Bias-corrected groundwater model prediction uncertainty analysis

YONAS DEMISSIE[1], ALBERT VALOCCHI[1], BARBARA MINSKER[1] & BARBARA BAILEY[2]

1 *Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, 205 N. Mathews, Urbana, Illinois 61801, USA*

ydemissi@uiuc.edu

2 *Department of Mathematics and Statistic, San Diego State University, 5500 Campanile Drive, San Diego, California 92182, USA*

Abstract The incomplete description of the subsurface processes by physically-based groundwater models often results in biased and correlated prediction errors, thus suggesting the need for systematic correction of errors before conducting prediction uncertainty analysis. In this work, error-mapping artificial neural networks (ANN) are used to correct the physically-based groundwater model (MODFLOW) prediction errors. The resulting prediction uncertainty of the coupled MODFLOW-ANN model is then assessed using three alternative methods. The first method establishes approximate confidence and prediction intervals using first-order least-squares regression approximation (also called first-order error analysis). The second method employs bootstrap approaches that involve resampling of the uncertain data with replacement and repeated model runs for constructing the confidence and prediction intervals. The third method relies on a Bayesian approach that uses analytical or Monte Carlo methods to derive the posterior distribution. The performance of these approaches is evaluated using a hypothetical case study developed based on a phytoremediation site at the Argonne National Laboratory. The results indicate that the three approaches yield comparable confidence and prediction intervals, thus making the computationally efficient first-order error analysis approach attractive for estimating the coupled model uncertainty. The results also demonstrate that the error-mapping ANN not only captures some of the local biases in the MODFLOW prediction, but also systematically reduces the prediction variance.

Keywords calibration; complementary modeling; bias correction; uncertainty analysis

INTRODUCTION

Natural hydrogeological complexity and the inability to extensively monitor the subsurface lead to uncertain and possibly inaccurate numerical simulations of groundwater flow and transport. As a consequence, a single deterministic prediction of the state variable is most often not sufficient to characterize the subsurface system. Instead, it is a common practice to assess standard errors or uncertainty associated with model prediction. Such analysis also provides useful information that can aid in decision-making. The standard uncertainty analysis techniques, such as the least-squares estimate of variance, are usually based on an unbiased model having random prediction error. However, many physically-based mathematical models in hydrology violate this assumption. For example, hydrological models that do not account for the necessary detail of the true system may have biased parameter estimates and yield substantially correlated prediction errors (Moore & Doherty, 2005; Cooley & Christensen, 2006). This, in return, results in prediction interval estim ates that either provide inadequate coverage relative to the observed data or are excessively wide.

In this paper, we use error-predicting artificial neural network (ANN) in a complementary fashion to handle the prediction bias and error correlation arising mainly from ignored or misrepresented processes in the physically-based groundwater model, MODFLOW. The ANN is trained to map the prediction errors using the MODFLOW head predictions and other related data as inputs (see Demissie *et al*. 2006, for details). The forecasted error is then simply superimposed upon the head prediction to get a more accurate result with significantly reduced bias and correlation in the error. Our main objective here is to present least-squares, bootstrap and Bayesian based approaches to compute standard error for the combined MODFLOW-ANN model. In all three cases, we consider the propagation of uncertainty due to the estimation variance of both the MODFLOW and ANN parameters. However, since the ANN model uses the estimated MODFLOW heads as an input, its predictive variance also incorporates the effect of input data uncertainty. The standard errors are presented as approximate 95% confidence and prediction intervals. The methods are applied to a hypothetical test case, developed based on a phytoremediation site at the Argonne National Laboratory, demonstrating improved estimation of groundwater head prediction intervals.

LEAST-SQUARES ESTIMATE APPROACH (FOEA)

We start by casting the combined MODFLOW-ANN model in regression form as:

$$h_i(\boldsymbol{\theta}, \boldsymbol{w}) = M_i(\boldsymbol{\theta}) + N_i\left(M_i(\boldsymbol{\theta}), \boldsymbol{w}\right) \quad \text{i = 1,..., n} \tag{1}$$

where $h_i(.)$ is the updated head prediction, $M_i(.)$ is the head predicted by MODFLOW which depends upon calibrated parameters $\boldsymbol{\theta}$, $N_i(.)$ is the head correction (residual) computed by ANN which depends upon calibrated weight parameters $\boldsymbol{w}$ and the MODFLOW head prediction. The combined uncertainty is computed by considering the variances from each model separately without explicit consideration of their covariance. In related work, van Horssen *et al*. (2002), evaluate the uncertainty of a logistic regression model that uses spatially interpolated data as inputs to identify wetland areas. They considered uncertainty from the regression model and ordinary block kriging while neglecting the covariance between the two models. In our case, however, since the ANN uses the estimated MODFLOW head as an input, the uncertainty of the MODFLOW prediction will be reflected in the resulting uncertainty of ANN by considering both parameter (weights and biases) and input variability. A modified method, originally presented by Ogden & Tarpey (2005) to quantify uncertainty of regression models with externally estimated parameters, is adopted here to estimate the ANN variance, $\boldsymbol{\sigma}_N^2$. The MODFLOW variance, $\boldsymbol{\sigma}_M^2$, is estimated using the least-squares theory (see Chapter 8 of Hill & Tiedeman, 2006). The decomposition of the combined prediction variance, $\boldsymbol{\sigma}_h^2$, is:

$$\boldsymbol{\sigma}_h^2 = \boldsymbol{\sigma}_M^2 + \boldsymbol{\sigma}_N^2 \tag{2}$$

$$\begin{aligned}\boldsymbol{\sigma}_h^2 = \boldsymbol{F}_\theta\left(\sigma_{\varepsilon M}^2(\boldsymbol{F}_\theta^t\boldsymbol{F}_\theta)^{-1}\right)\boldsymbol{F}_\theta^t + \boldsymbol{F}_w[\sigma_{\varepsilon N}^2(\boldsymbol{F}_w^t\boldsymbol{F}_w)^{-1} + (\boldsymbol{F}_w^t\boldsymbol{F}_w)^{-1}(\boldsymbol{F}_{\theta N}^t\boldsymbol{F}_w)^t\left(\sigma_{\varepsilon M}^2(\boldsymbol{F}_\theta^t\boldsymbol{F}_\theta)^{-1}\right)\ldots \\ (\boldsymbol{F}_{\theta N}^t\boldsymbol{F}_w)(\boldsymbol{F}_w^t\boldsymbol{F}_w)^{-1}]\boldsymbol{F}_w^t\end{aligned} \tag{3}$$

where $\boldsymbol{F}_\theta = \partial M(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is a sensitivity matrix for the MODFLOW prediction evaluated at the optimal $\boldsymbol{\theta}$ values, $\hat{\boldsymbol{\theta}}$; $\boldsymbol{F}_w = \partial N(M(\hat{\boldsymbol{\theta}}), \boldsymbol{w})/\partial\boldsymbol{w}$ is a sensitivity matrix for the ANN prediction evaluated at the optimal $\boldsymbol{w}$ values, $\hat{\boldsymbol{w}}$; $\boldsymbol{F}_{\theta N} = \partial N(M(\boldsymbol{\theta}), \hat{\boldsymbol{w}})/\partial\boldsymbol{\theta}$ is a sensitivity matrix for the ANN with respect to $\boldsymbol{\theta}$ and evaluated at $\hat{\boldsymbol{\theta}}$, $\sigma_{\varepsilon M}{}^2 = \phi_M / (n-p)$ and $\sigma_{\varepsilon N}^2 = \phi_N / (n-q)$ are the calculated error variance of MODFLOW and ANN with the corresponding Φ_M and Φ_N fitting objective function values and p and q number of parameters, respectively. Assuming Gaussian error for the combined model and using the Student-t distribution for critical values, confidence and prediction intervals are computed (see Chapter 8 of Hill & Tiedeman, 2006).

BOOTSTRAP AND MONTE CARLO APPROACH

The FOEA approximation of confidence and prediction intervals are sufficiently accurate for the cases where the model is linear, at least after first-order Taylor series expansion near the optimal parameter values, and sample sizes are large enough to have a Gaussian error. However, these conditions rarely apply to actual groundwater modelling application. Furthermore, based on the nature of the model, the computation of the sensitivity matrix could be complex. As an alternative to the FOEA approach, we can employ the bootstrap (Efron & Tibshirani, 1993) and Monte Carlo approaches to estimate the ANN and MODFLOW variances, respectively. First, using the estimated variance of $\boldsymbol{\theta}$ resulting from calibration, we generate multiple $\boldsymbol{\theta}$ parameter sets that will be used for Monte Carlo simulation to get the corresponding head distribution. Then, we resample from the head distribution with replacement and repeatedly train the ANN to get its prediction distribution. Finally, the confidence and prediction intervals for the combined model are computed by adding the percentiles from the two predictions' distributions.

BAYESIAN APPROACH

We also examine the Bayesian approach as a third alternative to account for input data uncertainty when computing the ANN prediction variance. This variance, together with the variance of MODFLOW from the least-squares estimate, is used to compute the combined model prediction variance. Unlike a standard ANN which is trained to minimize the least-squares model fit objective function, in this case we use the Bayesian ANN which is trained to maximize the posterior distribution of the ANN parameters (MacKay, 1992). Once the network is trained, the predictive distribution of the network output can be obtained by integrating the trained ANN over the optimal posterior distribution of the weights. Wright (1999) further extends the Bayesian ANN framework to incorporate input data uncertainty. Assuming Gaussian prior and model error, he derived a Gaussian ANN prediction with variance:

$$\boldsymbol{\sigma}_n^2 = \boldsymbol{F}_w^t \boldsymbol{H}^{-1} \boldsymbol{F}_w + \boldsymbol{\sigma}_m^2 \boldsymbol{F}_{mn}^t \boldsymbol{F}_{mn} \tag{4}$$

Where $\boldsymbol{H} = \partial^2 \Phi(\boldsymbol{w}) / \partial \boldsymbol{w}^2$ is a Hessian matrix evaluated at $\hat{\boldsymbol{w}}$, Φ is the training objective function, and $\boldsymbol{F}_{mn} = \partial N(M(\boldsymbol{\theta}), \hat{\boldsymbol{w}}) / \partial M(\boldsymbol{\theta})$ is the sensitivity of ANN to input data. The same procedures discussed for FOEA are followed to construct the confidence and prediction intervals.

CASE STUDY

A single-layer unconfined heterogeneous synthetic aquifer is numerically generated based on an existing groundwater flow model of a phytoremediation site at the Argonne National Laboratory

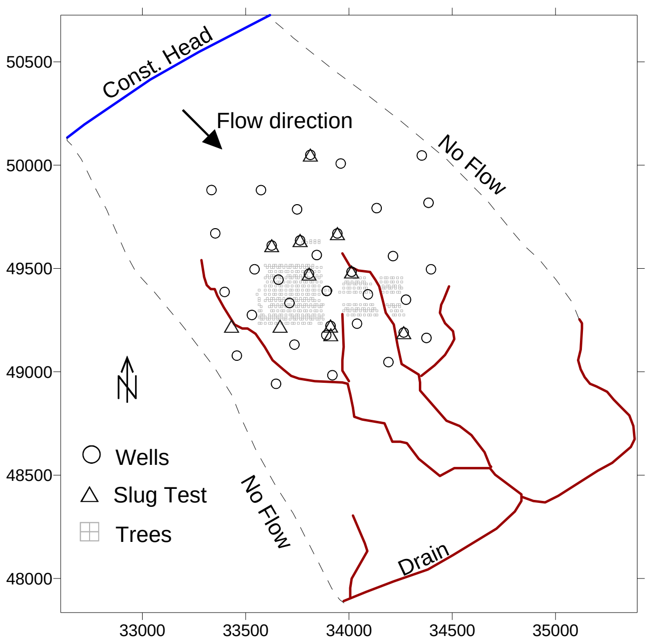

Fig. 1 MODFLOW model set-up for the study area.

(Quinn *et al*., 2001). In this study, instead of using a uniform hydraulic conductivity field as in the original model, a heterogeneous hydraulic conductivity field is generated using conditional SGSIM (Deutsch & Journel, 1998). The generated log-conductivity field has a geometrical mean of 0.7 cm/day and an isotropic exponential variogram with correlation length of 45.72 m and variance of 3.46 (Fig. 2). Hydraulic conductivity estimates from 11 slug tests are used to condition the simulation (see Fig. 1 for their locations). The aquifer is assumed to receive input from monthly varying boundary head to the northwest and from seasonally varying but spatially uniform infiltration recharge from above. The evapotranspiration rates from the trees also vary as a function of their maturity level as well as the season of the year. A transient MODFLOW model is constructed to generate the synthetic aquifer that would serve as a surrogate for the real site and provides observation datasets to develop and test the proposed methodologies. A simplified MODFLOW model is constructed for the synthetic aquifer using a uniform hydraulic conductivity field while keeping the remaining model inputs and boundary conditions identical to the synthetic model. We generate six years of monthly piezometric head measurements from 36 spatially scattered monitoring wells from the synthetic aquifer. The first four years data are used to calibrate the simplified model, and to train and test the ANN, while the subsequent two years data are used as prediction datasets. The simplified model uniform hydraulic conductivity is estimated using PEST, a nonlinear parameter estimation software developed by Doherty (2003). Further details about the case study are given by Demissie *et al.* (2006).

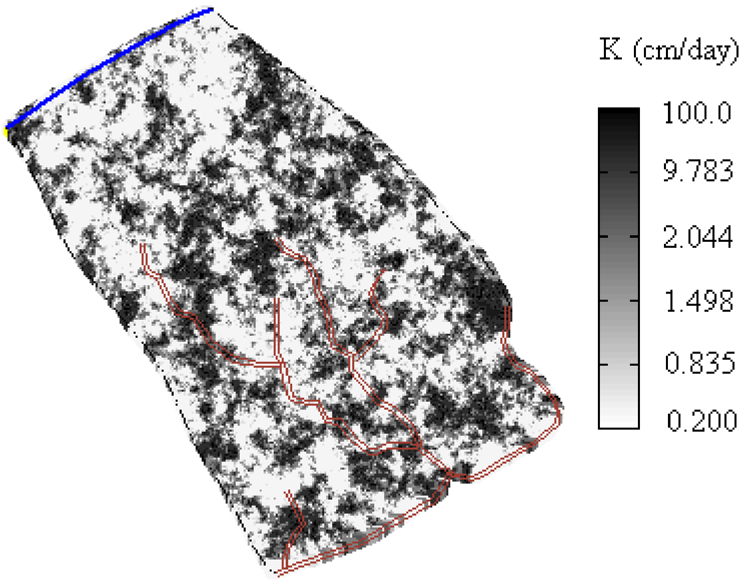

Fig. 2 The generated hydraulic conductivity field ("True" field).

RESULTS AND DISCUSSION

The histograms (Fig. 3(a),(b)) show the overall error distributions before and after the ANN update for calibration and prediction, respectively. These figures illustrate that the error-correcting ANN significantly reduced the prediction uncertainty of MODFLOW. Moreover, even though the calibrated MODFLOW model is globally unbiased (see the mean values in Fig. 3), it misses some of the localized behaviour derived from heterogeneities, which in return leads to biased predictions at the individual wells. The ANN provides improved prediction by recovering some of the errors that presumably result from the unaccounted system behaviours due to the overly simplified representation of the true hydraulic conductivity field. For example, compared to the MODFLOW model alone, the coupled model achieves 87% and 85% reduction in RMSE for the four-year calibration and two-year prediction datasets, respectively. Comparing the absolute mean errors

(AME) at the individual wells during calibration period, 19 out of 36 wells have AME above 1.0 ft before update. After the update, the AME value is reduced to less than 0.1 ft. Similarly, during

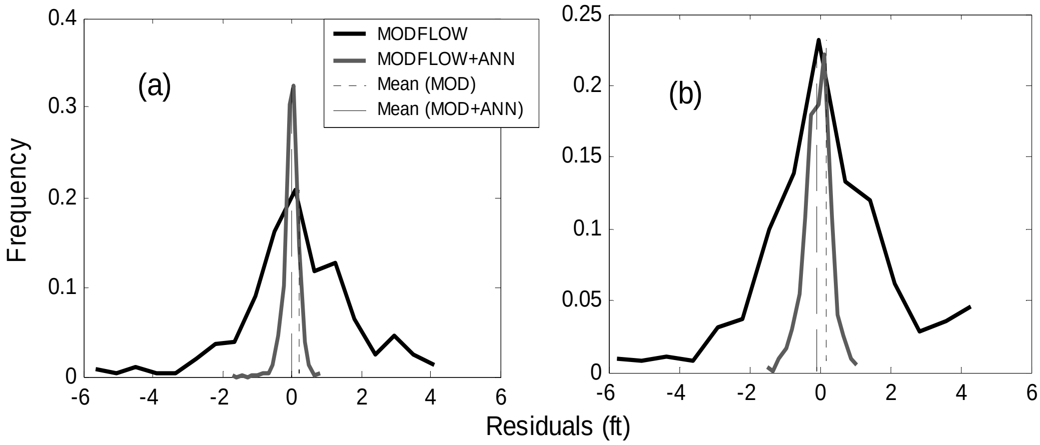

Fig. 3 Error distributions before and after update during calibration (a) and prediction (b) periods. The broken lines are mean values of the distributions.

prediction 19 out of 36 wells have AME above 1.21 ft before update which is reduced to 0.24 ft after the update.

The updated model confidence and prediction intervals accuracy (i.e. percentage of data enclosed within the intervals) is close to the desired value of 95% for most of the wells. For example, prior to the update, out of the 36 wells no single well has at least 90% of the head data fall within the 95% confidence intervals. However, after the update 15 wells for the FOEA, 16 wells for Bayesian, and 21 wells for Bootstrap have at least 90% of the head data enclosed within the 95% confidence intervals. In the case of prediction, prior to the update, 16 wells have at least 90% of the head predictions within the 95% prediction intervals compared to 22 wells for FOEA, 25 wells for Bayesian and 24 wells for Bootstrap after the update. In addition to this, the update also narrowed the prediction intervals. We select one typical well to demonstrate the results of the uncertainty analysis. Figures 4 and 5 show the 95% confidence and prediction intervals for both the pre and post-updated MODFLOW model during calibration and prediction periods, respectively. It can be seen from Fig. 4 that prior to the update, the model confidence interval is very narrow and does not contain most of the observed data. This is not a surprise since we have calibrated the model for a single hydraulic conductivity value using 1728 measured groundwater head data values; thus, a global optimal result is achieved which is reflected by narrow confidence intervals. However, because the predictions are locally biased as we discussed above, these intervals might not contain 95% of the data. The ANN corrects this by reducing the bias and increasing the confidence intervals. Thus, after the update, most of the observed groundwater head data are enclosed within the confidence interval (Fig. 4). The accuracy (percentage of data within the intervals) for most of the wells is very close to the desired value of 95%. On the other hand, the pre-updated model prediction interval (Fig. 5) contains most of the data but is very wide, which might reduce the model's potential to be used as a decision support tool. This is because of the relative large value of the calibration error variance, which is an explicit component of the prediction intervals to account for possible future measurement errors. The ANN corrects this by reducing the bias and decreasing calibration error variance. As shown in Fig. 5, most of the future observations are also enclosed within the prediction interval of the updated model. However, compared to the pre-updated model, the updated prediction has a more narrow prediction interval band.

The three approaches provide reasonably similar confidence and prediction intervals. As stated previously, the accuracy of the FOEA and the Bayesian approaches, examined in this study, highly depend on the characteristics of the residuals (that is, observed minus computed head values). The two approaches are expected to provide reasonably accurate intervals if the residuals are un-biased, independent and normally distributed. In the present study, since the updating techniques capture most of the observed biases and dependencies within the residuals, the two approaches are expected to provide equally accurate intervals. As long as the main source of prediction uncertainty is parameter error as in the present study, and provided enough parameter

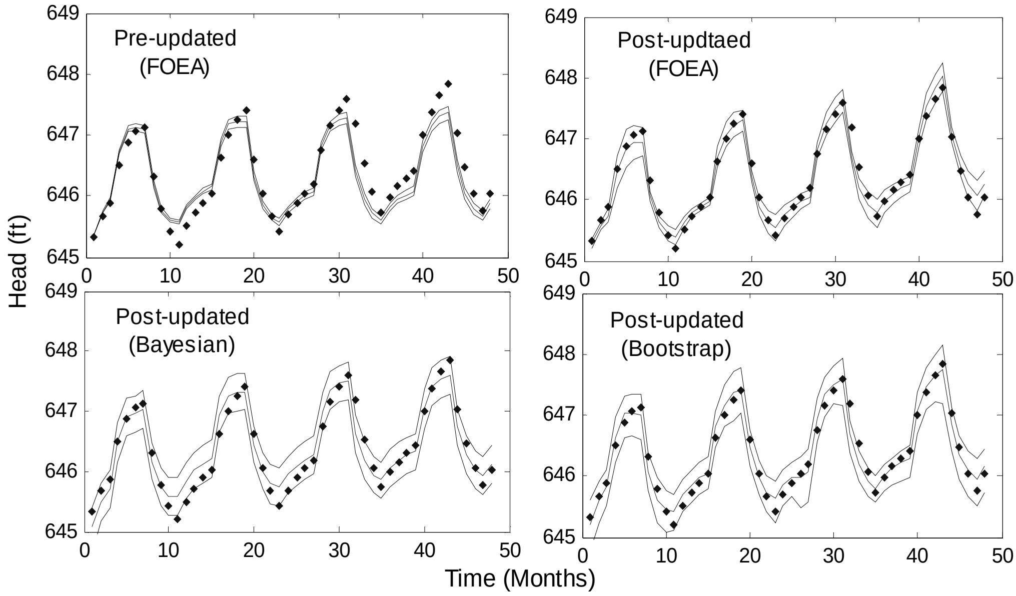

Fig. 4 The 95% confidence intervals before and after update during calibration periods for the selected example well.

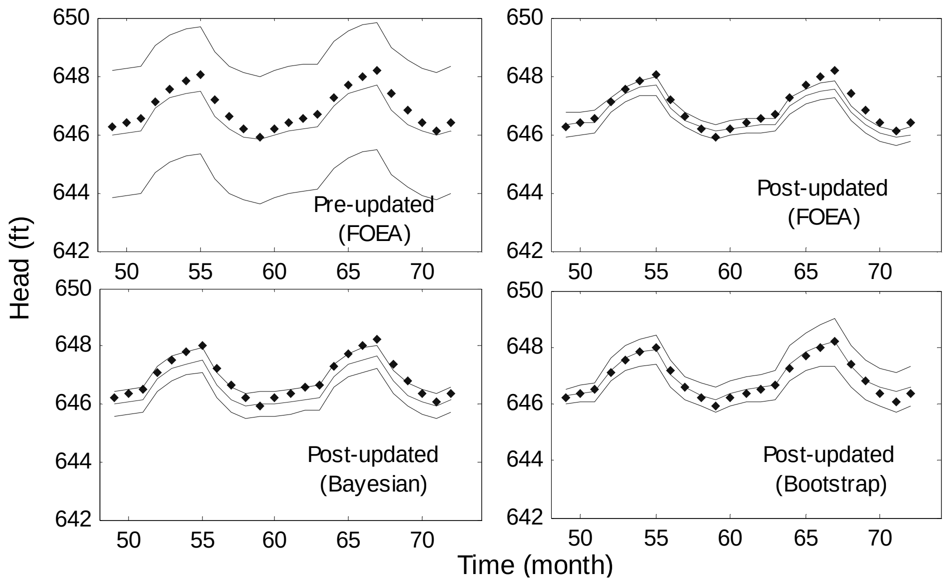

Fig. 5 The 95% prediction intervals before and after update during prediction periods for the example well.

realizations are used, the non-parametric Bootstrap approach is also expected to provide reasonably accurate intervals. Another plausible explanation for these results might be that the model is simple without strong non-linearity. Since the residuals characteristics depends on model complexity and degree of linearity, further study is required for complete assessment of the potential of computationally efficient FOEA and Bayesian approaches to estimate uncertainty intervals.

SUMMARY AND CONCLUSION

In this paper, we have presented the FOEA, bootstrap and Bayesian framework for propagating uncertainty through the coupled MODFLOW-ANN model. The approaches are evaluated using a hypothetical groundwater flow problem with the hydraulic conductivity field as the sole source of uncertainty. The results demonstrate that, by not properly accounting for the hydraulic conductivity heterogeneity, the calibrated MODFLOW model leads to locally biased predictions with narrow confidence and wide prediction intervals, which fail to include most of the observed values. However, updating these predictions with an error-predicting ANN model results in reduced model bias and prediction uncertainty. These enhance the potential use of the model for management practices. The three approaches also provide comparable results in this simple case study, but in a more complex case study they might provide different results. In the future, we will evaluate the approaches for real site applications.

Acknowledgements This material is based upon work supported by the US Department of Energy under Grant no. DE-FG07-02ER63502.

REFERENCES

Cooley, R. & Christensen, S. (2006) Bias and uncertainty in regression-calibrated models of groundwater flow in heterogeneous media. *Adv. Water Resour.* **29**(5), 639–656.

Deutsch, V. & Journel, G. (1998) *GSLIB Geostatistical Software Library and User's Guide*. Oxford University Press, New York, USA.

Demissie, Y., Valocchi, A., Minsker, B. & Bailey, B. (2006) Data-driven modeling approach to enhance MODFLOW head prediction. In: *Proc. MODFLOW and More 2006* (May 2006, Colorado, USA).

Doherty, J. (2002) PEST: Model-independent parameter estimation. In: *Watermark Numerical Computing*. Corinda, Australia.

Efron, B. & Tibshirani, R. (1993) *An Introduction to the Bootstrap.* Chapman & Hall, Florida, USA.

Hill, M. & Tiedeman, C. (2006) *Effective Groundwater Model Calibration*. John Wiley & Son, New Jersey, USA.

MacKay, D. (1992) A practical Bayesian framework for backprop networks. *Neural Computation* **4**(3), 448–472.

Moore, C. & Doherty, J. (2006) The cost of uniqueness in groundwater model calibration. *Adv. Water Resour.* 29(4), 605-623.

Ogden, R. & Tarpey, T. (2006) Estimation in regression models with externally estimated parameters. *Biostatistics* 7(1), 115 – 129.

Quinn, J., Negri, R. & Hinchman, L. (2001) Predicting the effect of deep-rooted hybrid poplars on the groundwater flow system at a large-scale phytoremediation site. *Int. J. Phytoremediation* **3**(1), 41–60.

van Horssen, P. W., Pebesma, E. & Schot, P. (2002) Uncertainties in spatially aggregated predictions from a logistic regression model. *Ecological Modelling* 154(1), 93-10.

Wright, W. (1999) Bayesian approach to neural-network modeling with input uncertainty. *IEEE Trans. Neural Networks* **10**, 1261–1270.