4/9/2002

General (Ground water, modeling)

LA-UR-02-726

Approved for public release; distribution is unlimited.

Title: Analysis of model sensitivity and predictive uncertainty of capture zones in the Espanola Basin regional aquifer, Northern New Mexico

Author(s): Velimir V. Vesselinov, EES-6, LANL
Elizabeth H. Keating, EES-6, LANL
George A. Zyvoloski, EES-6, LANL

Submitted to: ModelCare2002: 4th International Conference on Calibration and Reliability in Groundwater Modelling: A few steps closer to reality
Prague, Czech Republic, June 17-20, 2002
Acta Universitatis Carolinae Geologica

Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Form 836 (8/00)

13623

Analysis of model sensitivity and predictive uncertainty of capture zones in the Espanola Basin regional aquifer, Northern New Mexico

Velimir V. Vesselinov, Elizabeth H. Keating, George A. Zyvoloski
vvv@lanl.gov, ekeating@lanl.gov, gaz@lanl.gov,
Earth and Environmental Sciences Division, Los Alamos National Laboratory, NN 87545, USA

Abstract

Predictions and their uncertainty are key aspects of any modeling effort. The prediction uncertainty can be significant when the predictions depend on uncertain system parameters. We analyze prediction uncertainties through constrained nonlinear second-order optimization of an inverse model. The optimized objective function is the weighted squared-difference between observed and simulated system quantities (flux and time-dependent head data). The constraints are defined by the maximization/minimization of the prediction within a given objective-function range. The method is applied in capture-zone analyses of groundwater-supply systems using a three-dimensional numerical model of the Espanola Basin aquifer. We use the finite-element simulator FEHM coupled with parameter-estimation/predictive-analysis code PEST. The model is run in parallel on a multi-processor supercomputer. We estimate sensitivity and uncertainty of model predictions such as capture-zone identification and travel times. While the methodology is extremely powerful, it is numerically intensive.

Introduction

Models are an important and widely-used tool for analysis of natural hydrogeological systems. Typically, a model is calibrated against known system behavior by inversion and applied afterward to make some predictions. Any model prediction is undoubtably uncertain. Sources of uncertainty such as conceptualization, parameterization and discretization (grid-resolution) errors in the model are very important but are difficult to evaluate (cf. Caganis & Smith, 2001). Uncertainty in model parameter estimates (Carrera & Neuman, 1986) is also important; quantitative analysis of the significance of this source of model uncertainty to model predictions is the focus of this paper.

For a given model, due to uncertainty in the observations, discrepancy between observed and simulated system behavior, low model sensitivity of estimated parameters and correlations among their respective estimation errors, there may be multiple parameter sets that produce equally well-calibrated results (as calculated by some measure), but provide quite different predictions. Detailed exploration of the parameter space through Monte-Carlo analysis can be extremely computationally intensive and may, in fact, produce a very small set of parameter combinations that result in calibrated models. One way to circumvent this problem is through local sensitivity analysis of the predictions with respect to the model parameters for the calibrated model. This approach is valid only for a small portion of parameter space in the vicinity of calibration point due to correlations among parameter estimates and common model non-linearity. A better approach is to estimate uncertainty in the model predictions by constrained non-linear optimization of the inverse model, or the so-called prediction analysis (Vecchia & Cooley, 1987; Doherty, 2000; Keating et al. 2000; Vesselinov et al. 2001). The optimized objective function is the weighted squared-difference between observed and simulated system quantities. The constraints are defined by the maximization/minimization of a given prediction within an objective-function range which is defined by the excepted level of

uncertainty. This analysis allows an efficient, though computationally intensive, way to determine the impact of parameter uncertainty on the model predictions. This analysis does not depend on assumptions about model linearity and takes into account the uncertainty in the calibration targets and correlations among parameter estimates. The method is applied in capture-zone analysis of groundwater-supply systems using a 3D numerical inverse model of the Espanola Basin aquifer.

Hydrogeologic Setting and Model Development

The Espanola basin is an important source for municipal and agricultural groundwater supply in Northern New Mexico, USA. Los Alamos National Laboratory (LANL) is situated in the western margin of the basin (Fig.1). Due to concern over potential impact of present and past laboratory activities on the groundwater, LANL is conducting an extensive characterization program. A vital element in the program are model analyses, which have been used for conceptualization and parameterization of the aquifer, design and siting of new characterization wells, and prediction of the fate and transport of potential contaminants.

Details about the regional hydrogeology and the model development can be found in Keating et al. (2001). Topographic relief in the basin exceeds 2100 m. The Rio Grande, the Rio Chama, and the lower reaches of many tributaries comprise the regional groundwater discharge zone. The climate is semi-arid, with total precipitation ranging from 18 to 86 cm. Annual precipitation is strongly elevation dependent. Aquifer recharge is primarily from infiltration in the higher elevations—estimates range from 7–26% of total precipitation. Due to low precipitation rates and high evapotranspiration demand, little or no recharge occurs at lower elevations other than along stream channels. The aquifer is predominately comprised of weakly consolidated basin-fill sedimentary rocks over 3,000 m thick near the basin axis. Contours of water level data indicate that hydraulic gradients are generally towards the Rio Grande.

Three-dimensional inverse models have been developed using the finite-element simulator FEHM (Zyvoloski et al. 1997), grid generator LaGriT (Trease et al. 1996), parameter-estimation and predictive-analysis code PEST (Doherty, 2001). We have coupled a relatively coarse-grid basin-scale model with a high-resolution site-scale model via flux boundary conditions (Keating et al, 2002). Medium heterogeneity is defined by a 3D geologic basin model. We have identified 23 geologic units, including two fault zones; only some of these units exist within the site-scale model. Due to insufficient information about the connectivity between surface and subsurface water, the major rivers within the basin are simulated as specified head, even though this significant simplification of the actual hydrogeological conditions.

Although actual patterns of recharge in the basin are undoubtably complex, we use a very simple time-invariant recharge model whose parameters are estimated in the inverse process. We define two types of recharge: (1) areal recharge, which is spatially distributed based on the ground-surface elevation, and (2) canyon-focused recharge. Our model parameters are the total recharge flux, Q, the elevation below which no areal recharge occurs, Z_{min}, and the ratio between total canyon-focused and total areal recharge, κ.

Parameter estimation

We estimated the model parameters for both basin- and site-scale models independently (Keating et al., 2002) using the code PEST which is based on the Levenberg-Marquardt algorithm. The objective function subject to minimization is defined as:

(1) $\Phi = [\mathbf{c} - f(\mathbf{b})]^T \mathbf{W} [\mathbf{c} - f(\mathbf{b})]$

where **c** is a vector [N×1] of optimization targets, **b** is a vector [M×1] of model parameters, **W** is a diagonal weight matrix [N×M], and f is our model. By minimizing Φ, the algorithm searches for the maximum-likelihood parameter set **b** that provide the best fit between simulated $f(\mathbf{b})$ and measured **c** quantities. The vector of optimization targets includes (1) "pre-development" steady-state heads and fluxes to the rivers and (2) transient heads over a 50-year development period. The vector **b** includes (1) the recharge parameters (Q, Z_{min}, κ), (2) permeability (k) of the various hydrostratigraphic units, and (3) the globally uniform specific storage (Ss). The total recharge Q available to the site-scale model is determined by the basin model; during site-scale model calibration, we allow only redistribution of Q over the model domain by varying Zmin and κ. W represents the relative weight of each optimization target defined subjectively based on the measurement quality and spatial clustering of wells.

The details of our procedure and results are described in Keating et al. (2001). We use the parameter estimation process to determine the extent to which the recharge model and hydrogeological zonation are justified given the available calibration targets. The inverse results support the conceptual model that very or little no recharge occurs below certain elevations; in fact, the Z_{min} is determined with relatively low uncertainty for both models. All the hydrostratigraphic units and fault zones are assumed to be uniform; most are assumed to be isotropic. For some, we have tried to distinguish between horizontal and vertical permeability components, but the available data allow this only for three of the units. Analyses of parameter sensitivities and covariance eigenvectors of estimation errors suggest that the available data do not support the degree of detail present in the hydrostratigraphic model and so some units have been combined. Parameters successfully estimated by the site-scale inverse model and their respective 95% linear confidence limits are shown in Table 1.

Predictive analysis

Previously, using both forward and reverse particle tracking methods we have estimated the capture zones of the major water-supply wellfields in the region (Vesselinov & Keating, 2002). A particularly important result from these analyses was that potential contamination entering the saturated zone beneath Mortandad canyon at LANL would eventually be captured by a nearby municipal water-supply well (PM-5). According to our calibrated site-scale model, the median arrival time is approximately 5,600 years, and about 80% of the plume will be captured within 10,000 years. Since our model parameters are uncertain, however, it is reasonable to assume that this prediction is uncertain. We apply predictive analysis to determine the range of predictions possible, given our calibration criteria, and to determine which of the uncertain parameters most influence predictive uncertainty. The basis for this analysis is as follows: If we define a prediction p as

(2) $\quad p = f(\mathbf{b})$.

where f is our model under predictive conditions, and unknown **b** maximizes/minimizes p subject to

(3) $\quad [\mathbf{c} - f(\mathbf{b})]^T \mathbf{W} [\mathbf{c} - f(\mathbf{b})] = \delta\Phi_{min}$,

where Φ_{min} is defined for the maximum-likelihood estimates $\mathbf{b}_{ML}$. For the maximum-likelihood case (Bard 1974),

(4) $$\delta = \frac{N}{N-M} F_{\alpha}(N, N-M) + 1,$$

where F is the F distribution and α is the confidence level. The constrained optimization of **b** is solved using PEST as an iterative nonlinear Lagrangian problem as proposed by Vecchia & Cooley (1987).

The potential release of contamination from the canyon is simulated as an instantaneous source of 10,000 particles, released within a relatively small area (10×10 m2). We simulated conservative, advective-dispersive transport of these particles for 10,000 years, and estimated the model uncertainty about two predictions: (1) the proportion of total mass captured by PM-5; and (2) the median arrival time. For the purpose of this analysis we fix porosity and all other transport parameters (Vesselinov & Keating, 2002). We vary all the model parameters listed in Table 1.

Results and conclusions

Our results suggest that significant uncertainty exists with our predictions of total mass captured by PM-5 and also of mean travel time to this well. The results of the total-mass-captured predictive analysis are shown in Table 1 and Figure 1. Within the 95% confidence range of our best parameter estimates, we find that the contamination plume can be either entirely captured or entirely missed by the pumping well. A snapshot of the particle distribution 10,000 years after their input and the head [m] contours are presented in Fig.1. On the figure, the particles locations (small pluses) define the predictive estimate of the plume for the case of no contamination reaching the well, and solid lines represent the respective flow field. For the other extreme predictive case, all the contamination is already captured at the well and that is why no particles are shown; dashed lines define the head contours. This is achieved with relatively small changes in the model parameters (Table 1), which modify the flow directions substantially enough (Fig.1) but produce satisfactory match between observed and simulated calibration targets. Most of the model parameters are defined with relatively high estimation uncertainty; however, only the recharge parameter κ dictates the predictive uncertainty. This parameter explicitly defines the amount of water recharged along the canyon; the higher the κ, the higher the canyon recharge. This result allows us to concentrate on data collection to improve estimation of canyon recharge rates so that we will able to decrease the predictive uncertainty of our model. It is important to note, however, that this result is specific to this particular model prediction and cannot be generalized to other model applications.

In the total-mass-captured predictive analysis, median arrival times range from 1,700 to more than 10,000 years (no contamination reaches the well within 10,000 years). Similar predictive analysis about the median arrival time identified even lower value—1,300 years.

Our results demonstrate that uncertainty in the model predictions can be substantial and are important to analyze. The constrained nonlinear optimization of an inverse model, or the so-called predictive analysis in PEST, is an extremely powerful tool for analysis of predictive model uncertainty. A major disadvantage of the applied second-order search method for nonlinear optimization is that the global minimization of both estimation and predictive analyses is not guaranteed. Therefore, we plan the utilization of other, more robust, search algorithms. Future work will also include analyses of the impact of model errors (conceptualization, parameterization and discretization) on the uncertainty of our predictions.

Acknowledgements

This work has been supported by the Department of Energy, USA, and Los Alamos National Laboratory Hydrogeologic Characterization Program.

References

Bard, M., *Nonlinear Parameter Estimation*, Academic Press, New York, New York, 1974.

Caganis, P, Smith, L., A Bayesian approach to the qualification of the effect of model error on the predictions of groundwater models, Water Res. Research, 39, 2309-3222, 2001.

Carrera, J., Neuman, S. P., Estimation of Aquifer Parameters Under Transient and Steady State Conditions: 2. Uniqueness, Stability, and Solution Algorithms, *Water Res. Research*, 25, 211-227, 1986.

Doherty, J., *PEST: Model Independent Parameter Estimation v5.1*, Watermark Computing, Brisbane, Australia, 2000.

Keating, E.H., Doherty, J., Vesselinov, V.V., Approaches to Quantifying Uncertainty in Flow Directions in the Regional Aquifer Beneath the Pajarito Plateau, Northern New Mexico. *EOS Trans., AGU*, 81(48), 2000.

Keating, E., Kwicklis, E., Vesselinov, V., Idar, A., Lu, Z., Zyvoloski, G., Witkowski, M., A Regional Flow and Transport Model for Groundwater at Los Alamos National Laboratory, *Los Alamos National Laboratory Report*, LA-UR-01-2199, 2001.

Keating, E., Vesselinov, V.V., Kwicklis, E., Lu, Z., Coupling Large- and Local-Scale Inverse Models of the Española Basin, *Groundwater*, 2002, (submitted).

Trease, H., George, D., Gable, C.W., Fowler, J., Kuprat, A. and Khamyaseh, A., The X3D grid generation system, *Numerical grid generation system in computational fluid dynamics and related fields*, B.K. Soni, J.F. Thompson, H. Hausser and P.R. Eiseman (ed.), Mississippi State University, 1996.

Vecchia, A.V., Cooley, R.L., Simultaneous Confidence and Prediction Intervals for Nonlinear Regression Models with Application to a Groundwater Flow Model. *Water Res. Research*, 23(7), 1237-1250, 1987.

Vesselinov, V.V., Keating, E.H., Doherty, J., Analysis of uncertainty in model predictions of flow and transport in the Espanola Basin regional aquifer, Northern New Mexico. *Eos. Trans., AGU*, 82(47), 2001.

Vesselinov, V.V., Keating, E.H., Analysis of capture (well-head protection) zones of the Buckman wellfield and a potential new Ranney collector north of the Otowi bridge, *Los Alamos National Laboratory Report*, 2002, (in print).

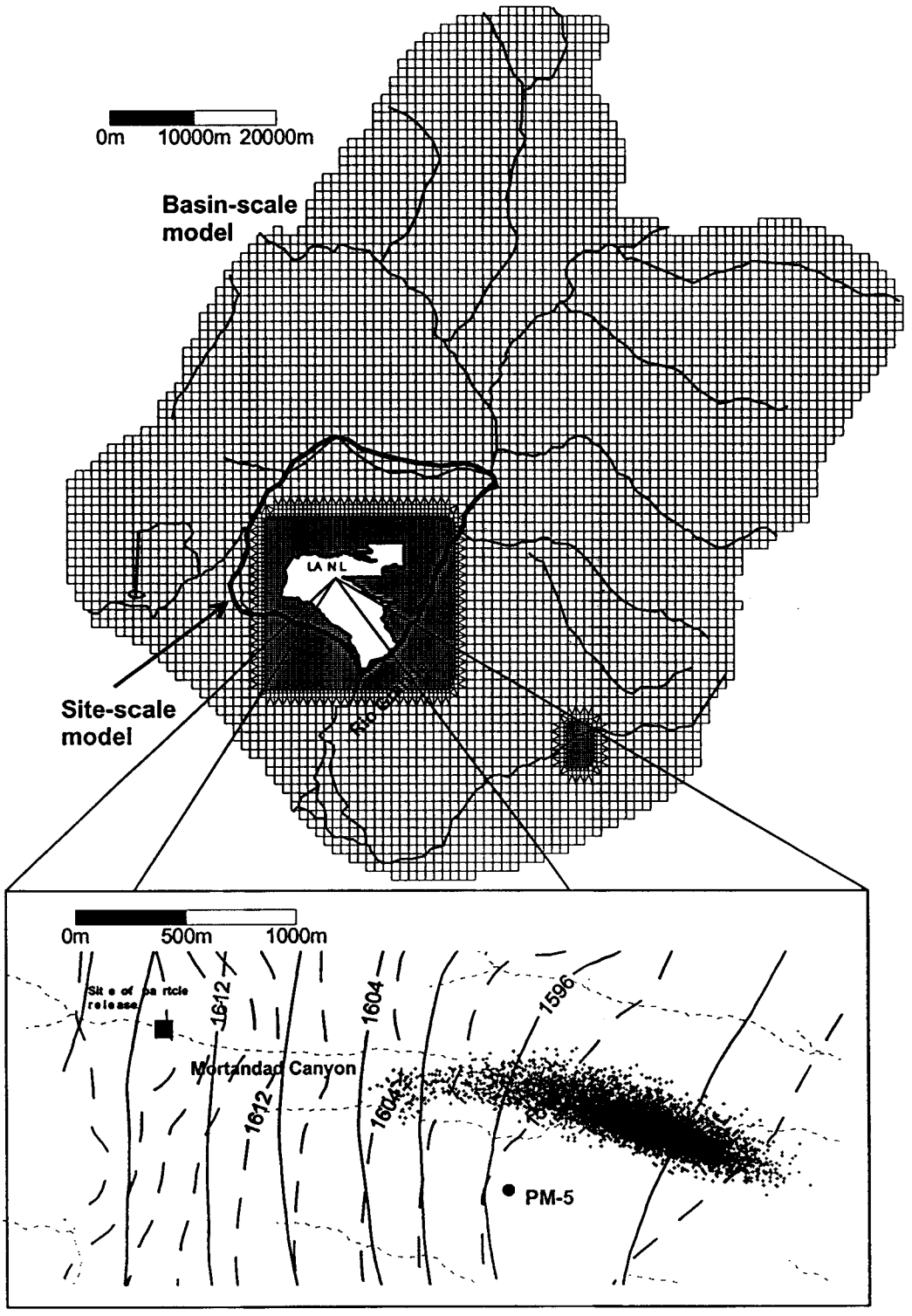
Zyvoloski, G.A., Robinson, B.A., Dash, Z.V., Trease, L.L., 1997. Summary of the models and methods for the FEHM application: A finite-element heat- and mass-transfer code, *Los Alamos National Laboratory Report*, LA-13306-MS, 1997.

Table 1. Selected parameter estimates derived from calibration and predictive analysis of the site scale inverse model about mass captured at PM-5.

Parameters		Calibration estimates		Predictive estimates	
		Estimates	Conf. limits	max	min
Recharge:					
• Z_{min} [m]		2214	362	2161	2229
• κ [–]		0.03	27.3	0.079	0.00076
Pemeabilities ($\log_{10}[m^2]$):					
• Cerros del Rio Basalts		-11.93	0.40	-11.95	-11.71
• Puye Fanglomerate	horizontal	-14.23	3.24	-14.24	-14.08
	vertical	-14.96	2.95	-14.93	-14.99
• Chaquehui Formation	horizontal	-13.31	0.29	-13.42	-13.27
	vertical	-15.61	1.44	-15.40	-15.84
• Santa Fe Group	horizontal	-13.08	0.17	-13.05	-13.09
	vertical	-15.36	0.24	-15.38	-15.39
Specific Storage $\log_{10}[m^{-1}]$ (S_s)		-3.70	0.43	-3.67	-3.73
Mass captured [%]		80	–	100	0
Mean travel time [years]		5600	–	1700	>10000

Figures

Figure 1. Planview of basin- and site-scale model domains and grids. Inset shows a portion of the site model where the capture-zone analysis is performed.


0m 10000m 20000m
Basin-scale model
LA NL
Site-scale model
0m 500m 1000m
Site of particle release
1612
1604
1596
Mortandad Canyon
PM-5