Predictive uncertainty analysis of a saltwater intrusion model using null-space Monte Carlo

Daan Herckenrath,[1] Christian D. Langevin,[2] and John Doherty[3,4]

Received 19 March 2010; revised 10 February 2011; accepted 1 March 2011; published 7 May 2011.

[1] Because of the extensive computational burden and perhaps a lack of awareness of existing methods, rigorous uncertainty analyses are rarely conducted for variable-density flow and transport models. For this reason, a recently developed null-space Monte Carlo (NSMC) method for quantifying prediction uncertainty was tested for a synthetic saltwater intrusion model patterned after the Henry problem. Saltwater intrusion caused by a reduction in fresh groundwater discharge was simulated for 1000 randomly generated hydraulic conductivity distributions, representing a mildly heterogeneous aquifer. From these 1000 simulations, the hydraulic conductivity distribution giving rise to the most extreme case of saltwater intrusion was selected and was assumed to represent the "true" system. Head and salinity values from this true model were then extracted and used as observations for subsequent model calibration. Random noise was added to the observations to approximate realistic field conditions. The NSMC method was used to calculate 1000 calibration-constrained parameter fields. If the dimensionality of the solution space was set appropriately, the estimated uncertainty range from the NSMC analysis encompassed the truth. Several variants of the method were implemented to investigate their effect on the efficiency of the NSMC method. Reducing the dimensionality of the null-space for the processing of the random parameter sets did not result in any significant gains in efficiency and compromised the ability of the NSMC method to encompass the true prediction value. The addition of intrapilot point heterogeneity to the NSMC process was also tested. According to a variogram comparison, this provided the same scale of heterogeneity that was used to generate the truth. However, incorporation of intrapilot point variability did not make a noticeable difference to the uncertainty of the prediction. With this higher level of heterogeneity, however, the computational burden of generating calibration-constrained parameter fields approximately doubled. Predictive uncertainty variance computed through the NSMC method was compared with that computed through linear analysis. The results were in good agreement, with the NSMC method estimate showing a slightly smaller range of prediction uncertainty than was calculated by the linear method.

Citation: Herckenrath, D., C. D. Langevin, and J. Doherty (2011), Predictive uncertainty analysis of a saltwater intrusion model using null-space Monte Carlo, *Water Resour. Res.*, *47*, W05504, doi:10.1029/2010WR009342.

1. Introduction

[2] Saltwater intrusion models are commonly used to predict the impact of groundwater withdrawals and sea level rise on coastal freshwater resources [*Lebbe et al.*, 2008; *Giambastiani et al.*, 2007; *Zhang et al.*, 2004]. These models require many input parameters, most of which are uncertain. Causes of this uncertainty are, for instance, the heterogeneity in aquifer properties, sea level rise projections, and the lack of observation data. To improve reliability in a model prediction, saltwater intrusion models are often calibrated in order to infer hydraulic properties by endowing them with the ability to replicate past system responses to historically known inputs. As calibration data are limited in space and time, a calibrated model is never a perfect representation of reality since a calibrated parameter field must, of necessity, represent the simplest parameterization that is compatible with observations. Consequently, model predictions are always in error, especially those that depend on hydraulic property detail that cannot be inferred through the calibration process [*Moore and Doherty*, 2005, 2006]. *Carrera and Neuman* [1986] discuss how calibration can be an underdetermined and nonunique inverse problem. In the present context this means that many different parameter sets can result in physically reasonable saltwater intrusion models that match the observations to a satisfactory level. Ideally, many different feasible (in terms of geological plausibility and the ability to represent historical system behavior) parameter sets should then be used

[1]Department of Environmental Engineering, Technical University of Denmark, Kongens Lyngby, Denmark.
[2]U.S. Geological Survey, Reston, Virginia, USA.
[3]Watermark Numerical Computing, Brisbane, Queensland, Australia.
[4]National Centre for Groundwater Research and Training, Flinders University, Adelaide, South Australia, Australia.

0043-1397/11/2010WR009342

to calculate model predictions of interest. In this manner, the uncertainty of each model prediction can be characterized by a probability density function with a mean and standard deviation. The mean of this distribution provides an approximation to the prediction of minimum error variance, while its standard deviation characterizes the uncertainty of the model prediction.

[3] Numerical models of variable-density groundwater flow and solute transport are routinely used to assess freshwater availability in coastal aquifers [e.g., *Giambastiani et al.* 2007; *Hughes et al.*, 2010; *Sanford and Pope*, 2010]; the outcomes of model-based studies can have a profound effect on costly coastal management practices. As coastal freshwater availability becomes limited because of climate change and population trends, for example, expensive management strategies will have to be evaluated. Unfortunately, the uncertainties that accompany saltwater intrusion predictions are rarely quantified using formal techniques, and thus, management decisions are often made without a full appreciation of the risks associated with undesirable environmental outcomes that may follow from an adopted management practice. *Carrera et al.* [2010] noted a related concern in that with the exception of the study by *Dausman et al.* [2010], automated parameter estimation techniques have not been used with saltwater intrusion models. To our knowledge, a formal uncertainty analysis for a saltwater intrusion model has never been documented in the literature despite the fact that quantification of the potential for model predictive error is recommended practice [*Dagan and Zeitoun*, 1998; *Sanz and Voss*, 2006]. There are many reasons why uncertainty analyses are not routinely performed; *Pappenberger and Beven* [2006] provide seven popular reasons and characterize each one as being indefensible. For saltwater intrusion models, which are inherently difficult to develop, the additional effort and computational demand is certainly one explanation, but there remains a clear lack of awareness of uncertainty and parameter estimation methods among groundwater modeling professionals. There is little dispute, however, that quantification of the uncertainty associated with model predictions is of paramount importance, particularly as these predictions continue to play an ever-increasing role in shaping coastal management practices.

[4] Several methods are available for evaluating the uncertainties associated with predictions made by a calibrated model. None of the methods are perfect, and all are computationally intensive, particularly in the highly parameterized context where recognition is paid to the hydraulic property heterogeneity that prevails in most real-world aquifers. This heterogeneity is often a significant or dominant contributor to the uncertainty associated with model predictions.

[5] *Moore and Doherty* [2005] and *Tonkin et al.* [2007] describe linear and nonlinear methodologies for quantification of the uncertainty in predictions made with a calibrated model. Linear methods are based on propagation of variance, as described by *Bard* [1974], *Seber and Wild* [1989], and *Draper and Smith* [1981] in the traditional overdetermined inverse modeling context and by *Moore and Doherty* [2005] and *Gallagher and Doherty* [2007] in the underdetermined context. Nonlinear methods of predictive uncertainty analysis described by *Vecchia and Cooley* [1987] and *Christensen and Cooley* [1999] have been extended for use in the highly parameterized context by *Tonkin et al.* [2007].

[6] Monte Carlo methods for predictive uncertainty analysis seek to generate many different parameter sets that enable the model to match the observed data within some level of misfit that is commensurate with measurement errors associated with that data. Classical examples of such approaches include the generalized likelihood uncertainty estimation (GLUE) method of *Beven and Binley* [1992] as well as Markov chain Monte Carlo (MCMC) methods [see, e.g., *Huisman et al.*, 2010, and references therein]. Unfortunately, these methods are difficult to use in highly parameterized contexts, especially in conjunction with solute transport models with long computer run times. Examples of calibration-constrained Monte Carlo analysis in the groundwater modeling contexts include the work of *Harvey and Gorelick* [1995], *Kitanidis* [1996], *Yeh et al.* [1996], *Woodbury and Ulrych* [2000], and *Carrera et al.* [2005], who use Bayesian formulations; *Rubin and Dagan* [1987], *Guadagnini and Neuman* [1999], *Hernandez et al.* [2006], and *Graham and McLaughlin* [1989a, 1989b], who employ stochastic equations; and *Lavenue and de Marsily* [2001] and *Gómez-Hernández et al.* [1997, 2003], who employ deformation methods. The number of model runs required and mathematical complexity associated with all of these methods is large. Furthermore, the performance of some of them suffers as a result of approximations that are required in their implementation. Consequently, reported applications of these methods are normally restricted to simple models. *Maxwell et al.* [2007] performed calibration-constrained Monte Carlo analysis for a more complicated, three-dimensional flow and transport problem, but the number of realizations was limited to 100.

[7] *Tonkin and Doherty* [2009] present a new subspace technique for calibration-constrained Monte Carlo analysis. In this paper the technique is referred to as the "null-space Monte Carlo" (NSMC) method. This technique provides the ability to efficiently calculate many different calibration-constrained parameter fields with a model run burden that is significantly reduced compared with that associated with most other methods. The method works by first calibrating the model using some form of mathematical regularization. On the basis of parameter sensitivities the null-space of the model's parameter field is then defined. This is comprised of parameter combinations that cannot be estimated using the available observation data. Next, many stochastic realizations of model parameter fields are generated. Null- and solution space projection of each stochastic parameter set is then undertaken; the null-space component of the parameter field is retained, while the solution space component is replaced by that of the calibrated model. If the model were linear, the resulting field would calibrate the model just as well as that achieved through the preceding regularized inversion process. However, model nonlinearity, and an indistinct boundary between solution and null-spaces, requires that recalibration of the stochastic parameter field, obtained in this manner, must be undertaken. As this is achieved through adjustment of solution space parameter field components only, it requires only as many model runs per iteration as there are dimensions in the solution space, with these often being few in number. Further computational savings are gained through reusing precalculated solution-space-projected parameter sensitivities for the first

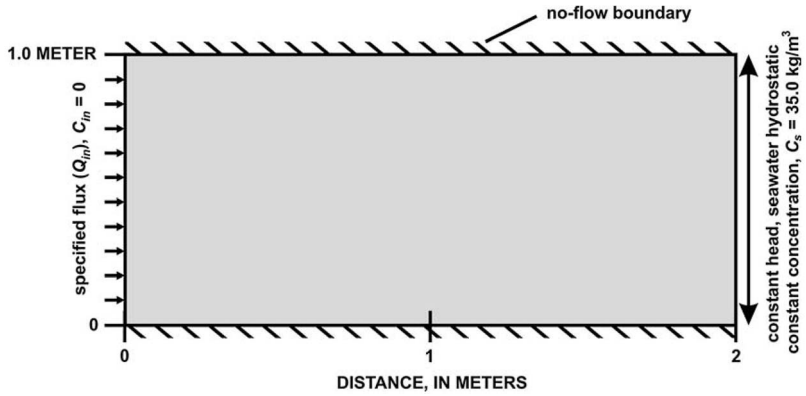

Figure 1. Schematization of the Henry problem. Variables are defined in Table 1.

(and often only) iteration of the recalibration process. The NSMC method of *Tonkin and Doherty* [2009] shares some commonalities with the uncertainty methods that stem from traditional geostatistical inversion techniques [*Zimmerman et al.*, 1998; *Kitanidis and Vomvoris*, 1983; *Vasco et al.*, 1997]. However, it has a number of advantages that, taken together, differentiate it from other techniques that have been used for generating calibration-constrained stochastic parameter fields. First, it is model independent, meaning that no code changes are required to be made to the models for which it is employed. Hence, it can be used with any model of any type, even where that model is comprised of a series of sequential programs executed through the agency of a batch or script file. Second, it is efficient, allowing imposition of rigid constraints imposed by the necessity for a model to replicate historical measurements of system state, with a run time burden that is light compared with that required by many other methods; these constraints can be met by parameter fields that are realistic, yet very different. Third, the methodology relies on few assumptions and approximations; hence, it is robust in a theoretical sense.

[8] In this paper the NSMC methodology is applied and tested for a saltwater intrusion model based on the Henry problem [*Henry*, 1964]. In order to provide maximum challenge to the method we first generated 1000 stochastic variants of the hydraulic conductivity parameterization associated with the Henry problem and selected the one giving rise to the most extreme prediction of saltwater intrusion as the synthetic model representing the "truth." Using the selected synthetic model, we created an observation data set by sampling heads and concentrations and adding measurement noise; to increase the challenge, the observation data set did not include point measurements of hydraulic conductivity. We then calibrated the model and evaluated the difference between the extreme prediction made by the synthetic model and that calculated using the calibrated model. The NSMC method was then applied and tested to determine if the uncertainty range for the prediction of interest encompassed the "true" prediction as made by the synthetic model.

[9] As far as we are aware, this paper is the first to apply formal uncertainty characterization to a saltwater intrusion model. The paper discusses and evaluates several different ways of applying the NSMC method in this context. It concludes with a comparison between uncertainty estimates obtained through NSMC analysis and those gained through linear uncertainty analysis.

2. Henry Saltwater Intrusion Model

[10] In addition to being widely used for benchmarking variable-density flow and transport computer codes [e.g., *Voss and Souza*, 1987; *Langevin and Guo*, 2006; *Oude Essink*, 2001], the well-known Henry problem [*Henry*, 1964] has also been used to test new concepts in saltwater intrusion modeling. For example, *Held et al.* [2005] tested an upscaling technique for homogenization of a heterogeneous Henry problem. *Abarca* [2006] and *Abarca et al.* [2007] converted the original Henry problem, which is based on diffusive mixing, into an anisotropic dispersive Henry problem to investigate factors controlling saltwater intrusion. *Abarca* [2006] also designed a heterogeneous version of the Henry problem to assess the effect of hydraulic conductivity variations on steady state saltwater concentrations. *Sanz and Voss* [2006] used the Henry problem to evaluate parameter sensitivity and correlation. In this paper the Henry problem is used as a test bed for evaluating the NSMC uncertainty analysis methodology of *Tonkin and Doherty* [2009].

[11] The setup of the Henry problem is shown in Figure 1; input and numerical solution parameters are given in Table 1. The SEAWAT computer program [*Langevin and Guo*, 2006; *Langevin et al.*, 2008] was used for all of the simulations. The configuration of the Henry problem is a specified freshwater flux Q_{in} at the left model boundary with a salt concentration C_0 of 0 kg/m^3 and hydrostatic seawater conditions with a constant concentration C_s of 35 kg/m^3 at the right boundary; no-flow conditions are specified for the top and bottom of the model. Porosity θ was held constant for all simulations at 0.35 and was not treated as a stochastic parameter. Longitudinal and transverse dispersivity, characterized respectively as α_L and α_T, were set to zero so that, consistent with the original Henry problem, mixing occurs solely through diffusion. Henry uses a diffusion coefficient D_m of 1.629 m^2/d; to increase sensitivity of the concentration field to heterogeneity in hydraulic conductivity, however, a lower diffusion coefficient of 0.1 m^2/d was used for the simulations reported herein. As will be discussed shortly, hydraulic conductivity *Kf* was parameterized stochastically on a cell-by-cell basis with a mean value of

Table 1. Input and Numerical Solution Parameters

Input		Numerical Solution	
Parameter	Value	Parameter	Value
Q_{in} (calibration)	7.1275 m^2/d	cell size (columns 1–101): dx, dz	0.02 × 0.02 m
Q_{in} (prediction)	5.7020 m^2/d	solver flow PCG	
C_{in}	0 kg/m^3	head stop criterion	10^{-7} m
Kf stochastic field		flow stop criterion	10^{-7} m^3/d
θ	0.35	solver advective transport TVD	
α_L, α_T	0 m	concentration stop criterion	10^{-6} kg/m^3
D_m	0.1 m^2/d	solver dispersion and source terms implicit finite difference	
C_s	0 kg/m^3	time step length	10^{-2} d
p_s	0 kg/m^3		
p_f	0 kg/m^3		

864 m/d, the same value as that used by Henry. The model was discretized with 101 columns and 50 layers, resulting in a cell size of 0.02 × 0.02 m. The transport equation was solved using an implicit finite difference scheme with upstream weighting. The lengths of transport steps were fixed at 5×10^{-3} days.

[12] Two stress periods were used for the simulations. The length of each stress period was 1 day, which is sufficient to allow the concentration field to reach equilibrium in each case. During the first stress period, the freshwater flux entering the left side of the model domain was specified as 7.1275 m^2/d, which is 25% larger than the value used by Henry. This was the period used for model calibration. For the second stress period, a freshwater flux of 5.702 m^2/d was employed, which corresponds to the flux value used by Henry. Reduction of the freshwater flux for the second stress period is intended to represent some inland hydrologic change, such as an increase in inland groundwater pumping or drought conditions. Reduction of the freshwater flux for the Henry problem has been shown to magnify the effect of density variations on the resulting concentration field, thus making the problem more nonlinear and hence more suitable for benchmark testing [*Simpson and Clement*, 2004]. This second stress period was used to make the single model prediction whose uncertainty is explored in this study. This prediction was chosen as the farthest inland position of the 10% seawater concentration (which normally occurs at the base of the model domain). This inland position is subsequently referred to as the "interface position".

3. Unconstrained Monte Carlo Analysis

[13] To evaluate the variability in the predicted interface position induced by variability of hydraulic conductivity, an unconstrained Monte Carlo analysis was first performed on the basis of 1000 stochastic realizations of hydraulic conductivity. The term "unconstrained" signifies that no attempt was made to calibrate the conductivity fields with measured data; thus, variability in model predictions is solely a function of variability between hydraulic conductivity realizations. This conductivity variability is expected to be larger than for a "calibration-constrained" Monte Carlo analysis (such as that discussed below), where variability is limited to hydraulic conductivity realizations that result in a good match of pertinent model outputs with corresponding members of the calibration data set. Each hydraulic conductivity field is heterogeneous and isotropic. Sequential Gaussian simulation, as implemented in the SGSIM code [*Deutsch and Journel*, 1992], was used for stochastic field generation. We specified a mean hydraulic conductivity value of 864 m/d and used an exponential variogram for the base 10 logarithm (log10) of the hydraulic conductivity, which is described by

$$\gamma(h) = c\left[1 - \exp\left(-\frac{h}{a}\right)\right]. \quad (1)$$

In equation (1), $\gamma(h)$ is the variogram value at a separation distance h, c is the variogram sill (equal to the maximum variance of the log hydraulic conductivity between two points), and a is a length parameter or integral scale, which for an exponential variogram, defines a range of approximately $3a$ [*Deutsch and Journel*, 1992; *Rubin and Journel*, 1991; *Rubin*, 2003]. In the present application, values for c and a were assigned as 0.1 and 0.15, respectively. The specified variance used here corresponds to a mildly heterogeneous medium. The relatively large variogram range (compared to the model domain) was intentionally selected to ensure that salt concentration patterns would be strongly dependent on the spatial variability of hydraulic conductivity and would therefore provide yet another challenging test of the NSMC method. Because of the highly diffusive nature of the Henry problem, variations in hydraulic conductivity do not significantly affect flow and transport patterns; smaller covariance ranges were employed.

[14] SEAWAT was used to simulate freshwater and saltwater movement using each of the 1000 hydraulic conductivity field realizations, and the resulting interface positions were recorded under both calibration and predictive conditions. A histogram of the interface positions under predictive conditions is shown in Figure 2. The mean interface position is 0.68 m from the left model boundary. The standard deviation of the interface position is about 0.19 m. Inspection of Figure 2 reveals that the interface is within 0.05 m of the left model boundary for one of the stochastic realizations. On the basis of the histogram of Figure 2, this prediction has a low probability of occurrence. Nevertheless, its occurrence is compatible with the variogram of equation (1) and the specified model characteristics; because of its extreme nature, such an occurrence would be likely to have adverse consequences in a real-world modeling counterpart to the synthetic case described herein. This extreme interface position is plotted in Figure 3. It is apparent from Figure 3 that for the particular stochastic hydraulic conductivity field on which this realization is based, by reducing the freshwater flux, the interface moves inland from 0.31 to 0.05 m.

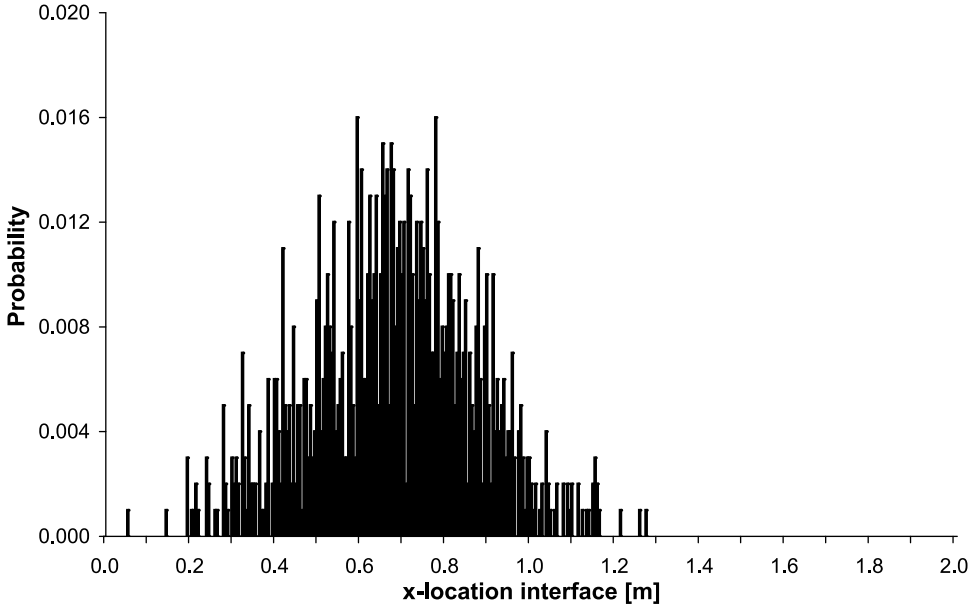

Figure 2. Histogram for the predicted location of the 10% seawater concentration at the farthest inland location of the model after stress period 2, based on 1000 SEAWAT runs with different stochastic fields used for hydraulic conductivity.

[15] The hydraulic conductivity field giving rise to the extreme case depicted in Figure 3 was selected as the "true" hydraulic conductivity field for the purpose of testing the NSMC technique for quantification of predictive uncertainty, as it leads to a prediction of low probability and therefore constitutes a challenging test of the NSMC technique. "Observations" were calculated on the basis of the true hydraulic conductivity field at the end of stress period 1; these comprised the calibration data set used in the calibration-constrained Monte Carlo exercise documented in section 5. Although we selected this one realization to represent the truth, during the course of the investigation we performed analyses similar to that documented herein using other realizations that yielded predictions closer to the mean. As the NSMC method had little difficulty in including these predictions within its calculated uncertainty interval for those realizations, we chose to focus the present discussion solely on the most challenging test, which arises from the extreme case. Furthermore, from a risk management point of view, the realization resulting in the most extreme case of saltwater intrusion is of most interest.

[16] It is apparent from Figure 3 that the 10% seawater line is close to the inland boundary for stress period 2 for the extreme prediction selected as our "reality." Since a specified freshwater flux is applied evenly along this boundary, the position of the 10% seawater line may be affected by the boundary. It is thus apparent that an unavoidable consequence of our desire to design a challenging test of the NSMC method by using a large variogram range relative to the size of the problem domain is that some hydraulic conductivity fields have mean values that differ from the specified mean value. It is possible that the 10% seawater line would have moved farther inland had the boundary been located farther inland. While the analysis is valid for the model design and boundary conditions employed here, the test may have been more challenging had the model domain been laterally extended. This does not invalidate our study; irrespective of whether or not the interface would have moved farther inland if the model domain had been larger, the prediction on which the present analysis is based is nevertheless extreme, as Figure 2 demonstrates.

4. Calibration

[17] A first step in applying the NSMC technique of *Tonkin and Doherty* [2009] is model calibration. As the extreme case from the previous analysis had been selected as the truth, simulated heads and concentrations from this model at the end of stress period 1 were used as observations. The true hydraulic conductivity field is shown in Figure 4. From Figure 4 it is apparent that a high-permeability area at the bottom of the model domain appears to be responsible for

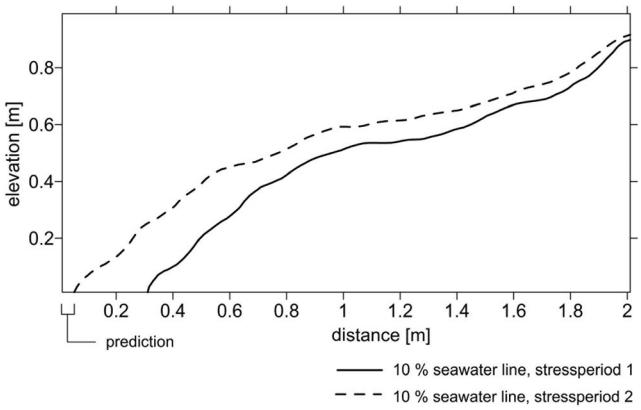

Figure 3. Location of 10% seawater line at the base of the aquifer before (stress period 1) and after (stress period 2) the freshwater flux (Q_{in}) is decreased by 20%. Results are for the hydraulic conductivity field that resulted in the most extreme case of saltwater intrusion following freshwater reduction.

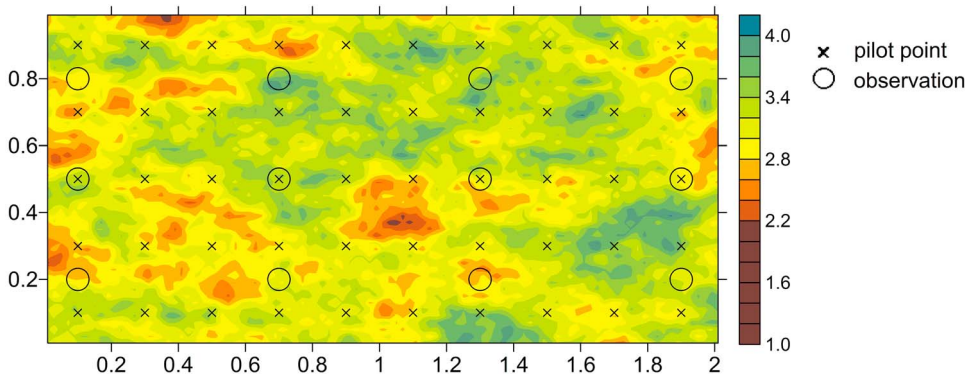

Figure 4. Hydraulic conductivity field selected as the "truth." This field resulted in the worst case of saltwater intrusion arising from unconstrained Monte Carlo analysis. The log value for the hydraulic conductivity (m/d) is shown.

the large inland interface movement. Model calibration was performed using the parameter estimation package PEST [*Doherty*, 2010] with pilot points [*Certes and de Marsily*, 1991; *Doherty*, 2003; *Alcolea et al.*, 2006] employed as a parameterization device.

[18] Tikhonov regularization [*Tikhonov and Arsenin*, 1977; *Doherty*, 2003] was employed to maintain numerical stability and to promulgate achievement of a minimum error variance solution to the inverse problem of model calibration. The hybrid, truncated singular value decomposition (TSVD) functionality in PEST was considered as an option for calibration, but the model runtimes were short enough and the parameters were few enough that the model could be calibrated without reducing the number of parameters to a limited number of linear combinations of estimable parameters (superparameters) as described by *Tonkin and Doherty* [2005].

[19] Figure 4 shows the location of the 12 observation points, for which an "observed" concentration and a head were available for inclusion in the calibration data set. Random error was applied to each of the head and concentration observations on the basis of a mean of zero and a standard deviation of 10% of the head range over the model domain and 10% of the concentration value. Observation weights were calculated as the inverse of the variance of the observation error [*Hill and Tiedeman*, 2007]. As some of the observed concentrations had values near zero, extremely large weights were calculated on this basis. For this reason, a weight limit of 100 (Table 2) was applied for concentration values less than 10^{-2} kg/m^3.

[20] PEST was used to estimate hydraulic conductivity at the 50 pilot point locations shown in Figure 4; prior to each forward SEAWAT run, hydraulic conductivities assigned to pilot points are interpolated to all cells of the model domain. Interpolation was implemented using kriging; the variogram employed for this purpose was the same as that used for the generation of the stochastic field that represents reality for this synthetic model. The initial hydraulic conductivity value assigned to each pilot point at the commencement of the regularized inversion process, through which model calibration was implemented, was specified as the mean field value used in stochastic field generation (864 m/d). Because of computational reasons, a smaller number of pilot points might be chosen for real-world models. Optimal placement of these pilot points could be based on sensitivity maps obtained with the approach described by *Sanz and Voss* [2006]. *Beckie* [1996] and *Rubin* [2003] quantify the relationship between the spacing of these pilot points and available observations relative to the assumed heterogeneity of hydrogeological subsurface structures and the impact on covering its uncertainty by the parameterization and conditioning network of the model. In our application we assume that the pilot points are dense enough to capture the calibration solution space appropriately.

[21] Tikhonov regularization, rather than including conditional hydraulic conductivity measurements, was

Table 2. Simulated and Observed Concentration and Head Values for the Calibrated Model, With Associated Observation Weights

x	z	Weight Heads	Observed Heads (m)	Modeled Heads (m)	Weight Concentrations	Observed Concentrations (kg/m^3)	Modeled Concentrations (kg/m^3)
0.1	0.8	3.33E + 02	1.017	1.021	1.00E + 02	4.96E-14	2.37E-15
1.9	0.8	3.33E + 02	1.006	1.006	2.72E + 00	3.47E + 00	3.49E + 00
1.3	0.8	3.33E + 02	1.002	1.013	1.00E + 02	4.35E-02	4.38E-02
0.7	0.8	3.33E + 02	1.018	1.016	1.00E + 02	3.82E-06	1.85E-07
0.1	0.5	3.33E + 02	1.030	1.022	1.00E + 02	2.74E-10	4.48E-12
1.9	0.5	3.33E + 02	1.001	1.000	2.99E-01	3.23E + 01	3.34E + 01
1.3	0.5	3.33E + 02	1.016	1.010	9.86E-01	8.66E + 00	8.51E + 00
0.7	0.5	3.33E + 02	1.008	1.016	4.97E + 01	2.00E-01	2.00E-01
0.1	0.2	3.33E + 02	1.019	1.022	1.00E + 02	2.79E-07	2.99E-09
1.9	0.2	3.33E + 02	1.000	1.000	2.86E-01	3.49E + 01	3.50E + 01
1.3	0.2	3.33E + 02	1.005	1.000	3.03E-01	3.20E + 01	3.21E + 01
0.7	0.2	3.33E + 02	1.007	1.008	5.81E-01	1.81E + 01	1.81E + 01

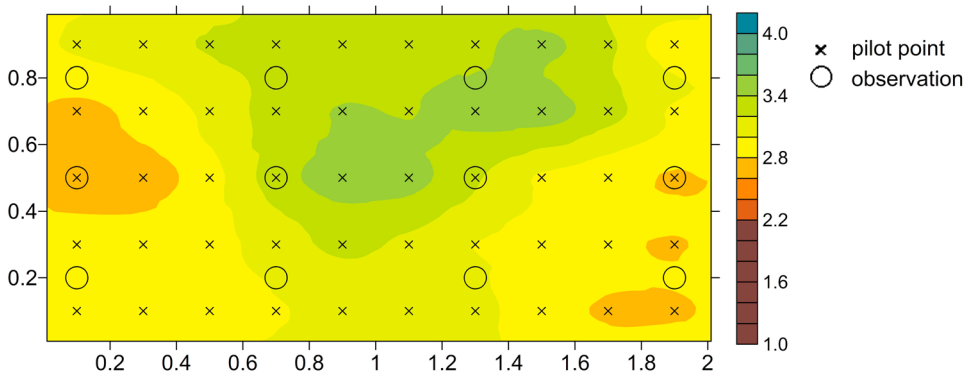

Figure 5. Calibrated hydraulic conductivity field. The plotted values are equal to the log of the hydraulic conductivity (m/d).

implemented as part of the regularized inversion process through which hydraulic conductivity values were assigned to pilot points. Tikhonov regularization defines a "default parameter condition" from which minimal departure is sought in obtaining a good fit between model outputs and field measurements. Examples of such information include specified relations between parameters or the assignment of preferred values to individual parameters based on expert knowledge of likely parameter values. We considered using a pervasive homogeneity constraint, which is implemented by assigning preferred spatial parameter value differences of zero throughout the model domain. This is a useful Tikhonov constraint in real-world modeling contexts where mean hydraulic conductivity values are unknown. For the present application, however, a preferred hydraulic conductivity value of 864 m/d was specified for each pilot point, which is in accordance with the mean values used in stochastic field generation. A weighting matrix applied to preferred pilot point value specifications was calculated as the inverse of the covariance matrix of parameter variability computed on the basis for the hydraulic property variogram [*Maurer et al.*, 1998]. An overall weight factor applied to this matrix is computed by PEST on an iteration-by-iteration basis as part of its implementation of the Tikhonov regularization methodology. This weight factor is determined as that required to achieve a user-specified level of model-to-measurement misfit that is commensurate with measurement noise.

[22] Misfit between the simulated head and concentration values and observed values at the observation points depicted in Figure 4 is reported by PEST as the "measurement objective function"; this is the sum of the weighted squared residuals between observed values and their model-generated counterparts. As stated, when implementing Tikhonov regularized inversion, the procedure will achieve a user-specified target level of model-to-measurement fit if it is able to but will not reduce the level of misfit below this. This strategy limits the departure of parameter values from those that adhere to the preferred parameter conditions specified in Tikhonov constraints at the same time as it prevents overfitting; departures are incurred through the necessity of the model to replicate observations of the historical system state. In the present synthetic case, measurement noise is known, as it was generated using a random number generator; exact specification of a suitable target measurement objective function is thus possible. The addition of measurement noise to the calibration data set resulted in a measurement objective function value of 37 for the true model (i.e., the model from which the calibration data set and model prediction were generated). This same value was then selected as the target measurement objective function in the Tikhonov calibration process.

[23] The calibrated hydraulic conductivity field obtained through the regularized inversion process is shown in Figure 5. As expected, spatial variability of this field is much smaller than that of the true field (Figure 4); this is an outcome of the fact that the information content of the calibration data set does not support unique estimation of a higher level of detail than this. In other words, the calibration process was able to attain the specified value of the measurement objective function without the need to introduce heterogeneity beyond that shown in Figure 5. In the calibrated hydraulic conductivity field, a zone with a relatively high conductivity value can be distinguished in the central part of the model domain; elevated conductivities are also present within the true hydraulic conductivity field in roughly the same area. In general, however, the spatial variability present in the true hydraulic conductivity field cannot be estimated from the limited observation data set despite the fact that observation locations are evenly distributed throughout the model domain.

[24] It is worth noting that the use of pilot points as a parameterization device also limits the level of spatial parameterization detail that can be represented in the calibrated model. However, the fact that this does not limit the capacity of the model to replicate the calibration data set to an extent that is compatible with measurement noise indicates that the detail that is lost through the use of pilot points is in the null-space and, hence, was always beyond the reach of the calibration process. While this precludes the need to represent inter-pilot-point variability in the calibrated parameter field, it does not necessarily preclude the need to represent this detail in a stochastic sense when undertaking predictive uncertainty analysis because the null-space, especially in the present context, may make a substantial contribution to prediction uncertainty. This matter will be further discussed in section 5.

[25] Simulated and observed heads and concentrations are provided in Table 2 and are shown in Figure 6. Despite a relatively large model misfit for three of the

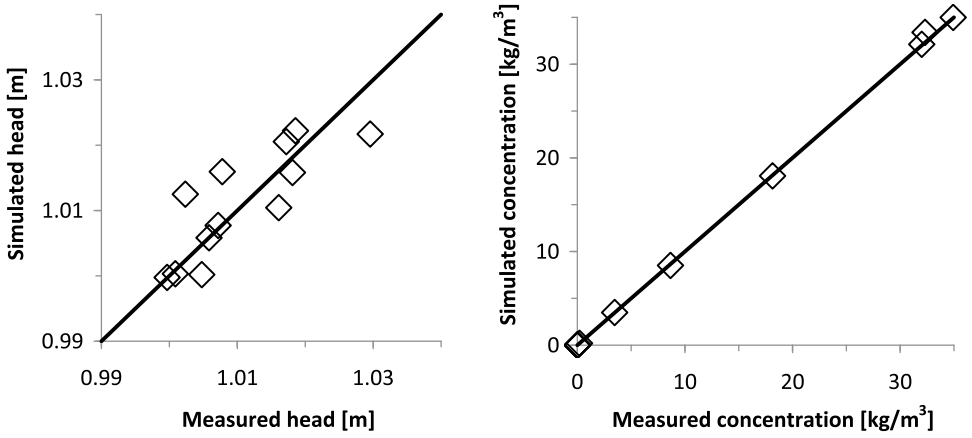

Figure 6. Simulated versus observed concentrations and heads (diamonds) for the calibrated model. These should be located on the 1:1 line (solid line) if the simulated values matched the observations perfectly.

head observations, the model replicated the calibration data set reasonably well, without any systematic over or underestimation of heads and concentrations.

[26] Of importance to the present investigation is the value of the prediction calculated by the calibrated model. The calibrated model predicts an interface position at 0.18 m from the left model boundary under the reduced freshwater inflow condition prevailing during the predictive stress period. This is much farther seaward than the known true value of 0.05 m. The prediction made by the calibrated model is thus substantially in error, a relatively good fit between outputs of this calibrated model and members of the calibration data set achieved during the calibration stress period notwithstanding. In the present case, most of this error is the result of the limited information provided by the calibration data set, which prevents unique estimation of a level of detail greater than that shown in Figure 5. The penchant for predictive error (and hence the level of predictive uncertainty) is exacerbated by the fact that the model domain size is relatively small compared to the variogram range used, resulting in stochastic fields with mean values that can deviate from the specified mean of 864 m/d, which is also the default value employed in Tikhonov regularization. This, of course, represents a calibration context that is typical of many that are faced in real-world modeling practice. As such, it is a problem that any method of uncertainty analysis must be capable of successfully addressing, with "successfully" in the present context meaning that the true value of the prediction must lie within the predictive uncertainty margins provided by the method. Note that further examples of the high level of uncertainty that can accompany predictions made by a "well-calibrated model" are provided by *Moore and Doherty* [2005, 2006]. We will now use the NSMC technique to explore the uncertainty associated with the prediction of interest in our current example.

5. Null-Space Monte Carlo Analysis

[27] As discussed, to apply the NSMC technique, random parameter sets are generated, from which null-space projections are retained while solution space projections are replaced by that of the calibrated parameter field. Details of the step-by-step procedure required for implementation of NSMC analysis are given by *Tonkin and Doherty* [2009] and *Doherty* [2010]. For a perfectly linear model, each parameter set processed in this manner would calibrate the model as well as that obtained through Tikhonov-based inversion of the calibration data set in the manner described in section 4; in the present case, however, adjustment of each processed stochastic parameter set is required in order to return the model to a calibrated state, thus requiring the implementation of a limited model recalibration process. An ensemble of stochastic parameter sets is thus achieved through this sequence of repeated stochastic parameter set generation, projection-based processing, and minor recalibration. If a prediction is then made on the basis of all of the parameter sets forthcoming from this process, the post-calibration uncertainty range of that prediction can be estimated. In the present synthetic case the true value of the prediction is not only known, but is known to be extreme. The effectiveness of the technique is judged by whether the postcalibration predictive uncertainty range as estimated using NSMC analysis includes the true value of the prediction, despite the extreme nature of this prediction. In sections 5.1–5.3 a more detailed description of the NSMC procedure as it is implemented for our present example is provided.

5.1. Generation and Null-Space Projection of Random Parameter Sets

[28] Each of the stochastic hydraulic conductivity fields that were used during the unconstrained Monte Carlo analysis was sampled at pilot point locations to generate a set of pilot point values. The sampling was undertaken through calculation of a pilot-point-based parameter field that provides the best fit with the original stochastic parameter field on a cell-by-cell basis in the least squares sense. Alternatively, hydraulic conductivity values could have been sampled directly at pilot point locations; however, the least squares approach tends to provide a slightly better representation of the stochastic field. It is this pilot-point-based parameter set that undergoes null- and solution space projection in the manner already described.

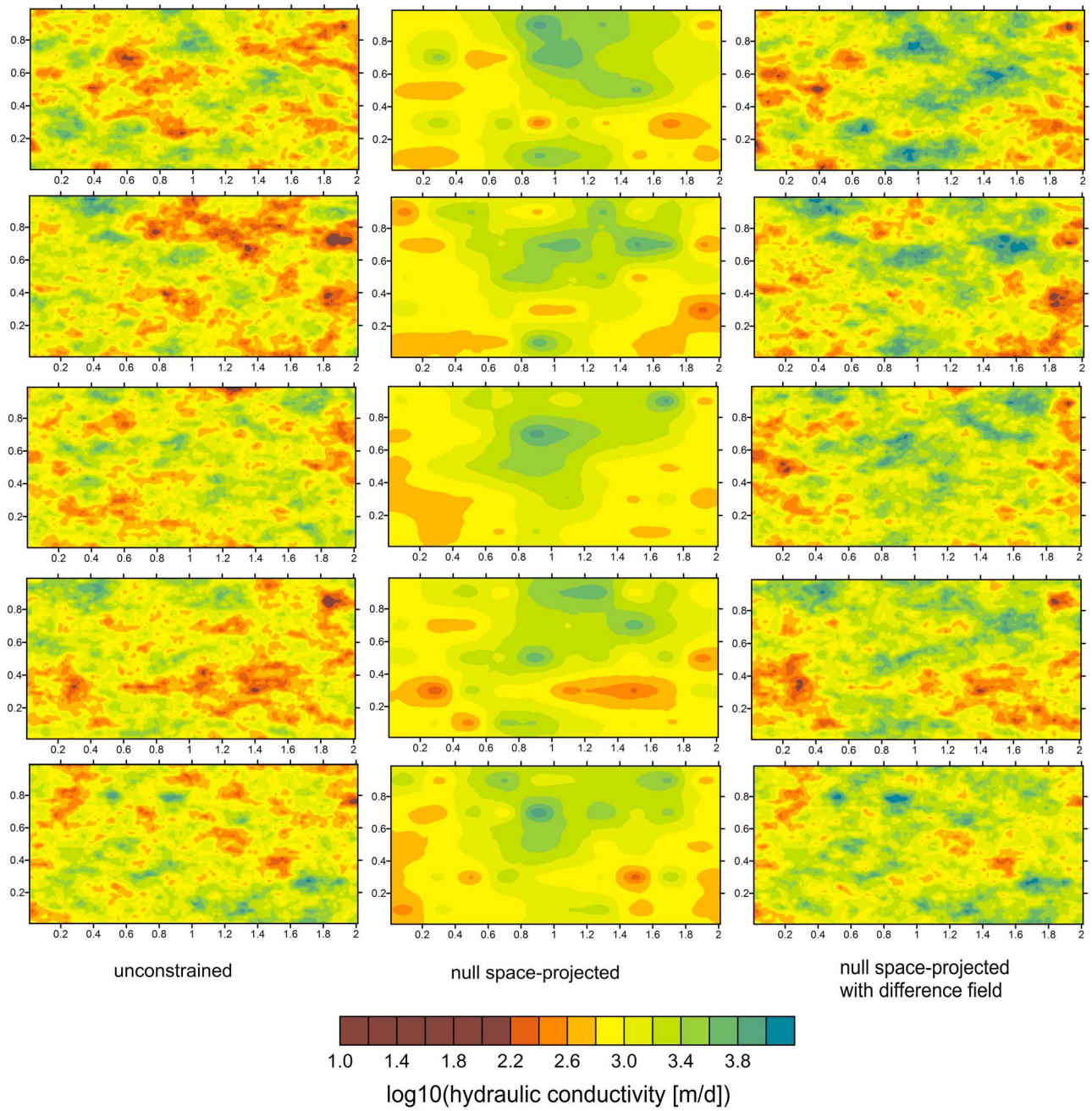

Figure 7. Hydraulic conductivity fields at three processing stages. (left) Unconstrained fields. (middle) The same parameter fields after removal of solution space components and replacement of these components by those of the calibrated model. (right) Cell-by-cell difference field returned to the pilot point field.

[29] As part of this pilot point sampling process, a "difference field" is created for each pilot point parameter set. This difference field represents the difference (on a cell-by-cell basis) between the best fit pilot-point-based field and the original stochastic hydraulic conductivity field from which it was derived. This difference field represents small-scale stochastic hydraulic property field heterogeneity that cannot be represented using a pilot-point-based parameter field; for this reason, the difference field is likely to lie within the calibration null-space, as mentioned. As will be discussed shortly, the difference field can be added back to the null-space-projected pilot point field during model recalibration so that the final calibration-constrained stochastic parameter field retains the same level of detail as the original stochastic parameter field from which it was derived. Including difference fields in the NSMC analysis is an optional step, and one whose efficacy is tested in sections 5.2–5.4.

[30] Figure 7 shows the progression of hydraulic conductivity field processing for five realizations. Figure 7 (left) contains the original stochastic fields. Figure 7 (middle) shows the equivalent pilot-point-based parameter fields after sampling, null-space projection, and solution space replacement by that of the calibrated parameter field to

achieve near calibration of the model. In Figure 7 (right) the cell-by-cell-based difference field is added onto the null-space-projected pilot-point-based field, resulting in fields that are in much closer visual correspondence with the original stochastic parameter fields.

[31] It is interesting to note that when compared to the original stochastic fields depicted in Figure 7 (left), all of the fields shown in Figure 7 (middle) show higher conductivities for the central part of the model domain; the calibrated field of Figure 5 also exhibits high conductivities in this area. Apparently, hydraulic conductivities in this part of the model domain are relatively well informed by the calibration data set.

5.2. Parameter Set Recalibration

[32] As mentioned, null-space-projected random field recalibration is undertaken solely through adjustment of solution space components of these fields. By definition, null-space components need no adjustment. They can thus be left alone to retain their variability, thereby accruing a considerable saving in the overall computational burden of random field recalibration. Furthermore, only the pilot-point-based component of each parameter field is adjusted because (as stated) the difference field is presumed to lie within the calibration null-space (as it represents small-spatial-scale parameter variability that only minimally affects model outputs at observation sites under calibration conditions).

[33] Subdivision of the pilot-point-based parameter set into solution and null-space components (the former requiring adjustment, with the latter remaining unchanged) was achieved using singular value decomposition of the Jacobian matrix of sensitivities of model outputs used in the calibration process to pilot point parameters adjusted through that process; in the present case this was based on sensitivities to pilot points from which the calibrated parameter field of Figure 5 was computed. Despite the use of 50 pilot points for parameterization of the model domain, the maximum possible dimensionality of the solution space in the present case is 24, which is equal to the number of observations used in calibration of the model. Solution space dimensionality is roughly equivalent to the number of significantly nonzero singular values obtained from the singular value decomposition process. In most calibration contexts the dimensionality of the calibration solution space is substantially smaller than the number of observations comprising the calibration data set as a result of the duplication of information within those observations.

[34] In selecting the number of dimensions with which to characterize the calibration solution space (referred to herein as the "solution space dimensionality," or SSD), a modeler is given some control over the speed with which the NSMC process can be implemented. However, as will be seen, greater recalibration speed is accompanied by a potential for underestimation of the full predictive uncertainty range. The "null-space dimensionality" (NSD) is the difference between the total number of parameters employed in the calibration process and the SSD. If a large number is used for the SSD in the parameter projection operation (and hence the null-space is decreed to be small), processed stochastic parameter fields will decalibrate the model to only a small extent. Hence, subsequent recalibration requires fewer model runs. On the other hand, if the value for SSD is chosen to be small, fewer "superparameters" require adjustment during the recalibration process; while this may prove beneficial in some circumstances, adjustment may not result in satisfactory recalibration of all stochastic parameter fields. Through providing the user with the ability to choose an SSD for use in a particular context, the NSMC process, as implemented in the PEST software suite, provides a "lever" for balancing the effects of SSD and NSD on overall calibration run time and on accuracy of uncertainty quantification. In some circumstances, selection of an appropriate value for SSD can have a large impact on the overall computational efficiency of the NSMC method.

[35] During each iteration of a model calibration process, the model is run as many times as there are adjustable parameters. When implementing the NSMC technique, only combinations of parameters spanning the calibration solution space are adjusted. Hence, the number of runs required per iteration of the stochastic field recalibration process is the same as the user-specified dimensions of the calibration solution space. For extra efficiency in recalibrating the stochastic parameter fields, sensitivities of solution space parameter combinations for the first iteration are computed from sensitivities obtained from the originally calibrated model. As initial parameter sensitivities do not then need to be recalculated for each new stochastic parameter field, the first recalibration iteration requires no model runs for Jacobian matrix calculation and then only requires one or two for testing the effectiveness of parameter upgrades. The overall computational effort required for recalibration of many stochastic parameter fields can thus be comparatively small. Nevertheless, the outcome of this process will be the attainment of many different stochastic parameter fields, all of which are geologically feasible (as they were generated on the basis of a suitable stochastic descriptor of geological reality) and all of which provide a good fit between model outputs and field measurements composing the calibration data set.

[36] For the study documented herein, the NSMC process was implemented in four different ways to test the effectiveness of that process in each of these ways and in order to test the relative efficiency of each of these variants.

[37] The first two implementation variants of the NSMC method are based on the chosen dimensionality of the calibration solution space. This tests the efficiency gained by implementing parameter subspace projection in a way that results in random parameter sets that decalibrate the model to only a small extent and therefore require very little computational effort to recalibrate the model as well as the costs of that efficiency in terms of reductions in the range of assessed postcalibration predictive variability. Values of 5 and 13 were tested for the SSD. An SSD value of 5 was selected using the methodology proposed by *Doherty and Hunt* [2009]. According to this method, the onset of the null-space as singular values increase is marked by an increase, rather than a decrease, in postcalibration error variance of an eigencomponent corresponding to a particular singular value when that eigencomponent is assigned to the solution space rather than being retained in the null-space. The alternative SSD of 13 was arbitrarily selected as a dimensionality that is much larger than 5 but small enough to allow substantial dimensionality of the null-space.

[38] The third and fourth variants of the NSMC test implementation reported here pertain to the effect of including the difference fields in the model recalibration process when using SSDs of 5 and 13. When these difference fields are included in the recalibration process to account for intra-pilot-point-scale heterogeneity, they are added to the null-space-processed pilot-point-based parameter fields so that the sum of the two is manipulated as pilot points are varied through the model recalibration process. Inclusion of the difference field results in a greater recalibration computational effort, as these difference fields do, in fact, retain some solution space components in spite of the fact that they are dominated by small-scale rather than large-scale heterogeneity. To the extent that a particular model prediction depends on this detail, inclusion of this (randomly variable) detail in these fields is likely to decrease the potential for underestimation of the uncertainty associated with this prediction.

5.3. Determination of Target Objective Function Value for Recalibration

[39] For the calibration procedure described in section 4, the target measurement objective function was "known" to be 37. In determining a suitable target measurement objective function for more general use during stochastic field recalibration exercises required as part of the NSMC process, a formula used by *Christensen and Cooley* [1999] for imposing constraints on postcalibration predictive variability can be employed:

$$\Phi_0 = \Phi_{\min}\left[\frac{t^2_{\alpha/2}(n-m)}{(n-m)} + 1\right]. \quad (2)$$

In equation (2) $\Phi_{\min}$ is the expected value of the measurement objective function (calculated on the basis of expected measurement noise statistics), t signifies Student's t distribution, and α signifies a level of confidence. In our case, we set α to 0.0005, which is equivalent to a 99.9% confidence level. The number of observations n is set to 24, and the number of parameters m is equal to 5 in accordance with the estimated dimensionality of the calibration solution space; in addition, $\Phi_{\min}$ was retained at the measurement objective function target used (and attained) through the Tikhonov inversion process, i.e., 37. With these settings, Φ_0 is calculated to be 1.79 times $\Phi_{\min}$, or 66.4 in the present case. The value of 66.4 was specified as the target measurement objective function value for subsequent recalibration of all stochastic parameter sets. For the SSD = 13 scenarios, the same target objective function value was used because the SSD value of 13 pertains only to the solution space dimensionality used for null-space projection and does not characterize the innate solution space dimensionality of the inverse problem.

[40] It should be pointed out that equation (2) was derived for use in conjunction with well-posed calibration problems and is not strictly applicable in the ill-posed calibration context that is the focus of the present paper. We use it in our study, however, as no equivalent expression is available for deployment where exploration of predictive uncertainty follows calibration undertaken using highly parameterized inversion employing Tikhonov regularization. Where inversion involves only singular value decomposition with a predetermined solution space dimensionality that guarantees problem well-posedness, then equation (2) is, in fact, exact. *Moore* [2005] demonstrates that the use of equation (2) in conjunction with Tikhonov regularization is conceptually justifiable and leads to predictive uncertainty outcomes that closely approximate those computed by other means.

5.4. Results

5.4.1. Objective Function Values

[41] To indicate the extent to which null-space processing of random parameter sets brings the random parameter sets closer to calibration, histograms for the measurement objective function are plotted in Figure 8 before and after null-space projection and solution space replacement (with no subsequent recalibration). The plots also illustrate the differences in effectiveness of this process with and without the addition of the difference fields. The histograms pertaining to the null-space-projected parameter sets show that most of these sets are calibrated or nearly calibrated. Furthermore, use of a larger SSD results in a higher number of calibrated parameter sets. Finally, use of a difference field lowers the effectiveness of null-space projection because of the small residual solution space content of the difference fields.

5.4.2. Scenario Efficiency and Effect on Prediction

[42] Table 3 lists the average number of model runs required for recalibration of the null-space-processed random sets. Table 3 indicates that a 36% increase in efficiency (without difference fields) can be gained through setting SSD to 13 instead of 5. Only a 14% gain is observed when the difference fields are included. This finding suggests that a computational benefit can be gained by using an expanded solution space and, hence, a diminished null-space for the purpose of null-space projection of the random parameter sets. Use of a difference field, however, substantially increases (by more than a factor of 2) the average number of model runs required to recalibrate stochastic parameter sets. As remarked above, however, although there is an additional cost in using the difference fields, their use may still be worthwhile in some cases if predictions are highly sensitive to fine-scale detail not included at the pilot point level; calculation of the uncertainty of these predictions would therefore demand recognition of this detail.

[43] Table 4 summarizes the results of NSMC trials in terms of the mean, standard deviation, minimum, and maximum values of the prediction (which is the position of the 10% seawater position from the left boundary at the end of the simulation). Table 4 also reports whether the uncertainty range, computed as an outcome of NSMC analysis of the pertinent type, is wide enough to encompass the extreme prediction. For both of the SSD = 5 variants, the NSMC-computed uncertainty range is broad enough to encompass the true prediction. This is not the case, however, for the SSD = 13 scenarios. This finding indicates that although some improvement in computational efficiency can be gained by increasing SSD in this case, the analysis no longer provides an accurate assessment of predictive uncertainty. This was an outcome of the fact that too many random field components were removed during the null-space processing operation.

[44] For SSD = 5, it appears that inclusion of a difference field does not have a significant effect on the computed

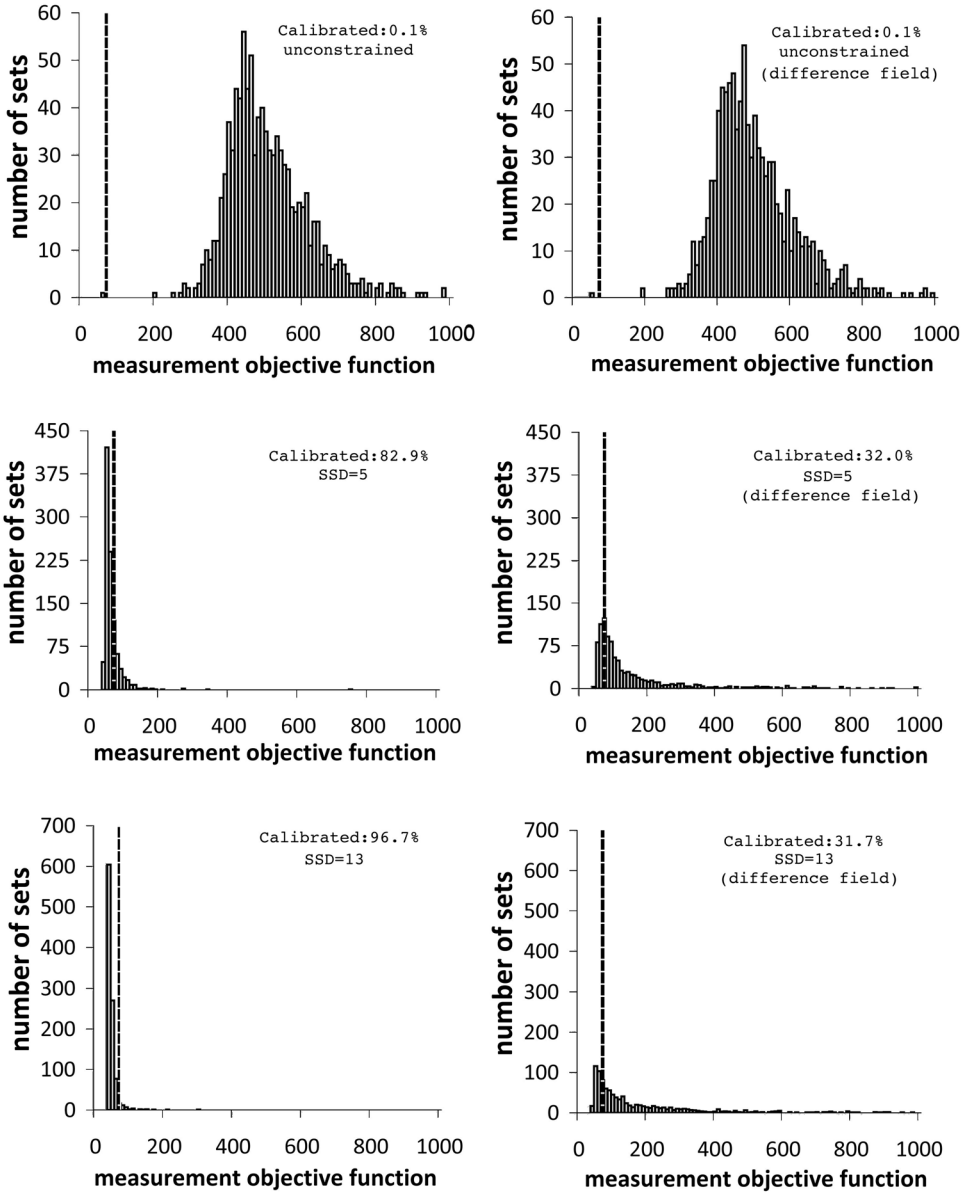

Figure 8. Histograms of the measurement objective function before and after null-space projection of the random parameter sets with SSDs of 5 and 13 (with and without the addition of the associated difference fields). The percentages of parameter sets which meet the calibration criterion of 66.4 (dashed line) are also indicated. Other than null-space processing, none of the parameter sets depicted here were altered through recalibration.

uncertainty interval; the truth is encompassed for both cases. In contrast, for SSD = 13, inclusion of the difference field has a noticeable effect on the prediction mean and standard deviation. It should be noted that all scenarios resulted in a maximum likelihood prediction estimate that is seaward of the true prediction. This is in accordance with the prediction made by the calibrated model and reinforces the conclusion that the limited information contained within the calibration data set is unable to constrain parameter fields sufficiently well to reduce the potential for significant predictive error.

5.4.3. Variograms

[45] To help explain the results of the analyses performed in our study, a suite of variograms was prepared on the basis of the cell-by-cell hydraulic conductivity fields forthcoming from these analyses. In Figure 9 the variogram of the model selected as the truth is compared with the variogram used to generate the 1000 hydraulic conductivity fields for the unconstrained Monte Carlo analysis. An average variogram

Table 3. Average Number of Model Runs to Recalibrate the Model After the Null-Space Projection

Scenario[a]	Model Runs
SSD = 5, without difference field	1.43
SSD = 5, with difference field	3.44
SSD = 13, without difference field	1.05
SSD = 13, with difference field	3.01

[a]SSD, solution space dimensionality.

Table 4. Descriptive Statistics for Each Null-Space Monte Carlo Variant and for the Linear Method[a]

Scenario	Mean (m)	Standard Deviation (m)	Minimum (m)	Maximum (m)	Captured Truth?
Unconstrained Monte Carlo	0.680	0.190	0.050	1.280	-
SSD = 5, without difference field	0.116	0.063	0.010	0.293	yes
SSD = 5, with difference field	0.119	0.065	0.010	0.333	yes
SSD = 13, without difference field	0.188	0.024	0.071	0.311	no
SSD = 13, with difference field	0.231	0.054	0.078	0.392	no
Linear predictive error variance method	0.180	0.080	0.020	0.340	yes

[a]The true prediction value was 0.05 m (minimum unconstrained Monte Carlo). The mean prediction value associated with the linear method is based on the originally calibrated model.

for the entire ensemble of the 1000 uncalibrated hydraulic conductivity fields is also shown; as expected, this exhibits a close match with the predefined variogram model. The variogram for the "true field" is slightly different from that used for its generation, showing slight fluctuations between lag distances of 0.2 and 0.8 m. The fact that the variogram of a stochastic field is slightly different from that used in its generation is not an unexpected result [*Deutsch and Journel*, 1992]; however, a high degree of similarity between these two variograms is, nevertheless, apparent. The variogram for the calibrated model is also shown in Figure 9; this variogram differs significantly from the defined and truth variograms. This is consistent with the findings of *Moore and Doherty* [2006], who demonstrate that a calibrated parameter field can often do little more than reflect a subdued form of the broad-scale hydraulic property heterogeneity exhibited by reality, which is an outcome of limited information content of the calibration data set. It is apparent that when using the uniformly distributed set of head and concentration observations composing the calibration data set used in the present study, the calibration process could capture only broad spatial variation of the real hydraulic conductivity field.

[46] Average variograms for the entire ensemble of 1000 calibration-constrained stochastic fields produced through NSMC analysis were calculated for SSD = 5 and SSD = 13 (Figure 10). This variogram analysis was only conducted for the implementations of the NSMC method where the difference fields were included, as pilot-point-based parameter fields cannot represent the same level of heterogeneity as that included in the real hydraulic conductivity field. The variograms pertaining to those parameter fields should look similar to the variogram of the calibrated parameter field shown in Figure 9. The average variogram for SSD = 5 is, indeed, similar to the variogram of the hydraulic conductivity field. Figure 10, however, exhibits a more distorted pattern for fields based on a SSD of 13. This result is an outcome of the fact that too many null-space dimensions were removed from the stochastic parameter fields as an outcome of projection into a null-space that was falsely diminished. The recalibration process then needed to introduce substantial solution space variability to compensate for this omission (as this is the only variability open to it) in order to bring the model back to a calibrated state. As solution space variability tends to be of broader spatial scale than null-space variability, this perturbs the variogram to a greater degree.

5.4.4. Comparison With Linear Uncertainty Analysis

[47] For comparison purposes a simple linear method was also used to calculate the uncertainty in the prediction. The linear method uses the following equation derived by *Christensen and Doherty* [2008] and employed by *Dausman et al.* [2010] for data worth analysis in the context of density-dependent flow where fluid density is a function of both solute concentration and temperature:

$$\sigma_s^2 = \mathbf{y}^t\mathbf{C}(\mathbf{p})\mathbf{y} - \mathbf{y}^t\mathbf{C}(\mathbf{p})\mathbf{X}^t[\mathbf{X}\mathbf{C}(\mathbf{p})\mathbf{X}^t + \mathbf{C}(\varepsilon)]^{-1}\mathbf{X}\mathbf{C}(\mathbf{p})\mathbf{y}. \quad (3)$$

In equation (3), $\sigma^2{}_s$ represents the postcalibration uncertainty of a prediction s whose sensitivities to model parameters are encapsulated in the vector $\mathbf{y}$ (the superscript t denotes the transpose operation), and $\mathbf{X}$ is the Jacobian matrix of sensitivities of model outputs used in the calibration process to all model parameters. $\mathbf{C}(\varepsilon)$ is the covariance matrix of measurement noise.

[48] The advantage of linear methods is that computation of predictive uncertainty can be accomplished with a much smaller computational burden than that required for more general methods such as NSMC. An obvious disadvantage is that the linear method cannot estimate higher statistical moments. Furthermore, the effect of model nonlinearity on estimation of the range of postcalibration predictive uncertainty is unknown.

[49] The standard deviation of predictive uncertainty calculated using equation (3) was found to be 0.08 m; this value should be compared with the value of 0.06 m calculated on the basis of the predictive histogram achieved through use of the SSD = 5 scenario discussed above. The 95.45% confidence intervals calculated on the basis of linear theory do, indeed, encompass the truth; however, the

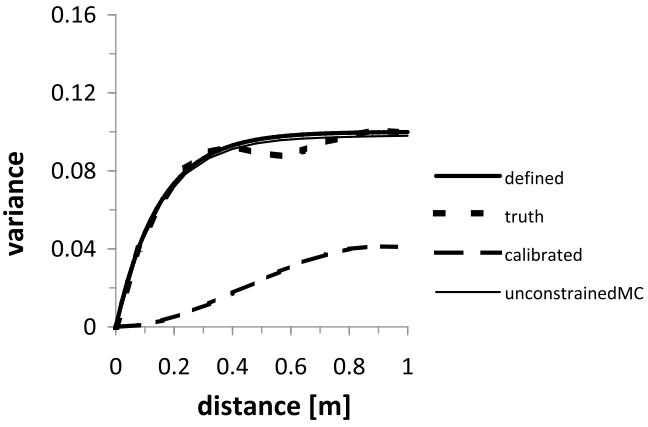

Figure 9. Comparison of the defined variogram that was used to generate the unconstrained Monte Carlo sets with the variograms of the true field and of the calibrated field.

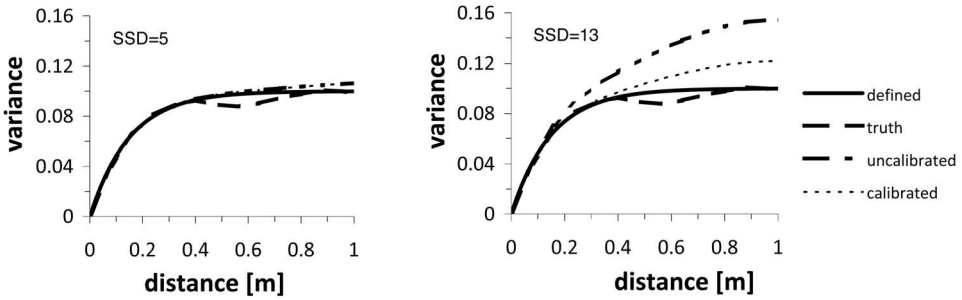

Figure 10. Comparison of the average variograms of the uncalibrated and calibrated null-space-projected fields (difference fields included) with the defined and the true variogram for SSDs of 5 and 13.

linear analysis shows a larger degree of postcalibration predictive uncertainty (Table 4).

6. Conclusions

[50] This paper describes an application of the *Tonkin and Doherty* [2009] NSMC method to a hypothetical saltwater intrusion model patterned after the Henry problem [*Henry*, 1964]. We demonstrate that the NSMC method is an efficient and robust method for quantifying uncertainty in predictions made using a saltwater intrusion model with a mildly heterogeneous parameter field. A comparison with an unconstrained MC analysis shows that use of the NSMC methodology significantly reduces the range of predictive uncertainty (as is expected as a result of the constraints imposed by the calibration data set) while not underestimating the degree of postcalibration uncertainty.

[51] For the case analyzed herein, modest gains in efficiency of the NSMC method were made by reducing the dimensionality of the null-space prior to null-space processing of random parameter fields. By artificially reducing the null-space, however, uncertainty estimates forthcoming from application of the NSMC method no longer encompassed the true range of predictive uncertainty. This was accompanied by the introduction of distortions into calibration-constrained hydraulic conductivity fields during the recalibration process.

[52] The use of difference fields as a way of retaining fine-scale parameter variability in calibration-constrained stochastic fields was also tested in the NSMC analysis. While the use of difference fields resulted in calibration-constrained hydraulic conductivity distributions with details that more closely resemble the true stochastic field, their use does not introduce a noticeable increase in the uncertainty of the prediction of interest in this particular study. However, the computational burden incurred by their use is substantial, more than a factor of 2.

[53] Predictive uncertainty calculated using the NSMC method compares well with the outcomes of linear analysis for this particular prediction. However, the uncertainty associated with the model prediction as calculated with the linear method is slightly greater than that calculated using the NSMC technique. A broader comparison of linear and nonlinear uncertainty analysis methods is presently underway and will form the content of a future publication.

[54] The present study represents, to the authors' knowledge, the first occasion on which the NSMC method has been applied in the context of saltwater intrusion modeling. This type of modeling can often be computationally demanding because of the need for inclusion of density effects in the simulation process, which, in turn, creates the need for representation of many layers in a numerical model. Notwithstanding its difficulties, the outcomes of saltwater intrusion modeling can often have important repercussions for the management of coastal aquifers. Furthermore, in many such modeling contexts a large amount of historical head and concentration data are available through which to calibrate these types of models. Critical components of the modeling process thus become (1) the extraction of as much information as possible from the historical data set through the calibration process and (2) quantification of the uncertainty that remains in critical model predictions after this information has been extracted so that the magnitude of this uncertainty, and the risks associated with this uncertainty, can be properly incorporated into the decision-making process.

[55] *Hunt et al.* [2007] demonstrate the importance of highly parameterized regularized inversion in achieving the first of these outcomes. This paper demonstrates that the NSMC method provides a computationally efficient and robust methodology for achieving the second of these important goals.

References

Abarca, E. (2006), Seawater intrusion in complex geological environments, Ph.D. thesis, 154 pp., Univ. Politèc. de Catalunya, Barcelona, Spain.

Abarca, E., J. Carrera, X. Sanchez-Vila, and M. Dentz (2007), Anisotropic dispersive Henry problem, *Adv. Water Resour.*, *30*, 913–926, doi:10.1016/j.advwatres.2006.08.005.

Alcolea, A., J. Carrera, and M. Medina (2006), Inversion of heterogeneous parabolic-type equations using the pilot-points method, *Int. J. Numer. Methods Fluids*, *51*, 963–980, doi:10.1002/fld.1213.

Bard, J. (1974), *Nonlinear Parameter Estimation*, Academic, New York.

Beckie, R. (1996), Measurement scale, network sampling scale, and groundwater model parameters, *Water Resour. Res.*, *32*(1), 65–76, doi:10.1029/95WR02921.

Beven, K., and A. Binley (1992), The future of distributed models—Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*(3), 279–298, doi:10.1002/hyp.3360060305.

Carrera, J., and S. P. Neuman (1986), Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information, *Water Resour. Res.*, *22*(2), 199–210, doi:10.1029/WR022i002p00199.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L. J. Slooten (2005), Inverse problem in hydrogeology, *Hydrogeol. J.*, *13*(1), 206–222, doi:10.1007/s10040-004-0404-7.

Carrera, J., J. J. Hidalgo, L. J. Slooten, and E. Vazquez-Sune (2010), Computational and conceptual issues in the calibration of seawater intrusion models, *Hydrogeol. J.*, *18*(1), 131–145, doi:10.1007/s10040-009-0524-1.

Certes, C., and G. de Marsily (1991), Application of the pilot point method to the identification of aquifer transmissivities, *Adv. Water Resour.*, *14*(5), 284–300, doi:10.1016/0309-1708(91)90040-U.

Christensen, S., and R. L. Cooley (1999), Evaluation of prediction intervals for expressing uncertainties in groundwater flow model predictions, *Water Resour. Res.*, *35*(9), 2627–2639, doi:10.1029/1999WR900163.

Christensen, S., and J. Doherty (2008), Predictive error dependencies when using pilot points and singular value decomposition in groundwater model calibration, *Adv. Water Resour.*, *31*(4), 674–700, doi:10.1016/j.advwatres.2008.01.003.

Dagan, G., and D. G. Zeitoun (1998), Seawater-freshwater interface in a stratified aquifer of random permeability distribution, *J. Contam. Hydrol.*, *29*(3), 185–203, doi:10.1016/S0169-7722(97)00013-2.

Dausman, A. M., J. Doherty, C. D. Langevin, and M. C. Sukop (2010), Quantifying data worth toward reducing predictive uncertainty, *Ground Water*, *48*(5), 729–740, doi:10.1111/j.1745-6584.2010.00679.x.

Deutsch, C. V., and A. G. Journel (1992), *GSLIB: Geostatistical Software Library User's Guide*, 340 pp., Oxford Univ. Press, New York.

Doherty, J. (2003), Ground water model calibration using pilot points and regularization, *Ground Water*, *41*(2), 170–177, doi:10.1111/j.1745-6584.2003.tb02580.x.

Doherty, J. (2010), PEST: Model-independent parameter estimation, *Watermark Numer. Comput.*, Brisbane, Queensl., Australia. [Available at http://www.pesthomepage.org..

Doherty, J., and R. J. Hunt (2009), Two statistics for evaluating parameter identifiability and error reduction, *J. Hydrol.*, *366*, 119–127, doi:10.1016/j.jhydrol.2008.12.018.

Draper, N. R., and H. Smith (1981), *Applied Regression Analysis*, 2nd ed., John Wiley, Hoboken, N. J.

Gallagher, M., and J. Doherty (2007), Parameter estimation and uncertainty analysis for a watershed model, *Environ. Model. Software*, *22*(7), 1000–1020, doi:10.1016/j.envsoft.2006.06.007.

Giambastiani, B. M. S., M. Antonellini, G. H. P. Oude Essink, and R. J. Stuurman (2007), Salt water intrusion in the unconfined coastal aquifer of Ravenna (Italy): A numerical model, *J. Hydrol.*, *340*, 91–104, doi:10.1016/j.jhydrol.2007.04.001.

Gómez-Hernández, J. J., A. Sahuquillo, and J. E. Capilla (1997), Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data: 1. Theory, *J. Hydrol.*, *203*, 162–174, doi:10.1016/S0022-1694(97)00098-X.

Gómez-Hernández, J. J., H. J. Hendricks Franssen, and A. Sahuquillo (2003), Stochastic conditional inverse modeling of subsurface mass transport: A brief review of the self-calibrating method, *Stochastic Environ. Res. Risk Assess.*, *17*, 319–328, doi:10.1007/s00477-003-0153-5.

Graham, W., and D. Mclaughlin (1989a), Stochastic analysis of nonstationary subsurface solute transport: 1. Unconditional moments, *Water Resour. Res.*, *25*(2), 215–232, doi:10.1029/WR025i002p00215.

Graham, W., and D. Mclaughlin (1989b), Stochastic analysis of nonstationary subsurface solute transport: 2. Conditional moments, *Water Resour. Res.*, *25*(11), 2331–2355, doi:10.1029/WR025i011p02331.

Guadagnini, A., and S. P. Neuman (1999), Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly nonuniform domains: 1. Theory and computational approach, *Water Resour. Res.*, *35*(10), 2999–3018, doi:10.1029/1999WR900160.

Harvey, C. F., and S. M. Gorelick (1995), Mapping hydraulic conductivity: Sequential conditioning with measurements of solute arrival time, hydraulic-head, and local conductivity, *Water Resour. Res.*, *31*(7), 1615–1626, doi:10.1029/95WR00547.

Held, R., S. Attinger, and W. Kinzelbach (2005), Homogenization and effective parameters for the Henry problem in heterogeneous formations, *Water Resour. Res.*, *41*, W11420, doi:10.1029/2004WR003674.

Henry, H. R. (1964), Effects of dispersion on salt encroachment in coastal aquifers, *U.S. Geol. Surv. Water Supply Pap.*, *1613-C*.

Hernandez, A. F., S. P. Neuman, A. Guadagnini, and J. Carrera (2006), Inverse stochastic moment analysis of steady state flow in randomly heterogeneous media, *Water Resour. Res.*, *42*, W05425, doi:10.1029/2005WR004449.

Hill, M. C., and C. R. Tiedeman (2007), *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions and Uncertainty*, Wiley-Interscience, Hoboken, N. J.

Hughes, J. D., C. D. Langevin, and L. Brakefield-Goswami (2010), Effect of hypersaline cooling canals on aquifer salinization, *Hydrogeol. J.*, *18*(1), 25–38, doi:10.1007/s10040-009-0502-7.

Huisman, J. A., J. Rings, J. A. Vrugt, J. Sorg, and H. Vereecken (2010), Hydraulic properties of a model dike from coupled Bayesian and multi-criteria hydrogeophysical inversion, *J. Hydrol.*, *380*, 62–73, doi:10.1016/j.jhydrol.2009.10.023.

Hunt, R. J., J. Doherty, and M. J. Tonkin (2007), Are models too simple? Arguments for increased parameterization, *Ground Water*, *45*(3), 254–262, doi:10.1111/j.1745-6584.2007.00316.x.

Kitanidis, P. K. (1996), On the geostatistical approach to the inverse problem, *Adv. Water Resour.*, *19*(6), 333–342, doi:10.1016/0309-1708(96)00005-X.

Kitanidis, P. K., and E. G. Vomvoris (1983), A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations, *Water Resour. Res.*, *19*(3), 677–690, doi:10.1029/WR019i003p00677.

Langevin, C. D., and W. Guo (2006), MODFLOW/MT3DMS-based simulation of variable-density ground water flow and transport, *Ground Water*, *44*(3), 339–351, doi:10.1111/j.1745-6584.2005.00156.x.

Langevin, C. D., D. T. Thorne Jr., A. Dausman, M. C. Sukop, and W. Guo (2008), SEAWAT version 4: A computer program for simulation of multi-species solute and heat transport, U.S. Geol. Surv., Reston, Va.

Lavenue, M., and G. de Marsily (2001), Three-dimensional interference test interpretation in a fractured aquifer using the pilot point inverse method, *Water Resour. Res.*, *37*(11), 2659–2675.

Lebbe, L., N. Van Meir, and P. Viaene (2008), Potential implications of sea-level rise for Belgium, *J. Coastal Res.*, *24*(2), 358–366, doi:10.2112/07A-0009.1.

Maurer, H., K. Holliger, and D. E. Boerner (1998), Stochastic regularization: Smoothness or similarity?, *Geophys. Res. Lett.*, *25*(15), 2889–2892, doi:10.1029/98GL02183.

Maxwell, R. M., C. Welty, and R. W. Harvey (2007), Revisiting the Cape Cod bacteria injection experiment using a stochastic modeling approach, *Environ. Sci. Technol.*, *41*(15), 5548–5558, doi:10.1021/es062693a.

Moore, C. (2005), The use of regularized inversion in groundwater model calibration and prediction uncertainty analysis, Ph.D. thesis, Univ. of Queensland, St. Lucia, Australia.

Moore, C., and J. Doherty (2005), Role of the calibration process in reducing model predictive error, *Water Resour. Res.*, *41*, W05020, doi:10.1029/2004WR003501.

Moore, C., and J. Doherty (2006), The cost of uniqueness in groundwater model calibration, *Adv. Water Resour.*, *29*(4), 605–623, doi:10.1016/j.advwatres.2005.07.003.

Oude Essink, G. H. P. (2001), Salt water intrusion in a three-dimensional groundwater system in the Netherlands: A numerical study, *Transp. Porous Media*, *43*(1), 137–158, doi:10.1023/A:1010625913251.

Pappenberger, F., and K. J. Beven (2006), Ignorance is bliss: Or seven reasons not to use uncertainty analysis, *Water Resour. Res.*, *42*, W05302, doi:10.1029/2005WR004820.

Rubin, Y. (2003), *Applied Stochastic Hydrogeology*, Oxford Univ. Press, New York.

Rubin, Y., and G. Dagan (1987), Stochastic identification of transmissivity and effective recharge in steady state groundwater flow: 1. Theory, *Water Resour. Res.*, *23*(7), 1192–1200.

Rubin, Y., and A. G. Journel (1991), Simulation of non-Gaussian space random functions for modeling transport in groundwater, *Water Resour. Res.*, *27*(7), 1711–1721, doi:10.1029/91WR00838.

Sanford, W. E., and J. P. Pope (2010), Current challenges using models to forecast seawater intrusion: Lessons from the eastern shore of Virginia, USA, *Hydrogeol. J.*, *18*(1), 73–93, doi:10.1007/s10040-009-0513-4.

Sanz, E., and C. I. Voss (2006), Inverse modeling for seawater intrusion in coastal aquifers: Insights about parameter sensitivities, variances, correlations and estimation procedures derived from the Henry problem, *Adv. Water Resour.*, *29*(3), 439–457, doi:10.1016/j.advwatres.2005.05.014.

Seber, G. A. F., and C. J. Wild (1989), *Nonlinear Regression*, John Wiley, New York.

Simpson, M. J., and T. P. Clement (2004), Improving the worthiness of the Henry problem as a benchmark for density-dependent groundwater flow models, *Water Resour. Res.*, *40*, W01504, doi:10.1029/2003WR002199.

Tikhonov, A. N., and V. Y. Arsenin (1977), *Solution of Ill-Posed Problems*, V. H. Winston, Washington, D. C.

Tonkin, M. J., and J. Doherty (2005), A hybrid regularized inversion methodology for highly parameterized environmental models, *Water Resour. Res.*, *41*, W10412, doi:10.1029/2005WR003995.

Tonkin, M., and J. Doherty (2009), Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques, *Water Resour. Res.*, *45*, W00B10, doi:10.1029/2007WR006678.

Tonkin, M., J. Doherty, and C. Moore (2007), Efficient nonlinear predictive error variance for highly parameterized models, *Water Resour. Res.*, *43*, W07429, doi:10.1029/2006WR005348.

Vasco, D. W., A. Datta-Gupta, and J. C. S. Long (1997), Resolution and uncertainty in hydrologic characterization, *Water Resour. Res.*, *33*(3), 379–397, doi:10.1029/96WR03301.

Vecchia, A. V., and R. L. Cooley (1987), Simultaneous confidence and prediction intervals for nonlinear-regression models with application to a groundwater-flow model, *Water Resour. Res.*, *23*(7), 1237–1250, doi:10.1029/WR023i007p01237.

Voss, C. I., and W. R. Souza (1987), Variable density flow and solute transport simulation of regional aquifers containing a narrow freshwater-saltwater transition zone, *Water Resour. Res.*, *23*(10), 1851–1866, doi:10.1029/WR023i010p01851.

Woodbury, A. D., and T. J. Ulrych (2000), A full-Bayesian approach to the groundwater inverse problem for steady state flow, *Water Resour. Res.*, *36*(8), 2081–2093, doi:10.1029/2000WR900086.

Yeh, T. C. J., M. H. Jin, and S. Hanna (1996), An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields, *Water Resour. Res.*, *32*(1), 85–92, doi:10.1029/95WR02869.

Zhang, Q., R. E. Volker, and D. A. Lockington (2004), Numerical investigation of seawater intrusion at Gooburrum, Bundaberg, Queensland, Australia, *Hydrogeol. J.*, *12*(6), 674–687, doi:10.1007/s10040-004-0333-5.

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 Resour. Res.*, *34*(6), 1373–1413, doi:10.1029/98WR00003.

J. Doherty, Watermark Numerical Computing, 336 Cliveden Ave., Brisbane, Qld 4075, Australia.

D. Herckenrath, Department of Environmental Engineering, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark. (daah@env.dtu.dk)

C. D. Langevin, U.S. Geological Survey, 12201 Sunrise Valley Dr., Reston, VA 20192, USA.