Hierarchical Multimodel Saltwater Intrusion Remediation and Sampling Designs: A BMA Tree Approach

Basic Information

<table>
<thead>
<tr>
<th>Title:</th>
<th>Hierarchical Multimodel Saltwater Intrusion Remediation and Sampling Designs: A BMA Tree Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Number:</td>
<td>2010LA76G</td>
</tr>
<tr>
<td>Start Date:</td>
<td>9/1/2010</td>
</tr>
<tr>
<td>End Date:</td>
<td>8/31/2013</td>
</tr>
<tr>
<td>Funding Source:</td>
<td>104G</td>
</tr>
<tr>
<td>Congressional District:</td>
<td>Louisiana</td>
</tr>
<tr>
<td>Research Category:</td>
<td>Ground-water Flow and Transport</td>
</tr>
<tr>
<td>Focus Category:</td>
<td>Groundwater, Management and Planning, Methods</td>
</tr>
<tr>
<td>Descriptors:</td>
<td>None</td>
</tr>
<tr>
<td>Principal Investigators:</td>
<td>Frank Tsai, Jeff Hanor</td>
</tr>
</tbody>
</table>

Publications

10. Tsai, F.T.-C. (2010), Scavenger Wells Stop Saltwater Intrusion in Baton Rouge, 2010 Louisiana
Water Quality Technology Conference, Alexandria and Baton Rouge, Louisiana, December 14-15,
2010. (invited)
Wells in the Baton Rouge Area, Louisiana Capital Area Ground Water Conservation Commission,
September 14, 2010. (invited)
SYNOPSIS

Title: Hierarchical Multimodel Saltwater Intrusion Remediation and Sampling Designs: A BMA Tree Approach  
Project Number: G10AP00136  
Start Date: 9/1/2010  
End Date: 8/31/2013  
Funding Source: 104G  
Research Category: Ground-water Flow and Transport  
Focus Categories: GW, M&P, MET  
Descriptors: Remediation Design, Sampling Design, Saltwater Intrusion, Optimization, Uncertainty  
Primary PI: Frank T.-C. Tsai  
Other PI: Jeffrey S. Hanor

Problem and Research Objectives

The water withdrawal in Baton Rouge, Louisiana is approximately 629,000 m³ per day (166 million gallons per day) out of which 88% is ground water and the rest is surface water (Sargent, 2007). Due to excessive ground water pumping, saltwater is intruding from the saline aquifers in the south part of the Baton Rouge fault. Thus, in the absence of any remediation measure, some of public supply water wells in East Baton Rouge Parish are under the threat of being abandoned in the near future. The project objective is to analyze the uncertainty of the ground water numerical models, which are used for the management and remediation of the ground water resources. The first step in this project is to develop conceptual models that capture the complexity and heterogeneity of the subsurface geology.

Subsurface models have a distinctive position since subsurface data are scarce due to economic reasons. For example, this study area is approximately 1005 km². To characterize the subsurface 55 resistivity and self-potential logs are used. Assuming that the radius of influence of the long normal resistivity is 5 m and given a total number of 55 wells, thus the presence or absence of sand lenses is sampled over an area of approximately 0.00043% of the total area. Since this could be a common scenario for many subsurface problems, significant body of the subsurface literature focuses on developing stochastic data analysis techniques that would improve the utility of the scarce subsurface data and thus improve the model prediction and provide an analysis of model uncertainty.

Thus, due to limited amount of data and since model uncertainty always exists, multiple models are usually developed. Model selection, model elimination, model reduction, and model discrimination are commonly used to select the best model. It is clear that modeling uncertainty is always underestimated if only the best model is used. One would ask why only the best model is used afterwards when so many efforts have been devoted to calibrating many models. This certainly wastes valuable resources and important information from other good models. Hierarchical Bayesian model averaging (HBMA) best utilize all possible models for model prediction and application under Bayesian statistical framework. HBMA presents several advantages over model selection: (1) Information from all possible models is used based on their
model importance (model weights). Calibration efforts are not wasted. (2) The model importance is based on the evidence of data, which avoids over-confidence in the best model that does not have a dominant model weight. And (3) model structure uncertainty is increased and is better presented than that by using a single model. Moreover, HBMA is able to distinguish model uncertainty arising from individual models and between models. HBMA is able to identify unfavorable models even though they may present small prediction uncertainty.

In this study, HBMA is used to estimate the sand-clay distribution in the Baton Rouge aquifer system. Indicator geostatistical techniques are used to analyze electrical resistivity logs and reconstruct the subsurface accordingly. The HBMA is applied to analyze the conceptual model structure uncertainty arising from the different sand-clay line cutoff values for the resistivity logs and the different sand-clay cutoff probabilities for the interpolated values.

**METHODOLOGY**

**Indicator Kriging**

Given the volumetric domain \( D \subset \mathbb{R}^n \), the indicator function \( \{ I(\mathbf{x}, \nu) : \mathbf{x} \in D \} \) is a random function. The indicator random variable \( \nu \) describes the spatial extension of a categorical variable \( \mathcal{C} \), which is the sand-clay distribution in aquifers under different sand-clay line cutoff as determined from the electrical resistivity logs. The random function of the indicator random variable of class \( \mathcal{C} \) is defined as:

\[
I(\mathbf{x}, \nu) = \begin{cases} 
1 & \nu \in \mathcal{C}, \quad \nu(\mathbf{x}) > \alpha \\
0 & \nu \notin \mathcal{C}, \quad \nu(\mathbf{x}) < \alpha 
\end{cases}
\]

The indicator function \( I(\mathbf{x}, \nu) \) is a random function of two variables in which \( \nu \) is an outcome of random variable at location \( \mathbf{x} \) in which the one and zero indicates the presence of sand and clay, respectively.

The indicator variogram has the same definition as the normal variogram except that the real random function is replaced by the indicator random function \( I(\mathbf{x}) \) as follows

\[
\gamma_{I,I}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [I(x_i) - I(x_i + h)]^2
\]

where \( N(h) \) is the number of pairs within the lag interval \( h \). In this case study, 55 observation boreholes are used in which 42 are located north of the fault Baton Rouge fault and 13 wells are located at the south. The main source of the sample data, which are used to generate the IK variograms, is from the electrical resistivity logs that are provided by the Baton Rouge Water Company. For each foot and for every resistivity log location, the resistivity values indicate either sand or clay depending on the sand-clay line cutoff as determined from the clay line in the resistivity curves. Another source of data is from the study of Wendeborn and Hanor (2008), in which they analyzed spontaneous potential (SP) curves to identify the sand-clay distribution along the Baton Rouge fault. The interpretation of these logs in terms of sand-clay sequences are amended to the main data to provide more sampling locations. All observation points are
amended together through linear interpolation over each foot. The number of samples in each observation point varies from 800 to 3000 depending on the depth of each borehole. Thus, over a depth \( z \) with an increment of 1 foot, an experimental variogram is generated. A pseudo 3D horizontal experimental variogram is obtained by averaging all the 2D experimental variograms for all depths.

The exponential model fits well with the geological process understudy since it is an indicative of a sharp transition occurring between blocks of different values (Rubin, 2003). The exponential model is formulated as:

\[
\gamma_{\text{Exp}}(h) = X_1 + X_2 \cdot \left(1 - \exp \left(-3 \cdot \frac{h}{X_3}\right)\right)
\]

where the correlation parameters are the nugget effect \( X_1 \) and the sill \( X_2 \), which the variance approaches at an effective range \( X_3 \). The theoretical variogram is fitted to the experimental variogram automatically through using the pattern search method of Hooke and Jeeves (1961), which performs a direct directional search for the values of \( X_1, X_2 \) and \( X_3 \), which would minimize the weighted squared root difference between the experimental and the theoretical variograms.

Under the basic assumption that the sample domain is stationary, ergodic and sufficient to reliable reproduce the statistics, the obtained theoretical variogram is used for the indicator kriging interpolation as a method for constructing the subsurface stratigraphy. The aim of kriging is to estimate the value of a random variable at unsampled points. Over larger defined grid size, which is 500 m \( \times \) 500 m in this case, kriging uses weighted average of the neighboring sample data points to estimate the value in each grid using the following equation:

\[ A\hat{\lambda} = b \]

such as

\[
\begin{bmatrix}
\gamma(x_1, x_1) & \gamma(x_1, x_2) & \cdots & \gamma(x_1, x_N) & 1 \\
\gamma(x_2, x_1) & \gamma(x_2, x_2) & \cdots & \gamma(x_2, x_N) & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma(x_N, x_1) & \gamma(x_N, x_2) & \cdots & \gamma(x_N, x_N) & 1 \\
1 & 1 & \cdots & 1 & 0
\end{bmatrix}
\begin{bmatrix}
\hat{\lambda}_1 \\
\hat{\lambda}_2 \\
\vdots \\
\hat{\lambda}_N \\
L
\end{bmatrix}
= 
\begin{bmatrix}
\gamma(x_1, x_0) \\
\gamma(x_2, x_0) \\
\vdots \\
\gamma(x_N, x_0) \\
1
\end{bmatrix}
\]

in which \( \gamma(x_i, x_j) \) is the variogram of \( v \) between the data points \( x_i \) and \( x_j \), and the \( \gamma(x_i, x_0) \) is the variogram between the data point \( x_i \) and the target point \( x_0 \). To guarantee that the estimates are unbiased, the sum of the weights \( \hat{\lambda}_i \) is one. The unbiased constrained is imbedded to the minimization problem through the use of the Lagrange multiplier \( L \). By multiplying the inverse of the matrix \( A \) with the vector \( b \), we obtain the weight for each data point. The last step is to calculate the expected value and the kriging variance by solving the following equations.
\[ v^*(x_0) = \sum_{i=1}^{N} \lambda_i I(x_i) \quad \text{and} \quad \sigma_{ix}^2 = b^T \lambda \]

The estimated values could be viewed as the conditional probability that a value is less than a certain sand-clay cutoff.

**Hierarchical Bayesian Model Averaging (HBMA)**

To cope with sources of uncertainty in ground water conceptual models a hierarchical Bayesian model averaging is adopted. An early study by Elshall and Tsai (2011) conducted a sensitivity analysis on different methods and parameters. First, the use by of 2D variogram model for each layer is compared to a pseudo 3D variogram model, which is a weighted average of all the 2D variograms. Also, different sand-clay line cutoff and sand-clay cutoff probabilities are investigated. The results are much more sensitive to the cutoff of the sand-clay line and sand-clay cutoff probabilities in comparison with the selection of different variogram models. Four sand-clay line cutoffs of 10, 11, 12 and 13 ohm-m are considered. For each of these four models a sand-clay cutoff of probabilities 0.4, 0.5 and 0.6 is considered.

The key issue of HBMA is the determination of the posterior model probability. Given a number of conceptual models \( M = \{ M^{(p)} ; p = 1 \ldots 4 \} \) for determining the sand-clay distribution \( v \) over the model domain with different cutoff probabilities \( \Theta^{(p)} = \{ \theta_{q}^{(p)} ; q = 1 \ldots 3 \} \), the posterior probability of predicted sand-clay distribution is obtained by using the twelve conceptual models through HBMA given a set of data \( D \), which is the boreholes data from the USGS indicating the sand lenses at different depths at 65 different locations. An HMBA method after Li and Tsai (2009) is adopted in this study. In HBMA in which multi-parameter uncertainty is considered, the posterior probability for the given data \( D \), sand-clay line cutoff \( M \) and sand-clay cutoff probability \( \Theta^{(p)} \) is given as

\[
\Pr(v|D) = E_M \left[ E_{\Theta} \left[ \Pr(v|M^{(p)}, \Theta_{q}^{(p)}, D) \right] \right] = \sum_p \sum_q \Pr(v|M^{(p)}, \Theta_{q}^{(p)}, D) \Pr(\Theta_{q}^{(p)}|M^{(p)}, D) \Pr(M^{(p)}|D)
\]

where \( \Pr(v|M^{(p)}, \Theta_{q}^{(p)}, D) \) is the posterior probability of the sand-clay distribution for a given data set \( D \) and sand-clay line cutoff \( M^{(p)} \) and sand-clay probability cutoff \( \Theta_{q}^{(p)} \). \( E_{\Theta} \) and \( E_{M} \) are the expectations. The joint posterior probability (model weight) according to the Bayes’ rule is:

\[
\Pr(\Theta_{q}^{(p)}, M^{(p)}|D) = \Pr(\Theta_{q}^{(p)}|M^{(p)}, D) \Pr(M^{(p)}|D) = \frac{\Pr(D|M^{(p)}, \Theta_{q}^{(p)})}{\sum_p \sum_q \Pr(D|M^{(p)}, \Theta_{q}^{(p)})}
\]

where \( \Pr(\Theta_{q}^{(p)}|M^{(p)}, D) \) is the marginal model likelihood function for a given sand-clay line cutoff \( M^{(p)} \), and sand-clay cutoff probability \( \Theta_{q}^{(p)} \). The total weight is \( \Pr(\Theta_{q}^{(p)}, M^{(p)}|D) = 1. \)
The marginal likelihood function \( \Pr(D | M^{(p)}, \theta_q^{(p)}) \) is commonly approximated using the Laplace approximation with the Bayesian information criterion (BIC)

\[
BIC_q^{(p)} = -2 \ln \Pr(D | M^{(p)}, \theta_q^{(p)}, \hat{\theta}_q^{(p)}) + m_q^{(p)} \ln n
\]

where \( \hat{\theta}_q^{(p)} \) are the maximum likelihood estimated parameters in a model structure given \( \theta_q^{(p)} \) and \( M^{(p)} \), \( m_q^{(p)} \) is the number of the parameters \( \hat{\theta}_q^{(p)} \) and \( n \) is the number of data \( D \). \( \Pr(D | M^{(p)}, \theta_q^{(p)}, \hat{\theta}_q^{(p)}) \) is the likelihood function value. The parameters \( \hat{\theta}_q^{(p)} \) are the variogram model parameters in this study. One can obtain the means of the sand-clay distribution using the law of total expectation:

\[
E(v | D) = E_M \left[ E_\theta \left[ E(v | M^{(p)}, \theta_q^{(p)}, D) \right] \right]
= \sum_p \sum_q E(v | M^{(p)}, \theta_q^{(p)}, D) \Pr(D | M^{(p)}, D) \cdot \Pr(M^{(p)} | D)
\]

with variance matrix of the predicated sand-clay distribution as:

\[
Var(v | D) = E_M E_\theta \left[ Var \left[ v | M^{(p)}, \theta_q^{(p)}, D \right] \right]
+ E_M Var_\theta \left[ Var \left[ v | M^{(p)}, \theta_q^{(p)}, D \right] \right]
+ Var_M E_\theta \left[ Var \left[ v | M^{(p)}, \theta_q^{(p)}, D \right] \right]
\]

The first term of the right side of the above equation is the within-model variance, which relates to the uncertainty of the predicated sand-clay distribution using combination of different sand-clay line cutoff \( M^{(p)} \) and sand-clay cutoff probability \( \theta_q^{(p)} \). The second term is the between-model variance, which relates to the uncertainty using different sand-clay probability cutoffs. The third term is between-model variance, which relates to the uncertainty of using different sand-clay line cutoffs.

![Figure 1. BMA tree of model weights in HMBA. The model weights in parentheses are \( \Pr(\theta_q^{(p)}, M^{(p)} | D) \). The model weights without parentheses are \( \Pr(\theta_q^{(p)} | M^{(p)}, D) \).](image)
Principal Findings and Significance
This section presents the results of studying the sand-clay distribution based on different sand-clay lines of 10, 11, 12 and 13 ohm-m and sand-clay cutoff probability of 0.4, 0.5 and 0.6. Model weights are calculated according to the BIC, which is more skewed to the best models as shown in Figure 1.

The models for different probability values are the third level models, which are 12 models corresponding to the three cutoff probabilities for each for the four sand-clay line cutoffs. When the third level models are averaged with their weights, they form the second level models, which are 4 models corresponding to the four different sand-clay lines as shown in Figure 2. Although the weights are calculated based on the results of 190 layers from a depth of 1460 to 1650 feet below msl, the results are only shown for a 2D plan at a depth of 1551 feet below msl.

![Figure 2. Sand-clay distribution for the second level models for different sand-clay line cutoffs.](image)

The BMA prediction variance includes the within-model variance and between-model variance with the total variance being the summation of these two components. Although the calculated sand-clay distribution can be significantly different when using different sand-clay lines and cutoff probabilities, yet the within-model variance is larger than the between-model variance.
Figure 3. Within-model variance (WMV) and between-model variance (BMV) for different sand-clay line cutoffs.
Similarly, one can obtain the first level model, which is a weighted average of the second level models as shown in Figure 4. For the between-model variance being smaller than the within-model variance does not suggest the unimportance of the sand-clay line. Rather it simply suggests that good models produce similar predictions close to the expectation of the BMA predictions. Bad models have little influence on the sand-clay distribution because their model weight is very small and thus their predictions are further apart from the BMA prediction.

![Figure 4](image)

**Figure 4.** Sand-clay distribution, within-model variance, between-model variance and total variance for the first level model.

**CONCLUSION**

This study focuses on model structure uncertainty. Using a single model will ignore the uncertainty arising from different model parameters. The HBMA applies a Bayesian statistical approach to quantify the overall sand-clay distribution uncertainty. In general, HBMA is successfully applied to study the uncertainty of stratigraphic models. Accordingly, the method can be readily extended to include other sources of structural uncertainty such as the morphology and dipping angle of Baton Rouge fault and borehole elevation uncertainty.

**REFERENCES**


PUBLICATION

1. Articles in Refereed Scientific Journals
N/A

2. Book Chapters
N/A

3. Theses and Dissertations
N/A

4. Water Resources Research Institute Reports
- Frank Tsai, 2011, Multimodel Uncertainty Analysis for Chance-Constrained Saltwater Intrusion Management, Louisiana State University, Baton Rouge, Louisiana, 10 pages. (USGS 104B)

5. Conference Proceedings

6. Other Publications (Presentations)

7. Student Support
• Nima Chitsazan, PhD student
• Elizabeth L. Chamberlain, MS student