

Quantifying Data Worth Toward Reducing Predictive Uncertainty

by Alyssa M. Dausman¹, John Doherty², Christian D. Langevin³, and Michael C. Sukop⁴

Abstract

The present study demonstrates a methodology for optimization of environmental data acquisition. Based on the premise that the worth of data increases in proportion to its ability to reduce the uncertainty of key model predictions, the methodology can be used to compare the worth of different data types, gathered at different locations within study areas of arbitrary complexity. The method is applied to a hypothetical nonlinear, variable density numerical model of salt and heat transport. The relative utilities of temperature and concentration measurements at different locations within the model domain are assessed in terms of their ability to reduce the uncertainty associated with predictions of movement of the salt water interface in response to a decrease in fresh water recharge. In order to test the sensitivity of the method to nonlinear model behavior, analyses were repeated for multiple realizations of system properties. Rankings of observation worth were similar for all realizations, indicating robust performance of the methodology when employed in conjunction with a highly nonlinear model. The analysis showed that while concentration and temperature measurements can both aid in the prediction of interface movement, concentration measurements, especially when taken in proximity to the interface at locations where the interface is expected to move, are of greater worth than temperature measurements. Nevertheless, it was also demonstrated that pairs of temperature measurements, taken in strategic locations with respect to the interface, can also lead to more precise predictions of interface movement.

Introduction

Kohout (1965) hypothesized that temperature and salinity variations affect groundwater flow in the Floridan aquifer system (Figure 1). Kohout's conceptual model consists of cold sea water entering the base of the Floridan aquifer system and then being warmed by geothermal

heating as the water flows toward the center of the platform. The less dense, warm water then circulates upward toward the surface (Figure 2). The numerical and laboratory experiments of Henry and Hilleke (1972), designed to test the feasibility of the Kohout hypothesis, suggest that large differences in temperature can affect overall flow patterns as well as equilibrium positions of the salt water–fresh water interface. This, in turn, suggests that inclusion of temperature measurements in calibration datasets will enhance the predictive ability of these models.

Models that simulate regional density-dependent flow and dispersive solute transport are often complex and therefore costly to develop. Hunt et al. (2007) argue that for such models to achieve maximum effectiveness in facilitating future system management, parameterization should be as complex as hydraulic property heterogeneity requires. It is important to note that use of complex, highly

¹Corresponding author: Florida Water Science Center, U.S. Geological Survey, 3110 SW 9th Avenue, Fort Lauderdale, FL 33315; (954) 377-5972; fax: (954) 377-5901; adausman@usgs.gov

²Watermark Numerical Computing and National Centre for Groundwater Research and Training, Australia; johndoherty@ozemail.com.au

³Florida Water Science Center, U.S. Geological Survey, 3110 SW 9th Avenue, Fort Lauderdale, FL 33315; langevin@usgs.gov

⁴Department of Earth Sciences, Florida International University, Miami, FL; sukopm@fiu.edu

Received December 2008, accepted December 2009.
Journal compilation © 2010 National Ground Water Association.
No claim to original US government works.
doi: 10.1111/j.1745-6584.2010.00679.x

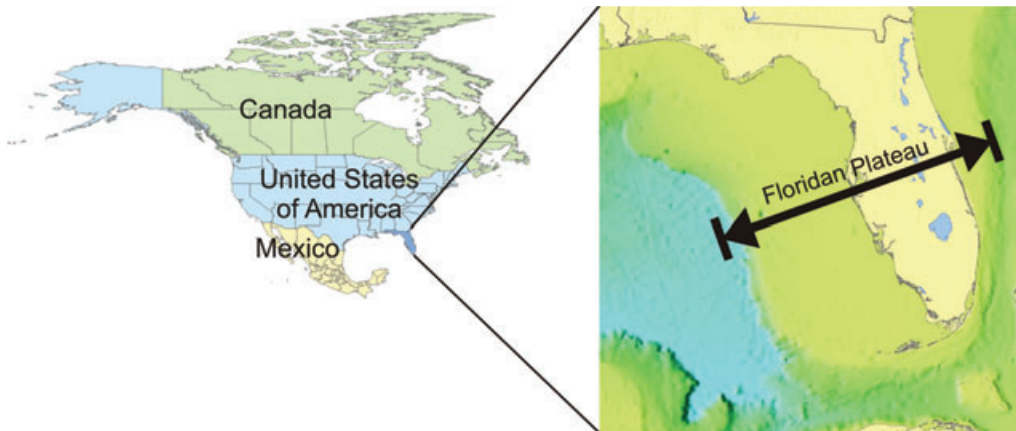


Figure 1. Floridan Plateau (modified from USGS CMG InfoBank Atlas: Florida EEA regions).

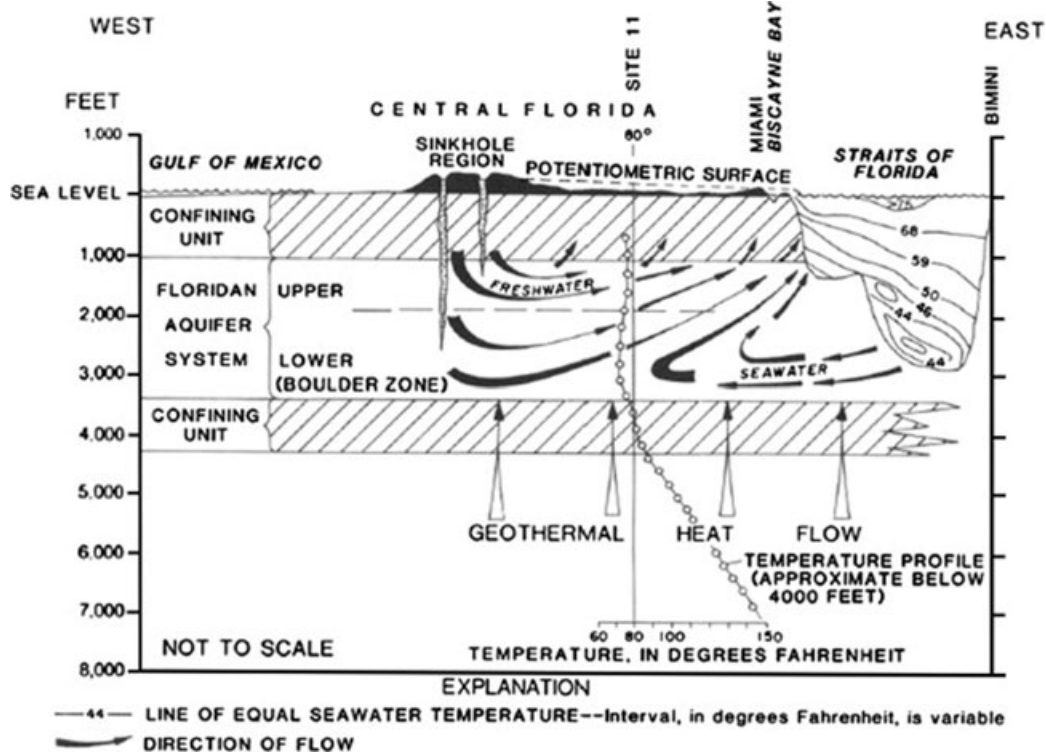


Figure 2. Kohout's conceptualization of convective flow in the Floridan aquifer system (from Kohout, 1965).

parameterized models does not guarantee model predictive accuracy. Hydraulic property complexity inevitably leads to ill-posedness of the inverse problem of model calibration, and hence to the existence of many parameter fields which allow the model to replicate historical observations of system behavior (Moore and Doherty 2005, 2006; Carter et al. 2006). What a highly parameterized model can guarantee, however, is that the inherent uncertainty of model predictions can be satisfactorily quantified. Use of such a model can also provide a mechanism for assessing the relative ability of different types of data to constrain predictive uncertainty, thereby providing a basis for assessing the worth of different types of data as targets for future acquisition. The use of models

in assessing the relative effectiveness of different data acquisition strategies has been discussed by a number of authors, including Moore and Doherty (2005), Gallagher and Doherty (2007), and Hill and Tiedeman (2007).

Different types of analyses have been used to calculate predictive error variance or predictive uncertainty as it pertains to both linear and nonlinear numerical models. The "error variance" of a prediction characterizes the potential for any prediction made by a calibrated model to be wrong. Error variance computation is normally based on the premise that a model has been calibrated against field data and is then used to make predictions of system behavior (Tonkin et al. 2007). The "calibrated model" is endowed with *one parameter field* that ideally should

possess a status such as that of minimum error variance which thereby guarantees its uniqueness. This is a similar but slightly different concept than that of “predictive uncertainty.” Predictive uncertainty explicitly or implicitly characterizes the potential variability of a prediction, through recognition of the existence of *multiple parameter fields* that can be employed by the model (Tonkin et al. 2007). The model is constrained, however, by the necessity for all such parameter fields to lead to model outcomes which correctly simulate historical system behavior. The concept of predictive uncertainty is therefore based on the premise that because model parameters (and inputs) are uncertain, then so too are model outputs. As Tonkin et al. (2007) have noted, “Predictive uncertainty analysis abandons the idea of parameter uniqueness by acknowledging that many different parameter sets enable the model to reproduce the available observations.” Therefore, models that are parameterized in many different ways can be considered to be “correct” (at various levels of confidence) and used to make predictions.

In the present study, model predictive uncertainty is examined using theory based on linear model behavior. For the purpose of this analysis, a model is referred to as “linear” when the relationship between its outputs and its parameters can be represented by a matrix whose coefficients are independent of parameter values. Moore and Doherty (2005) based their formulation of predictive error variance for highly parameterized models on an assumption of linear model behavior. This approach has its roots in the nonuniqueness of the inverse problem in modeling contexts where parameterization complexity is commensurate with real-world hydraulic property heterogeneity. To the extent that a prediction of interest depends on hydraulic property complexity that is unknown to the modeler, that prediction possesses a propensity for error. This propensity can be quantified as an adjunct to solution of the inverse problem of model calibration through regularized inversion. The error can then be reduced by acquiring additional data. The relative worth of one data acquisition strategy over another can be judged in terms of its ability to effect greater reduction in the error variance of a key model prediction than that of the other.

As in these previous studies, the present study also employs a method that is based on linear theory. As was done by Moore and Doherty (2005), a highly parameterized model is assumed. However, in this study, quantification of data worth is based on reduction of predictive uncertainty, rather than reduction of predictive error variance. This formulation was first presented by Doherty (2008a); it was also employed by Christensen and Doherty (2008). Using this methodology, the “worth” of one particular observation in the calibration dataset is ranked against that of others based on its effectiveness in reducing the uncertainty associated with one or more predictions of particular interest.

The present study employs a synthetic model to explore whether the construction of salt water intrusion models would benefit from the inclusion of temperature

measurements, in addition to salinity measurements, within their calibration datasets. Attention was also focused on determining the most effective locations for either type of measurement. The extent to which application of the methodology is compromised by its linear roots is explored by repeating the analysis using a number of different realizations of model hydraulic property fields.

This article is organized as follows. First, the theoretical underpinnings of the methodology employed for assessment of data worth are presented. Then the synthetic model employed in the current study is described. Following that, the methodology is applied to assess the relative merits of acquisition of temperature and concentration data, and to assess locations within the model domain at which acquisition of such data would be most effective. Next, the linear method is tested against heterogeneity-induced model nonlinearity. The article concludes with a discussion of results.

Theory

From a Bayesian viewpoint, measurements of system state that collectively comprise the calibration dataset of a model serve to constrain the values of parameters that can be potentially used by that model, such that only those which allow the model to replicate historical system behavior are admissible. Where model parameters, as well as the noise associated with historical measurements of system state, display Gaussian variability, the formulation of equations that describe the constraining effect of these measurements is relatively straightforward, as is demonstrated below.

Let \mathbf{x} be a vector of multi-Gaussian random variables with covariance matrix $\mathbf{C}(\mathbf{x})$. Let it be partitioned into two subvectors \mathbf{x}_1 and \mathbf{x}_2 such that:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \quad (1)$$

On the basis of this same partitioning, let:

$$\mathbf{C}(\mathbf{x}) = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} \quad (2)$$

Now, let it be supposed that the elements of \mathbf{x}_2 become known. Then \mathbf{C}'_{11} , the covariance matrix of \mathbf{x}_1 conditional on knowing \mathbf{x}_2 , is readily computed (Koch 1987) as:

$$\mathbf{C}'_{11} = \mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21} \quad (3)$$

This formula will now be used to characterize the reduction in uncertainty of model parameters, and predictions which depend on them, accrued through the acquisition of data that contains information that is pertinent to at least some model parameters. Let \mathbf{p} denote the set of parameters employed by a model; these are assumed to represent hydraulic properties distributed throughout

a model domain. It is assumed that the parameters \mathbf{p} are represented at a level of spatial detail that is commensurate with the sensitivity of predictions of interest to hydraulic property heterogeneity prevailing within a system; the effects of this heterogeneity on the uncertainty associated with model predictions can therefore be explored (and hopefully reduced through acquisition of pertinent data). Let the vector \mathbf{h} represent measurements of system state, and the vector $\boldsymbol{\varepsilon}$ represent the noise associated with these measurements. Let the matrix \mathbf{X} represent the means by which model outputs under calibration conditions are calculated from parameters employed by the model; use of this matrix in place of the model thus comprises an assumption of linear model behavior. Then (ignoring parameter and model output offsets without loss of generality):

$$\mathbf{h} = \mathbf{X}\mathbf{p} + \boldsymbol{\varepsilon} \quad (4)$$

Let the scalar s represent a model prediction of interest, and let the vector \mathbf{y} represent the sensitivity of that prediction to model parameters. Then (ignoring offsets once again):

$$s = \mathbf{y}^t \mathbf{p} \quad (5)$$

Combining (4) with (5) leads to:

$$\begin{bmatrix} s \\ \mathbf{h} \end{bmatrix} = \begin{bmatrix} \mathbf{y}^t & \mathbf{0} \\ \mathbf{X} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ \boldsymbol{\varepsilon} \end{bmatrix} \quad (6)$$

Then, using standard matrix relationships for propagation of covariance, the joint covariance matrix of s and \mathbf{h} is calculated as:

$$\begin{aligned} \mathbf{C} \left(\begin{bmatrix} s \\ \mathbf{h} \end{bmatrix} \right) &= \begin{bmatrix} \mathbf{y}^t & \mathbf{0} \\ \mathbf{X} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{C}(\mathbf{p}) & \mathbf{0} \\ \mathbf{0} & \mathbf{C}(\boldsymbol{\varepsilon}) \end{bmatrix} \begin{bmatrix} \mathbf{y} & \mathbf{X}^t \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{y}^t \mathbf{C}(\mathbf{p}) \mathbf{y} & \mathbf{y}^t \mathbf{C}(\mathbf{p}) \mathbf{X}^t \\ \mathbf{X} \mathbf{C}(\mathbf{p}) \mathbf{y} & \mathbf{X} \mathbf{C}(\mathbf{p}) \mathbf{X}^t + \mathbf{C}(\boldsymbol{\varepsilon}) \end{bmatrix} \quad (7) \end{aligned}$$

In Equation (7), $\mathbf{C}(\mathbf{p})$ is the covariance matrix of innate parameter variability, expressing the precalibration uncertainty of hydraulic properties within the model domain. Its diagonal elements express the extent to which the hydraulic property is unknown at any one point within the model domain. The off-diagonal elements of the covariance matrix express the extent to which a single hydraulic property is likely to show spatial continuity, or is likely to be correlated with properties of other types. $\mathbf{C}(\boldsymbol{\varepsilon})$ is the covariance matrix of measurement noise. The $\mathbf{C}(\boldsymbol{\varepsilon})$ matrix is often assumed to be diagonal in recognition of the statistical independence of individual measurements which collectively comprise a calibration dataset.

Application of Equation (3) to Equation (7) leads to:

$$\sigma_s^2 = \mathbf{y}^t \mathbf{C}(\mathbf{p}) \mathbf{y} - \mathbf{y}^t \mathbf{C}(\mathbf{p}) \mathbf{X}^t [\mathbf{X} \mathbf{C}(\mathbf{p}) \mathbf{X}^t + \mathbf{C}(\boldsymbol{\varepsilon})]^{-1} \mathbf{X} \mathbf{C}(\mathbf{p}) \mathbf{y} \quad (8)$$

Equation (8) expresses the predictive uncertainty variance of a prediction s . The first term on the right

side of this equation is the precalibration uncertainty of the prediction. The second term expresses the amount by which this precalibration uncertainty is reduced through the constraining effects of measurements comprising the model's calibration dataset.

An interesting characteristic of Equation (8) is that it contains neither parameter values, nor the values of model outputs; it only features the sensitivities of model outputs to parameters under calibration and predictive conditions, these being contained in the matrix \mathbf{X} and vector \mathbf{y} , respectively. The existing calibration dataset can theoretically be upgraded with new information by simply adding rows to the \mathbf{X} matrix. (Recall that this matrix comprises the sensitivities of corresponding model outputs to parameters employed by the model.) The reduction in resulting σ_s^2 is a measure of the worth of such additions to the calibration dataset; this metric for observation worth is employed in the present study.

Through appropriate transformation, it can be shown that Equation (8) can be rewritten as:

$$\sigma_s^2 = \mathbf{y}^t [\mathbf{X}^t \mathbf{C}^{-1}(\boldsymbol{\varepsilon}) \mathbf{X} + \mathbf{C}^{-1}(\mathbf{p})]^{-1} \mathbf{y} \quad (9)$$

Equation (9) (which like Equation 8 is strictly Bayesian if both measurement noise and innate parameter variability can be characterized by Gaussian probability distributions) can be shown to be equivalent to the *opr* statistic described by Hill and Tiedeman (2007) under the conditions that:

1. weights applied to measurements are the inverse of measurement noise uncertainties;
2. every parameter is accompanied by prior information in which that parameter is equated to its preferred, precalibration value; and
3. the weight matrix assigned to this prior information is equal to the inverse of the $\mathbf{C}(\mathbf{p})$ matrix of innate parameter variability.

Where parameter numbers are high and observation numbers are low, Equation (8) is numerically easier to use than Equation (9) because the matrix requiring inversion in Equation (8) is square with rank equal to the number of observations comprising the calibration dataset, whereas the matrix requiring inversion in Equation (9) has a rank equal to the number of parameters employed by the model. As Fienen et al. (2009) point out the use of a large number of parameters is important in applying either Equation (8) or (9) in the context of data worth analysis, this number normally far exceeding the number that can be uniquely estimated on the basis of currently available data. If parameterization density is reduced to a level at which parameters are uniquely estimable through the imposition of manual regularization devices, such as zones of piecewise constancy, data worth assessment would be compromised by the misinformation that is implicit in the use of such devices.

Central to both of Equations (8) and (9) are two matrices, namely, the matrix of innate parameter variability, $C(\mathbf{p})$, and the matrix that describes the stochastic properties of measurement noise, $C(\boldsymbol{\epsilon})$. The first is an outcome of local hydrogeological knowledge; the second is normally assessed from model-to-measurement fit. Neither can be known with certainty. The nature of spatial variability of subsurface hydraulic properties is often a matter of conjecture in groundwater study areas, even where geostatistical characterizations have taken place. Meanwhile, model-to-measurement misfit is often more reflective of structural noise than measurement noise, the former possessing an unknown (but typically non-diagonal) covariance matrix. Nevertheless, as Gallagher (2008) discusses, reasonable estimates of these matrices can be made at most modeling sites. Furthermore, errors in data worth assessment arising from the erroneous assignment of values to elements of these matrices are likely to be low enough in most modeling contexts to maintain their effective usage in Equations (8) and (9).

The Model

Henry and Hilleke (1972) describe a laboratory experiment and complimentary numerical modeling exercise designed to examine the effects of temperature and salinity variations on groundwater flow. As part of their study, Henry and Hilleke evaluated a variety of different aspect ratios. In the present study, SEAWAT Version 4 (Thorne et al. 2006; Langevin et al. 2008) was used to simulate the Henry and Hilleke problem with an aspect ratio of one.

The design of the Henry and Hilleke problem is shown in Figure 3. Model parameters and grid dimensions

are provided in Table 1. The cross section is 1.025 m by 1.025 m in the $x-z$ vertical plane and the domain defined by the centers of the four corner cells is 1 m by 1 m. The right boundary has a salinity concentration of 35.7 kg/m^3 and a fixed-head boundary, while the left boundary has a constant influx of fresh water. Constant temperature cells surround the domain (Figure 3). The right boundary has a constant temperature of 5°C , the lower left corner has a constant temperature of 50°C , and the upper left corner has a constant temperature of 38.75°C . Temperatures vary linearly along the boundaries. Equations used to calculate the constant temperature at each cell along the boundary are shown in Figure 3. A more detailed description of the problem can be found in Henry and Hilleke (1972) and in Thorne et al. (2006).

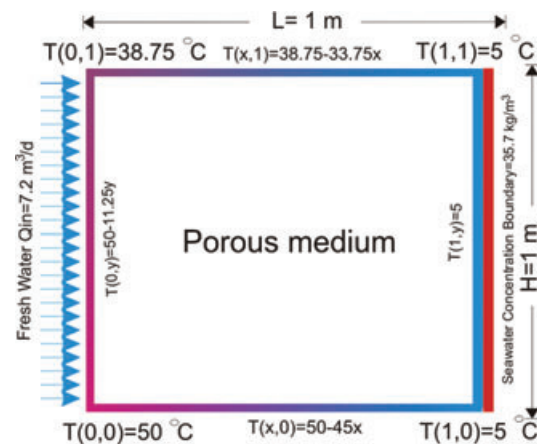


Figure 3. Boundary conditions and problem dimensions for the Henry and Hilleke (1972) problem.

Table 1
Input Parameters for the Henry–Hilleke Problem

Input Parameter	Value	Units	Comments
Columns and layers	41	—	Assigned
$\Delta x(\text{DELR})$ and $\Delta z(\text{DELZ})$	0.025	m	Assigned
K	864	m/d	Hydraulic conductivity
θ	0.35	—	Porosity
D_m	2.0571	m^2/d	Molecular diffusion coefficient
D_t	20.571	m^2/d	Bulk thermal diffusivity
S	0.01	1/m	Specific storage
ρ_{solid}	2710	kg/m^3	Density (calcite aquifer)
K_d	2E-04	m^3/kg	Thermal distribution factor
α	0	m	Dispersivity
ρ_f	1000	kg/m^3	Density of fresh water
ρ_s	1025	kg/m^3	Density of sea water
$\frac{\partial \rho}{\partial C}$	0.7	—	Density change with concentration
$\frac{\partial \rho}{\partial T}$	-0.375	$\text{kg}/(\text{m}^3^\circ\text{C})$	Density change with temperature
Q_{in}	7.2	m^3/d	Total flow in along the left-hand boundary, split into 41 nodes

Model Parameterization

For the simulation intended to represent the original Henry–Hilleke problem (stochastic variations of the original problem are also examined later in this article), hydraulic properties within the model are homogeneous, with spatially uniform values assigned to vertical and horizontal hydraulic conductivity, specific storage, and porosity. However an exception to hydraulic conductivity uniformity exists along the top and bottom layers of the model domain; these are constant temperature cells and are assigned a hydraulic conductivity of 432 m/d. Use of a reduced hydraulic conductivity in this manner minimizes convective heat flux from the temperature boundaries (see Thorne et al. [2006] for an explanation of this strategy). Table 1 lists hydraulic properties employed by the model and Figure 4 shows the salinity concentration and temperature fields computed using the model.

The following parameter types are assumed to be only approximately known at our synthetic study site and therefore are featured in Equation (8):

- dispersivity
- molecular and thermal diffusion coefficients
- thermal distribution factor (controlling the heat transfer between the water and matrix)
- density of the solid matrix
- salinity value for the right-hand boundary
- temperature values along the boundaries
- porosity
- specific storage
- vertical and horizontal hydraulic conductivities.

The first five of these were each assigned a single parameter for the entire model domain; that is, each

was described by a single element of the \mathbf{p} vector. Vertical and horizontal hydraulic conductivities, porosity, and specific storage were defined using the pilot point method (de Marsily et al. 1984). Using this method, hydraulic properties are first assigned to user-specified points within the model domain; these properties are then spatially interpolated to all active model cells (Doherty, 2003). Pilot points were distributed uniformly over the model grid at a spatial interval of 0.05 m; hence, 400 pilot points were assigned to each of these parameter types. Pilot points were also used along respective model boundaries to represent spatial variability of temperature boundary values; these were uniformly spaced at 0.025 m intervals. A total of 1808 parameters are thus contained in the vector \mathbf{p} .

Uniform property values were assigned to all pilot points of the same type (e.g., all 400 horizontal hydraulic conductivity pilot points are assigned a value of 864 m/d) for the purpose of computing *observation* and *prediction sensitivities* (the coefficients of the sensitivity matrices, \mathbf{X} and \mathbf{y}). Spatial variability of hydraulic properties is introduced to Equation (8) (and therefore to our analysis of data worth) through the $\mathbf{C}(\mathbf{p})$ matrix of innate parameter variability. The number of pilot points was chosen to allow the potential for a substantial amount of parameter heterogeneity to exist within the model domain, while also limiting the number of parameters for efficient calculation of the sensitivity matrix \mathbf{X} .

Observations

Temperature and salinity concentration observations were simulated at the locations depicted as black triangles in Figure 4. Thirty-two observations of each of these types were assumed, providing a total of 64 observations in all.

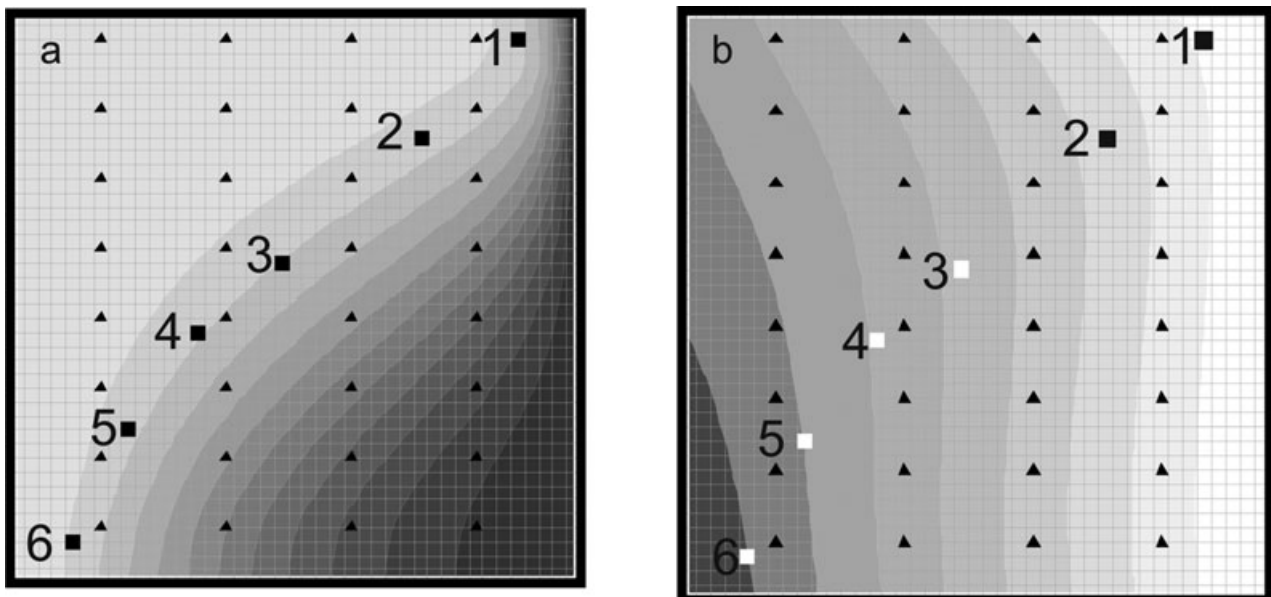


Figure 4. Henry–Hilleke problem showing model outputs computed using parameter values listed in Table 1. (a) Salinity contours (intervals of 3 kg/m^3), 0 kg/m^3 in light gray to 35.7 kg/m^3 in dark gray. (b) Temperature contours (intervals of 5°C) from 5°C in light gray to 50°C in dark gray. Observation locations are depicted as black triangles. Prediction locations are depicted as squares and labeled 1–6.

The \mathbf{X} matrix of Equation (8) thus contains 64 rows and 1808 columns.

Predictions

The ability of salinity concentration and temperature measurements to reduce the uncertainty of six model “predictions” was explored in this study. These predictions are situated at the locations labeled “1” to “6” in Figure 4; they are composed of steady-state concentrations after fresh water input along the left-hand side of the model is reduced by 25%. Sensitivities of each of these predicted concentrations to the 1808 parameters used by the model thus comprised the elements of six different \mathbf{y} vectors employed in Equation (8).

Covariance Matrices

Heterogeneity of model parameters is expressed through the $\mathbf{C}(\mathbf{p})$ matrix of innate parameter variability discussed earlier. In practical settings, this will normally be assessed on the basis of geological knowledge of local aquifer materials, including the range of possible hydraulic property values and the likely degree of hydraulic property continuity. For the present, synthetic, problem, however, elements of the $\mathbf{C}(\mathbf{p})$ matrix were calculated on the basis of a log-exponential variogram with a range of 0.45 m. Sill (i.e., variance) values of 0.07279, 1, and 0.030535 were employed for the logs of hydraulic conductivity, specific storage, and porosity, respectively. This range was chosen because the model domain has dimensions of only 1 m by 1 m. Variance for each parameter (log-transformed in Table 2) was chosen to reflect the expected variability of hydraulic properties in a typical aquifer. The $\mathbf{C}(\mathbf{p})$ matrix describing temperature variability perturbations (from a uniform gradient) along the boundary was based on a variogram with a sill of 0.25°C^2 (standard deviation of 0.5°C) and range of 0.3 m. All other parameters were assumed to possess log variability, with variances depicted in Table 2. Construction of the $\mathbf{C}(\boldsymbol{\epsilon})$ matrix of measurement noise for use in Equation (8) assumed a standard deviation of temperature measurement noise of 0.01°C and a standard deviation of salinity measurement noise equal to 0.3% of measured salinity values.

Application

The following method was employed to assess the importance of individual observations in reduction of uncertainty of predictions made by the numerical model. The sensitivity matrix (\mathbf{X}) was numerically computed by varying each parameter incrementally and computing the change in observation values. The predictive scenario was then run (by decreasing fresh water inflow by 25%) and the prediction sensitivities encapsulated in the vector \mathbf{y} were obtained in the same manner. The predictive uncertainty variance for a specific prediction was calculated using Equation (8). This variance calculation includes the effect of the 32 temperature and 32 salinity observations and the parameter $\mathbf{C}(\mathbf{p})$ matrices described previously.

Each observation in turn is then removed from the model, and the variance is recalculated for a prediction. The difference between the variance calculated before and after removal of a specific observation is the increase in uncertainty for the prediction incurred by removal of the observation; observation worth is assessed on that basis. Thus, for example, if removal of an observation does not increase the uncertainty of the prediction, then that observation is not worth collecting. However, if the uncertainty increases for the prediction when an observation is removed from the dataset, then that specific observation improves the accuracy with which the prediction can be made. This exercise was repeated for all six predictions.

Testing the Effects of Nonlinearity

As previously stated, the matrix \mathbf{X} represents the sensitivities of model outputs (\mathbf{h}) to parameters employed by the model. Equation (8) assumes that the parameter inputs and model outputs are linearly related. It is important to distinguish between the values of parameters employed by the model for the purpose of computation of the \mathbf{X} matrix and \mathbf{y} vector of Equation (8) and the type of parameter variability assumed for the purpose of filling the $\mathbf{C}(\mathbf{p})$ matrix used in this same equation. *Uniform* parameters were employed in the original Henry–Hilleke model for computation of sensitivities in the \mathbf{X} matrix. The $\mathbf{C}(\mathbf{p})$ matrices for Equation (8) are developed based on the natural spatial variability of hydraulic properties one would expect to encounter from field tests or geostatistical analyses.

To investigate the extent to which Equation (8) is valid for nonlinear model behavior, *nonuniform* parameter fields were generated for the purpose of computing sensitivities used in the \mathbf{X} matrix (as opposed to the uniform parameter field used in the original Henry–Hilleke simulation). Calculations of observation worth were then repeated using \mathbf{X} matrices and \mathbf{y} vectors computed on the basis of these nonuniform parameter fields.

Nonuniform parameter fields were created by generating a series of stochastic realizations for the temperature boundary conditions, vertical and horizontal hydraulic conductivity, specific storage, and porosity. These stochastic fields were created based on the same $\mathbf{C}(\mathbf{p})$ matrices discussed in the “Covariance Matrices” section of this article; these same $\mathbf{C}(\mathbf{p})$ matrices were used previously in Equation (8) to describe the spatial variability of each property. Separate random values were also generated for each of the parameters in Table 2, which do not vary spatially. These new parameters were generated from a normal distribution with standard deviations calculated from variances depicted in Table 2. The objective was to test the linear method in conjunction with a nonlinear model by repeating calculations of data worth based on a suite of parameter fields that displayed as much variability as those expressed (through the $\mathbf{C}(\mathbf{p})$ matrix) in the linear equations themselves. Data worth analysis was repeated for five such random parameter fields. This was in addition to the analysis based on the uniform parameter field already described. Hence, data

Table 2
Values Used to Calculate the Predictive Uncertainty Variance for the Parameters Used in the Predictive Uncertainty Exercise.

Parameter	Parameterization	Transformation	Variogram Range	Variance
Horizontal (m/d) ¹	Pilot points	Log	0.45	0.07279
Vertical K (m/d) ¹	Pilot points	Log	0.45	0.07279
Porosity ¹	Pilot points	Log	0.45	0.030535
Specific storage ¹	Pilot points	Log	0.45	1
Boundary temperature (°C) ¹	Pilot points	None	0.3	0.25
Boundary concentration (m ³ /kg)	Uniform	Log	na	0.000976
Density (solid) (kg/m ³)	Uniform	Log	na	0.014228
Thermal distribution factor (m ³ /kg)	Uniform	Log	na	1
Thermal diffusivity (m ² /d)	Uniform	Log	na	0.014228
Molecular diffusion coefficient (m ² /d)	Uniform	Log	na	0.014228
Dispersivity (m)	Uniform	Log	na	0.0625

¹Parameters that spatially vary.

worth assessment was made on the basis of six different X matrices and accompanying y vectors, all pertaining to the same observations and same predictions, but each calculated on the basis of a different parameter field. Differences between data worth rankings obtained through calculations based on these different parameter fields provide a measure of errors incurred through use of linear analysis based on Equation (8) in conjunction with a non-linear model.

Results

The worth of each of the 32 concentration observations and 32 temperature observations to each of the six predictions discussed earlier was calculated (using Equation 8), as the increase in error variance of that prediction incurred through omission of the respective observation from the calibration dataset. Figure 5 shows the results of these computations for predictions 2 and 3 in the original Henry–Hilleke problem; in this figure, the worth of salinity and temperature observations is contoured.

An inspection of Figure 5 reveals the following:

- The worth of salinity data in making predictions of salinity increases is an order of magnitude greater than that of temperature data in making predictions of salinity increases.
- Salinity concentration observations made close to the point at which predicted concentrations are required, and at locations to which the interface moves after diminution of fresh water inflow from the left boundary, are of much greater worth than those made at other locations.
- The requirement for proximity of observation location to prediction location appears to be of less importance for temperature observations than for concentration observations. Of just as much importance is the depth at which a temperature measurement is taken.

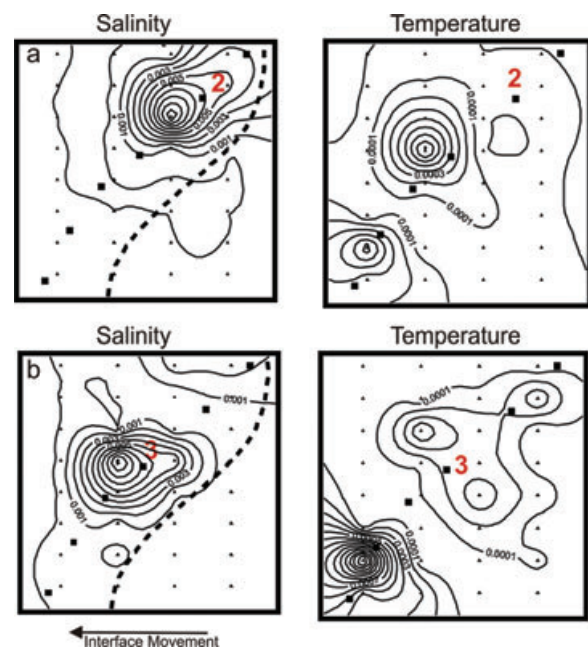


Figure 5. Increase in the predictive uncertainty variance of (a) prediction 2 and (b) prediction 3, incurred by omission of a salinity observation (in column 1) and a temperature observation (in column 2) from the calibration dataset in the Henry–Hilleke problem. The 50% salinity contour is shown dashed. The locations of “predicted” concentrations are shown as squares. Observation locations are shown as triangles.

Measurements made near the toe of the present and future interface (where the temperature is relatively large) seem to be particularly high in predictive information content relative to temperature measurements made elsewhere.

Figure 6 shows the outcomes of salinity concentration and temperature observation worth assessment where sensitivities employed in Equation (8) are computed by the model when supplied with five different stochastic

parameter fields. The horizontal hydraulic conductivity field is shown in each case, together with the position of the 50% sea water line. Conclusions that can be drawn from this figure are almost identical to those that can be drawn from Figure 5, namely:

- Salinity observations are of significantly greater worth to the making of predictions 2 and 3 than are temperature observations.

- Salinity observations in the vicinity of the prediction point, and somewhat to the left of that point where they can detect concentration changes as the interface approaches, are the most informative of the prediction.
- To the extent that a temperature observation can be guaranteed to be useful, it should be located ahead of the interface (in the direction in which it is moving), and near its toe.

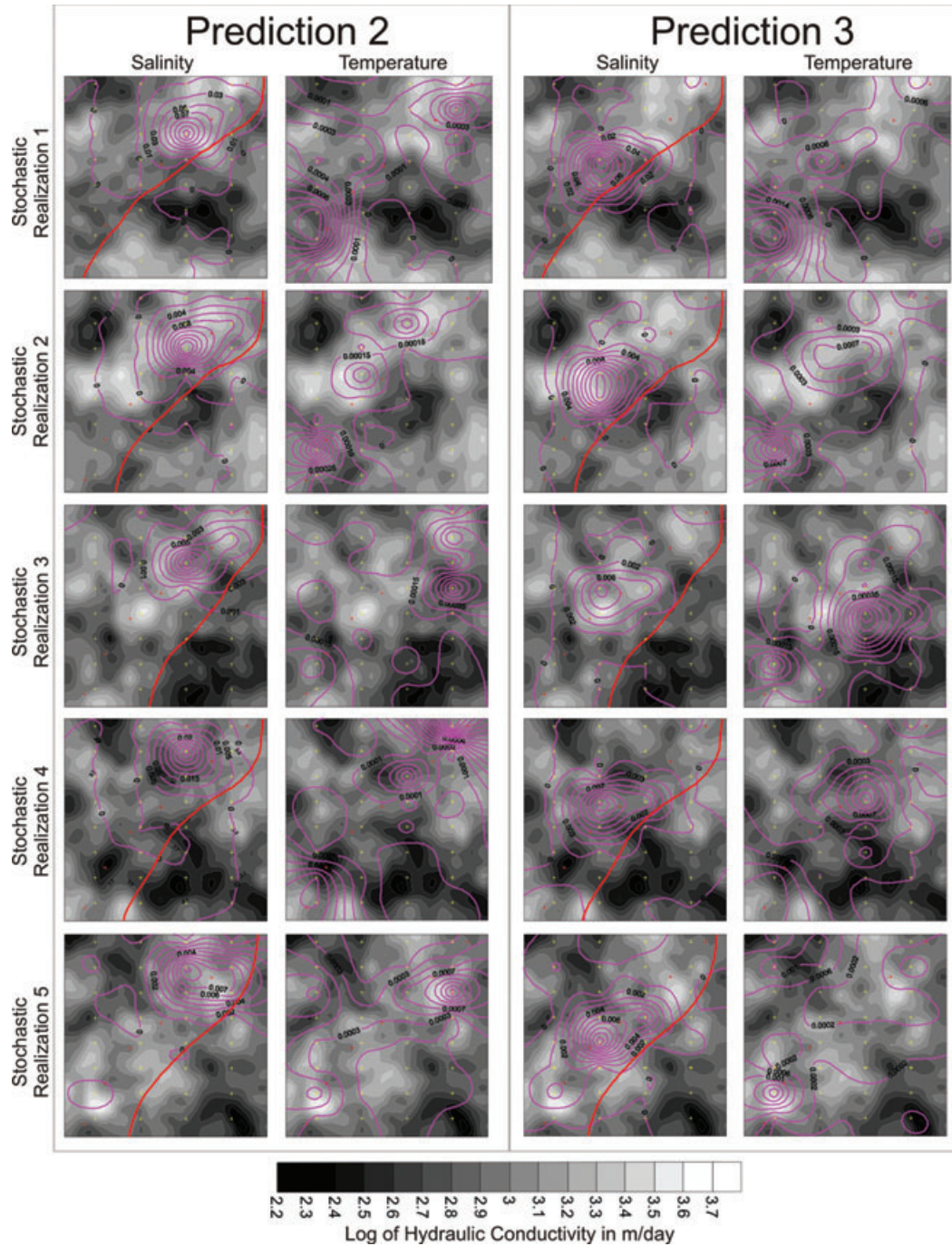


Figure 6. Increase in the predictive uncertainty variance of predictions 2 and 3 incurred by omission of salinity and temperature observations from the calibration dataset calculated for five different parameter set realizations in pink contours. In each case, the salinity contour denoting 50% sea water is shown in red. The hydraulic conductivity field is shown as background to each plot.

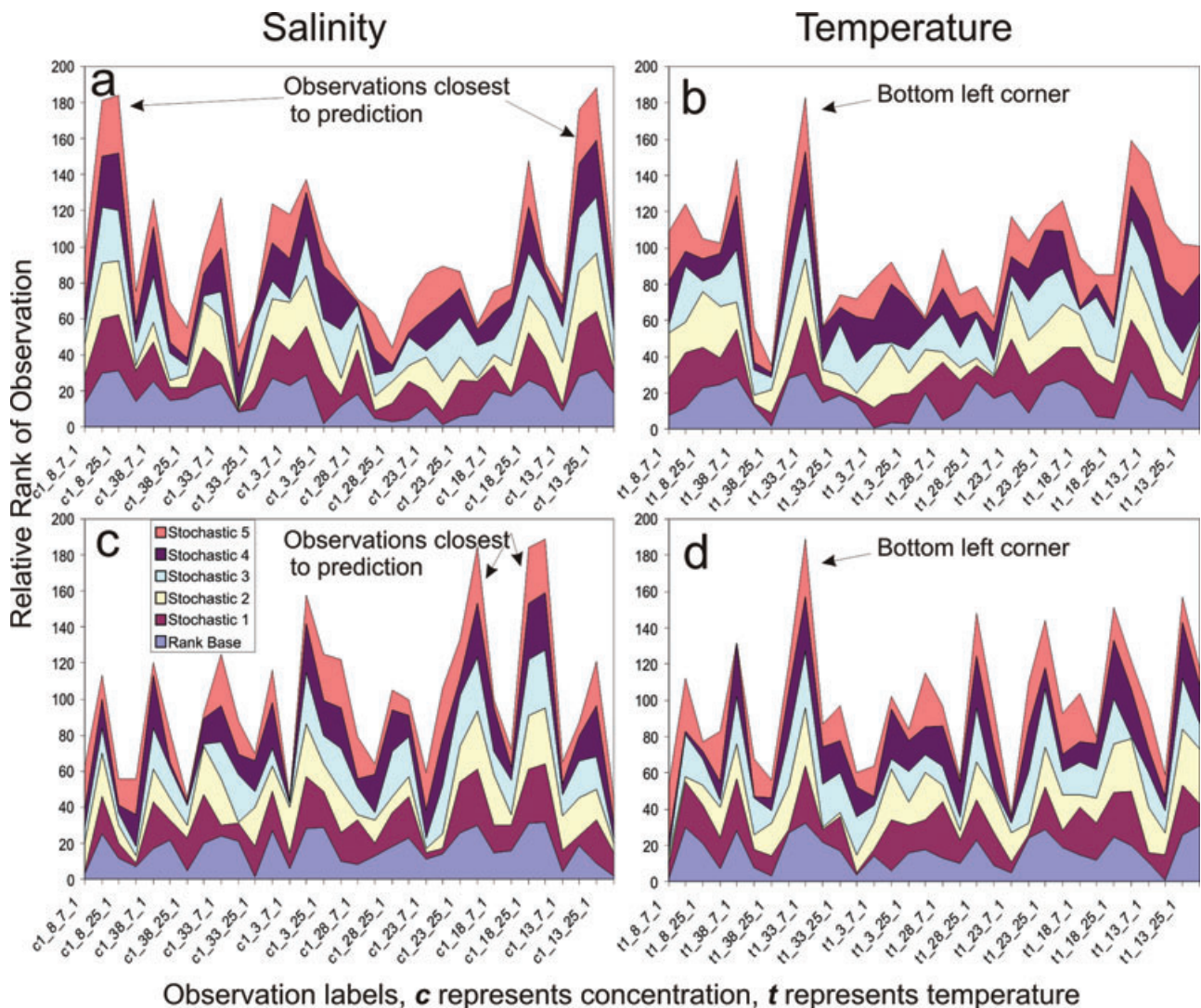


Figure 7. Stacked profiles of the relative ranking of the worth of concentration and temperature measurements for the making of (a) and (b) prediction 2 and (c) and (d) prediction 3. Sensitivities are calculated on the basis of six different hydraulic property sets—a uniform set denoting maximum likelihood parameter values from the original Henry–Hilleke problem (Rank Base) and five different realizations of parameter values generated on the basis of the same $C(p)$ matrices as that employed for the analysis in Equation (8).

The consistency of results plotted in Figure 6 is illustrated in Figure 7 where the rank of each observation in terms of its worth to each prediction is plotted as a series of stacked profiles. For each type of observation, the rank varies between 1 and 32, the last indicating that the respective observation is of most worth to the making of a respective prediction. Observations are named according to a protocol employed in this study whereby “c1” stands for “concentration” and “t1” stands for temperature, whereas the numbers following this prefix indicate the model grid layer, column, and row in which the observation is made. This naming convention is based on the assumption that layer 1 is at the top of the model and that layer numbers increase with depth while column numbers increase from left to right; the model has only one row as it is a cross-sectional model.

As stated earlier, Figures 6 and 7 demonstrate that, for salinity measurements, observations made closest to points at which predictions are required are consistently of greater worth than observations made elsewhere. In the case of temperature, a measurement near the bottom left corner of the model domain is consistently the most information rich for both of the investigated predictions. Results of analyses completed for the other four predictions shown in Figure 4 produced similar outcomes.

The dominance in importance of concentration measurements over temperature measurements is most likely an outcome of the fact that the prediction of interest in the present study is that of future groundwater concentration. It is not surprising, therefore, that because this prediction relates to salinity changes in the area surrounding salt water–fresh water interface, current concentration

information is most diagnostic of the future location and spread of the interface.

Multiple factors affect density-driven groundwater flow, including fluid temperature, and the concentration of solutes within a fluid, both of which affect its buoyancy. Other factors that affect density-driven fluid movement include local hydraulic conductivity and head gradients. To the extent that any or all of these are known, or can be inferred from measurements, the ability to predict the rate and direction of fluid flow, and hence the future position of the interface, is improved. However, because an interface is in fact a concentration gradient, measurements of concentration provide the best basis for prediction of the future position of the interface; this being an outcome of the fact that it is concentration that actually defines the interface. Knowledge of the location of the interface at any particular point of time is the primary source of information on its location at a future point in time.

The same cannot be said for temperature. The secondary worth of temperature data as an indicator of future interface position is evident in Figures 5 and 6. These figures also evince the interrelatedness of buoyancy with other local system properties (e.g., hydraulic conductivity) in their influence on local groundwater flow. They also indicate that temperature worth computations are more affected by nonlinearity than are computations of concentration worth. Figures 5 and 6 also suggest that optimal temperature measurements are best made at locations at which they are informative of buoyancy effects lower in the system than the point at which the actual prediction is required. However, an important feature of the temperature measurement worth plots of Figures 5 and 6 is that most of these plots exhibit two peaks rather than one; in all such cases, one of these peaks is situated below the point at which the interface prediction is required, whereas the other is situated above it. This suggests (not surprisingly) that knowledge of temperature gradient is as important as knowledge of temperature itself.

Discussion and Conclusions

A synthetic model has been employed to demonstrate the application of an easily implemented methodology through which the efficacy of data acquisition in reducing the uncertainty of important model predictions can be maximized, thereby ensuring maximum returns on investments in future data acquisition. The theoretical underpinnings of the methodology rest on an assumption of linear model behavior. This is both its strength and its weakness. Its strength is an outcome of the fact that it can be implemented at relatively light numerical cost, in conjunction with complex models in the highly parameterized contexts in which data worth assessment must necessarily take place. Furthermore, in implementing the methodology, the actual values of parameters, observations, and predictions do not need to be known, only their dependence on parameters used by the model. The assumption of linear model behavior, however, is also a potential weakness of the method, because outcomes of calculations

which rest on this assumption are only approximate, with the level of approximation increasing with the level of model nonlinearity.

A model designed to predict the location of a salt water–fresh water interface is nonlinear, for the sensitivities of model-calculated concentrations and temperatures (especially, those near a salt water–fresh water interface) to parameters employed by the model are dependent on the current position of the interface, which is in turn dependent on parameters employed by the model. This dependence is illustrated in Figure 6, where the use of five different sets of parameters results in five different interface locations.

The purpose of the present study was to first investigate the usefulness of Equation (8) as a basis for guiding acquisition of data that can most effectively illuminate predictions of the future position of a salt water–fresh water interface. Of particular importance in this regard is the extent to which its usefulness may be compromised by nonlinear model behavior. A secondary role (and one that depended on first establishing the usefulness of the methodology in a nonlinear context) was to inquire into the relative abilities of concentration and temperature measurements to inform future interface predictions and investigate the locations at which such measurements should be optimally made.

From the outcomes of this investigation, it can be concluded that the linearity assumption that underpins the methodology described herein does not invalidate its usefulness in evaluating data worth as it pertains to predictions of future interface positions. In particular, although observation worth outcomes varied to some extent when computed on the basis of widely varying parameter fields, the relative worth of measurements of the same types made at different locations (with respect to the current interface position) was typically consistent.

The greater utility of concentration measurements over temperature measurements in reducing the uncertainty of predictions pertaining to the future position of the salt water–fresh water interface was demonstrated. It was further demonstrated that these measurements need to be taken close to the interface, but offset from it in the direction that it is expected to move. Analyses presented herein suggest that the importance of this information rests more on its ability to describe the nature of the interface (in particular, its current location and concentration spread), than on its ability to provide information that is pertinent to local nuances of groundwater flow. Temperature measurements, on the other hand, provide less information on the current interface position or characteristics. Hence, they are of secondary importance to predictions of future interface location. Furthermore, the relationship between the point at which a concentration prediction is required and that at which temperature measurements should be most effectively made in order to support that prediction is not as clear as it is for concentration measurements. The study suggests that a pair of temperature measurements, one made below the predictive point and the other above it, would be most effective, as these measurements

thereby establish the local temperature gradient. If both measurements cannot be taken, then the lower one would be preferred.

The study documented herein could be extended to inquire into the efficacy of acquisition of data of other types. For example, the effectiveness of head measurements and/or of direct measurement of system hydraulic properties in reducing the uncertainty of concentration (or other) predictions could be calculated, thereby providing the modeler with extra dimensions of information on which to base his/her strategies for future data acquisition.

Although demonstrated on a synthetic case, the methodology discussed herein, through which yet-to-be-acquired data can be ranked in order of its relevance to the making of important predictions of future system behavior, is generally applicable. The authors wish to point out that this method has been used in practical settings in conjunction with models that have been built to address real-world environmental problems and calibrated against site-specific data. Use of the methodology at those sites has provided valuable guidance in selecting between alternative data acquisition strategies.

Software

Uncertainty calculations described in this study were implemented using programs from the PEST suite of software (Doherty 2008a, 2008b). Computation of observation worth, based on Equation (8), was implemented using the PREDUNC5 program from that suite. PEST and its utility support software are open source and public domain, and it can be downloaded from the following site: <http://www.pesthomepage.org>

References

- Carter, J.N., P.J. Ballester, Z. Tavassoli, and P.R. King. 2006. Our calibrated model has poor predictive value: An example from the petroleum industry. *Reliability Engineering and System Safety* 91, 1373–1381.
- Christensen, S., and J.D. Doherty. 2008. Predictive error dependencies when using pilot points and singular value decomposition in groundwater model calibration. *Advances in Water Resources* 31, 674–700.
- de Marsily, G.H., G. Lavedan, M. Boucher, and G. Fasanino. 1984. Interpretation of interference tests in a well field using geostatistical techniques. In *Geostatistics for Natural Resources Characterization*, eds. Verly et al., 831–850. Holland: D. Reidel Publishing.
- Doherty, J. 2008a. *PEST: Model-Independent Parameter Estimation*. Brisbane, Australia: Watermark Numerical Computing.
- Doherty, J. 2008b. *PEST Groundwater Data Utilities*. Brisbane, Australia: Watermark Numerical Computing.
- Doherty, J. 2003. Groundwater model calibration using pilot points and regularisation. *Ground Water* 41, no. 2: 170–177.
- Fiene, M.N., J. Doherty, R. Hunt, and H. Reeves. 2010. Using prediction uncertainty analysis to design hydrologic monitoring networks: Example applications from the Great Lakes Water Availability Pilot Project. U.S. Geological Survey Scientific Investigation Report in preparation.
- Gallagher, M.R. 2008. Predictive uncertainty analysis for surface water and groundwater flow models. Ph.D. thesis, University of Queensland, Australia.
- Gallagher, M., and J. Doherty. 2007. Predictive error analysis for a water resource management model. *Journal of Hydrology*, 34, no. 3–4: 513–533.
- Henry, H.R., and J.B. Hilleke. 1972. *Exploration of Multiphase Fluid Flow in a Saline Aquifer System Affected by Geothermal Heating*, 105 pp. University of Alabama Bureau of Engineering Research Contract No. 14-08-0001-12681. Washington DC: U.S. Geological Survey.
- Hill, M.C., and C.R. Tiedeman. 2007. *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty*. Hoboken, New Jersey: John Wiley and Sons Inc.
- Hunt, R.J., J.E. Doherty, and M.J. Tonkin. 2007. Are models too simple? Arguments for increased parameterization. *Ground Water* 45, no. 3: 254–262.
- Koch, K.R.. 1987. *Parameter Estimation and Hypothesis Testing in Linear Models*. Berlin, Germany: Springer-Verlag.
- Kohout, F.A. 1965. A hypothesis concerning cyclic flow of salt water related to geothermal heating in the Floridan Aquifer. *New York Academy of Science Transactions, Series 2* 28, 249–271.
- Langevin, C.D., D.T. Thorne, Jr., A.M. 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. Geological Survey Techniques and Methods Book 6, Chapter A22, 39 pp.
- Moore, C., and J.E. Doherty. 2006. The cost of uniqueness in groundwater model calibration. *Advances in Water Resources* 29, no. 4: 605–623.
- Moore, C., and J.E. Doherty. 2005. Role of the calibration process in reducing model predictive error. *Water Resources Research* 41, no. 5: W05050.
- Thorne, D., C.D. Langevin, and M.C. Sukop. 2006. Addition of simultaneous heat and solute transport and variable fluid viscosity to SEAWAT. *Computer and Geosciences* 32, 1758–1768.
- Tonkin, M., J. Doherty, and C. Moore. 2007. Efficient non-linear predictive error variance for highly parameterized models. *Water Resources Research* 43, W07429, DOI: 10.1029/2006WR005348.