Assessing Recharge and Hydrostratigraphic Model Uncertainty in the Climax Mine Area of the Nevada Test Site

Ming Ye$^1$, Karl F. Pohlmann$^2$, Jenny B. Chapman$^2$, Greg M. Pohll$^3$, and Donald M. Reeves$^3$

$^1$Florida State University, mingye@scs.fsu.edu, Tallahassee, FL, USA
$^2$Desert Research Institute, Karl_Pohlmann@dri.edu, Jenny_Chapman@dri.edu, Las Vegas, NV, USA
$^3$Desert Research Institute, Greg_Pohll@dri.edu, Matt_Reeves@dri.edu, Reno, NV, USA

ABSTRACT

Hydrologic analyses are commonly based on a single conceptual model. Yet hydrologic environments are open and complex, rendering them prone to multiple interpretations and conceptualizations. Considering conceptual model uncertainty is a critical process for the assessment of hydrologic uncertainty. This study assesses recharge and geologic model uncertainty for the Climax Mine area of the Nevada Test Site, Nevada, where recharge processes and site geology are described by five recharge models and five hydrostratigraphic framework models (HFMs). Combining the recharge models and HFMs, a total of 25 conceptual models are formulated and incorporated into the Death Valley Regional Flow System modeling framework developed by the U.S. Geological Survey. The conceptual model uncertainty is quantified using a model averaging method. The method estimates model relative plausibility (measured by model probabilities) by incorporating expert judgment through expert elicitation and field observations of head and flow through inverse modeling. Parametric uncertainty of each model is assessed using Monte Carlo simulation, and each model's mean predictions and associated predictive uncertainty for hydraulic head and flow are averaged using the model probabilities as weights. The posterior (averaged) variance, incorporating both parametric and conceptual model uncertainty, is compared with those of individual models. It is shown that model averaging provides larger uncertainty than individual models, indicating that more uncertainty is incorporated and model predictions are more scientifically defensible.

INTRODUCTION

Hydrologic analyses are commonly based on a single conceptual model. Yet hydrologic environments are open and complex, rendering them prone to multiple interpretations and conceptualizations. This is true regardless of the quantity and quality of available hydrologic information and data. Focusing on only one conceptual model may lead to a Type I model error, which arises when one rejects (by omission) valid alternative models. It may also result in a Type II model error, which arises when one adopts (fails to reject) an invalid conceptual model (Neuman, 2003). Indeed, critiques of hydrologic analyses, and legal challenges to them, typically focus on the validity of the underlying conceptual model. If a proposed model is found to be severely deficient, hydrologic analysis based on the single model may damage professional credibility of the work, result in the loss of a legal contest, and/or lead to adverse environmental, economic and political impacts (Neuman, 2003).

Our study is focused on assessing conceptual uncertainty of models describing the recharge process and site geology at the Climax Mine area (northern Yucca Flat) of the Nevada Test Site, Nevada (Figure 1), where three underground nuclear tests were conducted between 1962 and 1966 in the Climax Mine granite stock. An objective of groundwater flow modeling in this area is to provide hydraulic heads and groundwater flows to the downgradient Yucca Flat area for corrective action investigation. The flow modeling results are also used to simulate radionuclide flux from the three tests in the Climax stock to Yucca Flat (Pohlmann et al., 2007). The flow modeling is conducted within the Death Valley Regional Flow System (DVRFS) modeling framework (Belcher et al., 2004), developed according to detailed characterization of hydrogeological conditions in southwestern Nevada and the Death Valley area of California (the boundary of the DVRFS is shown in Figure 1). Conceptual model uncertainty in describing the recharge process and site geology at the Climax Mine area is significant and cannot be ignored. This paper summarizes the uncertainty assessment using a model averaging method that formally incorporates expert judgments by expert elicitation and on-site observations of head and flow by inverse modeling. Details of the uncertainty assessment are described by Pohlmann et al. (2007).
ALTERNATIVE RECHARGE AND GEOLOGICAL MODELS

Recharge in the Climax Mine area (and the entire DVRFS) can be described by five models: (1) the modified Maxey-Eakin (MME) model, (2) two net infiltration models (NIM), one with (NIM1) and one without (NIM2) a runon-runoff component, and (3) two chloride mass balance (CMB) models, each with different zero-recharge masks, one for alluvium (CMB1) and one for both alluvium and elevation (CMB2). These five models are based on different methodologies that incorporate different levels of complexity. The MME model is empirical and the simplest one; the NIM models are the most complicated, considering various processes controlling net infiltration and potential recharge. Recharge estimates (m/d) of the five models are plotted in Figure 1, and show significant differences between the estimates. These differences are viewed as a result of conceptual model uncertainty, rather than parametric uncertainty, since they are caused by simplification and inadequacy/ambiguity in describing the recharge process and not by uncertainty in recharge measurements themselves. Since recharge is one of several major hydrological processes controlling groundwater flow paths and radionuclide travel times, it is critical to evaluate the recharge model uncertainty.

Five alternatives of the hydrostratigraphic framework model (HFM) are available to describe site geology in the Climax Mine area. The first model was developed by the U.S. Geological Survey and represents the configuration of hydrogeologic units in the entire DVRFS model; the second model was developed for the Yucca Flat-Climax Mine Corrective Action Unit as part of the Underground Test Area (UGTA) program. As illustrated in north-south cross-section in Figures 2a and 2b, the two models differ in both the number of hydrostratigraphic units and their subsurface configuration. Furthermore, two alternative models were developed based on the UGTA base model to address uncertainty regarding particular features of the hydrostratigraphy that might be important to groundwater flow. 

Figure 1. (a) Boundaries of the Death Valley Regional Flow System, the Nevada Test Site, the proposed Yucca Mountain nuclear waste repository, and recharge rate estimates (m/d) of the five recharge models. 

Figure 2. Two-dimensional illustration of difference between (a and b) the DVRFS and UGTA base models, (c and d) the UGTA base model and the CP Thrust alternative, and (e and f) the UGTA base model and the hydrologic barrier alternative.
and contaminant transport in Yucca Flat. One alternative (Model 3) incorporates a different interpretation of the configuration of hydrostratigraphic units with respect to the CP thrust fault, and the other alternative (Model 4) postulates a barrier to groundwater flow on the east side of the Climax stock. Differences between the two alternatives and the base model are depicted in Figures 2c – 2f. The fifth model merges the two UGTA alternative models to form a single alternative. As discussed below, the HFMs are more important than the recharge models for controlling groundwater flow, and their uncertainty dominates the total predictive uncertainty, relative to the recharge models.

ASSESSMENT OF CONCEPTUAL MODEL UNCERTAINTY USING MODEL AVERAGING

Model Averaging Method

The model averaging method has been advocated recently to assess conceptual model uncertainty (Neuman, 2003; Ye et al., 2004; Poeter and Anderson, 2005; Ye et al., 2008). If \( \Delta \) is the desired predicted quantity given a set of \( K \) alternative models, then its posterior distribution, given a dataset \( D \), is

\[
p(\Delta | D) = \sum_{k=1}^{K} p(\Delta | M_k, D) p(M_k | D)
\]

where \( p(\Delta | M_k, D) \) is the posterior distribution of \( \Delta \) under model \( M_k \) and \( p(M_k | D) \) is posterior probability of \( M_k \) evaluated using the Bayes’ rule

\[
p(M_k | D) = \left[ p(D | M_k) p(M_k) \right] / \sum_{l=1}^{K} p(D | M_l) p(M_l)
\]

where \( p(D | M_k) \) is likelihood of model \( M_k \) (a measure of consistency between model predictions and site observations \( D \)) and \( p(M_k) \) is prior probability of \( M_k \). For each model \( M_k \), \( E[\Delta | D, M_k] \) and \( Var[\Delta | D, M_k] \) are mean and variance of \( \Delta \) due to model parameter uncertainty. The posterior mean and variance are

\[
E[\Delta | D] = \sum_{k=1}^{K} E[\Delta | D, M_k] p(M_k | D)
\]

\[
Var[\Delta | D] = \sum_{k=1}^{K} Var[\Delta | D, M_k] p(M_k | D) + \sum_{k=1}^{K} (E[\Delta | D, M_k] - E[\Delta | D])^2 p(M_k | D).
\]

The evaluation of each term in Equations 1 – 4 and the model averaging process are discussed below. The posterior mean and variance are final products of this study.

Prior and Posterior Model Probabilities

Prior probabilities of the models are elicited from two expert panels (one for the recharge models and the other for the HFMs) based on the panelists’ beliefs regarding relative plausibility of each model considering their apparent consistency with available knowledge and data. Assuming that recharge models and HFMs are independent, joint probability of each recharge-HFM combination is computed by multiplying their prior probabilities. The prior model probabilities of the 25 models are plotted in Figure 3, in which R and G represent recharge models and HFMs, respectively. For example, R1G1 is the combination of the MME recharge model and DVRFS HFM. The NIM recharge model with runon-runoff (R2) and the UGTA base model (G2) have the highest prior probabilities among the recharge models and HFMs, respectively. Therefore, model R2G2 has the highest prior probability of 8.4%, about twice as large as the average prior probability of 4%, recalling that summation of the prior probabilities is 100%.

However, Figure 3 shows that the prior model probabilities are more-or-less uniform, indicating that there is no justification to select one model and discard others based on expert judgment. This reflects the inherent uncertainty in the recharge and geological models, due to system complexity and sparse data characterizing the recharge process and site geology.

After incorporating the recharge and geological models into the regional DVRFS model, the regional flow models are calibrated to evaluate the alternative models using on-site observations of head and flow. Since observations are sparse (only 59) in the Climax Mine area, using the regional DVRFS model enables us to better constrain the flow system by honoring more numerous (total 4963) regional observations. A maximum of 32 hydraulic parameters most significant to flow and transport in the Climax
Mine area are selected for calibration; parameters located far and down-gradient from the Climax Mine area are not calibrated. Examination of the sum of squared weighted residuals (SSWR) for each calibrated model shows that the overall model fit is improved compared to the original calibration of the DVRFS model by Belcher et al. (2004). This is not surprising, because the calibration in this study can be viewed as further calibration of the DVRFS model. The model likelihood function \( p(D|M_k) \) in Equation 2 is calculated as the inverse of the SSWR of the 59 observations at the Climax Mine area, i.e., \( p(D|M_k) = 1/\text{SSWR} \). The smaller SSWR values indicate the larger likelihood of the model that the observations \( D \) can be simulated by the model. This calculation is different from the conventional way of calculating the model likelihood function using information criteria (e.g., AIC, AICc, BIC, and KIC) (Neuman, 2003; Poeter and Anderson, 2005; Ye et al., 2008). The reason for using the SSWR not the information criteria is that the alternative models at the Climax Mine area are evaluated by calibrating the regional DVRFS model. While using the SSWR does not penalize complex models as the information criteria, it should not be a problem, since values of the number of model parameters are similar for the 25 models (32 and 30 parameters).

Figure 3 plots two sets of posterior model probabilities. One set uses equal prior model probabilities, and the posterior model probabilities are determined solely by the model fit measured by SSWR; the other set uses the unequal prior model probability obtained from expert elicitation. Comparing the two sets of posterior model probabilities shows sensitivity of the posterior to prior model probabilities. The G3R2 model (combination of the NIM1 model (with runon-runoff) with the CP thrust alternative HFM) has the highest posterior probability among the 25 models. This was followed by the NIM1 model combined with the UGTA base model (G2R2) and the CMB2 model with the alluvium and elevation masks combined with the CP thrust model (G3R5). For the HFMs, models associated with the UGTA base and CP thrust alternatives have larger posterior probability; for the recharge models, the NIM1 model with the runon-runoff model generally provides the largest posterior probability, consistent with the prior probability. However, the MME model always provided the smallest posterior probability, different from the prior probability obtained from the expert elicitation. In general, the posterior model probabilities of various HFMs for each recharge model exhibited larger variation than the posterior probabilities of the various recharge models for each HFM. This indicates that simulation of groundwater flow in northern Yucca Flat is more sensitive to hydrostratigraphic conceptualization, as compared to the recharge model.

**Predictive Uncertainty Considering Parametric and Conceptual Method Uncertainty**

The posterior probabilities with unequal prior model probabilities are used to weight each model combination in the averaged flow model such that model combinations with the best calibration results carried the most weight in subsequent simulations. Parametric uncertainty arising from parameter estimation is assessed for each model by generating random values from a covariance matrix produced during model calibration. This form of parametric uncertainty results from measurement error of the calibration data, not from variability of data values measured in the field as is commonly done in the
assessment of parametric uncertainty. Sparse parameter measurements at the Climax Mine area, and in the larger DVRFS, prevent adequate and consistent characterization of all parametric distributions. Figure 4 plots a two-dimensional plane view of mean and standard deviation of simulated head for models G3R2 and G2R2, the two most plausible models with posterior model probabilities of 13.66% and 13.05%, respectively. The figure also plots the posterior mean and standard deviation obtained by the model averaging. The spatial pattern of the posterior mean head is similar to that of the single model, with high heads in the north and low heads in the south owing in part to the displacement in the area of the regional carbonate aquifer (LCA) by the upper clastic confining unit (UCCU) shown in Figure 2. A large gradient is predicted at the lower-left part of the area (between the green and blue colors). The posterior mean heads are relatively higher head at the northwest and southeast corners of the domain due to high heads predicted by the G3R5 and G2R5 models (Pohlmann), partly as a result of high recharge rates estimated by R5 (Figure 1). For the standard deviation of head, areas of high standard deviation appear at the northwest and southeast corners of the domain, where high heads are simulated. Comparing Figure 4 a-2 with Figures 4 b-2 and c-2 shows that the posterior standard deviation obtained from model averaging is significantly larger than the standard deviations simulated by individual models. This large posterior standard deviation is the result of between-model variance (the second term at right hand side of Equation 4) (Pohlmann et al., 2007). This indicates that focusing solely on parametric uncertainty, without considering conceptual model uncertainty, is likely to lead to an underestimation of predictive uncertainty. As shown in Pohlmann et al. (2007), the analysis of modeling results shows that uncertainty of HFMs dominates that of recharge models and that the calculation of posterior model probabilities is strongly influenced by several observations in the northern part of the domain where head observations are sparse.

REFERENCES


