Likelihood Function

The parameters **s** are related to observations **y** through a measurement equation

$$\mathbf{y} = \mathbf{h}(\mathbf{s}) + \mathbf{v} \tag{3.23}$$

where **y** is an $(n \times 1)$ vector of observations, such as hydraulic heads or solute concentrations, $\mathbf{h}(\mathbf{s})$ is a transfer function or numerical model that calculates predictions which are colocated spatially and temporally with the observation values, and **v** is an $(n \times 1)$ vector of epistemic uncertainty terms, modeled as a random process with zero mean and covariance matrix **R**. Epistemic uncertainty is the result of imperfect or sparse measurements and an incomplete or inappropriate conceptual model (Rubin, 2003, p. 4). The epistemic uncertainty terms are assumed to be independent and uncorrelated so

$$\mathbf{R} = \sigma_{\mathbf{R}}^{\mathbf{2}}\mathbf{W} \tag{3.24}$$

where σ_R^2 is the epistemic uncertainty parameter and **W** is an $(n \times n)$ diagonal weight matrix in which each element is $\mathbf{W}_{ii} = \frac{1}{\omega_i^2}$ where ω_i is the ith weight, specified by the user. The purpose of the values of ω is to allow for different confidence in different individual observations or groups of observations. In reality, the component of epistemic uncertainty due to measurement error is likely uncorrelated, but the component due to modeling and conceptual uncertainty is likely systematic and correlated (Gaganis and Smith, 2001). A significant portion of this uncertainty may be reduced by not lumping parameters into homogeneous zones (Gallagher and Doherty, 2007), and the means to characterize the structure of **R** are rarely available. If information about **R** is available, however, it could be included and equation 3.24 replaced by a more complicated matrix. This option is currently not available in bgaPEST, however. Proceeding with equation 3.24, the likelihood function, assumed to be multi-Gaussian, is

$$L(\mathbf{y}|\mathbf{s}) \propto \exp\left[-\frac{1}{2}\left(\mathbf{y}-\mathbf{h}(\mathbf{s})\right)^{\mathrm{T}}\mathbf{R}^{-\mathbf{1}}\left(\mathbf{y}-\mathbf{h}(\mathbf{s})\right)\right]. \tag{3.25}$$

The structural parameter for the likelihood function is σ_R^2 and is calculated along with θ by using restricted maximum likelihood.

Posterior Probability Density Function

Applying Bayes' theorem with the product of equations 3.20 and 3.25 yields the posterior pdf

$$p(\mathbf{s}|\mathbf{y}) \propto \exp\left[-\frac{1}{2}(\mathbf{s}-\mathbf{X}\beta^*)^{\mathrm{T}}\mathbf{G}_{\mathbf{ss}}^{-\mathbf{1}}(\mathbf{s}-\mathbf{X}\beta^*)-\frac{1}{2}\left(\mathbf{y}-\mathbf{h}(\mathbf{s})\right)^{\mathrm{T}}\mathbf{R}^{-\mathbf{1}}\left(\mathbf{y}-\mathbf{h}(\mathbf{s})\right)\right] \tag{3.26}$$

The best estimate of **s** maximizes the posterior pdf. A computationally efficient method to find the best estimates of **s** and β $\left(\hat{\mathbf{s}} \text{ and } \hat{\beta}\text{, respectively}\right)$ is through

$$\hat{\mathbf{s}} = \mathbf{X}\hat{\beta} + \mathbf{Q_{ss}}\mathbf{H}^{\mathbf{T}}\xi \tag{3.27}$$

which is the superposition of the prior mean (first term) and an innovation term that accounts for deviations of the model outputs from the observations (second term). **H** in the second term (often referred to as the Jacobian, sensitivity, or susceptibility matrix) is the sensitivity of observation values to parameter values where $H_{ij} = \frac{\partial \mathbf{h}(\mathbf{s})_i}{\partial \mathbf{s}_j}$ is calculated by using either finite-difference or adjoint-state methods. The values for $\hat{\beta}$ and

ξ are found by solving the $(n+p)\times(n+p)$ linear system of cokriging equations

$$\begin{bmatrix} \mathbf{Q_{yy}} & \mathbf{HX} \\ \mathbf{X}^T\mathbf{H}^T & -\mathbf{Q}_{\beta\beta}^{-1} \end{bmatrix} \begin{bmatrix} \xi \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ -\mathbf{Q}_{\beta\beta}^{-1}\beta^* \end{bmatrix} \tag{3.28}$$

where $\mathbf{Q_{yy}}$ is the auto-covariance matrix of the observations, defined as $\mathbf{HQ_{ss}H^T}+\mathbf{R}$.

Quasi-Linear Extension

As discussed by Kitanidis (1995), we must adjust calculations of the posterior pdf to account for nonlinearity. To do this, we expand the solution in a first-order Taylor expansion, resulting in an updated set of cokriging equations from equation 3.28

$$\begin{bmatrix} \mathbf{Q_{yy}} & \mathbf{HX} \\ \mathbf{X}^T\mathbf{H}^T & -\mathbf{Q}_{\beta\beta}^{-1} \end{bmatrix} \begin{bmatrix} \xi \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{y}-\mathbf{h}(\tilde{\mathbf{s}})+\mathbf{H}\tilde{\mathbf{s}} \\ -\mathbf{Q}_{\beta\beta}^{-1}\beta^* \end{bmatrix} \tag{3.29}$$

At each iteration (later referred to as inner iterations), the system in equation 3.29 is solved, resulting in an updated estimate of $\hat{\mathbf{s}}$ calculated through equation 3.27. At each iteration, the objective function, based on minimizing the negative logarithm of the posterior pdf (equation 3.26) is evaluated by using the current value of $\hat{\mathbf{s}}$: this is equivalent to finding the values of $\mathbf{s}$ that *maximize* the posterior probability. Switching to a minimization problem and taking the logarithm has computational advantages.

The objective function, then, is

$$\Phi_T = \Phi_M + \Phi_R \tag{3.30}$$

where Φ_T is the total objective function, Φ_M is the misfit objective function (also corresponding to the likelihood function) and Φ_R is the regularization objective function (also corresponding to the prior pdf). The components of equation 3.30 are

$$\Phi_M = \frac{1}{2}\left(\mathbf{y}-\mathbf{h}(\hat{\mathbf{s}})\right)^T \mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{h}(\hat{\mathbf{s}})\right) \tag{3.31}$$

and

$$\Phi_R = \frac{1}{2}(\hat{\mathbf{s}}-\mathbf{X}\beta^*)^T\mathbf{G}_{\mathbf{ss}}^{-1}(\hat{\mathbf{s}}-\mathbf{X}\beta^*) \tag{3.32}$$

where both the negative signs and exponentiation are obviated by taking the negative logarithm of $p(\mathbf{s}|\mathbf{y})$.

Implementation of Partitions into Beta Associations

The concept of beta associations is discussed above and details of their implementation are given here. First, the prior covariance matrix $\mathbf{Q_{ss}}$ is censored by assigning a value of zero to each element that characterizes covariance between cells of different regions or parameter types, as defined by beta associations. It is not required that the covariance model be the same for each beta association. If different covariance models are used for different zones, this is reflected in the appropriate parts of $\mathbf{Q_{ss}}$. Furthermore, in some applications, a single structural parameter, θ, may be estimated and applied to all of $\mathbf{Q_{ss}}$. In other cases, and necessarily if the covariance model differs in various beta association, multiple elements of θ are estimated.

A distinct prior mean parameter β^* is assigned for each beta association, and the matrix $\mathbf{X}$ (equation 3.13) is determined as explained above. In cases where the mean of each zone is completely unknown, no values for

β^* are provided, but the **X** matrix is constructed nonetheless and in both cases a value of $\hat{\beta}$ is calculated for each beta association.

Structural Parameters and Restricted Maximum Likelihood

A vital element to the method outlined above is proper selection of the structural parameters. Structural parameters—also called hyperparameters—or nuisance parameters, are the parameters that characterize the covariance structure of both the epistemic uncertainty related to the observations, and the inherent variability of the parameters. In bgaPEST, structural parameters may include the epistemic uncertainty term in equation 3.24 $\left(\sigma_R^2\right)$ and the prior pdf variogram parameters in equation 3.17 (θ). These parameters are estimated by using restricted maximum likelihood consistent with the approaches of Kitanidis and Vomvoris (1983), Kitanidis (1995) and Li and others (2007).

Applying Bayes' theorem to the structural parameters, given the measurements, we calculate

$$p\left(\theta|\mathbf{y}_{\mathbf{k}}^{'}\right) \propto L\left(\mathbf{y}_{k}^{'}|\theta\right) p\left(\theta\right) \tag{3.33}$$

The likelihood function evaluates how closely the observations and predictions match, given the current linearization and the current set of structural parameters

$$L\left(\mathbf{y}_{k}^{'}|\theta\right) \propto \det\left(\mathbf{G}_{\mathbf{yy}}\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\mathbf{y}_{k}^{'}-\mathbf{HX}\beta^*\right)^T \mathbf{G}_{\mathbf{yy}}^{-\mathbf{1}}\left(\mathbf{y}_{k}^{'}-\mathbf{HX}\beta^*\right)\right] \tag{3.34}$$

where $\mathbf{G}_{\mathbf{yy}}$ is the measurement autocovariance defined as

$$\mathbf{G}_{\mathbf{yy}} = \mathbf{Q}_{\mathbf{yy}} + \mathbf{HXQ}_{\beta\beta}\mathbf{X}^T\mathbf{H}^T. \tag{3.35}$$

Note that $\mathbf{Q}_{\mathbf{yy}}$ is intrinsically dependent upon the values of θ.

Prior information about the structural parameters may also be included, with prior mean θ^* and covariance matrix $\mathbf{Q}_{\theta\theta}$:

$$p\left(\theta\right) \propto \det\left(\mathbf{Q}_{\theta\theta}\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\theta-\theta^*\right)^T \mathbf{Q}_{\theta\theta}^{-1}\left(\theta-\theta^*\right)\right] \tag{3.36}$$

The posterior pdf is the product of equations 3.36 and 3.34

$$p\left(\theta|\mathbf{y}_{\mathbf{k}}^{'}\right) \propto \det\left(\mathbf{Q}_{\theta\theta}\right)^{-\frac{1}{2}} \det\left(\mathbf{G}_{\mathbf{yy}}\right)^{-\frac{1}{2}} \exp\left[\begin{array}{c} -\frac{1}{2}\left(\theta-\theta^*\right)^T \mathbf{Q}_{\theta\theta}^{-1}\left(\theta-\theta^*\right) \\ -\frac{1}{2}\left(\mathbf{y}_{k}^{'}-\mathbf{HX}\beta^*\right)^T \mathbf{G}_{\mathbf{yy}}^{-\mathbf{1}}\left(\mathbf{y}_{k}^{'}-\mathbf{HX}\beta^*\right) \end{array}\right]. \tag{3.37}$$

To find the most likely values for θ we minimize $-\ln\left(p\left(\theta|\mathbf{y}_{\mathbf{k}}^{'}\right)\right)$ resulting in the objective function

$$\Phi_S = \frac{1}{2}\ln\left(\det\left(\mathbf{G}_{\mathbf{yy}}\right)\right) + \frac{1}{2}\left[\left(\theta-\theta^*\right)^T \mathbf{Q}_{\theta\theta}^{-1}\left(\theta-\theta^*\right) + \left(\mathbf{y}_{k}^{'}-\mathbf{HX}\beta^*\right)^T \mathbf{G}_{\mathbf{yy}}^{-\mathbf{1}}\left(\mathbf{y}_{k}^{'}-\mathbf{HX}\beta^*\right)\right] \tag{3.38}$$

where unchanging quantities are absorbed into the constant of proportionality including $\det\left(\mathbf{Q}_{\theta\theta}\right)^{-\frac{1}{2}}$. The optimal values for θ are found byusing the Nelder-Mead simplex algorithm (for example, Press and others, 1992, p. 408-410). Non-negativity in the θ parameters can be enforced by using a power transformation (Box and Cox, 1964) discussed below. As indicated by Kitanidis (1995), nonlinearity requires that structural parameters to be estimated iteratively with the estimation of model parameters. This is accomplished through a sequence of coupled inversion as follows.

1. Initialize model parameters as $(\mathbf{s_0})$ and structural parameters $(\theta_\mathbf{0})$.
2. Solve for a new estimate of model parameters $(\hat{\mathbf{s}})$ holding θ constant.
3. Solve for a new estimate of structural parameters $\left(\hat{\theta}\right)$ holding $\mathbf{s}$ constant.
4. Repeat steps 2 and 3 until the change in θ in two consecutive outer iterations of steps 2 and 3 decreases below a specified tolerance.

Logarithmic and Power Transformations

In some cases, structural parameters and model parameters are best estimated in transformed space. A common reason is to enforce non-negativity. For model parameters either a logarithmic (base e) or a power transformation may be used. For structural parameters a power transformation is the only option.

The power transformation (Box and Cox, 1964; Fienen and others, 2004) is defined as:

$$\mathbf{s} = \alpha\left(\mathbf{p}^{\frac{1}{\alpha}} - 1\right) \tag{3.39}$$

where $\mathbf{s}$ is the vector of transformed parameters, $\mathbf{p}$ is the vector of non-transformed parameters, and α is a tuning variable that controls the strength of the transformation. The back-transformation is:

$$\mathbf{p} = \left(\frac{\mathbf{s}+\alpha}{\alpha}\right)^{\alpha} \tag{3.40}$$

At the limit, as α increases to infinity, the transformation and back-transformation converge on the natural logarithm and exponential function, respectively:

$$\mathbf{s} = \lim_{\alpha\to\infty} = \alpha\left(\mathbf{p}^{\frac{1}{\alpha}} - 1\right) = \ln(\mathbf{p}) \tag{3.41}$$

and

$$\mathbf{p} = \lim_{\alpha\to\infty} = \left(\frac{\mathbf{s}+\alpha}{\alpha}\right)^{\alpha} = \exp(\mathbf{s})\,. \tag{3.42}$$

Posterior Covariance

Calculation of the posterior covariance can be based on the inverse of the Hessian of the objective function (for example, Nowak and Cirpka, 2004). In closed form, the equation for the full posterior covariance matrix is:

$$\mathbf{V} = \mathbf{G_{ss}} - \mathbf{G_{sy}}\mathbf{G_{yy}^{-1}}\mathbf{G_{sy}^{\mathit{T}}} \tag{3.43}$$

where $\mathbf{G_{sy}} = \mathbf{G_{ss}}\mathbf{H}^T$ and $\mathbf{G_{yy}} = \mathbf{H}\mathbf{G_{ss}}\mathbf{H}^T + \mathbf{R}$. In the case where compression of $\mathbf{Q}_{ss}$ is not used, the full matrix $\mathbf{V}$ is calculated and reported. Where compression of $\mathbf{Q}_{ss}$ is used, however, the diagonal of $\mathbf{V}$ is returned as a vector of variances on parameters. This information is reported in a separate file but is also used to calculated posterior 95 percent confidence intervals. The full matrix, when reported, can be used to calculate conditional realizations (Kitanidis, 1995, 1996).