Li 2008b; Tsai and Li 2008c). The following section discusses the use of the BMA to integrate multiple parameterization methods.

7.6.1 Statistical Inference through Bayesian Model Averaging (BMA)

We consider a set of parameterization methods $\Theta = \left\{ M^{(p)}; p = 1, 2, \cdots \right\}$ in order to estimate the hydraulic conductivity distribution over the same region. Given available data ($\mathbf{D}$) and multiple parameterization methods (Θ), the conditional probability of a predicted quantity of interest (Δ) can be obtained through Bayesian model averaging (BMA), which is based on the law of total probability (Leamer 1978):

$$\Pr\left(\Delta \mid \mathbf{D}\right) = \sum\nolimits_p \Pr\left(\Delta \mid \mathbf{D}, M^{(p)}\right) \Pr\left(M^{(p)} \mid \mathbf{D}\right) \tag{7.47}$$

where $\Pr\left(\Delta \mid \mathbf{D}\right)$ is the conditional probability of the predicted quantity given data $\mathbf{D}$; and $\Pr\left(\Delta \mid \mathbf{D}, M^{(p)}\right)$ is the conditional probability of the predicted quantity given the data $\mathbf{D}$ and a parameterization method $M^{(p)}$. In addition, $\Pr\left(M^{(p)} \mid \mathbf{D}\right)$ is the posterior probability of a parameterization method given data $\mathbf{D}$, which represents posterior model weights. The models in the BMA refer to the parameterization methods.

According to Bayesian decision theory (Berger 1985), the posterior model probability of a parameterization method is

$$\Pr\left(M^{(p)} \mid \mathbf{D}\right) = \frac{\Pr\left(\mathbf{D} \mid M^{(p)}\right) \Pr\left(M^{(p)}\right)}{\sum\nolimits_j \Pr\left(\mathbf{D} \mid M^{(j)}\right) \Pr\left(M^{(j)}\right)} \tag{7.48}$$

where $\sum\nolimits_p \Pr\left(M^{(p)} \mid \mathbf{D}\right) = 1$. The prior probabilities of parameterization methods also represent prior model weights and $\sum\nolimits_p \Pr\left(M^{(p)}\right) = 1$.

The prior model probability is a subjective value and is based on the analysts' prior information and their philosophical beliefs. Choosing proper prior probabilities is a challenging practical issue, especially in the absence of substantial prior knowledge, and usually engenders philosophical debates. It could be argued that in one hand the posterior model probability is rather sensitive to the specification of the prior. On the other hand, the prior model probability should not dominate the likelihood as supported by data. At any rate, it is imperative to avoid using improper prior that substantially affects the analysis. A Kullback-Leibler (K-L) prior was suggested by Burnham and Anderson (2004) in the prior BIC weights, which lead to the AIC weights. Ye et al. (2005) suggested the identification of prior model probabilities through the entropy maximization method. If there is no informational support, it is

reasonable to have equal prior probabilities as a neutral choice (Hoeting et al. 1999; Wasserman 2000).

Considering equal prior probabilities and the KIC, we can approximate the posterior model probability as

$$\Pr\left(M^{(p)} \mid \mathbf{D}\right) \approx \frac{\exp\left(-\frac{1}{2}\mathrm{KIC}_D^{(p)}\right)}{\sum_j \exp\left(-\frac{1}{2}\mathrm{KIC}_D^{(j)}\right)} \,. \tag{7.49}$$

Equation (7.49) can be rewritten by considering the KIC difference to the minimum KIC value:

$$\Pr\left(M^{(p)} \mid \mathbf{D}\right) \approx \frac{\exp\left(-\frac{1}{2}\Delta\mathrm{KIC}_D^{(p)}\right)}{\sum_j \exp\left(-\frac{1}{2}\Delta\mathrm{KIC}_D^{(j)}\right)} \tag{7.50}$$

where $\Delta\mathrm{KIC}_D^{(p)} = \mathrm{KIC}_D^{(p)} - \mathrm{KIC}_{D,\min}^{(p)}$. Again, the KIC can be replaced by other informational criteria as needed.

According to the law of total expectation, the expectation of the predicted quantity is

$$\mathrm{E}\left[\Delta \mid \mathbf{D}\right] = \sum_p \mathrm{E}\left[\Delta \mid \mathbf{D}, \theta^{(p)}\right] \Pr\left(\theta^{(p)} \mid \mathbf{D}\right). \tag{7.51}$$

Similarly, the law of total (co)variance obtains the covariance matrix of the predicted quantity as

$$\begin{aligned}\mathrm{Cov}\left[\Delta \mid \mathbf{D}\right] = &\sum_p \mathrm{Cov}\left[\Delta \mid \mathbf{D}, M^{(p)}\right] \Pr\left(M^{(p)} \mid \mathbf{D}\right) \\ &+ \sum_p \left(\mathrm{E}\left[\Delta \mid \mathbf{D}, M^{(p)}\right] - \mathrm{E}\left[\Delta \mid \mathbf{D}\right]\right)\left(\mathrm{E}\left[\Delta \mid \mathbf{D}, M^{(p)}\right] - \mathrm{E}\left[\Delta \mid \mathbf{D}\right]\right)^T \Pr\left(M^{(p)} \mid \mathbf{D}\right)\end{aligned} \tag{7.52}$$

The first term on the right side of Equation (7.52) represents the covariance for individual parameterization methods (within covariance). The second term represents the covariance between different parameterization methods (between covariance).

We note that, in practice, the true groundwater model is impossible to obtain. All the results from model selection and model averaging are based on models that are imperfect. A 100% model weight in the BMA does not imply a 100% correct model, but merely indicates the single best model among candidate models. To be sure, the ideal situation is to obtain an exhaustive set of models. However, this is very expensive in groundwater modeling. Using a limited number of models is inevitable, but with an understanding that model weights calculated by posterior model

probabilities are the relative model weights among the selected models, not the true model weights over the model space.

7.6.2 Multigeneralized Parameterization (MultiGP)

Applications of the BMA depend on the data $\mathbf{D}$ and the predicted quantity Δ to be specified. In this section, we consider multiple generalized parameterization (GP) methods to estimate hydraulic conductivity. The data is the sample data $\mathbf{D}=\mathbf{p}^{data}$ and the predicted quantity is $\Delta=\mathbf{p}$.

In the multiGP method, we consider a set of zonal distributions $Z=\{\Omega^{(1)},\Omega^{(2)},\cdots\}$ and a set of interpolation methods $\Phi=\{\phi^{(1)},\phi^{(2)},\cdots\}$ to parameterize the same hydraulic conductivity field (Ω) using the same hydraulic conductivity data. Combinations of these zonal distributions and interpolation methods pose a multiGP scheme that involves many GP methods $\Theta=\mathrm{Z}\times\Phi=\{M^{(p)};p=1,2,\cdots\}$. Each GP method has its own weighting coefficients $\boldsymbol{\beta}^{(p)}=\left\{\beta_j^{(p)},j=1,2,\cdots,m\right\}$. The weighting coefficients $\boldsymbol{\beta}^{(p)}$ are the model parameters embedded in the posterior probability $\Pr\left(M^{(p)}\mid\mathbf{D}\right)$. The mean and covariance of predicted hydraulic conductivity using a GP method are

$$\mathrm{E}\left[\Delta\mid\mathbf{D},M^{(p)}\right]=\mathrm{E}\left[\mathbf{p}\mid\mathbf{p}^{data},M^{(p)}\right]=\mathbf{p}_{\mathrm{GP}}^{(p)} \tag{7.53}$$

$$\mathrm{Cov}\left[\Delta\mid\mathbf{D},M^{(p)}\right]=\mathrm{Cov}\left[\mathbf{p}\mid\mathbf{p}^{data},M^{(p)}\right]=\mathrm{C}_{\mathrm{GP}}^{(p)}. \tag{7.54}$$

The expectation of the BMA hydraulic conductivity is

$$\mathrm{E}\left[\mathbf{p}\mid\mathbf{p}^{data}\right]=\overline{\mathbf{p}}_{GP}=\frac{\sum_p\mathbf{p}_{GP}^{(p)}\exp\left(-\frac{1}{2}\Delta\mathrm{KIC}_D^{(p)}\right)}{\sum_j\exp\left(-\frac{1}{2}\Delta\mathrm{KIC}_D^{(j)}\right)}. \tag{7.55}$$

The covariance of the BMA hydraulic conductivity is

$$\mathrm{Cov}\left[\mathbf{p}\mid\mathbf{p}^{data}\right]=\frac{\sum_p\left[\mathrm{Cov}_{GP}^{(p)}+\left(\mathbf{p}_{GP}^{(p)}-\overline{\mathbf{p}}_{GP}\right)\left(\mathbf{p}_{GP}^{(p)}-\overline{\mathbf{p}}_{GP}\right)^T\right]\exp\left(-\frac{1}{2}\Delta\mathrm{KIC}_D^{(p)}\right)}{\sum_j\exp\left(-\frac{1}{2}\Delta\mathrm{KIC}_D^{(j)}\right)}. \tag{7.56}$$

7.6.3 BMA Groundwater Inverse Modeling

In BMA groundwater inverse modeling, the optimal weighting coefficients $\hat{\boldsymbol{\beta}}^{(p)}$ of individual GP methods need to be estimated using observation data. In this case, the data $\mathbf{D}=\mathbf{h}^{data}$ would be the observed groundwater heads and the predicted quantity $\Delta=\mathbf{h}$ would be the predicted groundwater heads at the observation space. One

cannot use the maximum likelihood of the posterior probability $\Pr(\Delta \mid \mathbf{D})$ in Equation (7.47) to estimate $\hat{\boldsymbol{\beta}}^{(p)}$ values because optimization will force 100% posterior probability to the best GP method that has the minimum KIC (or BIC) value and zero posterior probability for rest of the GP methods. Instead, one can implement maximum likelihood Bayesian model averaging (MLBMA) (Neuman 2003; Ye et al. 2004) to obtain optimal $\hat{\boldsymbol{\beta}}^{(p)}$ in $\Pr(\Delta = \mathbf{h} \mid \mathbf{D} = \mathbf{h}^{data}, M^{(p)})$ or use the inverse methods presented in Section 7.4. for individual GP methods. Then, $\Pr(\Delta \mid \mathbf{D}, M^{(p)})$ can be approximated by $\Pr(\Delta \mid \mathbf{D}, M^{(p)}, \hat{\boldsymbol{\beta}}^{(p)})$ to avoid calculation burden (Draper 1995) for BMA predictions.

Another approach that considers the maximum likelihood estimation in $\Pr(\Delta \mid \mathbf{D}, M^{(p)})$ is the following integration form:

$$\Pr(\Delta \mid \mathbf{D}, M^{(p)}) = \int_{\boldsymbol{\beta}} \Pr(\Delta \mid \mathbf{D}, M^{(p)}, \boldsymbol{\beta}^{(p)}) \Pr(\boldsymbol{\beta}^{(p)} \mid \mathbf{D}, M^{(p)}) d\boldsymbol{\beta}^{(p)} . \tag{7.57}$$

Because Δ is the predicted quantity at the observation space, $\Pr(\Delta \mid \mathbf{D}, M^{(p)})$ can be seen as a likelihood function. Using the Laplace approximation, one can obtain

$$\Pr(\Delta \mid \mathbf{D}, M^{(p)}) \approx \exp(-\frac{1}{2} \mathrm{KIC}_{\Delta}^{(p)}) \tag{7.58}$$

where $\mathrm{KIC}_{\Delta}^{(p)}$ is the KIC with respect to the predicted quantity Δ:

$$\mathrm{KIC}_{\Delta}^{(p)} = -2 g_{\Delta}^{(p)} - m^{(p)} \ln 2\pi + \ln \left| \mathbf{F}_{\Delta}^{(p)} \right| . \tag{7.59}$$

The Fisher information matrix is $\left[\mathbf{F}_{\Delta}^{(p)} \right]_{ij} = -\mathrm{E}\left[\partial^2 g_{\Delta}^{(p)} \Big/ \partial \beta_i^{(p)} \partial \beta_j^{(p)} \right]$, where $g_{\Delta}^{(p)} = \ln\left[\Pr(\Delta \mid \mathbf{D}, M^{(p)}, \hat{\boldsymbol{\beta}}^{(p)}) \Pr(\hat{\boldsymbol{\beta}}^{(p)} \mid \mathbf{D}, M^{(p)}) \right]$.

Then, the optimal $\hat{\boldsymbol{\beta}}^{(p)}$ can be obtained by minimizing the KIC:

$$\min_{\mathbf{0} \le \boldsymbol{\beta}^{(p)} \le \mathbf{1}} \quad \mathrm{KIC}_{\Delta}^{(p)} \quad , p = 1, 2, \cdots \tag{7.60}$$

7.6.4 $\mathrm{KIC}_{D}^{(p)}$ vs. $\mathrm{KIC}_{\Delta}^{(p)}$

In the general case, $\mathrm{KIC}_{D}^{(p)}$ and $\mathrm{KIC}_{\Delta}^{(p)}$ are different because the data in $\Pr(\Delta \mid \mathbf{D}, M^{(p)}, \hat{\boldsymbol{\beta}}^{(p)})$ and the data in $\Pr(\mathbf{D} \mid M^{(p)}, \boldsymbol{\beta}^{(p)})$ can differ. For example, one merely can use measured hydraulic conductivity data through the cross-validation

(CV) approach (Stone 1974) to estimate the $\boldsymbol{\beta}$ values. Using CV, $\mathbf{D} = \mathbf{p}^{data}$ are the measured hydraulic conductivity values and Δ are the cross-validated hydraulic conductivity values at the sample locations. This approach has been suggested in Ye et al. (2004). Both $\mathrm{KIC}_{\Delta}^{(p)}$ and $\mathrm{KIC}_{D}^{(p)}$ can be evaluated by the CV method. In groundwater inverse modeling, we still can choose $\mathbf{D} = \mathbf{p}^{data}$ to calculate $\mathrm{KIC}_{D}^{(p)}$ to obtain the GP weights in $\Pr\left(M^{(p)} \mid \mathbf{D}\right)$ through the CV method. However, $\mathrm{KIC}_{\Delta}^{(p)}$ is based on the predicted groundwater heads $\Delta = \mathbf{h}^{cal}$ at the head observation space. In a case where the same data are used in both $\Pr(\Delta \mid \mathbf{D}, M^{(p)}, \hat{\boldsymbol{\beta}}^{(p)})$ and $\Pr(\mathbf{D} \mid M^{(p)}, \boldsymbol{\beta}^{(p)})$, $\mathrm{KIC}_{D}^{(p)}$ and $\mathrm{KIC}_{\Delta}^{(p)}$ have the same forms for the inverse problem. Moreover, $\mathrm{KIC}_{D}^{(p)}$ and $\mathrm{KIC}_{\Delta}^{(p)}$ are different when the predicted quantity Δ is at the prediction space, not at the observation space. Furthermore, even though $\mathrm{KIC}_{D}^{(p)}$ and $\mathrm{KIC}_{\Delta}^{(p)}$ are the same in the inverse problem, $\mathrm{KIC}_{D}^{(p)}$ is mainly used to calculate the importance (weights) of individual GP methods in BMA. From a practical viewpoint, $\mathrm{KIC}_{D}^{(p)}$ can be modified or scaled as the empirical Bayesian inference in order to obtain reasonable GP weights. Recently, Tsai and Li (2008b) and Tsai and Li (2008c) introduced a variance window to scale $\mathrm{KIC}_{D}^{(p)}$ to cope with the narrow window size of Occam's window. However, scaling $\mathrm{KIC}_{\Delta}^{(p)}$ should not be considered.

As a consequence, the data $\mathbf{D}$ in BMA in Equation (7.47) actually represent two data sets: one data set $\mathbf{D}_{\Delta} \in \mathbf{D}$ for calculating $\mathrm{KIC}_{\Delta}^{(p)}$ and the other data set $\mathbf{D}_{D} \in \mathbf{D}$ for calculating $\mathrm{KIC}_{D}^{(p)}$. For the KIC case in the following analysis, we consider the prior $\boldsymbol{\beta}$ distribution in g_{Δ} and g_{D} to be independent within and between individual GP methods and to be uniformly distributed between [0,1].

7.6.5 Multi-Gaussian Distributions

Consider a subset of data $\mathbf{D}_{\Delta} \in \mathbf{D}$ is used to calculate $\Pr\left(\Delta \mid \mathbf{D}, M^{(p)}\right)$ and the prediction $\Delta(\boldsymbol{\beta}^{(p)})$ is at the $\mathbf{D}_{\Delta}$ space. For groundwater inverse modeling purposes, $\Pr\left(\Delta \mid \mathbf{D}_{\Delta}, M^{(p)}, \hat{\boldsymbol{\beta}}^{(p)}\right)$ describes the probability of errors between the predicted quantity and observation data, and is assumed to be a multi-Gaussian distribution with zero mean and covariance matrix $\mathbf{C}_{\Delta}$. Therefore, the $\mathrm{KIC}_{\Delta}^{(p)}$ is

$$\mathrm{KIC}_{\Delta}^{(p)} = Q_{\Delta}^{(p)} + \left(n_{\Delta} - m^{(p)}\right)\ln 2\pi + \ln\left|\mathbf{C}_{\Delta}\right| + \ln\left|\mathbf{F}_{\Delta}^{(p)}\right| \tag{7.61}$$

where n_{Δ} is the number of $\mathbf{D}_{\Delta}$ and

$$Q_{\Delta}^{(p)} = \left(\Delta(\boldsymbol{\beta}^{(p)}) - \mathbf{D}_{\Delta}\right)^{T} \mathbf{C}_{\Delta}^{-1}\left(\Delta(\boldsymbol{\beta}^{(p)}) - \mathbf{D}_{\Delta}\right) \tag{7.62}$$

is the fitting residual with respect to data, $\mathbf{D}_\Delta$. The Fisher information matrix, $\mathbf{F}_\Delta^{(p)}$, is (Kitanidis and Lane 1985)

$$\left[\mathbf{F}_\Delta^{(p)}\right]_{ij} = \frac{1}{2}\mathrm{Tr}\left[\mathbf{C}_\Delta^{-1}\frac{\partial \mathbf{C}_\Delta}{\partial \beta_i^{(p)}}\mathbf{C}_\Delta^{-1}\frac{\partial \mathbf{C}_\Delta}{\partial \beta_j^{(p)}}\right] + \frac{\partial \Delta^T}{\partial \beta_i^{(p)}}\mathbf{C}_\Delta^{-1}\frac{\partial \Delta}{\partial \beta_j^{(p)}}. \tag{7.63}$$

If the covariance matrix $\mathbf{D}_\Delta$ is constant or the trace is relatively small, the Fisher information matrix can be approximated to the inverse covariance matrix of the estimated weighting coefficients:

$$\left[\mathbf{F}_\Delta^{(p)}\right]_{ij} = \frac{\partial \Delta^T}{\partial \beta_i^{(p)}}\mathbf{C}_\Delta^{-1}\frac{\partial \Delta}{\partial \beta_j^{(p)}} \approx \mathrm{Cov}\left(\boldsymbol{\beta}^{(p)}\right)^{-1}. \tag{7.64}$$

Yeh and Yoon (1981) considered the groundwater heads $\Delta = \mathbf{h}$ in Equation (7.64).

Similarly, the subset of data $\mathbf{D}_D \in \mathbf{D}$ is used to calculate model weights $\Pr\left(\mathbf{D} \mid M^{(p)}\right)$, and $\mathbf{D}^{cal}(\boldsymbol{\beta}^{(p)})$ is the calculated data corresponding to the $\mathbf{D}_D$ space. The error between the calculated data and observation data is assumed to be multi-Gaussian with zero mean and covariance matrix $\mathbf{C}_D$ in the likelihood function $\Pr\left(\mathbf{D}_D \mid M^{(p)}, \hat{\boldsymbol{\beta}}^{(p)}\right)$. Accordingly, the $\mathrm{KIC}_D^{(p)}$ is

$$\mathrm{KIC}_D^{(p)} = Q_D^{(p)} + \left(n_D - m^{(p)}\right)\ln 2\pi + \ln\left|\mathbf{C}_D\right| + \ln\left|\mathbf{F}_D^{(p)}\right| \tag{7.65}$$

where n_D is the number of $\mathbf{D}_D$ and

$$Q_D^{(p)} = \left(\mathbf{D}^{cal} - \mathbf{D}_D\right)^T \mathbf{C}_D^{-1}\left(\mathbf{D}^{cal} - \mathbf{D}_D\right) \tag{7.66}$$

is the fitting residual with respect to data, $\mathbf{D}_D$. The Fisher information matrix, $\mathbf{F}_D^{(p)}$, is

$$\left[\mathbf{F}_D^{(p)}\right]_{ij} = \frac{1}{2}\mathrm{Tr}\left[\mathbf{C}_D^{-1}\frac{\partial \mathbf{C}_D}{\partial \beta_i^{(p)}}\mathbf{C}_D^{-1}\frac{\partial \mathbf{C}_D}{\partial \beta_j^{(p)}}\right] + \frac{\partial \mathbf{D}^{cal^T}}{\partial \beta_i^{(p)}}\mathbf{C}_D^{-1}\frac{\partial \mathbf{D}^{cal}}{\partial \beta_j^{(p)}}. \tag{7.67}$$

The multi-standard normal distribution is also applicable to BIC as follows:

$$\begin{aligned} \mathrm{BIC}_\Delta^{(p)} &= Q_\Delta^{(p)} + n_\Delta \ln 2\pi + \ln\left|\mathbf{C}_\Delta\right| + m^{(p)} \ln n_\Delta \\ \mathrm{BIC}_D^{(p)} &= Q_D^{(p)} + n_D \ln 2\pi + \ln\left|\mathbf{C}_D\right| + m^{(p)} \ln n_D \end{aligned} \tag{7.68}$$

7.7 Experimental Design

7.7.1 Experimental Design for Parameter Estimation

Experimental design deals with the selection of experimental conditions such that a specified criterion is optimized. The experimental conditions also are referred to as the decision variables. The decision variables of a groundwater experimental design generally consist of two groups: the excitation group and the observation group (Hsu and Yeh 1989; Sun 1994). In the excitation group, the decision variables may include the number and locations of pumping and injection wells, the pumping and injection rates, stress periods of pumping and injection, and artificial changes of boundary conditions. In the observation group, the decision variables may include the state variables to be observed, the number and locations of observation wells, and the observation frequency. When the decision variables in the excitation group are fixed, experimental design simplifies to observation network design. The optimization of experimental design is generally subject to a set of constraints. The constraints frequently encountered include: budget, allowable drawdown at selected locations, maximum pumping/recharge rates, duration of the experiment, allowable time interval between consecutive measurements, and reliability of the estimated parameters. To apply the optimal experimental design in practice requires that we: 1) establish a criterion (performance measure) so that different experimental designs can be compared, and 2) develop an algorithm so that the established criterion can be optimized over the proper choice of the decision variables. The formulated optimal experimental problem invaluably lends itself to a combinatorial optimization problem which, in principle, can be solved by a mixed integer nonlinear programming algorithm (Yeh 1992).

Steinberg and Hunter (1984) presented an extensive review of the classical criteria derived for linear statistical models. The most popular optimality criteria are:
(i) *D*-Optimality: A design is said to be *D*-optimal if it minimizes the determinant of the covariance matrix of the estimated parameters.
(ii) *A*-Optimality: A design is said to be *A*-optimal if it minimizes the trace of the covariance matrix of the estimated parameters.
(iii) *E*-Optimality: A design is said to be *E*-optimal if it minimizes the maximal eigenvalue of the covariance matrix of the estimated parameters.
(iv) *G*-Optimality: A design is said to be *G*-optimal if it minimizes $\max\ d(\mathbf{x})$, where $d(\mathbf{x})$ is the variance of the estimated response at $\mathbf{x}$, and the maximum is taken over all possible vectors $\mathbf{x}$ of predictor variables.
(v) I_λ-Optimality: A design is said to be I_λ-optimal if it minimizes $\int d(\mathbf{x})\lambda(d\mathbf{x})$, where λ is a probability measure on the space of predictor variables. This criterion, also called average integrated variance, belongs to a more general class of L-optimality criterion (Fedorov 1972). Among the proposed design criteria, the *A*-Optimality, *D*-Optimality and *E*-Optimality are the most widely used.

The inherent difficulty in experiment design is that the designs are predicated on

unknown parameters which themselves are to be estimated. However, in practice, initial estimates of parameters are determined from prior information. Using the initial estimates, an optimal experimental design is carried out and data are collected accordingly. With the collected data, the inverse problem of parameter identification is solved to update the parameter values. If necessary, another round of experimental design can be carried out with the updated parameter values. This procedure is called sequential design. Its convergence property has been investigated by Nishikawa and Yeh (1989) and Cleveland and Yeh (1990).

7.7.2 Experimental Design for Parameter Structure Identification

Traditional experimental design methods, such as *D*-optimal design or the *A*-optimal design, do not consider whether the information provided by a design is sufficient. As a result, under-sampling or over-sampling frequently occurs. Chang et al. (2005) utilized the GIP and presented a methodology for observation network design aimed at finding a minimum cost design that provides sufficient information for identifying both the parameter structure and parameter values. Sequential Gaussian simulation was used to generate different realizations for the unknown distributed parameter. For each possible realization, the number of observation wells was increased gradually during the design process until the information provided by the design was sufficient. The selection criterion for locating a new observation well was the maximization of the information content for parameter identification. The overall sufficiency of a design was assessed by Monte Carlo simulation.

To circumvent the need for initial estimates of parameters and a large number of Monte Carlo simulations, Sun and Yeh (2007b) proposed a robust experimental design procedure based on the worst-case parameter in the parameter admissible region. The worst-case parameter (WCP) is defined as one that requires the most information for its identification (Sun and Yeh 2007a). Therefore, if the data provided by a design are sufficient for identifying the WCP, then the design must be sufficient for identifying any other parameters with the same structure or a simplified structure. Sun and Yeh (2007b) presented the following procedure for constructing a sufficiency and robustness design D:

Step 1. Compile all available prior information including the objectives of model application and their accuracy requirements.

Step 2. Based on the prior information, estimate a structure $\boldsymbol{\theta}_A$ as the true structure and find its WCP, $\tilde{\mathbf{p}}_A$.

Step 3. Run the simulation model to generate a set of "observation data" $\mathbf{u}_D(\boldsymbol{\theta}_A, \tilde{\mathbf{p}}_A)$ according to the designed excitation strengths, observation locations and times.

Step 4. Run the application model to generate a set of "application results" $\mathbf{g}_E(\boldsymbol{\theta}_A, \tilde{\mathbf{p}}_A)$.

Step 5. Generate a series of structures one by one and calculate the following fitting residual RE_m and model application error AE_m for each structure $\boldsymbol{\theta}_m$:

$$RE_m = \min_{\boldsymbol{\theta}_m, \mathbf{p}_m} \left\| \mathbf{u}_D(\boldsymbol{\theta}_A, \tilde{\mathbf{p}}_A) - \mathbf{u}_D(\boldsymbol{\theta}_m, \mathbf{p}_m) \right\|_D \quad , \quad \text{s.t.} \quad \mathbf{p}_m \in Ad(\mathbf{p}_m) \tag{7.69}$$

$$AE_m = \left\| \mathbf{g}_{\mathbf{E}}(\boldsymbol{\theta}_A, \tilde{\mathbf{p}}_A) - \mathbf{g}_{\mathbf{E}}(\boldsymbol{\theta}_m, \tilde{\mathbf{p}}_m) \right\|_E \tag{7.70}$$

where m is the parameter dimension; $\|\cdot\|_E$ and $\|\cdot\|_D$ are L_2 norms defined in the objective space E and observation space D, respectively; $Ad(\mathbf{p}_m)$ is the admissible set of $\mathbf{p}_m$, and $(\boldsymbol{\theta}_m, \tilde{\mathbf{p}}_m)$ is the solution of Equation (5.69).

Step 6. If $AE_m > \varepsilon$ and $RE_m > 2\eta$, increase the parameter dimension m by 1 and repeat the above procedure to find $\boldsymbol{\theta}_{m+1}$, where ε is the specified accuracy requirement and η is the upper bound of the norm of observation error measured in the observation space.

Step 7. When m increases, the value of AE_m will decrease and the value of RE_m will decrease.

Thus, finally we must arrive at one of the following three situations: (i) $AE_m < \varepsilon$ but $RE_m \geq 2\eta$; (ii) $AE_m < \varepsilon$ for all $RE_m < 2\eta$; or (iii) $AE_m \geq \varepsilon$ but $RE_m < 2\eta$. If cases (i) and (ii) occur, we can conclude that the design is robust and sufficient. Otherwise, when case (iii) occurs, the design is insufficient.

As mentioned before, a robust design is a conservative design. It provides the maximum information for identifying the most difficult parameter in the admissible parameter region. Based on the WCP, Sun and Yeh (2007b) proposed a heuristic design procedure for finding a robust but sub-optimal experimental design.

7.8 Summary and Conclusions

i. We can classify parameter structure identification into the extended inverse problem (EIP) and the generalized inverse problem (GIP). The EIP considers the identification of parameter structure (parameter dimension, parameter pattern, and parameter values) based on maximum likelihood estimation. The GIP broadens the EIP by incorporating model prediction or management along with the maximum likelihood estimation. Model discrimination and model selection criteria, including the parsimony principle, parameter uncertainty, structure error, prediction error and statistical information criteria, are introduced to determine the best parameter structure.

ii. Parameterization is a necessary step to represent a distributed parameter (e.g., the hydraulic conductivity discussed in this chapter) for groundwater modeling and define parameter structure for parameter structure identification. We introduced a generalized parameterization (GP) method to integrate the zonation method and the

interpolation method to increase parameterization flexibility. This approach is important because a parameter estimate is not necessarily limited to a piece-wise constant or smooth distribution. Moreover, the GP method extends the traditional geostatistical framework to characterize spatially correlated parameters for stochastic inverse modeling.

iii. Interpolation point selection plays an important role in parameter structure identification. We introduced the indicator generalized parameterization (IGP) to determine the selection of sample data points for interpolation. The IGP focuses on an unsampled location's relevance to its neighboring sampled locations. This is more general than the traditional approach of focusing on the influential range of sample data on its neighboring locations where parameter value needs to be determined.

iv. In contrast to model selection and model discrimination, we introduced Bayesian model averaging (BMA) to address the non-uniqueness of parameterization methods. Using a single parameterization method for parameter structure identification is likely to underestimate estimation uncertainty. Based on the law of total probability, the BMA weighs candidate models by the evidence of data. Then, the spatial statistics of an estimated parameter using multiple parameterization methods can be obtained by the BMA expectation and covariance.

v. Experimental design complements the inverse theory. If the existing data are found to be insufficient for parameter structure identification, experimental design can be used to determine sampling strategies with the objective of reducing model uncertainty. We introduced several optimality criteria for optimizing experimental design for collecting informative data.

Acknowledgements

The first author was supported in part by the Department of the Interior, U.S. Geological Survey under Grant No. 05HQGR0142 and 06HQGR0088. The views and conclusions contained in the document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Government.

7.9 References

Akaike, H. (1973). "Information theory and an extension of the maximum likelihood principle." *Proceeding of the second international symposium on information theory*, B. N. Petrov and F. Csaki, eds., Akademiai Kiado, Budapest, 267-281.

Akaike, H. (1974). "A new look at statistical model identification." *IEEE Transactions on Automatic Control*, AC-19(6), 716-723.

Akaike, H. (1981). "Likelihood of a model and information criteria." *Journal of Econometrics*, 16(1), 3-14.