The Bayesian geostatistical approach is described in detail by Kitanidis and Vomvoris (1983), Hoeksema and Kitanidis (1984), Kitanidis (1995), and Nowak and Cirpka (2004) among others. The mathematics are reviewed here.

## The Bayesian Geostatistical Approach

In the Bayesian geostatistical approach, the posterior pdf is calculated as

$$p(\mathbf{s}|\mathbf{y}) \propto \underbrace{\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{h}(\mathbf{s}))^T \mathbf{R}^{-1}(\mathbf{y}-\mathbf{h}(\mathbf{s}))\right)}_{\mathrm{L}(\mathbf{y}|\mathbf{s})} \underbrace{\exp\left(-\frac{1}{2}(\mathbf{s}-\mathbf{X}\beta^*)^T \mathbf{G}_{\mathbf{ss}}^{-1}(\mathbf{s}-\mathbf{X}\beta^*)\right)}_{\mathrm{p}(\mathbf{s})} \tag{3.1}$$

where $\mathbf{s}$ is the $m \times 1$ vector of parameter values at distributed spatial locations in the model, $\mathbf{X}\beta^*$ is the prior mean, $\mathbf{G}_{\mathbf{ss}}$ is the $m \times m$ prior covariance of $(\mathbf{s}-\mathbf{X}\beta^*)$, $\mathbf{h}(\mathbf{s})$ is the $n \times 1$ vector of modeled forecasts colocated with observations $(\mathbf{y})$, and $\mathbf{R}$ is the $n \times n$ epistemic uncertainty covariance, modeled as $\sigma_R^2\mathbf{W}$ where $\sigma_R^2$ represents epistemic uncertainty, and $\mathbf{W}$ is an $n \times n$ diagonal matrix containing the observation weights. In general terms, the likelihood function $(L(\mathbf{y}|\mathbf{s}))$ characterizes the misfit between model forecasts and observations whereas the prior pdf $(p(\mathbf{s}))$ defines a characteristic (such as smoothness or continuity) that is assumed to apply to the parameter field. The prior pdf also serves the role of regularization. The best estimate of $\mathbf{s}$ maximizes the posterior pdf. A computationally efficient method to find the best estimates of $\mathbf{s}$ and $\beta$ $\left(\hat{\mathbf{s}} \text{ and } \hat{\beta}, \text{ respectively}\right)$ is through

$$\hat{\mathbf{s}} = \mathbf{X}\hat{\beta} + \mathbf{Q}_{\mathbf{ss}}\mathbf{H}^{\mathbf{T}}\xi \tag{3.2}$$

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). In this context, $\beta$ in the first term is not the prior mean, but is the best estimate of the mean (mapped onto the parameter field through the $m \times p$ $\mathbf{X}$ matrix) whereas the second term is fluctuations about the estimated mean. $\mathbf{H}$ in the second term (often referred to as the $n \times m$ Jacobian, sensitivity, or susceptibility matrix) is the sensitivity of observation values to parameter values where $H_{ij} = \frac{\partial \mathbf{h}(\mathbf{s})_i}{\partial \mathbf{s}_j}$, which can be calculated by using either finite difference or adjoint-state methods. In bgaPEST, finite-difference calculations for $\mathbf{H}$ are made by using PEST whereas adjoint-state calculations depend on the specific model being used and must be made by using an external program.

The values for $\hat{\beta}$ and $\xi$ are found by solving the $(n+p) \times (n+p)$ linear system of cokriging equations

$$\begin{bmatrix} \mathbf{Q}_{\mathbf{yy}} & \mathbf{HX} \\ \mathbf{X}^T\mathbf{H}^T & -\mathbf{Q}_{\beta\beta}^{-1} \end{bmatrix} \begin{bmatrix} \xi \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ -\mathbf{Q}_{\beta\beta}^{-1}\beta^* \end{bmatrix} \tag{3.3}$$

where $\mathbf{Q}_{\mathbf{yy}}$ is the $n \times n$ auto-covariance matrix of the observations, defined as $\mathbf{HQ}_{\mathbf{ss}}\mathbf{H}^{\mathbf{T}} + \mathbf{R}$, $n$ is the number of observations, and $p$ is the number of beta associations. In hydrogeologic applications, the numerical forward model is typically nonlinear. Further nonlinearity can be induced by using a logarithmic or power transformation, which is a convenient way to enforce non-negativity on parameters.

Provided that the nonlinearities introduced are not too extreme, a solution can be obtained through successive linearizations following the quasi-linear extension (Kitanidis, 1995). The forward model, $\mathbf{h}(\mathbf{s})$ is expanded about the current best estimate of the parameters $\tilde{\mathbf{s}}$

$$\mathbf{h}(\mathbf{s}) \approx \mathbf{h}(\tilde{\mathbf{s}}) + \tilde{\mathbf{H}}(\mathbf{s}-\tilde{\mathbf{s}}) \tag{3.4}$$

where $\tilde{\mathbf{H}}$, as a function of $\tilde{\mathbf{s}}$, is evaluated at each linearization. We assign the subscript $_k$ to indicate iteration