilot-point values fit the specified covariance structure to the extent possible. The estimation covariance between the estimate at pilot point i and pilot point j given by C_{pp}^{ij} is calculated by the equation below:

$$C_{pp}^{ij} = \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} \lambda_\alpha \lambda_\beta \mathrm{Cov}(u_0^\alpha - u_0^\beta) + \mathrm{Cov}(u_{pp}^i - u_{pp}^j) - \sum_{\alpha=1}^{n_i} \lambda_\alpha \mathrm{Cov}(u_0^\alpha - u_{pp}^j) - \sum_{\alpha=1}^{n_i} \lambda_\alpha \mathrm{Cov}(u_0^\alpha - u_{pp}^j) \quad (3)$$

where C_{pp}^{ij} is the covariance between the estimation error at the location of pilot point i (located at u_{pp}^i) and the estimate at pilot point j (at u_{pp}^j), Cov is the model covariance function used for kriging, λ_α is the kriging weight and u_0^α the location of the αth data point used for kriging at pilot point i, λ_β is the kriging weight and u_0^α the location of the βth data point used for kriging at pilot point j, and n_i and n_j are the number of data points used for kriging at pilot-point locations i and j. The derivation of this covariance equation is given in Appendix B. The first term in equation 6 gives the covariance between all of the data points used for kriging the priors for the two pilot points, the second term gives the expected covariance for the distance between the two pilot points, and the third and the fourth terms give the covariance of the data point used for kriging one pilot point with the location of the second pilot point. Note that this formula reduces to the familiar kriging estimation variance for $i = j$ [9]. Weighting the objective function with such a covariance matrix ensures that (a) pilot points located close to data points are highly correlated to those values and (b) pilot points are more correlated with nearby pilot points than with more distant points. Kriging assumes a geostatistical model for the aquifer field, which needs to be inferred from the field data for the conductivity.

Note that although this formulation assumes that there are measurements available of the conductivity field, it is applicable even when such data are not present. In such cases K'_{pp} could be based on a prior field obtained through expert elicitation. If such prior information is not available then the regularization scheme in Eq. (2) reduces to a smoothness objective, with K'_{pp} corresponding to a spatially uniform value taken to be the average of all of the pilot points (in this case K_{err} simply minimizes the variance in the pilot-point values). Also note that the specification of the estimation error covariance matrix is such that it is easy to incorporate uncertainties in the conductivity measurements – the errors would essentially show up as higher nugget values for the variogram model used to specify Cov in Eq. (3). In effect this would constrain the conductivity field less at the measurement locations and allow the head error to have relatively more dominance over the estimation process.

In addition to the two quantitative objectives (given by Eqs. (1) and (2)), there is an additional qualitative objective that is given by the human expert. The qualitative objective is obtained by showing the kriged conductivity fields and the predicted head field to the user along with information about the numerical objectives. The expert then ranks different solutions based on how realistic they seem based on his or her knowledge about the site. This process is discussed in further detail in the next section.

2.3. The IMOGA

To solve the formulation given in Section 2.2, the optimization process consists of two phases: non-interactive and interactive multi-objective optimization. The non-interactive phase is similar to other groundwater inverse methodologies such as Alcolea et al. [1] and Moore and Doherty [28] except for one important difference. Previous approaches have either combined the calibration objective with the regularization objective using a weighting scheme [1], or specified one as a constraint and used the other as an objective for optimization [28]. The constrained optimization problem is then solved using a non-linear optimization algorithm.

In our framework, the two objectives of regularization and calibration are treated *independently* within a Pareto-optimization framework. Such an approach converges to the non-dominated solution set (solutions that can not be improved on one objective without worsening the other), allowing the expert to evaluate and select the best trade-off between calibration error and physical plausibility of the conductivity fields. An added advantage of this multi-objective approach is that it does not require the prior specification of weights, penalties, or constraint levels for optimization.

For this framework we use an efficient multi-objective genetic algorithm called the elitist non-dominated sorting genetic algorithm (NSGA-II) [8] to find the Pareto optimal set of solutions. Genetic Algorithms (GAs) are heuristic search algorithms that work with a collection ('population') of possible designs. The designs in the population are evolved using an evolutionary process modeled after natural selection using evolutionary operators such as selection, crossover, and mutation to converge the population towards globally optimal solutions. GAs have been shown to effectively solve the groundwater inverse problem – examples include Solomatine et al. [38], Karpouzos et al. [18], Tsai et al. [40], among others.

Since GAs work with populations of individuals they are ideal for multi-objective (Pareto) optimization and have been used extensively for multi-objective water resources optimization applications (for example [6,14,16,32,34,36] among others). Multi-objective GAs use the same GA operators of selection, crossover, and mutation except that here instead of taking one objective as the driving 'fitness' for convergence, they take the Pareto optimality of a particular set of solutions as the GA's fitness function. The NSGA-II measures Pareto optimality using 'Pareto ranking' [12], in which each individual that has equal or higher fitness in all objectives is labeled a first-order Pareto-optimal (or rank 1) solutions. These are then removed from the ranking process and rank 2 solutions are defined as those that are strictly non-dominant with respect to the remaining solutions, and so on until all solutions in the population are ranked. To preserve diversity and ensure that the solution set has good coverage across the entire Pareto front, The NSGA-II algorithm then computes a metric called 'crowding distance,' which is defined as the Euclidean distance (in objective or decision space) between its two adjacent neighbors on the same Pareto front as the solution. During the selection process, the NSGA-II first uses the rank of the solution to select one individual over another. For solutions with the same rank, the solution with the larger crowding distance is preferred. The NSGA-II also utilizes elitism to preserve the best solutions in future generations and a convergence criterion of a maximum number of generations or when no new solutions are added to the population. Due to the large dimensionality of the search space this work also uses a real-coded parameter representation for the NSGA-II.

During this first non-interactive optimization phase, the IMOGA runs as a regular real-coded NSGA-II. At this stage a large population size and number of generations are used to effectively solve for the quantitative objectives. For this study, the crossover and mutation rates were set based on trial runs with small population sizes and a few generations. The population size for the non-interactive run was set using the iterative scheme given by Reed et al. [33], where the population size is successively increased and each successive NSGA run is seeded with solutions from the previous run, until larger populations give no further coverage across the optimal Pareto front. Details for the GA runs are provided in the results section.

Once the non-interactive session has converged, the expert then examines the estimated conductivity fields and the calibration targets for the optimal solutions. If the expert feels there are certain aspects of the estimated parameter fields that have not been captured or have been misrepresented, he or she initiates the next phase – the interactive multi-objective optimization. Fig. 2 gives

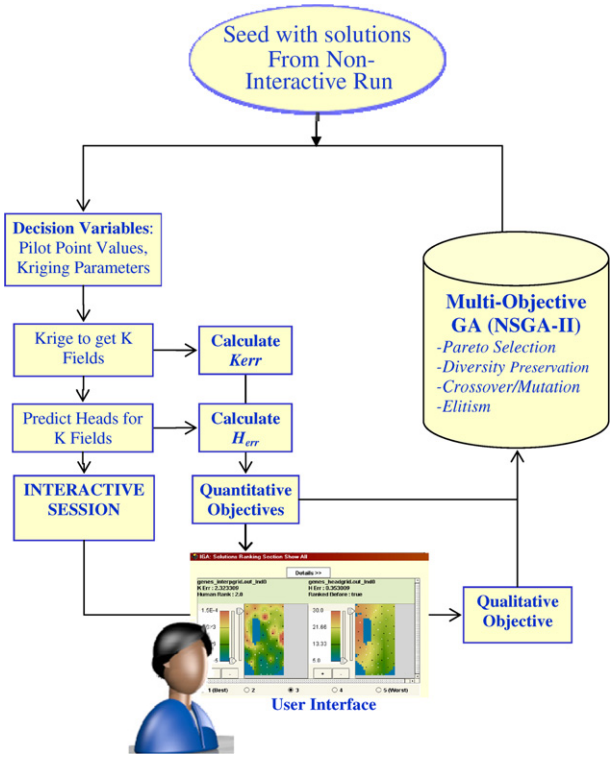

Fig. 2. IMOGA framework for interactive multi-objective groundwater inversion.

an overview of the IMOGA framework. The starting population of the IMOGA is initialized with solutions from the optimal trade-off front of the non-interactive run, as well as other randomly generated chromosomes. This ensures that the expert has a good starting set of solutions to search from, while the GA also has a diverse initial set of 'building-blocks' [12] to effectively explore the region around the quantitatively optimal solution space. It is noteworthy that this two-phase approach is consistent with the 'joint cognitive' philosophy of Brill et al. [3] who reasoned that *'if the most important dimension or dimensions are included in the model* (or mathematical formulation)*, then the problem solution*(s)*is likely to be in the region of the optimal solution'* of the quantitative formulation [3].

The operators for the IMOGA are the same as the NSGA-II (see box labeled 'Multi-Objective GA' in Fig. 2), except that in addition to the two quantitative objectives (given by K_{err} and H_{err} in this case) the IMOGA also uses an additional *qualitative* objective to find the best trade-off between both quantitative and qualitative criteria. The qualitative objective is initiated through an interactive session (Fig. 2) where the expert is allowed to visually inspect the current conductivity and head fields through a graphical user interface. Fig. 3 shows a typical user interface, in which the kriged conductivity and predicted hydraulic head field (corresponding to that conductivity field) are shown side by side. Additional information such as the numerical objectives (labeled as H_{err} and K_{err} in the grey panels above the images for conductivities and head fields), previous human rank of the solution (set to 0 if the solution has not been previously ranked), and other case-specific details are also shown to the expert. The user gives these solutions a rank from 1 (best) to 5 (worst). For this work, this interactive session was included in every IMOGA generation. However, for more complex field-scale application, requiring larger population sizes and more generations, user–fatigue due to prolonged interactions can become a challenge. While not presented in this paper, reducing the number of user evaluations is a major effort of this research, and has been discussed briefly in Section 5. The multi-objective

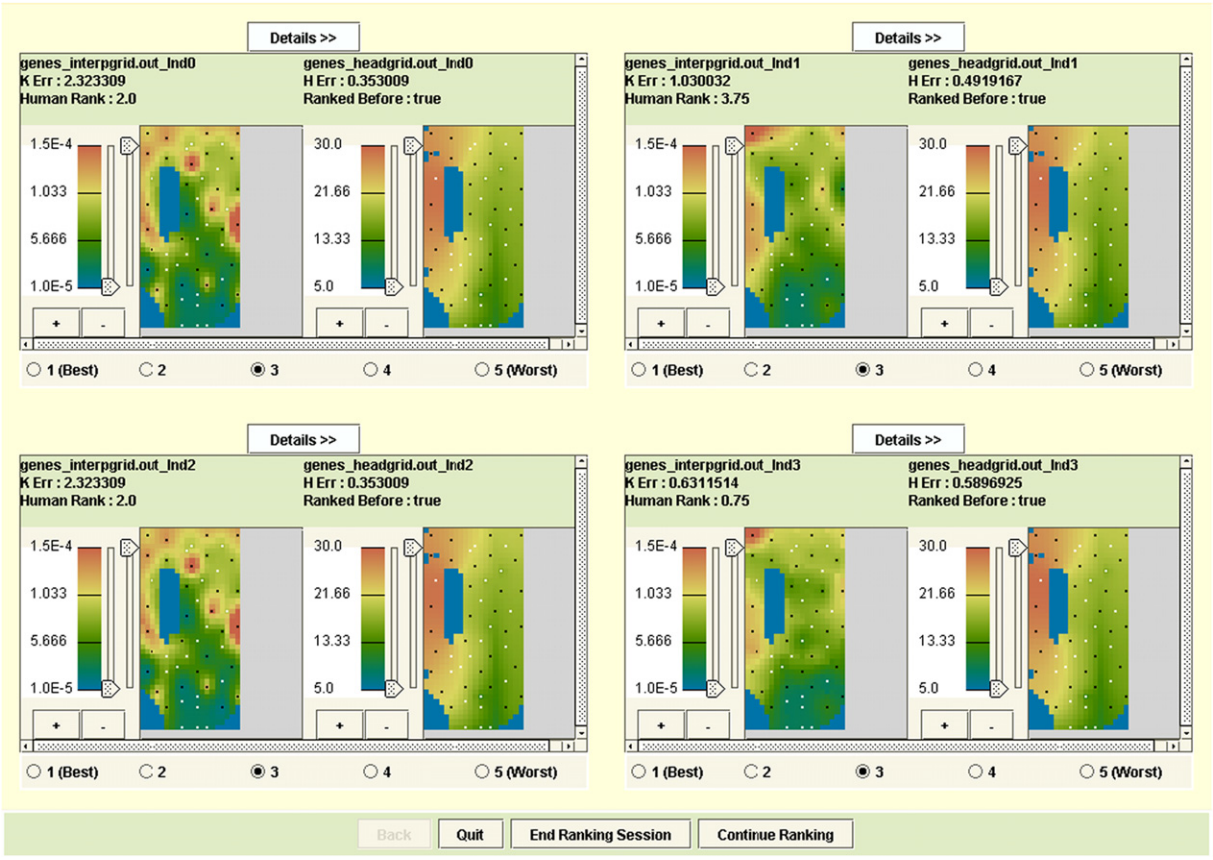

Fig. 3. Ranking panel for the IMOGA (each panel displays four unique IMOGA solutions with conductivity on the left and hydraulic head field on the right).

optimization algorithm (such as NSGA-II) then minimizes all three objectives – head error, regularization error, and human rank – to identify the best trade-off among all (quantitative as well as qualitative) criteria.

It is worth noting, that this multi-objective approach is different from earlier interactive optimization approaches that either considered only the human objective, or used it as an over-riding objective to select solutions from the quantitative trade-off curve. Both these approaches place the entire onus of the optimization search on the user, neglecting the quantitative objectives, which for the inverse problem are also very important.

The IMOGA framework has been built within the Data 2 Knowledge (D2K) system, supported by the Automated Learning Group (ALG) at the National Center for Supercomputing Applications (NCSA), IL, USA. See Welge et al. [43] for more details on the D2K software.

3. Case-study

The case study for this project is based on the example presented by Freyberg [11] and shown in Fig 4. This test case was chosen because Freyberg [11] has shown that the inversion of conductivity based on head observations is an ill-posed problem with head calibration error having possibly little correlation with the accuracy of the calibrated hydraulic conductivity field or with the accuracy of post-calibration predictions. Since numerical calibration led to unsatisfactory results this test case is suitable for testing the importance of expert interaction in improving the calibration process. The aquifer is considered as 'ground truth' and all observations are taken from this aquifer.

The hypothetical aquifer is a shallow, 2-D, unconfined aquifer 10 km long and 5 km wide underlain by an impervious stratum with no leakage. The aquifer is discretized with 20 columns and 40 rows with each cell measuring $250 \times 250\ m^2$. The hydraulic conductivity is assumed isotropic and steady-state conditions are considered. 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. The boundaries and outcrop are shown by the blue areas in Fig. 4. Uniform and constant recharge is assumed. There is a river with a steady, non-uniform stage that runs north to south across the aquifer (shown by the dashed lines in Fig. 4). The hydraulic conductivities and bottom elevations for the modeled aquifer were taken from the Freyberg [11] paper. As can be seen in the figure the hydraulic conductivities are highest in the northwest, with a decreasing trend to the southeast.

There are six wells pumping at constant rates, and 16 observation wells spread across the aquifer, giving a total of 22 observation points, shown by the white crosses in Fig. 4. The locations and values for the conductivities and heads are shown in Table 1. The hydraulic heads in this hypothetical aquifer were calculated using MODFLOW [23]. 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 research only the hydraulic conductivities were considered unknown, while all other parameters (such as layer thicknesses, elevations, and recharge) and the boundary conditions 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 problem more realistic the head measurements were contaminated with noise from a normal distribution with zero mean and variance of 0.25 (less than 1% measurement error on average). The diagonal elements of the C_{err} term were set equal to this variance in Eq. (1).

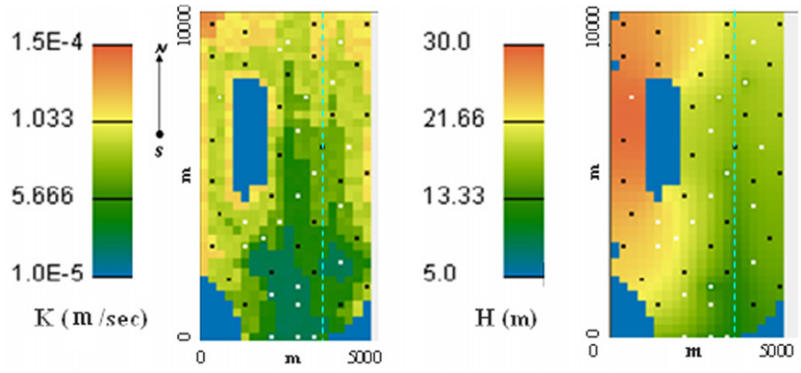

Fig. 4. Conductivities (K) and hydraulic heads (H) in the true aquifer. Blue areas indicate model boundaries and outcrops and the dashed line indicates a river running north–south in the model. The model is discretized with 20 columns and 40 rows with each cell measuring $250 \times 250\ m^2$.

Table 1
Locations and values of H and K measurements

Index	x (m)	y (m)	True H (m)	Measured H (m)	K (m^2/s)	Ln(K)
1	2625	9125	22.22158	22.12778	8.99625E − 05	−9.31612
2	4125	9125	19.69107	19.74297	8.70057E − 05	−9.34954
3	2375	8875	22.66976	22.47836	9.22049E − 05	−9.29150
4	3875	7875	16.29162	16.26502	8.98073E − 05	−9.31784
5	625	7375	28.83823	28.92293	8.56501E − 05	−9.36524
6	3125	7375	17.50686	17.76526	6.83011E − 05	−9.59158
7	2875	6375	18.44698	18.09578	6.93262E − 05	−9.57669
8	4375	5875	17.51641	17.25881	9.25927E − 05	−9.28730
9	3375	5125	15.22509	15.06429	7.02220E − 05	−9.56385
10	2875	4375	18.20994	18.25264	5.03435E − 05	−9.89664
11	4125	4125	16.07255	16.40875	6.83458E − 05	−9.59093
12	1375	3625	23.31628	23.80038	8.75577E − 05	−9.34321
13	2375	3625	20.02407	20.20937	7.33922E − 05	−9.51969
14	1875	3125	21.77313	21.97613	5.24470E − 05	−9.85571
15	1375	2875	22.78357	22.74787	7.24584E − 05	−9.53250
16	4125	2375	14.95987	14.93487	2.93851E − 05	−10.43502
17	2875	1625	11.85154	11.65154	2.88240E − 05	−10.45430
18	2125	1375	18.93550	19.05880	2.88149E − 05	−10.45462
19	2875	1125	14.96599	15.27539	3.07929E − 05	−10.38823
20	2125	125	15.60000	15.92400	2.94271E − 05	−10.43359
21	2875	125	13.00000	12.93050	3.05971E − 05	−10.39461
22	3375	125	12.00000	12.05430	2.99442E − 05	−10.41617

The conductivity values were log transformed before being used for inverse modeling. A total of 30 pilot-point locations were chosen (based on recommendations given by Doherty [10]) so that they had good coverage over the entire field. The pilot-point locations are shown by the black dots in Fig. 4. The formulation for the regularization error assumes a geostatistical model for the conductivity field. Details of the variographic analysis for the Freyberg dataset and the kriging parameters used for interpolation are provided by Singh [37].

The ultimate goal of inverse modeling is not just to minimize calibration error, but also to find realistic parameters so that the model has better predictive capabilities. To test and validate the calibrated model a predictive scenario was defined similar to the one used by Freyberg [11]. This consisted of predicting the aquifer response to reducing the conductance of the river bed by two orders of magnitude (this can be thought of as lining the river bed for a real site). Comparing the calibrated model's hydraulic head response with the 'true' response for this scenario provides information about the predictive capabilities of the calibrated model.

This test-case was solved using the two-phase multi-objective interactive approach discussed in Section 2. For the initial non-interactive NSGA-II a population size of 250 (the initial population size was set at 50 for the iterative population-sizing scheme [33] and increased up to 250, with increments of 50), tournament size of 2, crossover probability of 0.5, and mutation rate of 0.004 were used.

The interactive phase provides the expert an adaptive framework to incorporate his or her preferences and site-knowledge to improve these results. Since this was a hypothetical aquifer the expert (the first author) in some sense had complete knowledge of 'reality'. To make the human responses more realistic, the true conductivity field was only shown to the expert in advance (before both the non-interactive and interactive phases). The expert was not allowed to make direct comparisons between the true field and the conductivity fields shown by the IMOGA but rather was forced to work from memory. This is similar to real applications where experts may have secondary soft information in the form of maps and/or satellite images and also their own field experience to help them in decision making. Despite the fact that the case-study and expert-knowledge were 'synthetic,' this experiment serves as a proof of concept that subjective expert knowledge can be incorporated effectively with quantitative objectives.

For the interactive phase, small population sizes were necessary so that the user could evaluate the solutions without being overwhelmed by the amount of interaction. For this reason, we used a population size of 20 for 25 generations. Half the IMOGA population was initialized from the results of the non-interactive run and the other half was randomly initiated. This selection strategy was

chosen based on recommendations given by Babbar [2] on initializing interactive genetic algorithms. The IMOGA was run for 25 generations with the user evaluating every solution and ranking it on a scale of 1–5. To better explore the search space, the crossover and mutation rates were kept relatively high at 0.9 and 0.5, respectively. The total time required for the IMOGA is heavily dependent on the time taken by the user to inspect and rank solution images. In our experiments the entire IMOGA run could be completed in approximately 1 h (if the user was continuously ranking solutions). The results of these experiments are provided in Section 4.

4. Results and discussion

The results are given in four parts. Section 4.1 discusses the results obtained from the non-interactive multi-objective optimization. Section 4.2 then compares results from the interactive approach to the non-interactive results. Section 4.3 presents results for the predictive scenario (defined in Section 2) for both the non-interactive and interactive conductivity field estimates. Finally, to further explore the importance of user interaction for problems with sparse data, the observations for this case study were reduced from 22 to 17. The model was then calibrated with this reduced data set using both the interactive and non-interactive approach. These results are presented in Section 4.4.

However, before presenting these results it is important to consider the best inverse modeling solution possible for the case study considered in this paper. This solution, representing the closest approximation to the real conductivity field given the number and locations of pilot points available, is obtained from the true conductivity values at all of the pilot points. In essence this represents the solution we get if we have perfect knowledge of all of the decision variables for this optimization problem, which is the 'true' optimal solution to this groundwater calibration problem. Of course, this solution is only known for this case study because we are working with a hypothetical aquifer where the ground truth is known to us. The optimal pilot-point conductivity field is compared with the true conductivity field in Fig. 5. As expected, compared to the 'true' conductivity field, the optimal pilot-point field is smoother due to the effect kriging has on the interpolated field. However, the overall spatial characteristics of the true conductivity field are preserved. Since the optimal conductivity field in Fig. 5b is the best we can obtain with the current parameterization, it serves as a standard with which to compare the inverse modeling results.

4.1. Results for non-interactive optimization

Fig. 6 shows the optimal trade-off between the calibration error (as measured by H_{err}) and the regularization/plausibility objective (as measured by K_{err}) for the non-interactive optimization. It is worth noting that constrained single-objective calibration approaches using pilot points (such as [10]) would converge to one point on this trade-off front, depending on the constraint level set for one of the objectives. The graph also shows the calibration error and regularization objective for the optimal pilot-point solution shown in Fig. 5. The results shown in Fig. 6 give interesting insights into the tradeoff between model accuracy and plausibility. In general it is seen that highly regularized (more plausible) conductivity fields lead to higher calibration errors, while fields that are under-regularized have lower calibration errors (among others, [24] have discussed this problem of over-fitting for groundwater model calibration as well as the role regularization has in reducing it). Interestingly, the optimal pilot-point solution is not part of the final Pareto front, and is in fact dominated by the solutions found by the NSGA-II. This finding is not as surprising as it may seem at first and in fact forms the basis for the interactive approach. Recall that the head measurements taken from the aquifer had a small noise term (C_{err}) added to the head measurements. The dominance of the optimal pilot-point solution indicates that there are other solutions with the same level of regularization that have lower calibration errors. This is primarily due to the ill-posedness of the groundwater inverse problem where small variations in head can lead to very divergent conductivity estimates. In regions with good data support regularization constraints the conductivity values at their optimal levels, but in areas with little to no data support the optimization algorithm changes the conductivity field to better fit the noise in the head measurements, leading to (implausible) conductivity fields with lower calibration errors. Thus, with respect to the quantitative objectives of calibration error and regularization, the optimal pilot-point solution is perceived as sub-optimal and is not part of the final Pareto front. In such a case, optimizing the conductivity field based on the given objectives will not find the true optimal solution (recall this was the motivation for the 'joint cognitive approach' of Brill et al. [3] discussed in the introduction).

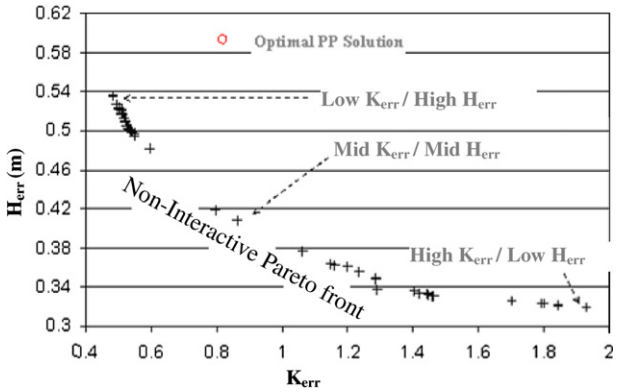

Fig. 6. Trade-off between calibration error and regularization objectives (for non-interactive optimization). The solutions visualized in Fig. 7 have been indicated on the Pareto front.

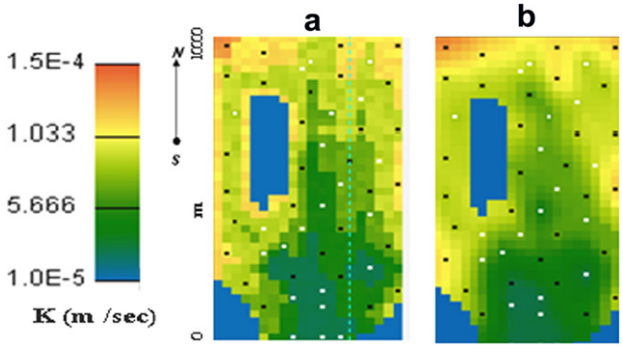

Fig. 5. True conductivity field (a) and optimal pilot-point conductivity field (b).

To better understand the results, it is necessary to investigate the actual conductivity fields found by non-interactive inversion. Fig. 7 compares conductivity fields for three solutions taken from the non-interactive trade-off front with the optimal pilot-point conductivity field. These represent solutions with low K_{err} and high H_{err} (high plausibility and high calibration error), medium K_{err} and H_{err}, and high K_{err} and low H_{err} (low plausibility and low calibration error). Some important issues related to inverse modeling are evident in Figs. 6 and 7. Notice that in terms of the spatial structure of the hydraulic heads, very different conductivity fields lead to almost identical head fields (the total range of head residuals is just 0.3 m from 0.25 to 0.55 m in Fig. 6) indicating the problem of non-

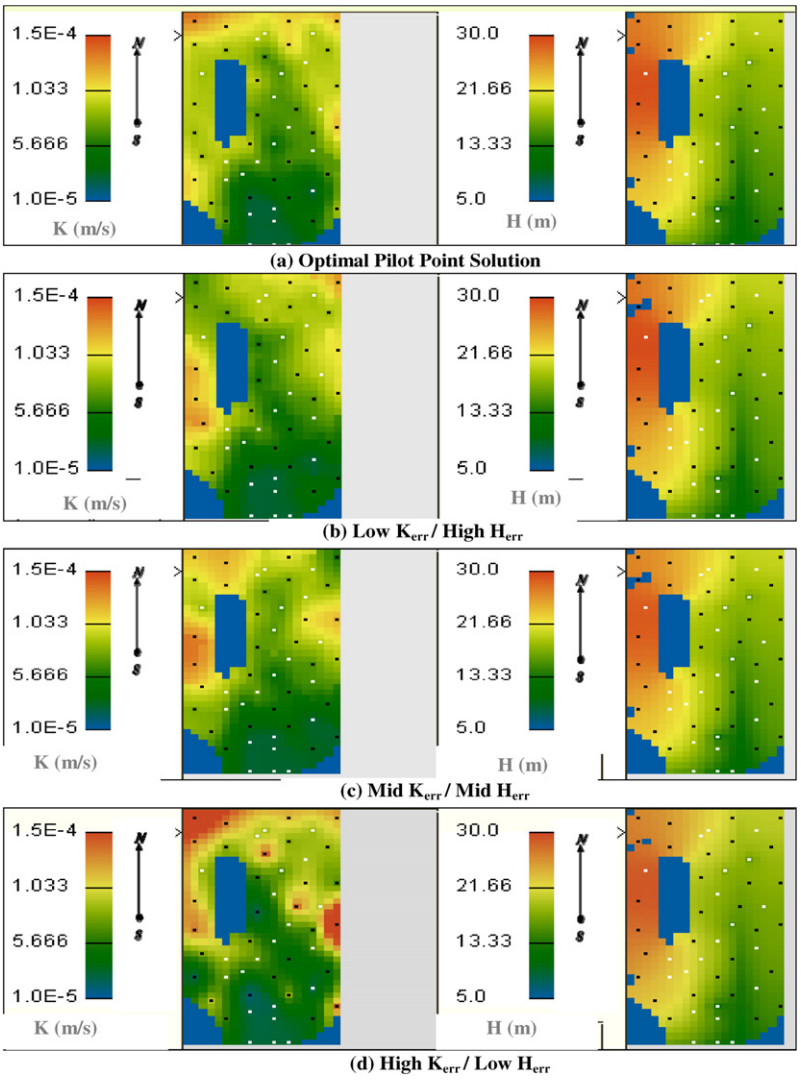

Fig. 7. Comparison of conductivity (left) and head (right) fields found by non-interactive optimization.

uniqueness for this example. We also see from Fig. 7 that as K_{err} increases (i.e. plausibility or regularization worsens) the conductivity fields become more heterogeneous because the pilot points are less constrained. However, the more 'calibrated' conductivity fields (i.e. solutions with low calibration error) are clearly different from the true conductivity field, indicating that for this test case (and expert), regularization and not calibration error is the more important objective. In this way, comparing the quality of different solutions along the trade-off front allows us to make decisions regarding the compromise between different objectives (something that we would not be able to do with the standard approach of using a fixed regularization or calibration error constraint). The solution closest to the true optimal field (Fig. 7a) is the most regularized solution with least K_{err} (Fig. 7b). However, even this conductivity field suffers from unrealistically high conductivities in the western portion of the site and low conductivities in the north-west portion of the aquifer. Unreasonably high conductivity values are observed in the west and north-west portions of the site for all other solutions along the non-interactive trade-off front. These are the areas with the least data support and thus the quantitative plausibility objective (as measured by K_{err}) is insufficient to ensure realistic conductivity values in these regions. The other difficulty is that the plausibility of the conductivity field is based on a two-dimensional spatial field. Any one-dimensional metric, such as the regularization term used here, is bound to lead to some loss of information.

4.2. Results from interactive optimization

The results from the non-interactive phase indicate that quantitative objectives can not guarantee sufficiently plausible conductivity fields. When the non-interactive results are fed into the IMOGA, Fig. 8 shows the resulting trade-off curve. In this case there are three objectives – the first two, K_{err} and H_{err}, being the same as before, and a human rank (from 1 to 5). The first two objectives are displayed on the x and y axis of the graph and the human rank is shown by different colored icons on the graph. Because IMOGA works with a much smaller population, the trade-off curve is sparser than the non-interactive results (Fig. 6). The graph also shows that there is some trade-off between the expert rankings and quantitative objectives. Thus, some low human rank (rank 3) solutions are on the quantitative Pareto front, and higher human-ranked solutions (ranks 2 and 1) are diverging from the quantitative Pareto front. In fact, rank 1 solutions (indicated by diamonds on the graph) are farthest from the quantitative Pareto front, and more interestingly the closest to the 'true' optimal solution, indicating that these conductivity fields are closest to 'reality'.

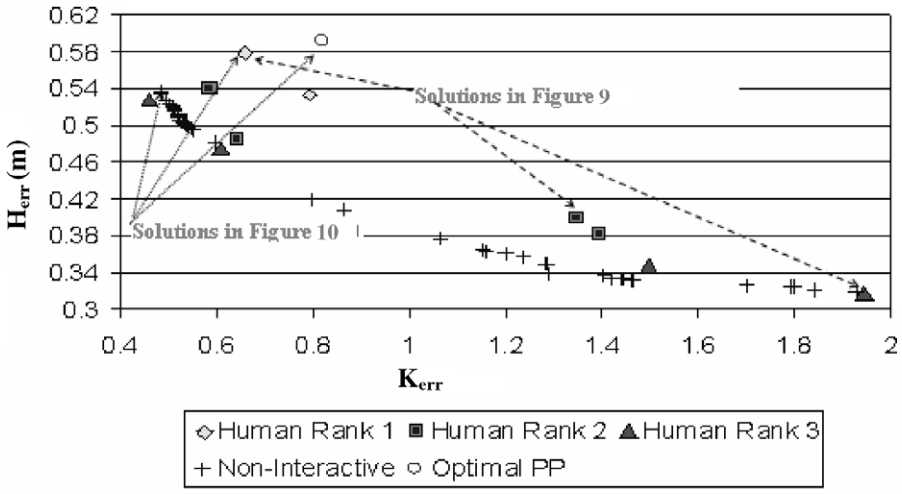

Fig. 8. Trade-off curve for IMOGA and non-interactive optimization.

As mentioned earlier (see Section 1), this issue has been pointed out by Brill et al. [3] when discussing the need for interactive systems. When solving multi-criteria problems, many solutions that are optimal in higher dimensional space are perceived as sub-optimal in low-dimension objective space and are lost when optimizing with the low-order objectives. It is important to note that the multi-objective approach considers both quantitative and qualitative objectives and thus some solutions on the quantitative Pareto front are also retained.

These differences in objective space are also reflected in the solution space of the conductivity fields. Fig. 9a–c shows three solutions from the interactive Pareto front with different human ranks (these are shown on the Pareto front in Fig. 8). It can be seen that while rank 1 solutions have the spatial structure preferred by the expert, rank 2 and rank 3 solutions deviate from the expert-preferred structure and are closer to the solutions found only by the non-interactive run. Thus, in one step the IMOGA has been able to find solutions that represent different levels of compromise between spatial structures found plausible by the expert and by the quantitative plausibility measure. This is important information as it gives the expert the relative value of the field data with respect to their own understanding of the site and model. It also safe-guards this approach from possible biases of the expert by establishing conflicts between expert knowledge and field data. In a situation such as this one, the expert is forced to make a conscious and considered decision about the type of conductivity structure appropriate for the model.

Finally, Fig. 10 compares the conductivity field of a Rank 1 solution (the solution found most plausible by the expert shown by Fig. 10c), the conductivity field with the lowest regularization/plausibility error from the non-interactive phase (Fig. 10b), and the true optimal pilot-point solution (shown earlier in Fig. 5b and shown again in Fig. 10a). Note that both (10b andc) these solutions have very similar calibration errors (0.58 m for the interactive solution and 0.54 m for the non-interactive solution). The interactive solution does not suffer from the high conductivity values seen in the west portion of the non-interactive solution. Moreover, unlike the non-interactive solution, interactive inversion is also able to capture the high conductivity values in the north-west region seen in the original conductivity field.

To compare the results from the interactive and non-interactive approaches more thoroughly, two additional metrics were calculated. The first was a measure of the root mean square difference between the 'true' conductivity field and the estimated conductivity field. The second was the root mean square difference between the 'true' hydraulic head field and the head field as predicted by the calibrated models. Note that unlike the calibration error objective that only used head data at measurement locations, this head difference is calculated for the entire spatial domain of the model. Fig. 11 shows the difference between the true conductivity field and the most plausible non-interactive (Fig. 10b) and interactive conductivity fields (Fig. 10c), respectively. As expected the figure shows the non-interactive conductivity field underestimates the conductivity (negative differences are shown in blue) in the north-west region and over-estimates (positive differences are shown in red) it in the west portion of the site. The interactive conductivity field, on the other hand, seems to underestimate the conductivity field in the south-west portion of the site but has some local high values (seen by the red and orange dots in Fig. 11b). Both fields match the true field exactly on the boundaries, outcrop, and the conductivity measurement locations (shown in white). Interestingly, the root mean square differences for the two fields were

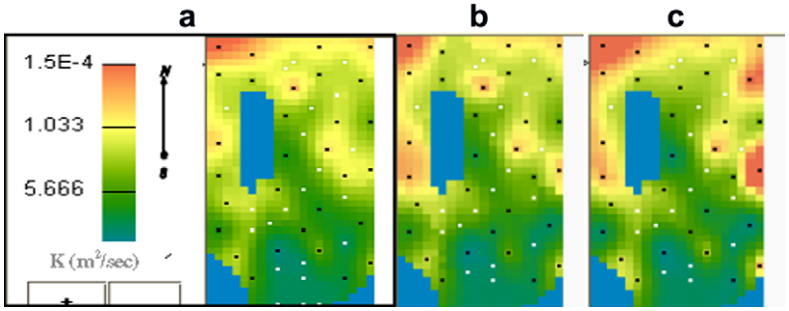

Fig. 9. (a) Rank 1, (b) Rank 2, (c) and Rank 3 solutions from interactive Pareto front Fig. 8.

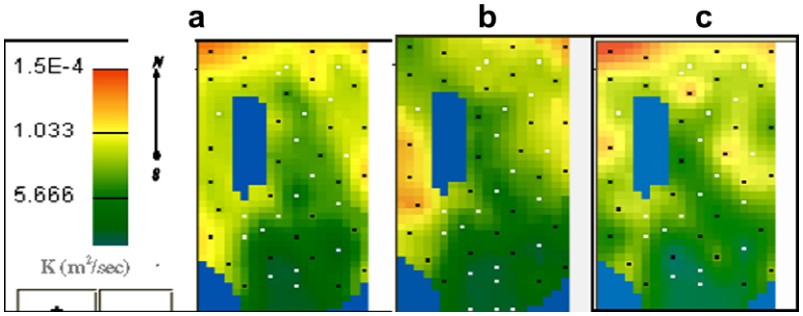

Fig. 10. Conductivity fields of (a) true optimal pilot point field, and most plausible, (b) non-interactive conductivity estimate, and (c) interactive conductivity estimate.

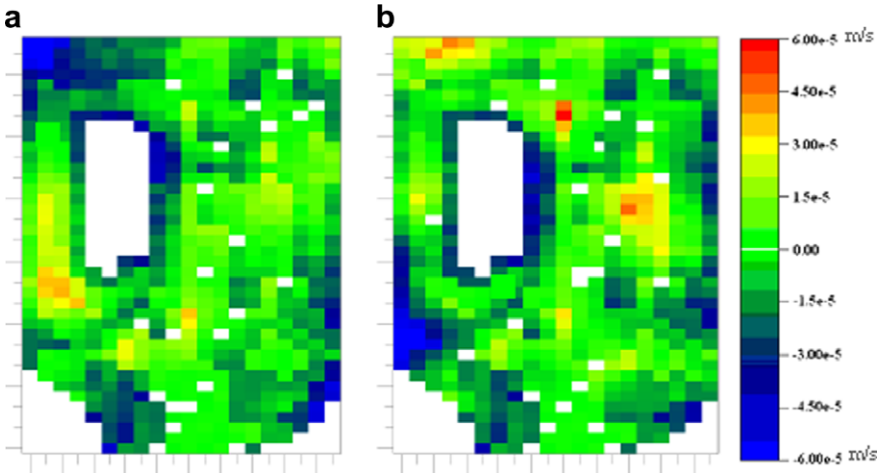

Fig. 11. Differences (in m/s) between (a) the non-interactive conductivity field and true conductivity field and (b) the interactive conductivity field and true conductivity field.

found to be almost identical – equal to 1.6×10^{-5} m^2/s. Thus, even though the overall spatial trend of the interactive estimate is more consistent with the real conductivity field, in terms of the RMSE metric the two fields are equivalent. This again demonstrates the fact that low-order summary statistics like RMSE can not fully capture spatial information.

The next metric compared is the difference in head predictions. For this metric, the known hydraulic heads from the true case are subtracted from the predicted heads for each of the two calibrated models. The root mean square error between the head fields was found to be 0.300 and 0.324 for the interactive and non-interactive solutions, respectively. Thus, on average the interactive solution does 8% better in predicting the head field compared to the non-interactive solution.

4.3. Results for the predictive scenario

Next, we examine the performance of the non-interactive and interactive calibrated models in predicting the effects of changing the bottom conductance of the river bed by two orders of magni-

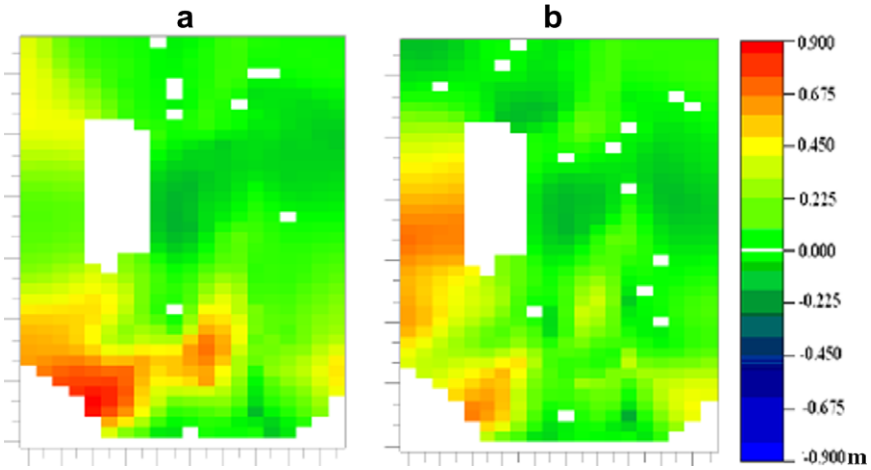

Fig. 12. Differences (in m) between (a) the non-interactive solution's head field and true head field with lined river bed and (b) the interactive solution's head field and true head field with lined river bed.

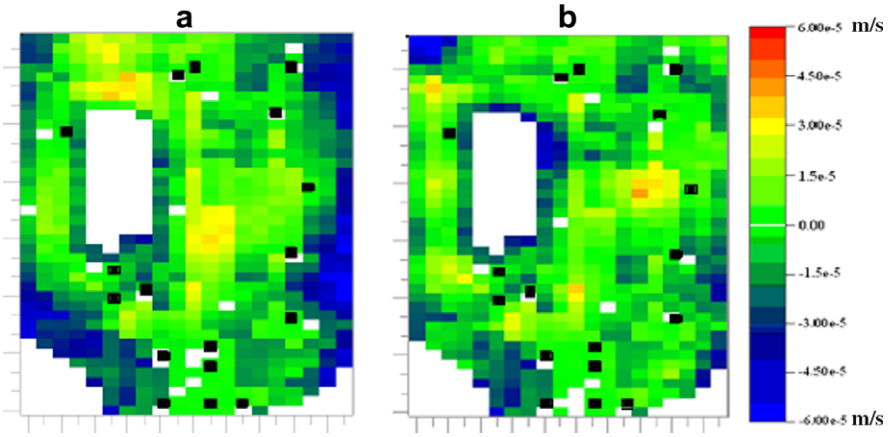

Fig. 13. Differences (in m/s) between (a) the non-interactive conductivity field and true conductivity field and (b) the interactive conductivity field and true conductivity field with reduced data. Locations of the reduced data set are shown by the rectangles.

tude and recalculating the hydraulic head field with the calibrated groundwater model [11]. Fig. 12 shows the difference between the head field of the calibrated models (same as in Fig. 10b and c) and the head field from the true model, both with lined river beds. This figure shows that both cases seem to over-predict the heads (as seen by the mostly positive differences in heads in Fig. 12). The magnitude of error for the non-interactive field is seen to be greater than the magnitude of errors for the interactive solution. In terms of the root mean square difference between the two calibrated models and the true model, the non-interactive field leads to a predictive RMS error of 0.377 while the interactive field has an RMS error of 0.334 – representing about 13% improvement for the latter.

4.4. Results with reduced data

The previous results showed that the plausibility of the estimated conductivity and the predictive performance of the calibrated model both improved when allowing the expert to incorporate his/her knowledge about the spatial characteristics of the site. The improvements were seen in regions with the least data support, indicating that the expert can compensate for sparse data with his/her knowledge about the site. To explore this idea further, data were selectively removed from the hypothetical aquifer to assess the importance of user interaction, especially in the face of sparse and skewed data sets. The rectangles in Fig. 13 show one such case with five measurement locations removed from the middle of the site – where earlier there was maximum data support (the original measurement locations are shown by the white crosses in Fig. 1).

The non-interactive and interactive optimization approaches were then tested with this reduced data case, with locations of the removed measurements used as pilot points. Fig. 13 shows the difference between the most plausible conductivity fields with and without interaction for the reduced data case and the true conductivity field. The non-interactive conductivity field (Fig. 13a) has higher errors than the interactive conductivity field. Compared to the results with more data (Fig. 11a and b), the errors around the edges of the model domain have been exacerbated and the non-interactive solution tends to underestimate the conductivity along the east and south-west edges of the site. Moreover, the conductivity errors in the middle are also higher (as seen by the patch of orange – positive errors) compared to the results with more data (Fig. 11a). In terms of the root mean square error of the difference between the estimated and true conductivity field, the interactive approach leads to an RMSE of 1.43×10^{-5} m^2/s, while the non-interactive solution has an error of 1.84×10^{-5} m^2/s. This represents a 29% mean improvement for the interactive approach compared to the non-interactive solution.

The root mean square error for the hydraulic head field (compared with the true head field) for the interactive solution is 0.188 m compared to 0.244 m for the non-interactive solution. This represents a 30% mean improvement for the interactive solution. Finally, the predictive scenario with the lined river bed was also run with and without interaction. The root mean square difference between the calibrated fields and the true field (both with lined river beds) was found to be 0.307 and 0.354 for the interactive and non-interactive approaches, respectively. This represents a 15% improvement of the interactive solution compared to the non-interactive solution for the predictive scenario. Subsequent experiments with reduced data from different locations showed similar results. In general, upon removal of data points, non-interactive inversion performed poorly. The IMOGA was able to partially compensate for the lack of data, improving results both in terms of the plausibility of the field and the predictive performance of the model.

5. Summary and conclusions

The methodology and results presented in this paper are part of an on-going effort to systematically integrate expert knowledge, secondary information, and hard data in the search for optimal parameters for groundwater models. This paper discusses a novel interactive multi-objective optimization framework called IMOGA that allows the user to adaptively interact with the search process and identify solutions with different levels of trade-offs between expert knowledge and field data.

To test this approach, we applied the IMOGA to a hypothetical aquifer using pilot-point-based regularization with the two quantitative objectives of head prediction error and regularization level, and an additional qualitative objective of 'user preference'. The results indicated that even without user interaction there is a significant trade-off between the two quantitative objectives of accuracy (as measured by the calibration error) and plausibility. The most accurate solutions were in fact those that differed most from the prior field (in other words, were most implausible with respect to field data) and this was ascribed to over-fitting of the measurement errors in the head data. It was seen that the objective values for the 'true optimal' pilot-point solution (which in a sense is the best approximation of reality and the solution we hope to find from inversion) was perceived as sub-optimal when only considering these quantitative objectives.

Moreover, the conductivity fields found with non-interactive optimization suffered from unrealistic conductivity values in regions with low data support. If the expert has some knowledge about these regions with low data support, then interactive optimization allows him/her to include this knowledge while still maintaining the (quantitative) optimality of the solutions in areas with high data support. Comparison of the most plausible interactive conductivity field with the most plausible non-interactive conductivity showed that the former had a spatial structure closer to the true optimal conductivity field. The IMOGA solution was seen to perform better than the non-interactive solution both in terms of the head response for the calibration scenario and the head response for a predictive scenario that served as validation for both models. Moreover, the improvements seen in the IMOGA increased as the data became sparser and less uniformly spread. Interaction with the expert could in fact provide information about areas where there was a lack of direct field measurements constraining the conductivity values, leading to more plausible fields. These findings indicate that the IMOGA is able to effectively incorporate expert knowledge in the search for optimal conductivity fields.

This paper is a first application of interactive approaches to groundwater model calibration and a number of issues remain to be addressed for improving this methodology. The greatest challenge to be addressed is referred to as 'user-fatigue' in the interactive optimization literature. Takagi [41] has discussed some of the issues of user fatigue and has pointed out how it can lead to deterioration of the solution quality of the interactive optimization by introducing errors and biases from the user. *Takagi* also discusses certain steps for reducing user fatigue – making interfaces simple, limiting the population size and the number of generations, and using 'surrogate functions' to model and supplement human interaction. Some of these recommendations (such as small population sizes, seeding from large population runs to reduce the computational burden, and simple interfaces and ranking scheme) have already been incorporated in the IMOGA. Currently, we are investigating methods to reduce user fatigue by using machine learning models to 'learn' user preferences from the human-ranked results and supplement user interaction when needed. This not only reduces user-fatigue, but it also allows the IMOGA to use larger population sizes because many of the solutions in a generation can be evaluated by the surrogate function. Initial results with such machine learning approaches have been promising (see [37] for details) and we expect to present these findings in subsequent papers.

The other significant direction for future research is to address uncertainty (both parametric and predictive) within the interactive framework. As the problem is formulated currently, the IMOGA is a deterministic calibration framework. However, in many applications, it is necessary to consider uncertainty in the calibrated parameters. To address this, we have extended the IMOGA framework by proposing a multi-level sampling approach that incorporates uncertainty in both large-scale trends and the small-scale geostatistical stochasticity. Using the large-scale conductivity fields estimated with IMOGA as well as geostatistical models of the small-scale variability, multiple realizations can be created that can be used to assess predictive uncertainty in the groundwater model.

The predictive uncertainty analysis stage of this research was also implemented on a highly heterogeneous, field-scale application based on the Waste Isolation Pilot Plant (WIPP) situated in Carlsbad, New Mexico [25]. This allowed us to test the interactive approach on a 'real' site using multiple experts (actual hydrogeologists who have knowledge of this site). Results show that the IMOGA *can* effectively include expert knowledge in the search for optimal parameter values leading to more robust decision-making. It can also effectively highlight conflicts in the knowledge and conceptualization of different experts – identifying areas where further collaboration and discussion may be required for consensus on the model. Details of this phase of research are given by Singh [37].

Acknowledgement

This research was funded by the Department of Energy – Grant No.: DE-FG07-02ER635302. We also wish to thank Dr. Barbara A. Bailey, for her insights on the mathematical formulation of the calibration problem and Dr. Meghna Babbar for her work on the D2K IMOGA code.

Appendix A

Given n data points and m pilot points the (log) conductivity field can be generated with ordinary kriging using the following equation [9]:

$$K(u_x)' = \sum_{\alpha=1}^{n_x} \lambda_\alpha K(u_0^\alpha) + \sum_{i=1}^{p_x} \lambda_i K(u_{pp}^i) + \overline{K}_x \left(1 - \sum_{i=1}^{p_x} \lambda_i - \sum_{\alpha=1}^{n_x} \lambda_\alpha\right) \quad n_x, p_x \in W_x \quad \forall x \tag{A1}$$

where $K(u_x)'$ is the kriging estimate for location u_x, $K(u_0^\alpha)$ is the value of the log conductivity and λ_α the corresponding kriging weight for the αth observation point located at u_0^α, $K(u_{pp}^i)$ is the value and λ_i the kriging weight for the ith pilot point located at u_{pp}^i, $\overline{K}_x$ is the expected value for the log conductivity for location u_x (for ordinary kriging this is unknown but assumed to be constant within the kriging window), n_x is the number of data points and p_x the number of pilot points contained within the kriging window W_x for location x. The kriging constraints are given below:

$$\begin{cases} \sum_{\beta=1}^{n_x} \lambda_\beta \mathrm{Cov}(u_0^\alpha, u_0^\beta) + \sum_{\gamma=1}^{p_x} \lambda_\gamma \mathrm{Cov}(u_0^\alpha, u_{pp}^\gamma) + \mu = \mathrm{Cov}(u_0^\alpha, u_x) & \alpha = 1, \ldots, n_x \\ \sum_{\delta=1}^{p_x} \lambda_\delta \mathrm{Cov}(u_{pp}^\gamma, u_{pp}^\delta) + \sum_{\alpha=1}^{n_x} \lambda_\alpha \mathrm{Cov}(u_{pp}^\gamma, u_0^\alpha) + \mu = \mathrm{Cov}(u_{pp}^\gamma, u_x) & \gamma = 1, \ldots, p_x \\ \sum_{\alpha=1}^{n_x} \lambda_\alpha + \sum_{i=1}^{p_x} \lambda_i = 1 \end{cases} \tag{A2}$$

where $\mathrm{Cov}(u^x, u^y)$ is the model covariance between locations u^x and u^y (these could be locations for the data points – subscript 0 – or pilot points – subscript pp), and μ is the Lagrange multiplier used to impose the constraints given by Eq. (A2) for the estimate given by Eq. (A1). Note that due to the third constraint, the value of $\overline{K}_x$ is not required for ordinary kriging.

Once the log-conductivity field has been generated using Eqs. (A1) and (A2), it is then back-transformed using the following relation from [35], to give the conductivity field to be used for model prediction:

$$K(u_x)' = \exp\langle \mathrm{Ln}[K(u_x)'] + \sigma_x^2/2 - \mu_x \rangle \tag{A3}$$

where $K(u_x)'$ is the back-transformed kriged estimate for location u_x, $\mathrm{Ln}(K_x)'$ is the kriging estimate in the log-transformed space, σ_x is the kriging estimation variance for location x (also calculated by the kriging algorithm), and μ_x is the Lagrange multiplier (from Eq. (A2)) used for ordinary kriging at location u_x.

Appendix B

The covariance between pilot point i and pilot point j is given by:

$$C_{pp}^{ij} = E\langle (K(u_{pp}^i)' - K(u_{pp}^i))(K(u_{pp}^j)' - K(u_{pp}^j)) \rangle \tag{B.1}$$

where C^{ij}_{pp} is the covariance between estimation errors for two pilot points located at u^i_{pp} and u^j_{pp}, $K(u^i_{pp})'$ is the kriging estimate for the ith pilot point (given by Eq. (A1)), $K(u^i_{pp})$ is the true value for the ith pilot point, and $E()$ is the expected value for the given variable. We can write (B.1) in terms of residual from the mean value.

$$C^{ij}_{pp} = E\langle [K(u^i_{pp})' - \overline{K}_i - (K(u^i_{pp}) - \overline{K}_i)][K(u^j_{pp})' - \overline{K}_j - (K(u^j_{pp}) - \overline{K}_j)]\rangle \quad \text{(B.2)}$$

Expanding (B.2) we get:

$$C^{ij}_{pp} = E(K(u^i_{pp})' - \overline{K}_i)(K(u^j_{pp})' - \overline{K}_j) + E(K(u^i_{pp}) - \overline{K}_i)(K(u^j_{pp}) - \overline{K}_j) - E(K(u^i_{pp})' - \overline{K}_i)(K(u^j_{pp}) - \overline{K}_j) - E(K(u^j_{pp})' - \overline{K}_j) \times (K(u^i_{pp}) - \overline{K}_i) \quad \text{(B.3)}$$

Substituting Eq. (A1) for $K(u^i_{pp})'$ and $K(u^i_{pp})'$ into (B.3):

$$= E\left(\sum_{\alpha=1}^{n_i} \lambda_\alpha (K(u^\alpha_0) - \overline{K}_i) \sum_{\beta=1}^{n_j} \lambda_\beta (K(u^\beta_0) - \overline{K}_j)\right) + E(K(u^i_{pp}) - \overline{K}_i) \times (K(u^j_{pp}) - \overline{K}_j) - E\left(\sum_{\alpha=1}^{n_i} \lambda_\alpha (K(u^\alpha_0) - \overline{K}_i)(K(u^j_{pp}) - \overline{K}_j)\right) - E\left(\sum_{\alpha=1}^{n_i} \lambda_\alpha (K(u^\alpha_0) - \overline{K}_i)(K(u^j_{pp}) - \overline{K}_j)\right) \quad \text{(B.4)}$$

Simplifying (B.4) we get Eq. (B.5), which is the same as Eq. (3) in Section 2.2.

$$C^{ij}_{pp} = \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} \lambda_\alpha \lambda_\beta \mathrm{Cov}(u^\alpha_0 - u^\beta_0) + \mathrm{Cov}(u^i_{pp} - u^j_{pp}) - \sum_{\alpha=1}^{n_i} \lambda_\alpha \mathrm{Cov}(u^\alpha_0 - u^j_{pp}) - \sum_{\alpha=1}^{n_i} \lambda_\alpha \mathrm{Cov}(u^\alpha_0 - u^j_{pp}) \quad \text{(B.5)}$$

References

[1] Alcolea A, Carrera J, Medina A. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. Adv Water Resour 2005(29):1678–89.

[2] Babbar M. Interactive genetic algorithms for adaptive decision making in groundwater monitoring design, PhD thesis. University of Illinois at Urbana-Champaign; 2006.

[3] Brill Jr ED, Flach JM, Hopkins LD, Ranjithan S. MGA: a decision support system for complex, incompletely defined problems. IEEE Trans Systems Man Cybernet 1990;20(4):745–57.

[4] Carrera J, Neuman SP. Estimation of aquifer parameters under transient and steady state conditions, maximum likelihood method incorporating prior information. Water Resour Res 1986;22(2):199–210.

[5] Carrera J, Alcolea A, Medina A, Hidalgo J, Slooten LJ. Inverse problem in hydrogeology. Hydrogeol J 2005;3:206–22.

[6] Cieniawski SE. An investigation of the ability of genetic algorithms to generate the tradeoff curve of a multi-objective groundwater monitoring problem, Masters thesis. University of Illinois, Urbana, IL; 1993.

[7] Coello CAC. 20 Years of evolutionary multi-objective optimization: what has been done and what remains to be done. In: Yen GY, Fogel DB, editors, Computational Intelligence: Principles and Practice, Chapter 4. IEEE Computational Intelligence Society; 2006. p. 73–88. Available from: http://delta.cs.cinvestav.mx/~ccoello/2006.html.

[8] Deb K, Agrawal S, Pratap A, Meyarivan T. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimisation: NSGA-II. In: Proceedings of the parallel problem solving from nature VI conference; 2000. p. 849–58.

[9] Deutsch CV, Journel AG. GSLIB: geostatistical software library and user guide. New York: Oxford University Press; 1998.

[10] Doherty J. Groundwater model calibration using pilot points and regularization. Ground Water 2003;41(2):170–7.

[11] Freyberg DL. An exercise in ground-water model calibration and prediction. Ground Water 1988;26(3).

[12] Goldberg DE. Genetic algorithms in search, optimization & machine learning. Reading, MA: Addison-Wesley; 1989.

[13] Gomez-Hernandez JJ, Sahuquillo A, Capilla JE. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data: 1 – theory. J Hydrol 1997;203:162–74.

[14] Halhal D, Walters GA, Ouazar D, Savic DA. Multi-objective improvement of water distribution systems using a structure messy genetic algorithm approach. ASCE J Water Resour Planning Manage 1997;123(3):137–46.

[15] Hernandez AF, Neuman SP, Guadagnini A, Carrera J. Conditioning mean steady state flow on hydraulic head and conductivity through geostatistical inversion. Stochas Env Res Risk Assess 2003;17(5):329–38.

[16] Khu ST, Madsen H. Multiobjective calibration with Pareto preference ordering: an application to rainfall–runoff model calibration. Water Resour Res 2005;41:W03004.

[17] Kowalsky MB, Finsterle S, Rubin Y. Estimating flow parameter distributions using ground-penetrating radar and hydrological measurements during transient flow in the vadose zone. Adv Water Resour 2004;27:583–99.

[18] Karpouzos DK, Delay F, Katsifarakis KL, de Marsily G. A multipopulation genetic algorithm to solve the inverse problem in hydrogeology. Water Resour Res 2001;37(9):2291.

[19] LaVenue AM, Pickens JF. Application of a coupled adjoint sensitivity and kriging approach to calibrate a groundwater flow model. Water Resour Res 1992;28(6):1543–70.

[20] LaVenue AM, RamaRao BS, de Marsily G, Marietta MG. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: part 2 – application. Water Resour Res 1995;31(3):495–516.

[21] de Marsily G. De l'identification des systèmes hydrologiques. Thèse de Doctorat d'Etat, Université Paris VI; 1978.

[22] de Marsily G, Delhomme JP, Coudrain-Ribstein A, LaVenue MA. Four decades of inverse problems in hydrogeology. In: Zhang D, Winter CL, editors. Theory, modeling, and field investigation in hydrogeology: a special volume in honor of Shlomo P. Neuman's 60th birthday, vol. 348. Boulder, Colorado: Geological Society of America Special Paper; 2000. p. 1–17.

[23] McDonald MG, Harbaugh AW. A modular three-dimensional finite-difference ground-water flow model. Techniques of water resources investigations 06 A1, United States Geological Survey; 1988.

[24] McKenna SA, Doherty J, Hart DB. Non-uniqueness of inverse transmissivity field calibration and predictive transport modeling. J Hydrol 2003;281(4):265–80.

[25] McKenna SA, Hart D. Conditioning of base T fields to steady-state heads, Analysis report, Task 3 of AP-088: Analysis plan for evaluation of the effects of head changes on calibration of Culebra transmissivity fields. Task number 1.3.5.1.2.1, Sandia National Laboratories, WIPP Records Center, May 13; 2003.

[26] McLaughlin D, Townley LR. A reassessment of the groundwater inverse problem. Water Resour Res 1996;32(5):1131–61.

[27] Moore C. The use of regularized inversion in groundwater model calibration and prediction uncertainty analysis. PhD Thesis. University of Queensland, Australia; 2005.

[28] Moore C, Doherty J. The cost of uniqueness in groundwater model calibration. Adv Water Resour 2006;29(4):605–23.

[29] Neuman SP. Calibration of distributed parameter groundwater flow models viewed as a multiple-objective decision process under uncertainty. Water Resour Res 1973;9(4):1006–21.

[30] Raghuwanshi MM, Kakde OG. Survey on multiobjective evolutionary and real coded genetic Algorithms, Proceeding of the 8th Asia Pacific Symposium on Intelligent and Evolutionary Systems; 6–10 Dec. 2004 p 150–161.

[31] Ramarao BS, Lavenue AM, de Marsily GH, Marietta MG. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields, 1. Theory and computational experiments. Water Resour Res 1995;31(3):475–93.

[32] Reed P, Minsker BS, Goldberg DE. A multi-objective approach to cost effective long-term groundwater monitoring using an elitist nondominated sorted genetic algorithm with historical data. J Hydroinform 2001;3(2):71–90.

[33] Reed P, Minsker B, Goldberg DE. Simplifying multi-objective optimization: an automated design methodology for the non-dominated sorted genetic algorithm-II. Water Resour Res 2003;39(7):TNN 2.1–2.5, 1196. doi:10.1029/2002WR001483 [July 2003].

[34] Ritzel BJ, Eheart JW, Ranjithan S. Using genetic algorithms to solve a multiple objective groundwater remediation problem. Water Resour Res 1994;30(5):1589–603.

[35] Roth C. Is lognormal kriging suitable for local estimation? Math Geol 1998;30(8):999–1009.

[36] Seibert J. Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. Hydrol Earth Syst Sci 2000;4(2):215–24.

[37] Singh A. An interactive multi-objective framework for groundwater inverse modeling, PhD thesis. University of Illinois at Urbana-Champaign; 2007.

[38] Solomatine DP, Dibiki YB, Kukuric N. Automatic calibration of groundwater models using global optimization techniques. Hydrol Sci J 1999;44(6):879–94.

[39] Sun N-Z. Inverse problems in groundwater modeling. The Netherlands: Kluwer Academic Publishers; 1995.

[40] Tsai FT-C, Sun NZ, Yeh WG. Global–local optimization methods for the identification of three dimensional parameter structure in groundwater modeling. Water Resour Res 2003;39(2):1043.

[41] Takagi H. Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation. Proc IEEE 2001;89(9):1275–96.

[42] Vesselinov VV, Neuman SP, Illman WA. Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff 1. Methodology and borehole effects. Water Resour Res 2001;37(12):3001–17.

[43] Welge M, Auvil L, Shirk A, Bushell C, Bajcsy P, Cai D et al. Data to knowledge (D2K) – an automated learning group report. National Center for

Supercomputing Applications, University of Illinois at Urbana-Champaign; 2003. Available at <http://alg.ncsa.uiuc.edu>.
[44] Wijns C, Boschetti F, Moresi L, Takagi H. Inversion in geology by interactive evolutionary computation. In: IEEE systems, man, and cybernetics conference, 7–10 October 2001, Tucson, USA. Proceedings of the IEEE international conference on systems, man, and cybernetics; 2001. p. 1053–7.
[45] Woodbury AD, Ulrych TJ. A full-Baysian approach to the groundwater inverse problem for steady state flow. Water Resour Res 2000;36(8):2081–93.
[46] Woods DD, Roth EM, Benett K. Explorations in joint human–machine cognitive systems. In: Robertson S, Zachary W, Black JB, editors. Cognition, computing and cooperation. Ablex; 1990. p. 123–58.
[47] Yakowitz S, Duckstein L. Instability in aquifer identification: theory and case study. Water Resour Res 1980;16(6):1045–64.
[48] Yeh WW-G. Review of parameter identification procedures in groundwater hydrology: the inverse problem. Water Resour Res 1986;22(1):95–108.
[49] Zimmerman DA et al. A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. Water Resour Res 1998;34(6):1373–413.