bgaPEST: A Bayesian Geostatistical Inverse Implementation with PEST–Documentation and Instructions

Michael N. Fienen[*], Marco D'Oria[†], John E. Doherty[‡], and Randall J. Hunt[*]

October 28, 2011

Acknowledgements

New USGS style has acknowledgements posted at the front. There will be many sources of support to acknowledge, but I'll start a list here.

This work was supported by several important sources including: The National Research Council Postdoctoral Research Associateship, The US Geological Survey Office of Groundwater, the USGS Water and Environmental Biogeochemical Budgets (WEBB) Program, The USGS Exchange Visitor Program and the USGS Wisconsin Water Science Center for supporting a six-month exchange in which Marco D'Oria was hosted in Wisconsin, The Italian AIPO (Agenzia Interregionale per il fume Po : The Interregional Agency for the Po River), and the University of Parma, Italy, Department of Civil Engineering.

Introduction

Environmental modeling serves many purposes for managing natural resources. In most cases, models are made to represent physical processes such as groundwater flow, con-

[*]US Geological Survey, Wisconsin Water Science Center, Middleton, Wisconsin USA
[†]Department of Civil Engineering, University of Parma, Parma, Italy
[‡]Watermark Numerical Computing, Brisbane, Australia

taminant transport, surface water flood routing, and others. The challenge to simulate the governing equations and concepts such as conservation of mass and momentum is only part of the challenge. Once a model adequately represents the physical characteristics of the problem, it is still a general tool until it is calibrated to observations at the specific site to be managed. Parameter estimation and calibration are both terms describing the process of incorporating the site-specific information into a model making it suitable to be applied to management decisions.

This concept is so important that one may interpret a model as a data-processing tool as much as a simulator of physical processes; the data processing being a pipeline from field observation to model parameters and ultimately to predictions. Another interpretation is a Bayesian one in which the model is a vehicle for updating an *a priori* understanding of the system to a conditional, *a posteriori* understanding. This is the approach adopted in this work and will be discussed in more detail. First, however, the concept of conditionality must be emphasized.

Bayes' theorem is at the heart of the techniques in this report:

$$p\left(\mathbf{s}|\mathbf{y}\right) \propto L\left(\mathbf{y}|\mathbf{s}\right) p\left(\mathbf{s}\right) \tag{1}$$

where: $\mathbf{s}$ is a vector of parameter values, $\mathbf{y}$ is a vector of observations, $p\left(\cdot\right)$ indicates a probability density function (pdf), $L\left(\cdot\right)$ indicates a likelihood function, and $|$ indicates conditionality. In words, Bayes' theorem states that the posterior probability of parameters conditional on the observations $p\left(\mathbf{s}|\mathbf{y}\right)$, is proportional to the prior probability of the parameters $p\left(\mathbf{s}\right)$ updated with the likelihood function $L\left(\mathbf{y}|\mathbf{s}\right)$ that expresses how well $\mathbf{y}$ is estimated using the model and a candidate parameter set $\mathbf{s}$. The pdfs in all cases are assumed to follow Gaussian distributions. This assumption is important and somewhat restrictive, but it is made for reasons of computational simplicity. Active research is ongoing on alternatives to this approach, but the Gaussian assumption is adopted in this report.

In the Bayesian context, by expressing the parameters and the likelihood function as probability distributions, they formally incorporate an estimate of their uncertainty rather than being treated as perfectly known values. All *a posteriori* (also called posterior) distributions are conditional upon the specific data used in the calibration process. Implicitly, they are also conditional on all other modeling and data decisions that go into formulating the problem: which model and what model options are chosen, nu-

merical considerations such as discretization and solver convergence criteria, boundary conditions that may or may not be considered static and known, variance values and weights given to individual observations and parameters, and others. The point is, if any of these things change, it is expected, in the Bayesian context, that the parameters estimated will also change.

As discussed in the previous paragraph, the conditionality includes *all* decisions made in the process of constructing a model and incorporating data. Not all of this information is explicit, however. In fact, the only *explicit* conditionality is on the data. The information contained in the prior pdf ($p(\mathbf{s})$) in eq. 1 is critical as well because this represents the state of knowledge about the parameters in the system prior to updating through the calibration process. In this report, the assumptions made prior to calibration are limited to an assumption of the mean value of each distinct parameter type and region within the model domain (beta association–described below), and a characteristic about continuity or smoothness of the parameter field. The degree to which this characteristic is enforced is dictated by the observations and the subsequent performance of the model. This is very similar (and in some cases, mathematically equivalent) to Tikhonov regularization. Limiting the information assumed *a priori* is similar to assuming an ignorance prior (e.g. *Jaynes and Bretthorst* 2003, chapter 12) such that the driver of the problem is more the observations than prior assumptions. Additionally, so-called structural parameters that enforce the characteristic are estimated using restricted maximum likelihood. All of these aspects of the algorithm, taken together, may be considered an Empirical Bayes approach.

The likelihood function in eq. 1, $L(\mathbf{y}|\mathbf{s})$, expresses the correspondence of model outcomes with field observations collocated in space and time. This correspondence is expressed as the sum of the squared differences between outcomes and observations, weighted by a covariance matrix. This is effectively equivalent to the weighted measurement objective function in PEST.

Purpose and Scope

This Report is intended to serve two main purposes. First, a Bayesian approach to parameter estimation–expressed in the context of the Bayesian Geostatistical Approach (BGA)–is introduced to provide a framework for moving a model from a general simulator of a physical process to a specific tool, based on a set of calibration information,

that can be used for management of a particular site and issue. Second, a computer code–bgaPEST–is introduced and documented in which BGA is deployed in a framework based on the free and open-source PEST suite of software *Doherty* (2010a). The input framework is consistent with the JUPITER project *Banta et al.* (2006). This marks the first implementation of BGA available for general use.

This initial implementation is written in FORTRAN-90 with limited capabilities in Python. For users on the Windows operating system, installation of Python is optional as the Python codes are compiled into executables using py2exe that can be called by the main program. For users on Macintosh or Linux systems, all the code must be compiled for the native platform and Python should already be installed so the Python scripts may be called directly without need to compile them separately. The use of external derivatives calculation with PEST and Python will be replaced by integration of a general parallel run management suite (GENIE) in the near future.

Part of the approach, as discussed beta in this report, is based on a distributed parameterization scheme meant to be flexible in that certain regions of the model (called beta associations) are assigned a unique parameter value to be estimated for each model node or cell. This leads to a large number of parameters which is accompanied by computational challenges. Some of these challenges can be alleviated by restricting beta associations in which distributed parameterization schema are implemented to have constant spacing in both lateral directions and must have consistent layer thickness. These three values need only be constant in their respective dimensions but they may differ from dimension to dimension. This restriction allows for implementation of a Toeplitz compression making it possible to use many more parameters than would otherwise be possible. The partitioning into beta associations makes it easy to combine distributed parameters in areas of most interest and where data are more available, and homogeneous zones in areas surrounding the main focus. The use of a Toeplitz compression is optional and, in cases where sufficient computational resources allow it, compression may be skipped and the restrictions on even spacing are not invoked.

The remainder of this report provides some details on the mathematical theory behind BGA, followed by detailed instructions for using the computer program. Conventions and assumptions for using the program are included in the discussion.

Bayesian Geostatistical Approach Background

The Bayesian geostatistical approach is described in detail in *Kitanidis and Vomvoris* (1983); *Hoeksema and Kitanidis* (1984); *Kitanidis* (1995); *Nowak and Cirpka* (2004) among others. In this Section, we provide an overview of the method.

The core of the Bayesian geostatistical inverse method is Bayes' theorem, which states

$$p(\mathbf{s}|\mathbf{y}) \propto L(\mathbf{y}|\mathbf{s})p(\mathbf{s}) \tag{2}$$

where $\mathbf{y}$ are the measured data, $\mathbf{s}$ are the unknown parameters, $p(\mathbf{s}|\mathbf{y})$ is the posterior probability density function (pdf) of $\mathbf{s}$ given $\mathbf{y}$, $L(\mathbf{y}|\mathbf{s})$ is the likelihood function, and $p(\mathbf{s})$ is the prior pdf of $\mathbf{s}$. Details of these pdfs are explained below.

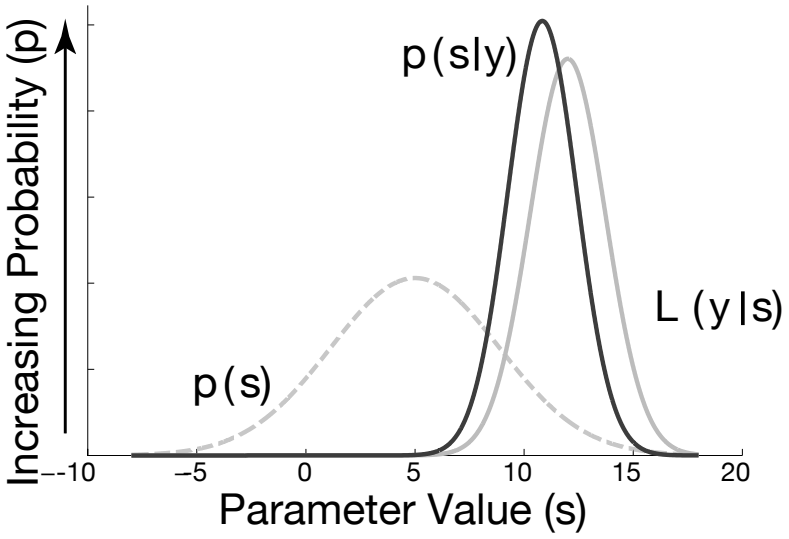

Figure 1: Graphical illustration of Bayes' theorem.

Figure 1 depicts one-dimensional distributions graphically illustrating Equation 2. In this example, the prior distribution ($p(\mathbf{s})$) is diffuse, meaning the variance is relatively high and, correspondingly, commitment to a particular value is low. The likelihood function ($L(\mathbf{y}|\mathbf{s})$) , on the other hand, has lower variance, suggesting a process that brings a higher level of certainty to the estimation of the parameters ($\mathbf{s}$) than is indicated by the prior distribution only. The resulting posterior distribution ($p(\mathbf{s}|\mathbf{y})$) is a convolution of the prior and likelihood functions. The peak is shifted significantly from the prior toward the likelihood and is higher. The posterior distribution also has lower

variance than the prior as illustrated by the spread in the figure. Since the likelihood function is more focused (less spread) and higher than the prior, the attributes of the likelihood contribute more to the posterior than the prior does.

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

$$p(\mathbf{s}|\mathbf{y}) \propto \underbrace{\exp\left(-\frac{1}{2}\left(\mathbf{y}-\mathbf{h}(\mathbf{s})\right)^T \mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{h}(\mathbf{s})\right)\right)}_{\mathrm{L}(\mathbf{y}|\mathbf{s})} \underbrace{\exp\left(-\frac{1}{2}(\mathbf{s}-\mathbf{X}\beta^*)^T \mathbf{G}_{\mathbf{ss}}^{-1}(\mathbf{s}-\mathbf{X}\beta^*)\right)}_{\mathrm{p}(\mathbf{s})} \tag{3}$$

where $\mathbf{s}$ is the vector of parameter values at distributed spatial locations in the model, $\mathbf{X}\beta^*$ is the prior mean, $\mathbf{G_{ss}}$ is the prior covariance of $(\mathbf{s}-\mathbf{X}\beta^*)$, $\mathbf{h}(\mathbf{s})$ is the modeled forecasts collocated with observations $(\mathbf{y})$, and $\mathbf{R}$ is the epistemic uncertainty covariance, modeled as $\sigma_R^2\mathbf{I}$ where σ_R^2 represents epistemic uncertainty, and $\mathbf{I}$ is an identity matrix. In general terms, the likelihood function $(L(\mathbf{y}|\mathbf{s}))$ characterizes the misfit between model forecasts and observations while 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 β $\left(\hat{\mathbf{s}}\text{ and }\hat{\beta}\text{, respectively}\right)$ is through

$$\hat{\mathbf{s}} = \mathbf{X}\hat{\beta} + \mathbf{Q_{ss}}\mathbf{H^T}\xi \tag{4}$$

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

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

cokriging equations

$$\begin{bmatrix} \mathbf{Q_{yy}} & \mathbf{HX} \\ \mathbf{X}^T\mathbf{H}^T & -\mathbf{Q}_{\beta\beta}^{-1} \end{bmatrix} \begin{bmatrix} \xi \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ -\mathbf{Q}_{\beta\beta}^{-1}\beta^* \end{bmatrix} \tag{5}$$

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

In hydrogeologic applications, the numerical forward model is typically nonlinear. Further nonlinearity can be induced by using a logarithmic transformation, which is a convenient way to enforce non-negativity on parameters and is often an appropriate transformation for hydraulic conductivity parameters in groundwater models.

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, $\mathbf{h(s)}$ is expanded about the current best estimate of the parameters $\tilde{\mathbf{s}}$

$$\mathbf{h(s)} \approx \mathbf{h}(\tilde{\mathbf{s}}) + \tilde{\mathbf{H}}(\mathbf{s} - \tilde{\mathbf{s}}) \tag{6}$$

where $\tilde{\mathbf{H}}$, as a function of $\tilde{\mathbf{s}}$, is evaluated at each linearization. Linearizations are performed until the change in $\tilde{\mathbf{s}}$ in successive iterations does not change over a specified tolerance. We assign the subscript $_k$ to indicate iteration number, and correct the measurements for the k^{th} linearization as

$$\mathbf{y}_k^{'} = \mathbf{y} - \mathbf{h}(\tilde{\mathbf{s}}_k) + \tilde{\mathbf{H}}_k\tilde{\mathbf{s}}_k. \tag{7}$$

Then the cokriging equations (Equation 5) are updated

$$\begin{bmatrix} \tilde{\mathbf{Q}}_{\mathbf{yy},k} & \tilde{\mathbf{H}}_k\mathbf{X} \\ \mathbf{X}^T\tilde{\mathbf{H}}_k^T & -\mathbf{Q}_{\beta\beta}^{-1} \end{bmatrix} \begin{bmatrix} \xi_k \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \mathbf{y}_k^{'} \\ -\mathbf{Q}_{\beta\beta}^{-1}\beta^* \end{bmatrix} \tag{8}$$

where $\tilde{\mathbf{Q}}_{\mathbf{yy},k} = \tilde{\mathbf{H}}_k\mathbf{Q_{ss}}\tilde{\mathbf{H}}_k^T + \mathbf{R}$. From this set of equations, the next estimate of $\mathbf{s}$ is

$$\tilde{\mathbf{s}}_{k+1} = \mathbf{X}\hat{\beta}_k + \mathbf{Q_{ss}}\tilde{\mathbf{H}}_k^T\xi_k. \tag{9}$$

This can be iterated until there is no difference in the parameter estimates, or when there is no further improvement in the objective function. In many cases, numerical instability makes convergence difficult. A line search is implemented in which a linear a

search is performed between the most recent best estimate of the parameters ($\hat{\mathbf{s}}$) and the current linearization of the parameters ($\widetilde{\mathbf{s}}$), seeking a parameter value that minimizes an objective function. The objective function to minimize is $-\ln p\left(\mathbf{s}|\mathbf{y}\right)$ which is equivalent to maximizing Equation 3

$$\mathcal{L} = -\ln p\left(\mathbf{s}|\mathbf{y}\right) = \left(\mathbf{s} - \mathbf{X}\beta^*\right)^T \mathbf{G}_{\mathbf{ss}}^{-\mathbf{1}} \left(\mathbf{s} - \mathbf{X}\beta^*\right) + \left(\mathbf{y} - \mathbf{h}(\mathbf{s})\right)^T \mathbf{R}^{-\mathbf{1}} \left(\mathbf{y} - \mathbf{h}(\mathbf{s})\right) \tag{10}$$

The linesearch optimizes a single parameter, ρ, along a linear dimension between $\hat{\mathbf{s}}$ and $\widetilde{\mathbf{s}}$ as

$$\mathbf{s}_{opt} = \hat{\mathbf{s}}\rho + \widetilde{\mathbf{s}}\left(1 - \rho\right) \tag{11}$$

where $\mathbf{s}_{opt}$ minimizes the objective function, $\mathcal{L}$, using a Nelder-Mead simplex (see, e.g., *Press et al.*, 1992), which guarantees monotonic decrease in $\mathcal{L}$ over successive iterations. It is recommended to limit the number of linesearch iterations to a relatively low number, as the goal of handling weak linearity is balanced against the computations required to perform the linesearch.

Empirical Bayes

We adopt an empirical Bayes (*Robbins*; *Casella*, 1985) approach to inference in this method, meaning that the general characteristics of the prior and epistemic covariances introduced above are provided in the model setup, but the values of structural parameters that control the balance between smoothness and misfit are estimated from the data. In equation 3 the structural parameter for $L\left(\mathbf{y}|\mathbf{s}\right)$ is the epistemic uncertainty parameter (σ_R^2) and the structural parameters for $p\left(\mathbf{s}\right)$ are covariance parameters (θ). Estimating the structural parameters (σ_R^2 and θ) from the data is performed using Restricted Maximum Likelihood. This approach is consistent with the principle of maximum entropy such that the smoothest solution is chosen, based on the structural parameters estimated from the data. For a discussion of subtle formal differences from minimum relative entropy, see *Rubin* (2003, p. 333-342).

An extension to this approach is the inclusion of information about the prior mean (*Nowak and Cirpka*, 2004). The mean is estimated in the solution, but a prior value and covariance can be supplied to constrain the estimate. Typically, a relatively high covariance magnitude is used so the information about the mean is "diffuse" and princi-

pally serves the role of providing numerical stability rather than compelling the solution to adhere closely to prior values.

The forward model is constructed to provide outputs of values collocated in space and time with the measured observations. The likelihood function brings the observed data into the calculation by quantifying the difference (misfit) between the model forecasts and collocated observations. In all modeling, we must acknowledge that perfect correspondence between forecasts and observations is neither attainable nor desirable. The observations themselves are corrupted by measurement errors and there is usually a lack of perfect correspondence between the exact nature of the measurements and the simulated counterparts. This corruption is due to uncertainty from sources including the paucity of observations, imperfections in the conceptual model, and approximations made to codify the physics of the phenomena into a numerical model framework. All of these sources of uncertainty are described by the term "epistemic uncertainty" (*Rubin*, 2003, p. 4). This epistemic uncertainty characterizes the expected misfit between forecasts and observations, and is expressed through a covariance function. As a result, the likelihood function can be characterized by a Gaussian distribution with zero mean and covariance defined by the epistemic uncertainty.

With both the prior pdf and likelihood function expressed as Gaussian distributions, the posterior pdf (discussed below) is also Gaussian. The values of the parameters $\mathbf{s}$ that result in the maximum value of the posterior pdf are therefore the most likely solution on a point-by-point basis. The solution as a whole is always a somewhat smoothed version of reality, but the small-scale variability can be characterized through conditional realizations. The balance between the strength of smoothing and the closeness of correspondence between forecasts and observations is found through calculation of optimal values for the structural parameters, including a value to quantify the epistemic uncertainty.

Beta Associations

In an idealized problem, a single covariance model is flexible enough to encompass the entire variability of the hydraulic parameters. However, in many hydrogeologic applications, lithologic contacts and unconformities can create discontinuities in parameter values that a single covariance model cannot characterize. Partitioning the field based on either the data (e. g. *Fienen et al.*, 2004) or through interrogation of preliminary

solutions (e. g. *Fienen et al.*, 2008) can greatly improve results. This partitioning is implemented by imposing discontinuities in the stochastic field that censor correlation among all cells that do not occur in the same partition. In this context, "stochastic" refers to the entity being partitioned (namely the correlation structure of the parameter field) but we emphasize here that the locations of the imposed discontinuities are themselves considered deterministic and certain. This partitioning is consistent with zonal boundaries in models made up of homogeneous zones. Furthermore, multiple types of parameters (for example, hydraulic conductivity and recharge) may be estimated. Similar censoring of correlation among difference types of parameters is also necessary in most applications through partitioning.

For hydrogeologic applications, we adopted the term "facies association" from the facies architecture field to describe these partitions (*Fienen et al.*, 2009). The term "facies association" typically refers to descriptive properties of a subset of a medium in the field or at least for a specific project. "Architectural elements" is used in the broader case where the characteristics are more formally defined, (see *Collinson* (1969); *Walker* (1984, 1992); *Swift et al.* (2003)). It is appropriate to use the less restrictive and less transferable term "facies association" in this work because when we subdivide the correlation structure of the medium, we often base the stochastic discontinuities (bounding surfaces, or contacts) on perceived hydraulic properties. These properties will often coincide with differences in age, provenance, or depositional environment, but such coincidence is not required for or by their use. In all cases, partitioning into facies associations must be based on readily observable hydrologic or lithologic attributes.

For bgaPEST to be a more general tool (not limited to hydrogeologic modeling), we have revised this term to be "beta association." As shown in Equation 4, the Greek letter β stands for the mean of a region of distributed parameters. Beta associations can, therefore, delineate regions of a distributed parameter field that have similar statistical properties and correspond to the same mean value. However, importantly, beta associations can also refer to completely different parameter types (for example, hydraulic conductivity and recharge).

To clarify our terminology in this work, partitions delineated by stochastic discontinuity within a distributed parameter field are referred to as "beta associations," whereas zones of piecewise continuity are referred to as "homogeneous zones." The beta associations delineate sub-regions of the model domain that share correlation characteristics and are uncorrelated from neighboring beta associations; they are usually delineated by features

that are easily identified in measured data or geologic conceptualizations of a given site area. In beta associations, variability of parameter values within each cell is allowed and constrained by the *a priori* covariance structure, whereas in homogeneous zones, a single parameter value represents the property for the entire zone. Beta associations also delineate regions in the model (whether defined by one or more parameter values) that are of different types.

Beta Associations and Zones: ***Why aren't "beta associations" just called "zones"?*** Beta associations are a term specific to bgaPEST. As discussed in the main text, this term evolved from the term "facies association," which describes partitioning of parameter fields based on hydrogeologic characteristics. This term was used in place of "zones" because of a long history of zones referring to regions of piecewise homogeneity. Beta associations are *not* homogeneous, so a distinct term was sought that described the characteristic of regions partitioned due to their correlation characteristics. To be more general than only hydrogeologic applications, and also to account for the fact the distinct parameter types require distinct partitions, "beta associations" was the term chosen. Each of these partitions has a distinct prior mean value (β) to be estimated, so the associations are made based on different β values resulting in the term "beta associations."

Parameter Field Anisotropy

In distributed parameter fields, such as hydraulic conductivity in groundwater models, it is common to encounter anisotropy along an axis that may or not be aligned with the coordinate (x, y, z) axes. bgaPEST allows the definition of anisotropy in a horizontal plane at any angle from the x-axis and also in the vertical direction. The general layout of horizontal anisotropy is illustrated in Figure 2. The angle from the x-axis (specified in degrees) is defined by `horiz_angle` and the amount of anisotropy is defined by `horiz_ratio`. `p_max` refers to the direction with maximum parameter values and `p_min` refers to the direction of minimum parameter values. The ratio is used to adjust the effective distance (and thereby the covariance values) along that principal direction. The user supplies values for `horiz_angle` and `horiz_ratio` for each Beta association. If some Beta associations are not meant to exhibit anisotropy, the user may simply set `horiz_ratio=1.0`. If none of the Beta associations exhibit anisotropy, the entire `parameter_anisotropy` block can be eliminated by setting the algorithmic control variable `par_anisotropy=0` which means the block, if present, is ignored.

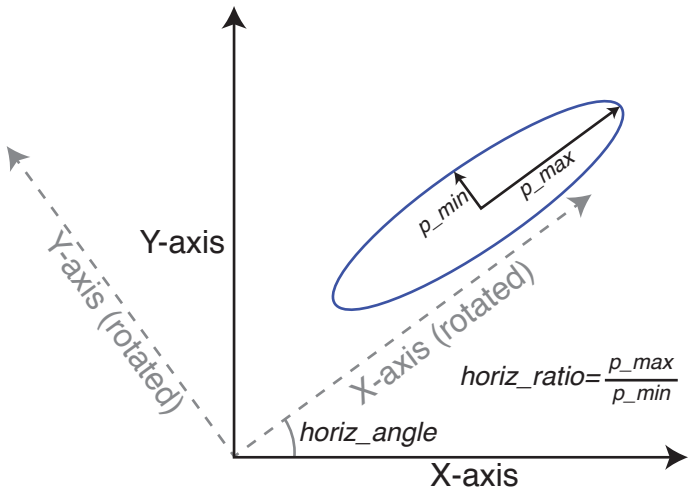

Figure 2: Schematic diagram of horizontal anisotropy defining conventions for bgaPEST.

Anisotropy is introduced in the calculation of distances that are, then, used in the calculation of the prior covariance matrix $\mathbf{Q_{ss}}$ discussed below. For every pair of points, they must first be rotated into the principal direction orientation. This is accomplished using a rotation matrix:

$$\begin{bmatrix} x_{rot,i} \\ y_{rot,i} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{12}$$

where i indicates the i^{th} point of the pair ($i = 1, 2$), θ is the angle (in degrees) specified by `horiz_angle`, x and y are the point coordinates in the original coordinate system, and x_{rot} and y_{rot} are the location projected into the coordinate system corresponding to the orientation of horizontal anisotropy.

Once this projection is made, the horizontal distance is calculated as

$$distance = \sqrt{(x_{rot,1} - x_{rot,2})^2 + \texttt{horiz\_ratio} \times (y_{rot,1} - y_{rot,2})^2} \tag{13}$$

For three-dimensional parameter fields, a second anisotropy ratio may be specified as `vertical_ratio`. In the vertical direction, no angle is specified, so the rotation step is

not required and distance is calculated as

$$distance = \sqrt{\begin{array}{c}(x_{rot,1} - x_{rot,2})^2 \\ +\texttt{horiz\_ratio}\times (y_{rot,1} - y_{rot,2})^2 \\ +\texttt{vertical\_ratio}\times (z_{rot,1} - z_{rot,2})^2\end{array}} \tag{14}$$

Prior Probability Density Function

The prior pdf of $\mathbf{s}$ can be characterized as multi-Gaussian through its mean and covariance. The $(m \times 1)$ unknown parameter vector, $\mathbf{s}$, is modeled as a random process with mean

$$E[\mathbf{s}] = \mathbf{X}\beta \tag{15}$$

where $E[\cdot]$ indicates expected value, m is the number of parameters, β is a $(p \times 1)$ vector of drift coefficients, and $\mathbf{X}$ is an $(m \times p)$ matrix of base functions. In the absence of prior drift, the β are the zone mean values and $\mathbf{X}$ is a selection matrix mapping each value in $\mathbf{s}$ and β into their appropriate zones. $\mathbf{X}$ contains all zeros except for $\mathbf{X}_{ij}^{th}$ element, which maps the i^{th} parameter to the j^{th} zone, that contains the value of one. Zones within distributed parameter domains have been critical to success in hydrogeologic settings which include strong contrasts in parameter values indicative of geologic contacts [see e.g. *Fienen et al.*, 2004 and *Fienen et al.*, 2008]. Prior drift is accounted for in $\mathbf{X}$ through trends expressed in the nonzero terms, although this is not currently implemented in bgaPEST.

The prior covariance ($\mathbf{Q_{ss}}$) of $\mathbf{s}$ for a known β is

$$\mathbf{Q_{ss}}(\theta) = E[(\mathbf{s}-\mathbf{X}\beta)(\mathbf{s}-\mathbf{X}\beta)^T] \tag{16}$$

where $\mathbf{Q_{ss}}$ is a covariance function with structural parameters θ. In this work, we use the exponential covariance function

$$\mathbf{R}(\mathbf{d}) = \sigma^2 \exp(-\frac{|\mathbf{d}|}{\ell}) \tag{17}$$

where $|\mathbf{d}|$ is separation distance, σ^2 is variance, and ℓ is integral scale. If the integral scale is set such that $\ell > \max(|\mathbf{h}|)$ we can substitute $\sigma^2 = \theta\ell$ and restate Equation 17

as

$$\mathbf{Q_{ss}}(\mathbf{h},\theta) = \theta\ell\exp(-\frac{|\mathbf{d}|}{\ell}) \tag{18}$$

We can also set $\ell = 10 \times \max(|\mathbf{d}|)$ so the behavior of the covariance function will be as a linear variogram (*Fienen et al.*, 2008) which enforces continuity at a scale determined by the single free structural parameter θ. The motivation for this covariance function choice is to impart minimal assumptions about parameter structure onto the solution. The appropriate value of θ is calculated through restricted maximum likelihood. For the remainder of this derivation, θ is assumed known. In bgaPEST, as discussed below in the input instructions, either the exponential or linear variogram models may be used.

Assembling the mean and covariance, the prior pdf is

$$p(\mathbf{s}|\beta) \propto \exp\left[-\frac{1}{2}(\mathbf{s}-\mathbf{X}\beta)^{\mathrm{T}}\mathbf{Q_{ss}^{-1}}(\mathbf{s}-\mathbf{X}\beta)\right] \tag{19}$$

In the case of no knowledge about the prior mean, the prior pdf of β can be modeled as uniform over all space as $p(\beta) \propto \mathbf{1}$ and both $\mathbf{s}$ and β are estimated together, so that the conditional distribution in Equation 19 is replaced by a joint distribution

$$p(\mathbf{s},\beta) \propto \exp\left[-\frac{1}{2}(\mathbf{s}-\mathbf{X}\beta)^{\mathrm{T}}\mathbf{Q_{ss}^{-1}}(\mathbf{s}-\mathbf{X}\beta)\right]. \tag{20}$$

Frequently, at least diffuse knowledge about the prior mean is available and can be modeled as multi-Gaussian with mean β^* and covariance $\mathbf{Q}_{\beta\beta}$. Typically, $\mathbf{Q}_{\beta\beta}$ is modeled as a diagonal matrix with variance values on the diagonal indicating independence among the β^*. Incorporating the prior information yields a prior pdf for $\mathbf{s}$

$$p(\mathbf{s}) \propto \exp\left[-\frac{1}{2}(\mathbf{s}-\mathbf{X}\beta^*)^{\mathrm{T}}\mathbf{G_{ss}^{-1}}(\mathbf{s}-\mathbf{X}\beta^*)\right] \tag{21}$$

where $\mathbf{X}\beta^*$ is the prior mean, and $\mathbf{G_{ss}} = \mathbf{Q_{ss}} + \mathbf{X}\mathbf{Q}_{\beta\beta}\mathbf{X^T}$ is the prior covariance (*Nowak and Cirpka*, 2004).

The incorporation of prior mean information, even assuming very high variance values in $\mathbf{Q}_{\beta\beta}$, can provide numerical stability without overly biasing the results.

Prior Covariance Matrix Storage Issues

In underdetermined problems suitable for bgaPEST, the number of parameters can be very large. The prior covariance matrix discussed above can, therefore, grow to such large dimensions that it cannot be practically stored in computer memory. However, two techniques are provided to alleviate some of this storage stress: compression and Toeplitz transformation.

Compression takes advantage of the fact that values in the $\mathbf{G_{ss}}$ matrix relating parameters in different beta associations, by definition, have the value of zero. As a result, a general $\mathbf{G_{ss}}$ matrix can be viewed as a partitioned matrix of nonzero blocks ($\mathbf{G_{ss},}_{\beta i}$) and zero blocks

$$\mathbf{G_{ss}} = \begin{bmatrix} \mathbf{G_{ss},}_{\beta 1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G_{ss},}_{\beta 1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G_{ss},}_{\beta p} \end{bmatrix}. \tag{22}$$

There is no need to store the zero elements provided that accommodations are made to avoid multiplications that involve the zeros. These accommodations have been made in bgaPEST and compression is, therefore, allowed.

In cases where spacial spacing is constant in respective directions, $\mathbf{G_{ss}}$ is a Toeplitz matrix (*Gray*, 2005). A square, symmetric $j \times j$ matrix is Toeplitz in form if it has diagonals that all have the same value as in this example

$$\mathbf{T} = \begin{bmatrix} t_0 & t_1 & \dots & t_{j-2} & t_{j-1} \\ t_1 & t_0 & t_1 & \ddots & t_{j-2} \\ \vdots & t_1 & t_0 & \ddots & \vdots \\ t_{j-2} & \ddots & \ddots & \ddots & t_1 \\ t_{j-1} & t_{j-2} & \dots & t_1 & t_0 \end{bmatrix}. \tag{23}$$

This matrix has the properties that there are only j unique values, and these values occur in a regular order such that only a vector of length j needs to be stored from which individual rows can be constructed to perform matrix multiplication operations. The spacing, as indicated above, must be constant. For example, in a spatial model such as properties in a groundwater model, Δx, Δy, and Δz must be constant, but these values need to be equal to each other. While this is restrictive in the sense that

it implies a regular grid that may not correspond to geometry in the field. However, the regular grid required to take advantage of Toeplitz storage and operations can be assigned to one beta association with a surrounding, irregular grid put in another beta association with fewer parameter values.

In order to use Toeplitz structure in three dimensions, there must be a three-level embedding of Toeplitz matrices (*D'Oria*, 2010). The first level corresponds to the model layers, the second level corresponds to model rows, and the third level corresponds to model columns. Inspection of the schematic example in Fig. 3 shows that every distinct value represented in the entire matrix is found in the first (leftmost) column. Cycling of rows or columns, relative to the single stored vector, can be used to reconstruct any row or column of the main matrix to be used in multiplication operations. In bgaPEST, a combination of Toeplitz and complete blocks can make up the $\mathbf{G_{ss}}$ matrix in Compressed form, as discussed above.

Detailed input instructions for bgaPEST are presented in Appendix A. It is important to note, however, that if Toeplitz compression is invoked, parameters must be listed in the `.bgp` input file in order, sorted first by layer, then by column, and finally by row.

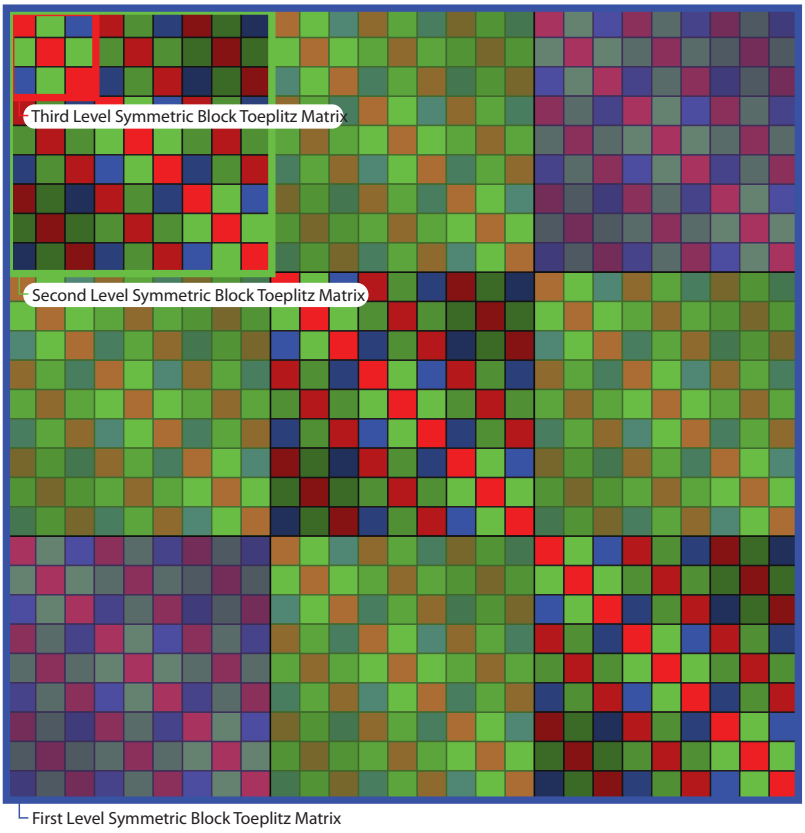

Figure 3: Schematic graphical visualization of a three-level embedded set of Toeplitz blocks in a covariance matrix (modified from *D'Oria* (2010)). The smallest squares represent individual matrix entries (values) and the colors correspond to distinct values. In this synthetic example, it is assumed that there are three rows, three columns, and three layers in the underlying model.

Likelihood Function

The parameters $\mathbf{s}$ are related to observations $\mathbf{y}$ through a measurement equation

$$\mathbf{y} = \mathbf{h}(\mathbf{s}) + \mathbf{v} \tag{24}$$

where $\mathbf{y}$ 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 collocated spatially and temporally with the observation values, and $\mathbf{v}$ is an $(n \times 1)$ vector of epistemic uncertainty terms, modeled as a random process with zero mean

and covariance matrix $\mathbf{R}$. 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 independent and uncorrelated so

$$\mathbf{R} = \sigma_{\mathbf{R}}^{\mathbf{2}}\mathbf{W} \tag{25}$$

where σ_R^2 is the epistemic uncertainty parameter and $\mathbf{W}$ 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 i^{th} 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 $\mathbf{R}$ are rarely available. If information about $\mathbf{R}$ is available, however, it could be included and Equation 25 could be replaced by a more complicated matrix. Proceeding with Equation 25 the likelihood function, assumed multi-Gaussian, is

$$L(\mathbf{y}|\mathbf{s}) \propto \exp\left[-\frac{1}{2}\left(\mathbf{y} - \mathbf{h}(\mathbf{s})\right)^{\mathrm{T}}\mathbf{R}^{-\mathbf{1}}\left(\mathbf{y} - \mathbf{h}(\mathbf{s})\right)\right]. \tag{26}$$

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

Posterior Probability Density Function

Applying Bayes' theorem with the product of Equations 21 and 26 yields the posterior pdf

$$p(\mathbf{s}|\mathbf{y}) \propto \exp\left[-\frac{1}{2}(\mathbf{s} - \mathbf{X}\beta^*)^{\mathrm{T}}\mathbf{G}_{\mathbf{ss}}^{-\mathbf{1}}(\mathbf{s} - \mathbf{X}\beta^*) - \frac{1}{2}\left(\mathbf{y} - \mathbf{h}(\mathbf{s})\right)^{\mathrm{T}}\mathbf{R}^{-\mathbf{1}}\left(\mathbf{y} - \mathbf{h}(\mathbf{s})\right)\right] \tag{27}$$

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

$$\hat{\mathbf{s}} = \mathbf{X}\hat{\beta} + \mathbf{Q}_{\mathbf{ss}}\mathbf{H}^{\mathbf{T}}\xi \tag{28}$$

which is the superposition of the prior mean (first term) and an innovation term which considers deviations of the model outputs from the observations (second term). $\mathbf{H}$ 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 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

$$\begin{bmatrix} \mathbf{Q_{yy}} & \mathbf{HX} \\ \mathbf{X}^T\mathbf{H}^T & -\mathbf{Q}_{\beta\beta}^{-\mathbf{1}} \end{bmatrix} \begin{bmatrix} \xi \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ -\mathbf{Q}_{\beta\beta}^{-\mathbf{1}}\beta^* \end{bmatrix} \tag{29}$$

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 we expand the solution in a first-order Taylor expansion, resulting in an updated set of cokriging equations from Equation 29

$$\begin{bmatrix} \mathbf{Q_{yy}} & \mathbf{HX} \\ \mathbf{X}^T\mathbf{H}^T & -\mathbf{Q}_{\beta\beta}^{-\mathbf{1}} \end{bmatrix} \begin{bmatrix} \xi \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{y} - \mathbf{h}\left(\tilde{\mathbf{s}}\right) + \mathbf{H}\tilde{\mathbf{s}} \\ -\mathbf{Q}_{\beta\beta}^{-\mathbf{1}}\beta^* \end{bmatrix} \tag{30}$$

At each iteration (later referred to as inner iterations), the system in Equation 30 is solved, resulting in an updated estimate of $\hat{\mathbf{s}}$ calculated through Equation 28. At each iteration, the objective function, based on minimizing the negative logarithm of the posterior pdf (Equation 27) is evaluated 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

$$\mathbf{\Phi}_T = \mathbf{\Phi}_M + \mathbf{\Phi}_R \tag{31}$$

where $\mathbf{\Phi}_T$ is the total objective function, $\mathbf{\Phi}_M$ is the misfit objective function (also corresponding to the likelihood function) and $\mathbf{\Phi}_R$ is the regularization objective function

(also corresponding to the prior pdf). The components of Equation 31 are

$$\mathbf{\Phi}_M = (\mathbf{y} - \mathbf{h}(\hat{\mathbf{s}}))^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{h}(\hat{\mathbf{s}})) \tag{32}$$

and

$$\mathbf{\Phi}_R = (\hat{\mathbf{s}} - \mathbf{X}\beta^*)^T \mathbf{G}_{\mathbf{ss}}^{-1} (\hat{\mathbf{s}} - \mathbf{X}\beta^*) \tag{33}$$

where the $\frac{1}{2}$ terms are absorbed into the constant of proportionality and 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 explained here. First, the prior covariance matrix $\mathbf{Q}_{\mathbf{ss}}$ is censored by assigning a value of zero to each element which characterizes covariance between cells of different zones. It is not required that the covariance model be the same for each zone. If different covariance models are used for different zones, this is reflected in the appropriate parts of $\mathbf{Q}_{\mathbf{ss}}$. Furthermore, in some applications, a single structural parameter, θ, may be estimated and applied to all of $\mathbf{Q}_{\mathbf{ss}}$. 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 15) is determined as explained in above. In cases where the mean of each zone is completely unknown, no values for β^* are provided, but the $\mathbf{X}$ 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 this work, structural parameters may include the epistemic uncertainty term in

Equation 25 (σ_R^2) and the prior PDF variogram parameters in Equation 18 (θ). These parameters are estimated using Restricted Maximum Likelihood consistent with the approaches of *Kitanidis and Vomvoris* (1983), *Kitanidis* (1995) and *Li et al.* (2007).

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

$$p\left(\theta|\mathbf{y_k'}\right) \propto L\left(\mathbf{y}_k'|\theta\right) p\left(\theta\right) \tag{34}$$

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

$$L\left(\mathbf{y}_k'|\theta\right) \propto \det\left(\mathbf{G_{yy}}\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\mathbf{y}_k' - \mathbf{HX}\beta^*\right)^T \mathbf{G_{yy}^{-1}}\left(\mathbf{y}_k' - \mathbf{HX}\beta^*\right)\right] \tag{35}$$

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

$$\mathbf{G_{yy}} = \mathbf{Q_{yy}} + \mathbf{HXQ}_{\beta\beta}\mathbf{X}^T\mathbf{H}^T. \tag{36}$$

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

$$p\left(\theta\right) \propto \det\left(\mathbf{Q}_{\theta\theta}\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\theta - \theta^*\right)^T \mathbf{Q}_{\theta\theta}^{-1}\left(\theta - \theta^*\right)\right] \tag{37}$$

The posterior pdf is the product of Equations 37 and 35

$$p\left(\theta|\mathbf{y_k'}\right) \propto \det\left(\mathbf{Q}_{\theta\theta}\right)^{-\frac{1}{2}} \det\left(\mathbf{G_{yy}}\right)^{-\frac{1}{2}} \exp\left[\begin{array}{c} -\frac{1}{2}\left(\theta - \theta^*\right)^T \mathbf{Q}_{\theta\theta}^{-1}\left(\theta - \theta^*\right) \\ -\frac{1}{2}\left(\mathbf{y}_k' - \mathbf{HX}\beta^*\right)^T \mathbf{G_{yy}^{-1}}\left(\mathbf{y}_k' - \mathbf{HX}\beta^*\right) \end{array}\right]. \tag{38}$$

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

$$\mathbf{\Phi}_S = \ln\left(\det\left(\mathbf{G_{yy}}\right)\right) + \left[\left(\theta - \theta^*\right)^T \mathbf{Q}_{\theta\theta}^{-1}\left(\theta - \theta^*\right) + \left(\mathbf{y}_k' - \mathbf{HX}\beta^*\right)^T \mathbf{G_{yy}^{-1}}\left(\mathbf{y}_k' - \mathbf{HX}\beta^*\right)\right] \tag{39}$$

where unchanging quantities are absorbed into the constant of proportionality including $\det\left(\mathbf{Q}_{\theta\theta}\right)^{-\frac{1}{2}}$ and the $-\frac{1}{2}$ terms. The optimal values for θ are found using the Nelder-Mead simplex algorithm (e.g. *Press et al.*, 1992, p. 408-410). Non-negativity in the θ parameters is enforced using a power transformation (*Box and Cox*, 1964). As indicated

by *Kitanidis* (1995), nonlinearity requires that structural parameters are estimated iteratively with the estimation of hydraulic parameters. This is accomplished through a sequence of coupled inversion as follows.

1. Initialize hydraulic parameters as ($\mathbf{s_0}$) and structural parameters ($\theta_\mathbf{0}$).
2. Solve for a new estimate of hydraulic 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.

Posterior Covariance

The posterior covariance can be calculated 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:

$$\mathbf{V} = \mathbf{G_{ss}} - \mathbf{G_{sy}}\mathbf{G_{yy}^{-1}}\mathbf{G_{sy}^{T}} \tag{40}$$

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 $\mathbf{Q}$ is not used, the full matrix $\mathbf{V}$ is calculated and reported. Where compression of $\mathbf{Q}$ 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 also used to calculated posterior 95% confidence intervals. The full matrix, when reported, can be used to calculate conditional realizations (*Kitanidis*, 1995, 1996).

Running bgaPEST

The bgaPEST program uses a single input file and generates several output files. These files are discussed in the context of progression of the bgaPEST program in the remainder of this section. Figure 4 shows the general progression of a bgaPEST parameter estimation run. The entire process is controlled by variables in the input `.bgp` file discussed below.

Outer iterations (also called bga iterations) are wrapped around the entire parameter estimation process with the exception of calculating posterior covariance. Inner iterations are performed to account for the restricted maximum likelihood estimation of structural parameters. If structural parameters are not chosen to be estimated, then a single outer iteration is performed using the initial values of structural parameters and inner iterations are performed until convergence or until the number of iterations reaches `it_max_phi`. If a linesearch is requested, this is performed within the inner iterations. If structural parameter optimization is requested, it is performed after convergence or maximum inner iterations are reached. Then, restricted maximum likelihood is performed to estimate a new set of structural parameters. The interdependence between structural parameters and model parameters requires reiteration of the inner iterations and structural parameter estimation until both have converged or the maximum number of outer iterations has been reached.

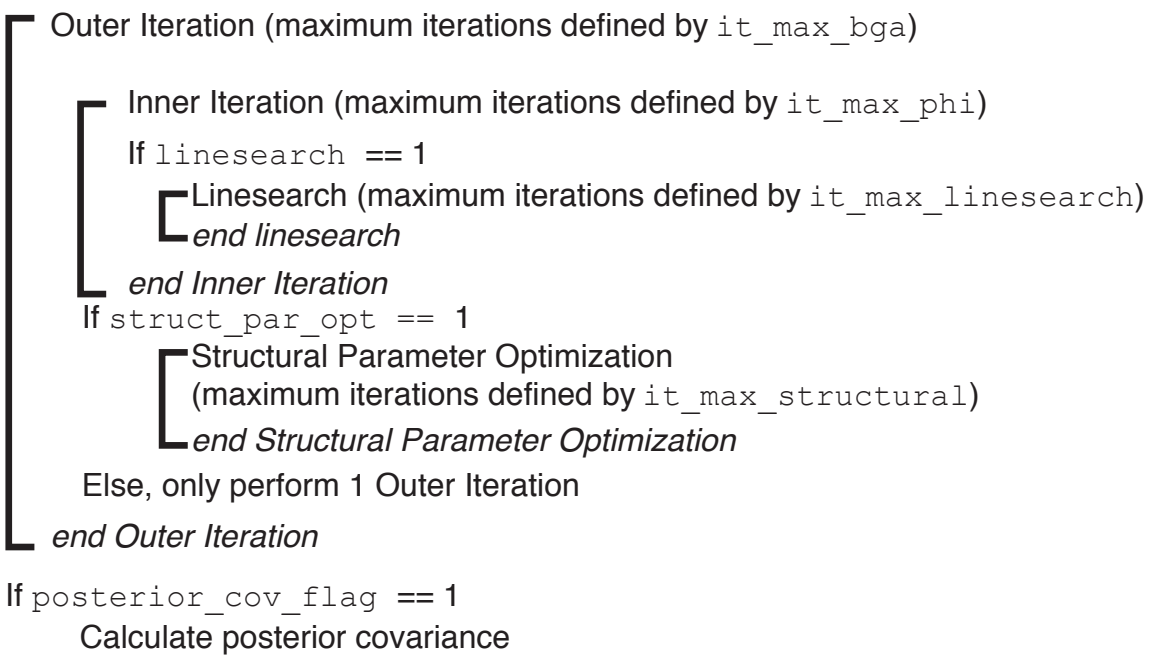

Figure 4: Abbreviated flowchart showing the progression of the major bgaPEST procedures.

Control Variables

Two types of variables are used in bgaPEST: control variables and data variables. Data variables are values such as model parameters, observations, file names, and other data that are needed by the bgaPEST program. Control variables, on the other hand, are in a sense abstracted one level from data and are meta-parameters that impact either the reading/writing of data or control the progression of the algorithm. Many control variables are straightforward (for example, `beta_cov_form`, an integer, is the form of the prior mean (β) covariance matrix ($\mathbf{Q}_{\beta\beta}$). [0] = none, [1] = diagonal, [2] = full matrix. *default = 0).* Such variables are defined in the context of the input instructions of in Appendix A. However, other control variables are accompanied by important conventions regarding their impact on the performance of the algorithm. For these cases, in this section, more detail is provided about certain control variables.

`it_max_bga` *integer* Total number of outer iterations allowed for the entire algorithm. *default=10*

`it_max_phi` *integer* Total number of inner iterations allowed in each quasi-linear estimation optimization. *default=10*

`it_max_structural` *integer* Total number of iterations allowed in structural parameter optimization. *default=10*

`structural_conv` *float* Convergence criterion for structural parameter convergence. Positive or negative values can be used to trigger two different measures of convergence, as noted below. In either case, however, convergence is compared to the absolute value of `structural_conv`.

If positive, convergence is based on the absolute difference in structural parameter objective function over consecutive iterations.

$$\text{conv} = \text{abs}\left(\mathbf{\Phi}_{S,i} - \mathbf{\Phi}_{S,i-1}\right) \tag{41}$$

where i is the current structural parameter optimization iteration, $i-1$ is the previous structural parameter optimization iteration, and $\mathbf{\Phi}_S$ is the structural parameter objective function.

If negative, convergence is based on the norm of the difference between consecutive

structural parameter values.

$$\text{conv} = \sqrt{(\theta_i - \theta_{i-1})^T (\theta_i - \theta_{i-1})} \tag{42}$$

where i and $i-1$ are as defined above, and θ is a vector containing all structural parameters currently being estimated (may include epistemic uncertainty, if requested). *default=0.001*

`phi_conv` *float* Convergence criterion for objective function convergence. The convergence at each inner iteration is evaluated as the absolute difference from one inner iteration to the next. This is evaluated as

$$\text{conv} = \text{abs}\left(\mathbf{\Phi}_{T,i_{in}} - \mathbf{\Phi}_{T,i_{in}-1}\right) \tag{43}$$

where i_{in} is the current inner iteration, and $\mathbf{\Phi}_T$ is the total objective function (Equation 31). *default=0.001*

`bga_conv` *float* Convergence criterion for objective function outer iterations. The convergence at each outer iteration is evaluated as the absolute difference from one outer iteration to the next. This accounts for both convergence of $\mathbf{\Phi}_T$ and $\mathbf{\Phi}_S$. The convergence is evaluated as

$$\text{conv} = \text{abs}\left(\mathbf{\Phi}_{T,i_{out}} - \mathbf{\Phi}_{T,i_{out}-1}\right) \tag{44}$$

where i_{out} is the outer iteration, and $\mathbf{\Phi}_T$ is the total objective function (Equation 31). *default*=10×`phi_conv`

`it_max_linesearch` *integer* Total number of outer iterations allowed for the line search. *default=4*

`Q_compression_flag` *integer* Flag to determine how to calculate Q_0 [0] = no compression– calculate full Q_0 matrix, [1] = Calculate separate Q_0 matrix for each beta association. In addition to controlling the behavior of prior covariance compression, this flag also determineds whether a full posterior covariance matrix or only the diagonal are calculated. *default=0*

`posterior_cov_flag` *integer* Flag to determine whether posterior covariance matrix should be calculated. [0] = do not calculate posterior covariance matrix, [1] =

calculate posterior covariance. If `Q_compression_flag=1`, only the diagonal of the posterior covariance matrix is calculated. If `posterior_cov_flag=0` then 95% confidence intervals are not calculated and the output file `<casename>.bpp.fin` discussed below does not include confidence intervals.

Input Files

The bgaPEST program is run from the command line by typing `bgaPEST.exe <casename>.bgp` where `<casename>` is a filename containing input instructions. Detailed input instruction are in Appendix A.

Output Files

Several output files are generated throughout the progression of a single bgaPEST run. These files are summarized in this section.

bgaPEST Record File `<casename>.bpr`

The main output file for bgaPEST is called `<casename>.bpr`. Initial values of bgaPEST input are repeated to form a record for the bgaPEST run. After each inner iteration, as defined above, the objective function is reported and external files are written that include current parameter values and observation values. After each outer iteration, structural parameter values are also reported for each beat association in which structural parameter estimation was requested and for the epistemic uncertainty term, if requested.

bgaPEST Parameter Value Files `<casename>.bpp.<#Oi>_<#Ii>`

The parameter values are written to files called `<casename>.bpp.<#Oi>_<#Ii>` where `<#Oi>` is the outer iteration number and `<#Ii>` is the inner iteration number. These ascii files are printed in columns with the following headers: `ParamName`; `ParamGroup`; `BetaAssoc`; `ParamVal`. At the beginning of a bgaPEST run, a file `<casename>.bpp.0` is written in the same format to record the initial parameter values used. This is done to avoid cluttering the `<casename>.bpr` file with what is often a very long list

of parameters and their values. Parameter values that were subjected to logarithmic transformation are reported in their linear space, *not* log-transformed space.

Another special case of parameter value files is written at the end of a bgaPEST run and called `<casename>.bpp.fin`. This file contains the final parameter values estimated as optimal by bgaPEST. Furthermore, if posterior covariance calculation was requested, two additional columns are added: `95pctLCL` and `95pctUCL` which are the 95% lower and upper confidence limits, respectively. These confidence limits are obtained by applying the subtraction and addition, respectively, of $2 \times \sqrt{\mathbf{V}_{ii}}$ to $\mathbf{s}_i$–the i^{th} optimal parameter value. In this case, $\mathbf{V}$ is the posterior covariance, so $\sqrt{\mathbf{V}_{ii}}$ is the standard deviation of the i^{th} parameter. The 95% confidence limits are reported in linear space, *not* log-transformed space, so for log-transformed parameters, the upper and lower 95% percent confidence limits are not symmetrical about the parameter value.

bgaPEST Observation Value Files `<casename>.bre.<#Oi>_<#Ii>`

The observation values obtained by running the forward model with the currently estimated parameters are written to files called `<casename>.bre.<#Oi>_<#Ii>` following a similar convention as with the `<casename>.bpp` files above. The ascii files are printed with the following headers: `ObsName`; `ObsGroup`; `Modeled`; `Measured`. These files may be easily copied into a spreadsheet or read with a plotting program to calculate and plot residuals.

Posterior Covariance File `<casename>.post.cov`

If the input variable `posterior_cov_flag=1`, then posterior covariance of the parameters $\mathbf{s}$ is calculated. In addition to this information being used to report 95% confidence limits as described above, the posterior covariance matrix is also written to the file `<casename>.post.cov`. If the variable `Q_compression_flag=1`, then compression is used for saving the prior covariance matrix. This is done when a large number of parameters is used and, thus, the full covariance matrices are unwieldy. Based on this choice, the posterior covariance is either reported as the diagonal of the posterior covariance matrix ($diag\,(\mathbf{V})$) if `Q_compression_flag=1` or the full covariance matrix $\mathbf{V}$ if `Q_compression_flag=0`. The output formats are discussed at the end of Appendix A.

Posterior Covariance and Parameter log-Transformations In this section, it was indicated that in the `<casename>.bpp.fin` file, parameter values and 95% confidence intervals are reported in linear (untransformed) space while in the `<casename>.post.cov` file, posterior covariance values are reported in estimation (log-transformed) space. Why the difference? The two files are serving slightly different purposes. The parameter output file presents values in the units they are entered in (and, persumably, the units "seen" by the forward model). As a result, 95% confidence intervals are reported in the same way. Furthermore, the addition and subtraction of $2 \times \sqrt{\mathbf{V}_{ii}}$ must be applied to the parameters before back-transformation, which explains the asymmetry of the confidence limits. On the other hand, the full posterior covariance matrix is intended for other analysis (propagation of variance through to predictions, conditional realizations, and others) in which the information should be retained in estimation (log-transformed) space. In the end, the decision of how to report these values is one of convention, and this side box is intended to make clear which was chosen in each case.

A Input Instructions

In this appendix, the general strategy for input instructions is described. The input is arranged in a file called `<casename>.bgp`, which is made up of input blocks, as discussed below. Following a discussion of more detail of the general input protocols, subsections in which specific input blocks are discussed including variables and data that can be input in them are presented.

General Structure of Input

The general input structure is designed on a subset of the JUPITER protocol (*Banta et al.*, 2006). The advantage of this protocol over XML or the previous input format for PEST is that annotations that are easily read by humans are part of the input protocol. The full JUPITER protocol, however, has memory and computational overhead that can become a problem for large and complicated data sets. The protocol used here, therefore, is simplified but should be easily recognizable to users of other JUPITER-compatible programs.

The strategy for input is designed to use `BLOCKS` that are made up of either `KEYWORDS`

for individual variables or TABLES for a series of data. The specification of whether a given block uses KEYWORDS or TABLES is preordained and the input blocks defined below indicate which is required.

Blocks

Input Blocks are allowed to take one of two forms: either KEYWORDS or TABLES. All input blocks are delineated by the words BEGIN and END. The header line also includes the name and type of the block and the final line contains the name of the block. For example:

```
BEGIN prior_mean_cv KEYWORDS
prior_betas=1
beta_cov_form = 0
END prior_mean_cv
```

Keywords

Keyword variables correspond to single values identifies with an "=" sign. Multiple KEYWORDS can be entered on each line in an input file but no spaces are allowed in KEYWORDS names or variable values. An example is: `prior_betas=1.`

Tables

Table variables are used for tabular data series that have multiple values in categories. Tables are identified by listing the number of rows (nrow), number of columns (ncol), and providing the keyword columnlabels. This is followed by nrow rows of data, with values arranged in ncol columns, corresponding to the same order as the columnlabels, and delimited by one or more spaces. For example:

```
BEGIN Q_compression_cv TABLE
nrow=2 ncol=5 columnlabels
BetaAssoc Toep_flag Nrow Ncol Nlay
1 0 21 21 1
2 1 21 21 1
END Q_compression_cv
```

Files

A user may want to shorten the length of the main input file by redirecting certain input to external text files. This can be done by signaling an input block with the word `FILES`, to read a file containing the entire set of information for the block. Regardless of whether the external text file contains a `KEYWORDS` or `TABLE` block, a block definition must be in place directing the program to the external file. For example:

```
BEGIN Q_compression_cv FILE
compression.txt
END Q_compression_cv
```

In this example, the contents of the file compression.txt would be:

```
BEGIN Q_compression_cv TABLE
nrow=2 ncol=5 columnlabels
BetaAssoc Toep_flag Nrow Ncol Nlay
1 0 21 21 1
2 1 21 21 1
END Q_compression_cv
```

bgaPEST Input Blocks

The specific input blocks used in bgaPEST are discussed, in order of appearance in the `<casename>.bgp` file. For each block, data types are identified either as *float*, *integer*, or *string*. Values entered as *float* can include scientific/engineering notation, but in all cases should contain a "." even if no fractional detail is included. Conversely, *integers* must not contain ".". Variables identified as *string* may not include spaces because whitespace is used as the delimiter for rows in tables and separating keywords.

Each block is also defined with a suffix of "cv" for "control variables," or "data" for data. Control variables are those that govern the behavior of the algorithm as a whole as opposed to data points (such as parameter values, structural parameter values, etc.).

A note on default variable values In the input instructions below, some variables list a default value. Part of the design strategy of this software was to not burden users with determining appropriate values for each and every variable that controls the algorithm. As a result, default values are provided for some variables. In those cases, input by the user in the `.bgp` file is optional. If no value is provided by the user, the default value will be used by bgaPEST. If a variable not listed with a default value in these input instructions is omitted by a user, bgaPEST will return with an error indicating that the variable is not present.

Algorithmic Control Variables `algorithmic_cv` KEYWORDS

The following `KEYWORDS` variables are in the `algorithmic_cv` block.

`structural_conv` *float* Convergence criterion for structural parameter convergence. If positive, convergence is based on the absolute difference in structural parameter objective function over consecutive iterations. If negative, convergence is based on the norm of the difference between consecutive structural parameter values. *default=0.001*

`phi_conv` *float* Convergence criterion for objective function inner iterations. *default=0.001*

`bga_conv` *float* Convergence criterion for objective function outer iterations. *default*=10×`phi_conv`

`it_max_structural` *integer* Total number of iterations allowed in structural parameter optimization. *default=10*

`it_max_phi` *integer* Total number of iterations allowed in each quasi-linear estimation optimization. *default=10*

`it_max_bga` *integer* Total number of outer iterations allowed for the entire algorithm. *default=10*

`linesearch` *integer* Flag to determine whether a line search should be conducted. [0] = do not use linesearch, [1] = use linesearch. *default=0*

`it_max_linesearch` *integer* Total number of outer iterations allowed for the line search. *default=4*

`theta_cov_form` *integer* Form of the theta covariance matrix. [0] = none, [1] = diagonal, [2] = full matrix. [0] means no prior covariance on theta provided and it is assumed totally unknown. *default = 0*

`Q_compression_flag` *integer* Flag to determine how to calculate Q_0 [0] = no compression– calculate full Q_0 matrix, [1] = Calculate separate Q_0 matrix for each beta association. *default=0*

`par_anisotropy` *integer* Flag to determine whether parameter anisotropy should be considered when making the **Q** matrix: [0] = do not consider anisotropy, [1]=consider anisotropy. If anisotropy is considered, a `parameter_anisotropy` block should be considered, as defined below. *default=0.*

`deriv_mode` *integer* Flag to determine whether sensitivities are calculated using an external call to PEST or using a user-supplied program (such as adjoint state). [0] = use PEST, [1] = use external program identified below in the `model_command_lines` block. *default=0*

`posterior_cov_flag` *integer* Flag to determine whether posterior covariance matrix should be calculated. [0] = do not calculate posterior covariance matrix, [1] = calculate posterior covariance matrix. If `Q_compression_flag=1`, only the diagonal of the posterior covariance matrix is calculated. *default=0*

`jacobian_file` *string* Name of the file generated by an external program if deriv _mode=1. *default="scratch.jco"*

`jacobian_format` *string* Format of the file indicated in jacobian_file. [binary] indicates a binary file formatted as a JCO file from PEST, [ascii] indicates a file of a standard PEST matrix format, discussed below in this documentation. *default="binary"*

Prior Mean Control Variables `prior_mean_cv` KEYWORDS

The following `KEYWORDS` variables are in the `algorithmic_cv` block.

`prior_betas` *integer* Flag indicating whether information about prior mean (β) will be supplied. [0] = no, [1] = yes.

`beta_cov_form` *integer* Form of the prior mean (β) covariance matrix ($\mathbf{Q}_{\beta\beta}$). [0] = none, [1] = diagonal, [2] = full matrix. *default = 0*

Prior Mean Data `prior_mean_data` TABLE

This table must contain the same number of rows as there are beta associations to be defined. The rows must be in ascending order of beta association numbers. This is also the block where beta associations are defined, even if prior means are not defined.

`BetaAssoc` *integer* Identifier of each beta association (one per row). These should be sequential integers.

`Partrans` *string* Transformation indicator. Acceptable values are `log` and `none`.

`beta_0` *float* Value of prior mean value (β_0) for the row's beta association.

`beta_cov_#` *float* The number of values provided is based on the value of beta_cov_form specified above:

If beta_cov_form=1, one value is provided.

If beta_cov_form=2, `nrow` values are provided, corresponding to the current row of the beta covariance matrix ($\mathbf{Q}_{\beta\beta}$).

Structural Parameter Control Variables `structural_parameter_cv` TABLE

This table must contain the same number of rows as there are beta associations to be defined. The rows must be in ascending order of beta association numbers.

`BetaAssoc` *integer* Identified of each beta association (one per row). These should be sequential integers.

`prior_cov_mode` *integer* Flag to indicate whether prior covariance of parameters ($\mathbf{Q}_{\mathbf{ss}}$) is supplied or calculated. This is reserved for future use - currently $\mathbf{Q}_{\mathbf{ss}}$ is always calculated, so this value is ignored if present.

`var_type` *integer* This is a flag to indicate which variogram type is used to express the prior covariance ($\mathbf{Q}_{\mathbf{ss}}$). Acceptable choices are [0] = pure nugget, [1] = linear, [2] = exponential. Note that choosing linear is equivalent to choosing exponential with fixed integral scale 10 times the maximum distance between nodes. *default=1*

`struct_par_opt` *integer* Flag for whether structural parameters are meant to be optimized or not. This can be chosen for each structural parameter individually. [0] = do not optimize (hold at initial value), [1] = optimize using a marginal distribution. *default=1*

`trans_theta` *integer* Flag for whether a power transformation should be applied to the structural parameters in the current row. [0] = do not transform, [1] = transform. *default=0*

`alpha_trans` *float* Exponent of the power transformation if `trans_theta` = 1. *default = 50*

Structural Parameter Data `structural_parameter_data` TABLE

This table must contain the same number of rows as there are beta associations to be defined. The rows must be in ascending order of beta association numbers.

`BetaAssoc` *integer* Identified of each beta association (one per row). These should be sequential integers.

`theta_0_1` *float* Initial value of $\theta_{1,0}$ which is the starting value of the first structural parameter for prior covariance.

`theta_0_2` *float* Initial value of $\theta_{2,0}$ which is the starting value of the second structural parameter for prior covariance. If using a linear or nugget variogram, an arbitrary negative value should be entered here.

Structural Parameter Covariance Data `structural_parameter_cov` TABLE

The only covariance model currently supported is diagonal, so there must be one covariance value for each θ parameter.

`theta_cov_1` *float* Variance of the current row's θ parameter. If using an exponential variogram, then a single beta association will have two structural parameters. To handle this possibility, the variance values should be listed, one per line, in the order of beta associations, then in order θ_1 then θ_2 .

Epistemic Error Term `epistemic_error_term` KEYWORDS

`sig_0` *float* Initial value of the epistemic error (σ_o).

`sig_opt` *integer* Flag indicating whether epistemic error should be optimized for or not. [0] = do not optimize, [1] = optimize. If `sig_opt` = 0, then the value of `sig_0` is used throughout the inversion.

`sig_p_var` *float* Prior variance on σ. `sig_p_var=0.` means no prior variance on epistemic error is provided and it is assumed totally unknown. *default=0.*

`trans_sig` *integer* Flag for whether a power transformation should be applied to the epistemic error. [0] = do not transform, [1] = transform. *default=0*

`alpha_trans` *float* Exponent of the power transformation if `trans_sig` = 1. *default = 50*

Parameter Control Values `parameter_cv` KEYWORDS

`ndim` *integer* Number of dimensions over which parameters span.

Prior Covariance Compression Control Variables `Q_compression_cv` TABLE

This table must contain the same number of rows as there are beta associations to be defined. The rows must be in ascending order of beta association numbers.

`BetaAssoc` *integer* Identified of each beta association (one per row). These typically are sequential integers.

`Toep_flag` *integer* This is a flag to determine whether a Toeplitz transformation should be applied to the prior covariance matrix ($\mathbf{Q_{ss}}$) . [0] = do not use Toeplitz transformation, [1] = do use Toeplitz transformation.

`Nrow` *integer* Number of rows in the current Beta association (only read if `Toep_flag = 1`).

`Ncol` *integer* Number of columns in the current Beta association (only read if `Toep_flag = 1`).

`Nlay` *integer* Number of layers in the current Beta association (only read if `Toep_flag = 1`).

Parameter Groups `parameter_groups` TABLE

Each row of this table corresponds to one of the parameter groups. These groups are used to group together parameters by type and are not the same as beta associations.

`groupname` *string* Name of the group in the current row. Note that these cannot contain spaces.

Parameter Data `parameter_data` TABLE

`ParamName` *string* Name for the parameter.

`StartValue` *float* Staring parameter value.

`GroupName` *string* Name of the group to which the parameter belongs. This name must be defined in the `parameter_groups` block.

`BetaAssoc` *integer* Beta association to which this parameter belongs.

`SenMethod` *integer* Senstivity method used for this parameter type. This parameter may now be arbitrary–it is reserved for future use and currently is ignored.

`x1` *float* Location in the first dimension.

`x2` *float* Location in the second dimension. Only read if `ndim >= 2`.

`x3` *float* Location in the third dimension. Only read if `ndim = 3`.

Observation Groups `observation_groups` TABLE

Each row of this table corresponds to one of the observation groups. These groups are used to group together observations by type and are used to report portions of the objective function.

`groupname` *string* Name of the group in the current row. Note that these cannot contain spaces.

Observation Data `observation_data` TABLE

One observation is presented on each line.

`ObsName` *string* Name of an observation.

`ObsValue` *float* Value of the observation.

`GroupName` *string* Name of the group to which the observation belongs. This name must be defined in the `observation_groups` block.

`Weight` *float* A relative weight that gets applied to the epistemic error.

Model Command Lines `model_command_lines` KEYWORDS

Currently a single forward model command and an option derivative model command can be supplied here. These string keywords can include path information if the command line batch files or shell scripts are not located in the current working directory, but spaces are not allowed.

`Command` *string* This is the batch file or shell script that runs the forward model.

`DerivCommand` *string* This is the optional batch file or shell script that is used to calculate derivatives. This is only used if `deriv_method = 1` in the `algorithmic_cv` block.

Model Input Files `model_input_files` TABLE

Each row of this table includes a matched template file and model input file. This allows the program to create the correct input files for the model.

`TemplateFile` *string* Name of a template file for making model input. Must end in `.tpl`.

`ModInFile` *string* Name of the model input file corresponding to the TemplateFile identified on the same row.

Model Output Files `model_output_files` TABLE

Each row of this table includes a matched instruction file and model output file. This allows the program to read the results of model runs correctly.

`InstructionFile` *string* Name of an instruction file for reading model output. Must end in `.ins`.

`ModOutFile` *string* Name of the model output file corresponding to the InstructionFile identified on the same row.

Parameter Anisotropy `parameter_anisotropy` TABLE

Each row of this table contains information for parameter anisotropy for a beta association. This block is read only if the variable `par_anisotropy` =1 in the `algorithmic_cv` block.

`BetaAssoc` *integer* Identifier of a beta association.

`horiz_angle` *float* Angle, in degrees of the principal direction of anisotropy in a horizontal plane. See Figure 2 for details.

`horiz_ratio` *float* Ratio of maximum to minimum principal property values in the horizontal plane. See Figure 2 for details.

`vertical_ratio` *float* Ratio of maximum to minimum principal property values in the vertical direction. See Figure 2 for details.

PEST Matrix Formats for Jacobian and Posterior Covariance

On two occasions in `bgaPEST` a matrix text file format from `PEST` is used to store matrices: when posterior covariance output from `bgaPEST` is specified as a full matrix; and when Jacobian sensitivity matrix information is exchanged from an external code with `bgaPEST`.

The posterior covariance matrix may take two forms: a full matrix; or a diagonal matrix. These options are discussed below. Two options are available for Jacobian sensitivity matrices (**H**) to be read by `bgaPEST`. If `deriv_mode=0`, `PEST` is used, external

to bgaPEST, to calculate the Jacobian matrix resulting in a binary file with the extension .jco. If deriv_mode=1 then an external program is used to calculate **H** and a file written by the external program must be communicated to bgaPEST. This file can either be a .jco file or a .jac file which is an ASCII file following the format of a standard matrix file used by PEST, as described in *Doherty* (2010b), Section 4.4.3. An example and description of the format of a standard PEST matrix follow, quoting from *Doherty* (2010b).

Figure A depicts an example matrix file holding a matrix with three rows and four columns.

```
   3   4   2
3.4423    23.323    2.3232   1.3232
5.4231    3.3124    4.4331   3.4442
7.4233    5.4432    7.5362   8.4232
* row names
apar1
apar2
apar3
* column names
aobs1
aobs2
aobs3
aobs4
```

Figure 5: Example of a standard PEST matrix file, adapted from *Doherty* (2010b).

The first line of a matrix file contains 3 integers. The first two indicate the number of rows (NROW) and number of columns (NCOL) in the following matrix. The next integer (named ICODE) is a code, the role of which will be discussed shortly. Following the header line is the matrix itself, in which entries are space-separated and wrapped to the next line if appropriate. The maximum line length is 500 characters, so wrapping to the next line must occur within 500 characters. It is recommended to wrap lines after 8 values and to maintain maximum possible precision.

In use for Jacobian matrices by bgaPEST, ICODE is set to 2, so the string "* row names" is printed next, followed by NROW names (of 20 characters or less in length), containing the names associated with rows of the matrix. NCOL column names follow in a similar format, following the string "* column names".

Other options for ICODE are described in *Doherty* (2010b) and are used bgaPEST for output of the posterior covariance matrix. The two options for posterior covariance output both refer to square matrixes that have the same names of columns and rows.

As a result, only one list of names follows the data following the string "`* row and column names`".

If compression is used in the prior covariance matrix, `bgaPEST` only outputs the diagonal elements of the posterior covariance. In this case, `ICODE=-1` and only the diagonal entries are listed, one per line, after the header line. If compression is not used, the entire posterior covariance matrix is printed using `ICODE=1` with 8 values per line.

B A computationally efficient method to calculate the objective function

Unknown Mean

We need to establish equality between the following two objective function calculation methods, labelled J_{std} for the standard version, and J_{eff} for the efficient version.

$$\begin{aligned}\mathbf{\Phi}_{std} &= \frac{1}{2}\mathbf{s}^T\mathbf{G}\mathbf{s} + \frac{1}{2}\left(\left(\mathbf{y}-\mathbf{h}\left(\mathbf{s}\right)\right)^T\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{h}\left(\mathbf{s}\right)\right)\right)\\ \mathbf{\Phi}_{eff} &= \frac{1}{2}\xi^T\left(\mathbf{H}\mathbf{Q}\mathbf{H}^T\right)\xi + \frac{1}{2}\left(\left(\mathbf{y}-\mathbf{h}\left(\mathbf{s}\right)\right)^T\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{h}\left(\mathbf{s}\right)\right)\right)\end{aligned}$$

For these two formulations to be equivalent, it is only necessary for the following to hold (removing the $\frac{1}{2}$ terms)

$$\xi^T\left(\mathbf{H}\mathbf{Q}\mathbf{H}^T\right)\xi = \mathbf{s}^T\mathbf{G}\mathbf{s} \tag{45}$$

Recalling that

$$\mathbf{G} = \mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{X}\left(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Q}^{-1} \tag{46}$$

and

$$\mathbf{s} = \mathbf{X}\beta + \mathbf{Q}\mathbf{H}^T\xi \tag{47}$$

We can expand the right-hand term of Equation 45 incorporating both Equations 46 and 47 into four terms

$$\mathbf{s}^T\mathbf{G}\mathbf{s} = \beta^T\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}\beta$$

$$
\begin{aligned}
&+ \quad \xi^T \mathbf{HQQ}^{-1}\mathbf{QH}^T\xi \\
&- \quad \beta^T \mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}\left(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}\beta \qquad (48)\\
&- \quad \xi^T \mathbf{HQQ}^{-1}\mathbf{X}\left(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{QH}^T\xi
\end{aligned}
$$

(49)

Simplifying all inverses, immediately the first and third terms cancel. Furthermore, the second term

$$\xi^T \mathbf{HQQ}^{-1}\mathbf{QH}^T\xi = \xi^T \mathbf{HQH}\xi \qquad (50)$$

which is our goal from Equation 45.

Finally, then, for Equation 45 to hold, the fourth term from Equation 48 must completely cancel. To evaluate this, we first simplify through inverses resulting in our new goal

$$\xi^T \mathbf{HX}\left(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{H}^T\xi = 0 \qquad (51)$$

Now, we must expand this using the definition

$$\xi = \mathbf{Q}_{yy}^{-1}\left(\mathbf{y} - \mathbf{HX}\beta\right) \qquad (52)$$

where $\mathbf{Q}_{yy}^{-1}$=$\left(\mathbf{HQH^T} + \mathbf{R}\right)^{-1}$ is substituted because early inspection indicates that $\mathbf{Q}_{yy}^{-1}$ cannot be reduced further.

Substituting Equation 52 into Equation 51 is accomplished by first making some temporary definitions, noting the property that for an arbitrary symmetix matrix $\mathbf{A}$, $\left(\mathbf{A}^{-1}\right)^T = \mathbf{A}^{-1}$.

First

$$
\begin{aligned}
\xi &= \mathbf{Q}_{yy}^{-1}\mathbf{y} - \mathbf{Q}_{yy}^{-1}\mathbf{HX}\beta \\
\xi^T &= -\beta^T\mathbf{X}^T\mathbf{H}^T\mathbf{Q}_{yy}^{-1} + \mathbf{y}^T\mathbf{Q}_{yy}^{-1}
\end{aligned}
$$

(53)

With these definitions, we can assign

$$A = \mathbf{Q}_{yy}^{-1}\mathbf{y} \qquad (54)$$

$$
\begin{aligned}
B &= \mathbf{Q}_{yy}^{-1}\mathbf{H}\mathbf{X}\beta \\
C &= \beta^T\mathbf{X}^T\mathbf{H}^T\mathbf{Q}_{yy}^{-1} \\
D &= \mathbf{y}^T\mathbf{Q}_{yy}^{-1} \\
E &= \mathbf{H}\mathbf{X}\left(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{H}^T
\end{aligned}
$$

Returning to Equation 51, substituting the assignations in Equation 54

$$
\begin{aligned}
\xi^T\mathbf{H}\mathbf{X}\left(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{H}^T\xi &= (-C+D)\,E\,(A-B) \\
&= (-CE+DE)\,(A-B) \\
&= -CEA-DEA+DEA+DEB \\
&= DEB-CEA
\end{aligned} \tag{55}
$$

To conclude, we can expand Equation 55 using Equation 54

$$
\begin{aligned}
DEB-CEA &= \mathbf{y}^T\mathbf{Q}_{yy}^{-1}\mathbf{H}\mathbf{X}\left(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{H}^T\mathbf{Q}_{yy}^{-1}\mathbf{H}\mathbf{X}\beta \\
&\quad -\beta^T\mathbf{X}^T\mathbf{H}^T\mathbf{Q}_{yy}^{-1}\mathbf{H}\mathbf{X}\left(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{H}^T\mathbf{Q}_{yy}^{-1} \\
&= 0
\end{aligned} \tag{56}
$$

Equation 56 is correct by virtue of the fact that both terms are equal scalars.

Diffuse Mean

To expand the above simplification of the objective function to cases with diffuse information about the mean, we make the following substitution

$$
\mathbf{Q} = \mathbf{Q}_{ss} + \mathbf{X}\mathbf{Q}_{\beta\beta}\mathbf{X}^T. \tag{57}
$$

By virtue of this new version of $\mathbf{Q}$ having the same dimensions as in the previous derivation, by direct substitution, the equation for J_{eff} can be used directly as

$$
\mathbf{\Phi}_{eff} = \frac{1}{2}\xi^T\left(\mathbf{H}\left(\mathbf{Q}+\mathbf{X}\mathbf{Q}_{\beta\beta}\mathbf{X}^T\right)\mathbf{H}^T\right)\xi + \frac{1}{2}\left(\left(\mathbf{y}-\mathbf{h}\left(\mathbf{s}\right)\right)^T\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{h}\left(\mathbf{s}\right)\right)\right) \tag{58}
$$

References

Banta, E., E. Poeter, J. Doherty, and M. Hill (2006), JUPITER: Joint Universal Parameter IdenTification and Evaluation of Reliability– An application programming interface (API) for model analysis, *Tech. Rep. U.S. Geological Survey Techniques and Methods Book 6, Section E, Chapter 1*, U.S. Geological Survey Techniques and Methods Book 6, Section E, Chapter 1.

Box, G., and D. R. Cox (1964), An analysis of transformations, *Journal of the Royal Statistical Society. Series B (Methodolodical)*, *26*(2), 211–252.

Casella, G. (1985), An introduction to empirical Bayes data-analysis, *American Statistician*, *39*(2), 83–87.

Collinson, J. D. (1969), Sedimentology of grindslow shales and kinderscout grit - a deltaic complex in namurian of northern england, *Journal of Sedimentary Petrology*, *39*(1), 194–221.

Doherty, J. (2010a), PEST, model-independent parameter estimation—user manual (5th ed., with slight additions).

Doherty, J. (2010b), PEST, model independent parameter estimation. Addendum to user manual: 5th edition.

D'Oria, M. (2010), Characterization of Aquifer Hydraulic Parameters: from Theis to Hydraulic Tomography, phd.

Fienen, M., P. Kitanidis, D. Watson, and P. Jardine (2004), An application of Bayesian inverse methods to vertical deconvolution of hydraulic conductivity in a heterogeneous aquifer at Oak Ridge National Laboratory, *Mathematical Geology*, *36*(1), 101–126.

Fienen, M., R. Hunt, D. Krabbenhoft, and T. Clemo (2009), Obtaining parsimonious hydraulic conductivity fields using head and transport observations: A Bayesian geostatistical parameter estimation approach, *Water Resour. Res.*, *45*.

Fienen, M. N., T. M. Clemo, and P. K. Kitanidis (2008), An interactive Bayesian geostatistical inverse protocol for hydraulic tomography, *Water Resources Research*, *44*.

Gaganis, P., and L. Smith (2001), A bayesian approach to the quantification of the effect of model error on the predictions of groundwater models, *Water Resources Research*, *37*(9), 2309–2322.

Gallagher, M. R., and J. Doherty (2007), Parameter interdependence and uncertainty induced by lumping in a hydrologic model, *Water Resources Research*, *43*(5).

Gray, R. (2005), Toeplitz and circulant matrices: a review, *Tech. rep.*, Now Publishers, Inc.

Hoeksema, R. J., and P. K. Kitanidis (1984), An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling, *Water Resources Research*, *20*(7), 1003–1020.

Jaynes, E. T., and G. L. Bretthorst (2003), *Probability theory : the logic of science*, xxix, 727 p. pp., Cambridge University Press, Cambridge, UK ; New York, NY.

Kitanidis, P. K. (1995), Quasi-linear geostatistical theory for inversing, *Water Resources Research*, *31*(10), 2411–2419.

Kitanidis, P. K. (1996), Analytical expressions of conditional mean, covariance, and sample functions in geostatistics, *Stochastic Hydrology and Hydraulics*, *10*(4), 279–294.

Kitanidis, P. K., and E. G. Vomvoris (1983), A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations, *Water Resources Research*, *19*(3), 677–690.

Li, W., A. Englert, O. A. Cirpka, J. Vanderborght, and H. Vereecken (2007), Two-dimensional characterization of hydraulic heterogeneity by multiple pumping tests, *Water Resources Research*, *43*(4).

Nowak, W., and O. A. Cirpka (2004), A modified Levenberg-Marquardt algorithm for quasi-linear geostatistical inversing, *Advances in Water Resources*, *27*(7), 737–750.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. O. Flannery (1992), *Numerical Recipes in C*, second ed., Cambridge University Press, Cambridge.

Robbins, H. (), An empirical Bayes approach to statistics, in *Proceeding of the Third Berkeley Symposium on Mathematical Statistics*, vol. 1, edited by J. Neyman, pp. 157–163, University of California Press.

Rubin, Y. (2003), *Applied stochastic hydrogeology*, xix, 391 pp., Oxford University Press, Oxford ; New York.

Swift, D. J. P., B. S. Parsons, A. Foyle, and G. F. Oertel (2003), Between beds and sequences: stratigraphic organization at intermediate scales in the quaternary of the Virginia coast, USA, *Sedimentology*, *50*(1), 81–111.

Walker, R. G. (1984), *General Introduction: Facies, facies sequences and facies models.*, chap. 1, pp. 1 – 9, 2nd ed., Geological Association of Canada, Toronto, edited by Roger G. Walker. A fully rewritten version of the first edition, with several new contributions. ill. ; 28 cm. Geoscience Canada, Reprint series ; 1. "Reprinted with revisions from a series of papers in Geoscience Canada, 1976-1979". Includes bibliographies.

Walker, R. G. (1992), *Facies, facies models and modern stratigraphic concepts.*, chap. 1, pp. 1 – 14, Geological Association of Canada, St. John's, Nfld., cn 92090412 C92-90412-2 edited by Roger G. Walker and Noel P. James. ill. ; 28 cm. Includes bibliographical references.