Approaches in Highly Parameterized Inversion: bgaPEST, a Bayesian Geostatistical Approach Implementation With PEST—Documentation and Instructions

By Michael N. Fienen, Marco D’Oria, John E. Doherty, and Randall J. Hunt

Techniques and Methods, Book 7, Section C9

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia 2012

For product and ordering information:
World Wide Web: http://www.usgs.gov/pubprod
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS

Suggested citation:
Fienen, M.N., M. D'Oria, J.E. Doherty, and R.J. Hunt, 2012, Approaches in highly parameterized inversion: bgaPEST, a Bayesian geostatistical approach implementation with PEST—documentation and instructions: U.S. Geological Survey Techniques and Methods, Book 7, Section C9 , 86 p. [http://pubs.usgs.gov/tm/tm7c9]

Any use of trade, product, or firm names in this publication is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Typeset on November 16, 2012

Contents

Figures

Tables

Conversion Factors

Multiply	By	To obtain
foot (ft)	0.3048	meter (m)
gallon per minute (gal/min)	0.06309	liter per second (L/s)
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

$$^\circ C = (^\circ F - 32)/1.8$$

Blank page

Approaches in Highly Parameterized Inversion: bgaPEST, a Bayesian Geostatistical Approach Implementation With PEST—Documentation and Instructions

By Michael N. Fienen[1], Marco D'Oria[2], John E. Doherty[3], and Randall J. Hunt[1]

Abstract

The application bgaPEST is a highly parameterized inversion software package implementing the Bayesian Geostatistical Approach in a framework compatible with the parameter estimation suite PEST. Highly parameterized inversion refers to cases in which parameters are distributed in space or time and are correlated with one another. The Bayesian aspect of bgaPEST related to Bayesian probability theory in which prior information about parameters is formally revised on the basis of the calibration dataset used for the inversion. Conceptually, this approach formalizes the conditionality of estimated parameters on the specific data and model available. The geostatistical component of the method refers to the way in which prior information about the parameters is used. A geostatistical autocorrelation function is used to enforce structure on the parameters to avoid overfitting and unrealistic results. Bayesian Geostatistical Approach is designed to provide the smoothest solution that is consistent with the data. Optionally, users can specify a level of fit or estimate a balance between fit and model complexity informed by the data. Groundwater and surface-water applications are used as examples in this text, but the possible uses of bgaPEST extend to any distributed parameter applications.

Introduction

This report documents the theory and computer code of a Bayesian Geostatistical Approach (BGA[4]). "Bayesian" refers to theory used to solve the inverse problem; "Geostatistical" refers to the use of an autocorrelated spatial or temporal function to provide prior information about the parameters. This BGA approach has been coded by using the protocol and approach of PEST (Doherty, 2010a,b), the most widely used computer code of its type. Though BGA and PEST have been applied primarily to the environmental modeling field, a general view of the inverse problem discussed herein covers most classes of problems where measurements of a system are used to infer system properties that create the measured value. Thus, this bgaPEST formulation is meant to be applicable to any class of inverse problem that adheres to the concepts described herein. In this introductory section, we outline the conceptual framework of the method; theory, implementation, and instructions for use of the bgaPEST computer code version 1.0 follow this conceptual discussion.

Environmental modeling can facilitate informed management of natural resources. In most cases, models represent physical processes such as groundwater flow, contaminant transport, surface water flood routing, and so forth. Simulating the overarching physics and chemistry with governing equations and concepts such as conservation of mass and momentum is only part of what is required. Once

[1] U.S. Geological Survey, Wisconsin Water Science Center, Middleton, Wisconsin

[2] Department of Civil and Environmental Engineering and Architecture, University of Parma, Parma, Italy.

[3] Watermark Numerical Computing, Brisbane, Australia and Australian National Centre for Groundwater Research and Training, Adelaide, Australia.

[4] Whereas the Bayesian Geostatistical Approach is referenced with the all capital abbreviation "BGA," lower-case is used in the software name ("bgaPEST") to visually differentiate "BGA" and "PEST."

a model adequately represents the processes of the problem, it is still only an abstract tool until estimates of the physical characteristics are tuned to observations from the specific area to be managed.

Input values that are used by processes are what control the behavior of a model at a given site; we refer to these input values as “parameters.” They represent specific characteristics (for example, hydraulic conductivity, recharge rate, chemical decay rate) of the natural system at an area of interest. Representative values appropriate for parameters in natural systems are often difficult to measure directly, however. Observations of system state (for example, water levels, streamflow magnitudes, chemical concentrations) are often easier to measure, and the model is chosen so that the observations of the system correspond to output values from the model. The outputs of observation data are dependent to varying degrees on the values of parameters. Therefore, changes in the simulated system state that align with observations allow inference of parameter values not available by direct measurement. “Parameter estimation” and “calibration” are both terms describing the process of incorporating site-specific observations to inform parameters and salient processes of a model in order to improve the representativeness and predictive ability of the model.

A model, therefore, can be thought of as a data-processing tool that quantitatively tests conceptualizations of a system as well as a simulator of physical processes; metaphorically, data processing is a pipeline from field observation to model parameters and ultimately to more representative predictions for supporting management decisions.

Put in terms of a Bayesian description, this process is one in which the model is a vehicle for updating soft-knowledge or expert-knowledge of the system (called a priori understanding in the Bayesian context). This initial vehicle is filtered by a measure of its ability to simulate the natural world informed, or updated, by site-specific observations (called a posteriori understanding). In the Bayesian approach, the ability to simulate features of the natural world “conditions” or narrows the wide range of possible outcomes that result from general expert-knowledge and soft-knowledge alone. A key benefit of the Bayesian approach is that it provides a theoretically rigorous way to continually incorporate new information and, in turn, update a posteriori understanding.

Purpose and Scope

This report is intended to serve two purposes. First, a Bayesian approach to parameter estimation—expressed in the context of the Bayesian Geostatistical Approach (BGA)—is described to provide an accessible and general tool for moving a model from a general simulator of a physical process to a more optimal tool, one that is tuned to a set of calibration information, which, in turn, can be used for improved prediction and decision-making. Second, a computer code—bgaPEST version 1.0—is introduced and documented in which BGA is deployed by means of the protocols and input/output concepts developed in the free and open-source PEST suite of software Doherty (2010a). This report gives details on the mathematical theory behind BGA, followed by detailed instructions for using the computer program. Conventions and assumptions for using the program also are included in the discussion. To our knowledge, this application marks the first implementation of a general BGA code available for widespread use.

The bgaPEST input framework is consistent with the input block and keyword concepts described by the JUPITER project Banta and others (2006). Although the relation of design concepts is beyond the scope of this report, the input block and keywords needed to run bgaPEST are described fully in appendix 1. A full description of the format of the general approach of template and instruction files are omitted here; detailed descriptions are provided in the PEST documentation (Doherty, 2010a, chapter 3). All options implemented in template and instruction files in PEST are available in bgaPEST. A distributed parameterization scheme discussed in this report facilitates the introduction of flexibility to the model. This parameterization scheme also can be applied to any region of interest and at the extreme, where sufficient data are available, can allow a modeler to estimate a unique parameter values for each model node or cell. This level of detail leads to a large number of parameters, a condition that poses computational challenges; alleviating these computational challenges is an active area of ongoing research and thus is not covered in detail here.

This report includes an overview of theory and use of the bgaPEST code in the main text. Detailed input instructions for bgaPEST version 1.0 are in appendix 1, and quick-start instructions are in appendix 2. A detailed mathematical derivation of the BGA methods is in appendix 3, and example problems are in appendixes 5 through 7.

Obtaining the Software

The software for bgaPEST Version 1.0 is available for download at `http://pubs.usgs.gov/tm/tm7c9`. This location includes a copy of this document and both executables and source code. As development of the code continues, a repository at `http://github.com/mnfienen-usgs/bgaPEST` provides a link to revisions in progress and provides a collaborative open-source space where users may submit revisions for consideration by the authors. As further development takes place, new code releases will be posted at `http://pubs.usgs.gov/tm/tm7c9`.

The Bayesian Geostatistical Approach

The presentation of BGA is in two parts, first primarily as narrative then later as a detailed mathematical approach in appendix 3. Those most interested in simply applying bgaPEST to their specific problem will likely spend most time with the narrative. In both presentations, the concept of conditionality is fundamental. This concepts is expressed here as Bayes' theorem,which forms the foundation of the techniques described in the rest of this report:

$$p(\mathbf{s}|\mathbf{y}) \propto L(\mathbf{y}|\mathbf{s})\, p(\mathbf{s}) \qquad (1)$$

where: $\mathbf{s}$ is an $m \times 1$ vector of m parameter values, $\mathbf{y}$ is $n \times 1$ vector of n observations, $p(\cdot)$ indicates a probability density function (pdf), $L(\cdot)$ indicates a likelihood function, and $|$ indicates conditionality. Put into words, Bayes' theorem states that the posterior probability of parameters conditional on the observations $p(\mathbf{s}|\mathbf{y})$ (often referred to simply as "the posterior probability of $\mathbf{s}$ given $\mathbf{y}$") is proportional to the prior probability of the parameters $p(\mathbf{s})$ updated with the likelihood function $L(\mathbf{y}|\mathbf{s})$ that expresses how well $\mathbf{y}$ is estimated by using the model and a candidate parameter set $\mathbf{s}$. The pdfs in all cases are assumed to follow or at least be well-approximated by Gaussian distributions. This assumption is important and somewhat restrictive, but is made for computational simplicity. Active research is ongoing on alternatives to this approach, but the traditional Gaussian assumption is adopted in this report and is still considered a practical and useful assumption for many cases.

In the Bayesian context, expressing the parameters and the likelihood function as probability distributions formally incorporates an estimate of their uncertainty rather than treating the parameters as perfectly known values. All a posteriori (also called posterior) distributions are conditional upon the specific data used in the calibration process. Perhaps less obviously, posterior distributions are also conditional on all other modeling and data assumptions and decisions that go into formulating the problem: which model and what model options are chosen, numerical 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. As a result, if any of these underlying assumptions and decisions change, it is expected in the Bayesian context that the parameters estimated and the associated posterior uncertainty also will change.

The conditionality includes *all* decisions made in the process of constructing a model and incorporating data, however, not all of this information is explicitly addressed by the modeler. In fact, the only *explicit* conditionality is on the observation data. The information contained in the prior pdf ($p(\mathbf{s})$) in equation 1 is critical because it represents the state of knowledge about the parameters in the system prior to updating through the calibration process. In this report, assumptions made prior to calibration are intentionally limited and are restricted to the assignment of a mean value (unknown) 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 (implemented as a variogram, also described below). The degree to which this continuity characteristic is enforced is dictated by the observations included and the subsequent performance of the model. This construction is similar to (and in some cases, mathematically equivalent to) Tikhonov regularization (Tikhonov, 1963a,b; Aster and others, 2005). Limitation of the information assumed a priori is similar to assuming a low or diffuse level of a priori soft knowledge (called an ignorance prior in Jaynes and Bretthorst, 2003, chapter 12). In this case, the resulting model is driven more by information obtained from site-specific observations than from prior assumptions based on soft-knowledge. A goal of this approach is to limit the subjective information and to favor instead an objective and repeatable result based on observation data. Additionally, so-called structural parameters that enforce the characteristic smoothness are estimated. An algorithm that encompasses all of these aspects is considered an Empirical Bayes approach.

A Note on Parameters

The term “parameters” refers to discrete values of system state. When we describe “parameter type” we mean a group of parameters that belong to the same class of system state (for example, hydraulic conductivity or recharge). In applications appropriate for bgaPEST, there must be multiple parameter values of a given type that are spatially or temporally distributed. When a “parameter” is listed, the meaning is restricted to a single value of system model input that is to be estimated.

The concept of a beta association is important and is a concept that is revisited throughout this report. Fienen and others (2009) describe the need to to represent the generalized mean value for a specific parameter type (for example, hydraulic conductivity in a groundwater model) in a specific region of a model referred to as a “facies association.” To be more general, this concept is incorporated here by the use of the term “beta,” derived from the mathematical symbol used. In the methodology described in this report, parameter values are estimated by estimating a mean value (termed β in the mathematics) and the fluctuations about that mean. Each parameter, therefore, must be associated with a mean value.

It would be tempting to use another term such as “zone” or “facies” to describe this concept, but the term “beta association” was selected specifically to highlight the flexibility of the concept. The important idea is that the method described in bgaPEST depends on being able to associate each parameter with a mean value. In the case of distributed parameters (for example, hydraulic conductivity or recharge being distributed throughout a region in a model in which each model cell or node is assigned a unique value), the subdivision of the entire model domain into beta associations accounts for hydrogeologic contacts or facies to be delineated. This delineation assumes that there is little or no correlation across these natural divisions. Similarly, parameters of one type are typically not correlated with parameters of a different type. Beta associations allow the inclusion of multiple parameter types and the delineation of important geologic features in distributed parameter sets. In the BGA algorithm, parameters in different beta associations are assigned zero correlation.

The likelihood function in equation 1, $L(\mathbf{y}|\mathbf{s})$, expresses the correspondence of model outcomes with field observations colocated in space and time. This correspondence is expressed as the sum of the squared differences between outcomes and observations, weighted by a covariance matrix, which expresses the relative certainty of each observation. This is equivalent to the weighted measurement objective function in PEST (Doherty, 2010a). The advantages to the Bayesian approach stem from the conceptual framework, the ability to use a probability density function to represent parameterization rather than single values, and the empirical nature of the balance between prior information and likelihood. These elements of Bayes’ equation form the fundamental basis for the bgaPEST software described here.

The *geostatistical* aspects of the method are expressed in the prior pdf $(p(\mathbf{s}))$ of the parameter values. Geostatistics is a form a interpolation that uses a spatial function called a variogram to fill in information between data points. The most common technique of geostatistics is called kriging. In this section, we discuss geostatistics in a conceptual way using a photographic image as an example. The mathematics relevant to BGA are discussed in later sections. More details about the history and use of geostatistics in general, including software for using the technique are found in Isaaks and Srivastava (1989), Deutsch and Journel (1992), Kitanidis (1997), and Remy and others (2009).

The photograph in the left panel of figure 1 is a JPG image taken on a digital camera. As a grayscale image, the information of the image can be stored in a matrix with the number of rows and columns, 320 rows and 240 columns in this case, indicating the number of pixels in the vertical and horizontal dimensions, respectively. The values at each pixel are a brightness value, in this case normalized to a maximum value of 64.0. In the original image, there are $320 \times 240 = 76,800$ pixels, each of which may be considered a discrete packet of information. To illustrate the kriging process, the photograph was first subsampled on an evenly spaced grid of 30 rows and 20 columns with the brightness value retained at each location. This subsampling results in a greatly reduced set of information containing $30 \times 20 = 600$ pixels. In the photograph in the right panel of figure 1, the faint impression of the subsampling grid is visible in some areas as the subsampled values are depicted at those locations. Using the geostatistical technique of kriging with an appropriate variogram, it is possible to “fill in” the missing data between subsampled data points to present a full image of the matrix but with substantially less detail than the original.

The main role of the variogram in geostatistics and, indeed, in BGA, is to act as a constraint, controlling the shape of the interpolated values filling in between the known data values. In BGA, this connection is not quite as direct as in the photograph interpolation example, but it is useful to think of the variogram (as a quantification of the prior pdf) as a control on the shape of the estimated parameters.

Various other interpolation techniques could be used to fill in the missing data and each would have its own degree of information loss or smoothing relative to the original. Kriging has a long history of use in earth science applications and, although the interpolated photographic image in the example is much smoother than the original image, there is more shape and information than if, for example, linear interpolation had been used to fill the values between each data point.

The variogram used in kriging is an empirical function that characterizes the difference between a property as a function of separation distance. To determine a variogram appropriate for a problem, the first step is to plot a variogram function (a function of difference in property value) against separation distance (depicted as red “x” marks in figure 2). Next, a function type is selected from a family of valid variogram model types. In the photograph example an exponential variogram is used. Later in this report, more mathematical details about variograms and variogram choice are presented.

Returning to the geostatistical aspect of BGA, a variogram model is used as the prior information in the Bayesian construction. As discussed above, a goal of BGA is to specify little information in the prior and to allow the information contained in the calibration data set to inform the results as much as possible. This is accomplished by specifying only the family of variogram model used rather than specifying its specific shape. Using the example in figure 2, the family of variogram (in this case exponential) indicates only that the function will assume a curvilinear shape; the specific parameters or the variogram function dictate the rate of curvature. In terms of using a variogram for the prior distribution in BGA, specifying the variogram type informs only the most general characteristic of the field (for example, the field must be continuous and “smooth”). The degree to which this characteristic is enforced is controlled by the calibration dataset.

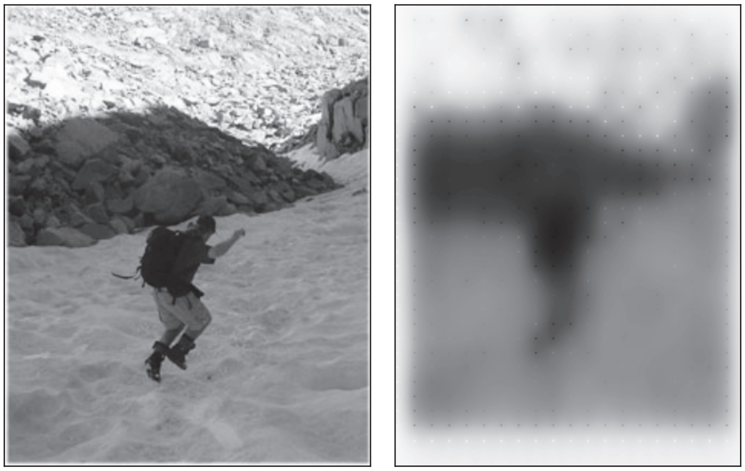

Figure 1. Photographic images illustrating the geostatistical interpolation process. The image on the left is a JPG format image with 320×240 pixels. The image on the right is also a JPG image, but it was created by first sampling a subset of the pixels in the original image (30 in the vertical direction and 20 in the horizontal) and then using the geostatistical technique of kriging to interpolate values for the other pixels. The interpolation was done by using the sGeMS software package (Remy and others, 2009).

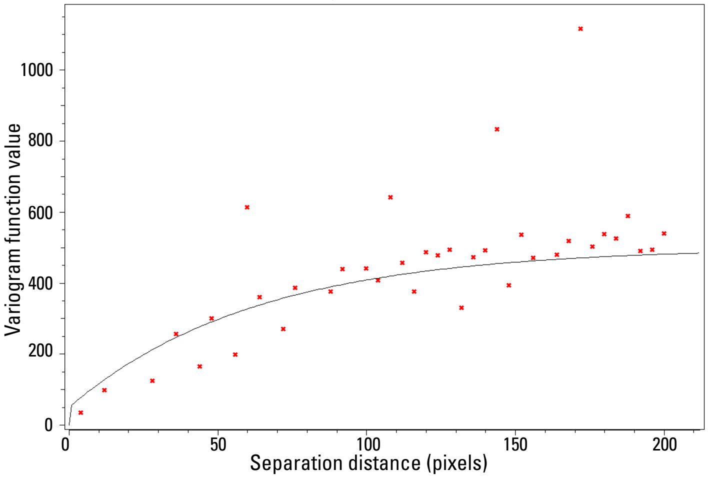

Figure 2. Exponential variogram fit to empirical variogram for the image processing example. The empirical variogram binned values are depicted by red “x” marks whereas the continuous black line shows the analytical variogram fit to the empirical values. The fit was done manually by using the sGeMS software package (Remy and others (2009)).

Overview

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. This section is a conceptual overview of the method. A more detailed description, including mathematical details, is in appendix 3.

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

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

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

Figure 3 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 narrower, representing less uncertainty.

In bgaPEST, an empirical Bayes perspective (Robbins, 1956; Casella, 1985) is adopted. Empirical Bayes means that the general characteristics of the prior and (optionally) epistemic covariances introduced above are provided in the model setup, but the values of "structural" parameters that control the structure of the system—the balance between smoothness and misfit—are estimated from the observation data. In other words, the level of roughness in the solution is dictated by the information content of the observation data rather than specified by the user ahead of time.

The prior distribution is the main mechanism by which soft-knowledge about the parameter field is imparted on the parameter estimation process. In the Empirical Bayes perspective, this soft-knowledge is intentionally limited such that significant flexibility is available to the algorithm and a specific practitioner's preconceived notions, which are more subjective, are replaced by the objective power of the site-specific observations. This idea is also inspired by Chamberlin's concept of multiple working hypotheses (Chamberlin, 1890). Chamberlin warned of scientists falling victim to a "paternalistic affection" for their initial explanation of a phenomenon such that they are blind to other explanations that may be more appropriate. This is not to discount the value of soft knowledge—indeed, the general characteristics imparted through specification of the prior information and the interpretation of the results of using BGA rely deeply on expertise—but it highlights a goal of leaving as much flexibility as possible in the process.

In bgaPEST, then, the practitioner specifies a type of variogram (nugget, linear, or exponential) that is used to control the variability—smoothness or roughness—of parameters within a beta association, but the *degree* to which this characteristic is enforced is determined by a Bayesian adaptation of restricted maximum likelihood (RML). In RML, the value of structural parameters that control the variogram behavior is treated as a probability distribution and the most likely values resulting in either the best possible fit (if the epistemic error term is estimated) or a user-specific level of fit (if the epistemic error term is fixed) are estimated. The Bayesian adaptation to RML in this report is through the inclusion of prior information and uncertainty, which is not strictly possible in traditional RML. "Fit," in this context, refers to the correspondence between observation data and model outputs colocated in space and time with the measured observations. Fit and epistemic error are discussed in more detail in the next section. A danger of providing a model with substantial flexibility, is an "overly complex" model that is "overfit" (for example, Draper and Smith, 1966; Hill, 2006). To mitigate this issue, the RML approach is consistent with the principle of maximum entropy such that the smoothest solution is chosen, an approach 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).

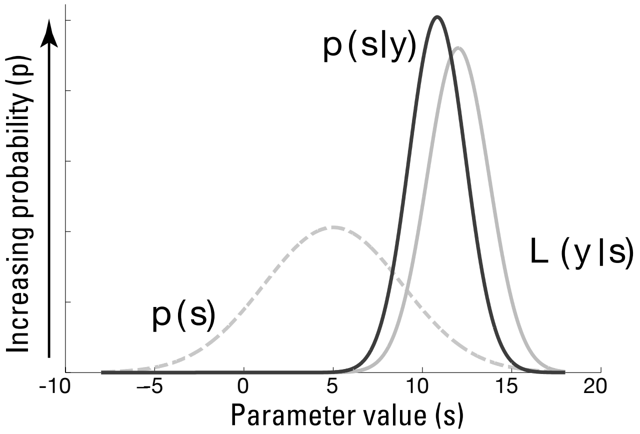

Figure 3. Graphical illustration of Bayes' theorem.

An extension of this approach is the inclusion of information about the prior mean (Nowak and Cirpka, 2004). Although the mean is estimated in the solution, a prior value and covariance can be supplied to constrain the estimate. Typically, a relatively high covariance magnitude is used so that the constraint on the estimated mean is weak or "diffuse." Thus the prior mean principally serves the role of providing numerical stability rather than compelling the solution to adhere closely to prior values. Similarly, prior information and covariance can be supplied on the structural parameters to constrain the estimated values to more closely follow an initial conception of the parameter field variability.

The forward model is constructed to generate outputs of values colocated in space and time with measured observations. The likelihood function quantifies the difference (misfit) between the model simulated outputs and associated observations. In all modeling, perfect correspondence between forecasts and observations is neither attainable nor desirable. The observations themselves are corrupted by measurement errors and perfect correspondence between the exact nature of the measurements and the simulated counterparts usually is lacking. 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 overarching term "epistemic uncertainty" (Rubin, 2003, p. 4). This epistemic uncertainty characterizes the expected misfit between simulated and observed equivalents, 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.

Structural Parameters

The term "structural parameters" used here has a specific meaning. Similar to the more general term "parameters," structural parameters are variable values that are estimated in the bgaPEST algorithm. Unlike typical parameters, however, structural parameters do not *directly* control physical aspects of the system in the way that, for example, hydraulic conductivity or stream roughness do in hydrologic models. Instead, structural parameters control the structure of the general parameters. For example, the variogram values (for example variance, slope, and correlation length) that control the roughness of distributed parameter fields are structural parameters, as is the value of variance controlling epistemic uncertainty. Because these parameters must be estimated but are not directly connected to the physics of the problem, they are also referred to by other authors as "nuisance" parameters or "hyperparameters." We adopt the term "structural" to highlight the fact that the impact these parameters has on the solution is control of the shape or structure of the distributed parameter fields.

With both the prior pdf and likelihood function expressed as Gaussian distributions, the resulting posterior pdf also is Gaussian. The values of the parameters **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 influence of small-scale variability can be approximated through conditional realizations. The balance between the strength of smoothing and the level of fit between simulated and observed equivalents is found through calculation of optimal values for the structural parameters. Optionally, this can include a value to quantify the epistemic uncertainty. The result will favor smoothness, but it may achieve a level of fit corresponding to an unrealistically low level of epistemic uncertainty. Hence, it is generally most appropriate to fix the level of epistemic uncertainty but allow the other structural parameters to be estimated.

Beta Associations

In an idealized problem, a single covariance model (for example, a single variogram) is flexible enough to encompass the entire variability of the hydraulic parameters. In many hydrologic applications, however, lithologic contacts and unconformities can create discontinuities in parameter values that a single covariance model cannot characterize. Partitioning the field either on the basis of data (for example, Fienen and others, 2004) or through interrogation of preliminary solutions (for example, Fienen and others, 2008) can greatly improve the parameter estimation 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 concept of partitioning is consistent with zonal boundaries in models made up of homogeneous zones but it allows more flexibility by allowing properties within the zone to vary. Furthermore, multiple types of parameters (for example, hydraulic conductivity and porosity in a flow and transport model) are commonly estimated. Although these parameters may be related at the physical level, they must correspond to different mean values, so similar censoring of correlation among different types of parameters also is necessary in most applications through partitioning.

For hydrogeologic applications, the term “facies association,” from the facies architecture field, is an apt description for these partitions (Fienen and others, 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 and others (2003)). It would be appropriate to use the less restrictive and less transferable term “facies association” in hydrogeologic applications 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 is most effective when based on readily observable hydrologic or lithologic attributes.

For bgaPEST to be a more general tool (not limited to hydrogeologic modeling), we have broadened this concept by adopting the term “beta association.” As shown in equation 3.2, 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 also can refer to completely different parameter types (for example, hydraulic conductivity and recharge).

To clarify our terminology, partitions delineated by stochastic discontinuity within a distributed parameter field are referred to as “beta associations,” whereas zones of piecewise continuity are referred to herein 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 correspond to different mean values (β).

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 on the basis of hydrogeologic characteristics. This term was used in place of "zones" because of a long history of zones referring to regions of piecewise constant homogeneity (one parameter value applied to every node within a region). Beta associations are *not* homogeneous, so a distinct term was sought that describes the characteristic of regions partitioned according to their characteristics and the way these characteristics correlate to regions around them. To generalize beyond hydrogeologic applications, and to account for the fact the distinct parameter types require distinct partitions, "beta associations" was the term chosen. Each of these parameter partitions has a distinct mean value (β) to be estimated within the region, so the partitioning of the problem results in different β values; because the parameter type and/or region must be associated with a mean value (β), we use the term "beta associations."

Overview of bgaPEST

The use of BGA concepts described previously has been restricted to primarily academic and/or custom applications, owing to the case-by-case nature of the BGA coding. The BGA formulation used in bgaPEST is meant to make the approach generally available to a wider class of modeling problems. This generality is achieved by way of the following design considerations. The input/output design of bgaPEST follows that of the widely used PEST software (Doherty, 2010a,b). This approach has two primary restrictions. First, input provided to the model, and output derived from a model, uses an ASCII text file format. This restriction can be relaxed, however, provided that a translation utility can be deployed for converting data of another format—for example, binary—to or from ASCII, as appropriate. Second, the model must run in "batch" mode where many model runs can be called by PEST without user intervention. Therefore, the only kind of model that PEST cannot easily accommodate is one in which any changes to model input or the reading of model output must take place in a graphical user interface. This generality of model compatibility is a powerful capability that bgaPEST is able to exploit by virtue of efficient open-source modules that make this external control of a model possible using the same protocols as PEST.

As discussed below, bgaPEST must control the model for two purposes: to evaluate the likelihood function (assessing the correspondence between model output and colocated observation data, given a candidate set of parameter values) and to calculate the "Jacobian" or "sensitivity" matrix that is required for solving the calibration equations. To enable PEST (and bgaPEST) to write input for a model, template files are created that map named parameters into their proper place in input files for the model. More than one template file can be used corresponding to multiple model input files. To enable reading of output files, instruction files are created that contain a set of instructions (including locating specific line numbers or searching for specific text) that enable extraction of output values to be compared with site observation data. Leveraging the modules that implement the PEST input/output protocols takes advantage of the flexibility and generality of PEST. It also makes it possible to take advantage of certain utility programs already created to be compatible with the PEST suite of software. Programs created using the JUPITER program employ a very similar set of protocols by virtue of the PEST modules having been provided to the JUPITER project. As a result, template and instruction files created to work with a model are largely interchangeable among projects implemented in PEST, bgaPEST, and programs created using JUPITER. A full description of the format of template and instruction files is not within the scope of this report: detailed descriptions are provided in the PEST documentation (Doherty, 2010a, chapter 3). All options implemented in template and instruction files in PEST are available in bgaPEST.

This initial implementation of bgaPEST is written in Fortran-90. The calculation of the Jacobian (sensitivity or derivatives) matrix can be implemented either using a script written by the user or employing a Python[5] script provided with bgaPEST. The Python script depends on several utilities that are standard with PEST and available for download at http://www.pesthomepage.org. The necessary executable files are also provided with bgaPEST. For users on the Windows[6] operating system, installation of Python is optional because the Python codes are compiled into executables using py2exe that can be called by the main program. For users on Macintosh[7] or Linux[8] 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 (sensitivity) calculation with PEST and Python can be replaced by using the parallel external derivatives capabilities described in appendix 4. Alternatively, it would be possible to implement the general parallel run management suite (GENIE, Muffels and others (2012)).

Running bgaPEST

The bgaPEST program uses a single input control file in combination with template and instruction files to control the underlying model, and it 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.

To obtain an optimal solution of the parameter estimation problem, multiple iterations are necessary. An iteration is defined as a single run of the entire estimation process with a particular set of values. Multiple iterations are required because of the nonlinearity of the problem and the necessity of estimating structural parameters separately from model parameters. Appendix 3 gives more detail about the methods used to obtain a solution for a set of optimal parameters and structural parameters in bgaPEST.

Outer iterations (also called BGA iterations) are wrapped around the traditional parameter estimation process with values of the structural parameters held constant. Inner parameter estimation 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 line search (discussed below) is requested, this is performed within the inner iterations. If structural parameter optimization is requested, it is performed after convergence has occurred or maximum inner iterations have been 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 either both have converged or the maximum number of outer iterations has been reached. At the end of both inner and outer iteration convergence, or exceedance of maximum iterations, posterior covariance is calculated, if requested.

Control Variables

Two types of variables are used in bgaPEST: control variables and data variables. Whereas data variables are values such as model parameters, observations, file names, and other data that are needed by the bgaPEST program, control variables, drive the actions that are performed on these data elements. As such, control variables operate on a level different level data variables and control either

[5] "Python®" is a registered trademark of the Python Software Foundation.

[6] "Windows®" is a registered trademark of the Microsoft group of companies.

[7] "Macintosh®" is a registered trademark of Apple, Inc. in the U.S. and other countries.

[8] "Linux®" is the registered trademark of Linus Torvalds in the U.S. and other countries.

```
Outer Iteration (maximum iterations defined by it_max_bga)
    The outer iterations iterate until convergence of both
    PHI and Structural Parameters (optional)

    Inner Iteration (maximum iterations defined by it_max_phi)
        The inner iteration progresses until convergence of PHI conditional
        on the current values of Structural Parameters
    end Inner Iteration
    If linesearch == 1
        Linesearch (maximum iterations defined by it_max_linesearch)
            The linesearch is an optional correction that can get optimization
            for PHI back on track if it strays
        end linesearch

    If struct_par_opt ==1
        Structural Parameter Optimization
        (maximum iterations defined by it_max_structural)
            If, optionally, Structural Parameter Optimization is invoked,
            iterations progress until convergence conditional on
            current process parameter values
        end Structural Parameter Optimization
end Outer Iteration

If posterior_cov_flag == 1
    Calculate posterior covariance
        posterior covariance is only calculated for the final parameter values
```

Figure 4. Abbreviated flowchart showing the progression of the major bgaPEST procedures. Text in blue italics is interpretive, summarizing the more programmatic language represented in black plain type.

the reading/writing of data or the progression of the algorithm. Many control variables are straightforward (for example, `it_max_phi`, an integer, is the total number of iterations allowed in each quasi-linear inner estimation optimization; *default=10).* Such variables are defined in the context of the input instructions listed in appendix 1. Other control variables, however, are accompanied by important conventions regarding their impact on the performance of the algorithm. More detail is given about certain control variables in this section for these cases.

`structural_conv` *float*, *default=0.001*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.

$$\text{conv} = \text{abs}\left(\Phi_{S,i} - \Phi_{S,i-1}\right) \tag{3}$$

where i is the current structural parameter optimization iteration, $i-1$ is the previous structural parameter optimization iteration, and Φ_S is the structural parameter objective function.
If negative, convergence is based on the norm of the difference between consecutive structural parameter values.

$$\text{conv} = \sqrt{\left(\frac{\theta_{i-1}-\theta_i}{\theta_{i-1}}\right)^T \left(\frac{\theta_{i-1}-\theta_i}{\theta_{i-1}}\right)} \tag{4}$$

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).

`phi_conv` *float, default=0.001* 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

$$\text{conv} = \text{abs}\left(\Phi_{T,i_{in}} - \Phi_{T,i_{in}-1}\right) \qquad (5)$$

where i_{in} is the current inner iteration, and Φ_T is the total objective function (equation 3.30).

`bga_conv` *float, default*=10×`phi_conv` 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 convergence of Φ_T and Φ_S. The convergence is evaluated as

$$\text{conv} = \text{abs}\left(\Phi_{T,i_{out}} - \Phi_{T,i_{out}-1}\right) \qquad (6)$$

where i_{out} is the outer iteration and Φ_T is the total objective function (equation 3.30).

`Q_compression_flag` *integer, default=0* Flag to determine how to calculate Q_{ss} [0] = no compression—calculate full Q_{ss} matrix, [1] = calculate separate Q_{ss} matrix for each beta association. In addition to controlling the behavior of prior covariance compression, this flag also determines whether a full posterior covariance matrix or only the diagonal is calculated.

`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 percent 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 1.

Output Files

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

Record File

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 also are reported for each beta association in which structural parameter estimation was requested and for the epistemic uncertainty term, if requested. After the final outer iteration, all structural parameter values—including those which were not estimated—are reported to make a complete record.

Parameter Value Files

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 or power transformation are reported in their linear space, *not* log-transformed or power-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 percent 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 percent confidence limits are reported in linear space, *not* log-transformed or power-transformed space, so for log-transformed or power-transformed parameters, the upper and lower 95 percent percent confidence limits are not symmetrical about the parameter value.

Observation Value Files

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 can be easily copied into a spreadsheet or read with a plotting program to calculate and plot residuals.

Posterior Covariance File

If the input variable `posterior_cov_flag=1`, then posterior covariance of the parameters **s** is calculated. In addition to this information being used to report 95 percent 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 many parameters are used and, thus, the full covariance matrices are unwieldy. On the basis of this choice, the posterior covariance is reported either 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 1.

Posterior Covariance and Parameter Transformations

In this section, it was indicated that in the `<casename>.bpp.fin` file, parameter values and 95 percent confidence intervals are reported in linear (untransformed) space, whereas in the `<casename>.post.cov` file, posterior covariance values are reported in estimation (log-transformed or power-transformed) space. Why the difference? The two files serve slightly different purposes. The parameter output file presents values in the units they are entered in (and, presumably, the units "seen" by the forward model). As a result, 95 percent 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 or power-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.

Suggestions and Guidelines for Initial Use

The Bayesian Geostatistical Approach is a highly parameterized method that is appropriate for some, but not all applications. In this section, we outline a few considerations to aid in the decision about whether to use bgaPEST on a given problem given the history and characteristics of the method. We also offer a few guidelines to help users avoid potential pitfalls in the application of bgaPEST.

This report documents the first release of bgaPEST and, to our knowledge, the first implementation of BGA available in a generalized package. As a result, users of this version will be among the first to apply this software outside of academia where custom programs have been the rule.

Nonetheless, the method has a 20-year history. The majority of applications have been to groundwater modeling projects including (but not limited to): pumping test analysis (Snodgrass and Kitanidis, 1998); hydraulic tomography (Li and others, 2007; Fienen and others, 2008; Li and others, 2008; Cardiff and Kitanidis, 2009; Cardiff and others, 2012); borehole logging (Fienen and others, 2004); contaminant source identification (Snodgrass and Kitanidis, 1997; Michalak and Kitanidis, 2002, 2003); and nonparametric tracer test analysis (Fienen and others, 2006). The main application to date that does not involve groundwater is in atmospheric modeling (Michalak and others, 2004; Mueller and others, 2008).

Characteristics of Appropriate bgaPEST Uses

The characteristics that unite these applications form a solid guide when deciding whether bgaPEST is appropriate for a given application. First and foremost, there must be a parameter set that varies in either space or time; for example, a time series of chemical concentrations (a breakthrough curve), a hydraulic conductivity field, a recharge field, or surface flux of atmospheric gases. These parameters should vary continuously over reasonably substantial areas so that a variogram serves as an adequate descriptor of the shape of the parameter field. Subareas delineated by geologic contacts—or in the case of time series, punctuated by known events—can be partitioned into beta associations, as discussed throughout this report. Another consideration is a more practical one: model run time.

The nature of bgaPEST is that many parameters are to be estimated. Throughout the parameter estimation process, a Jacobian sensitivity matrix must be calculated, requiring one model run per parameter. This computational burden must be considered and, potentially mitigated. In academic settings, many researchers have taken advantage of adjoint-state techniques to make the calculation of the Jacobian matrix more efficient in the case where parameters greatly outnumber observations. Adjoint-state versions of commercial and government codes are not typically available, however, but bgaPEST is equipped to handle Jacobian matrices calculated outside of bgaPEST so that users who are able to write such codes can make use of them. Similarly, parallelization is supported to a limited degree using Python scripts and Condor (Condor Team, 2012) for run management. This parallel implementation is documented in appendix 4.

Adjoint-state Jacobian calculation is an attractive method to mitigate high computational expense of this method; however, production codes for adjoint state calculations are rare. For more information on the technique, see Townley and Wilson (1985), Sykes and others (1985), Samper and Neuman (1986), RamaRao and others (1995), and Neupauer and Wilson (1999) and references therein.

A common occurrence in groundwater modeling applications is that parameters far exceed observations in number. This, of course, can change in transient simulations where, if each measurement in time at a single measurement location is considered an observation, the numbers of observations and parameters may equalize. Use of bgaPEST is most appropriate for the former case—where parameters outnumber observations, typically by a large margin. Several programming and mathematical accommodations are made to enable the number of parameters to grow very large (testing has been performed with 90,000 parameters). If the number of observation grows significantly, however, computer memory will become a limitation in many cases. For transient problems, one should consider the information content of each measurement point in time. Often, the number of observation points can be effectively reduced by considering moments rather than discrete points (Li and others, 2005) or by other time-series processing such as methods available in R (R Development Core Team, 2011) or TSPROC (Westenbroek and others, 2012).

Guidelines

The number of applications of bgaPEST thus far is limited. Because bgaPEST is new software implementing a relatively novel technique, it will take time for users to get a feel for the behavior and characteristics of the tool. In this section, we provide a few guidelines that we hope will help users avoid pitfalls. In future releases, building on the experience of a larger user base, more guidelines will be available.

Run Times For a typical groundwater model, somewhere between 5 and 15 outer iterations will often be required. For each outer iteration, it is likely that about 5 inner iterations will be necessary. This means as many as 75 calculations of the Jacobian matrix may be required. Without parallelization or adjoint state, users should carefully consider how many parameters can be accommodated as run times grow in length. For planning, assume that the time required for each Jacobian calculation will be the number of parameters (m) × Run Time.

Beta Associations Beta associations provide the ability to include knowledge about contacts and other partitions in the parameter fields. Some beta associations can have a separate parameter value in each node, whereas others can be treated as homogeneous zones. Specification of either alternative is accomplished through the design of the template file. In addition to allowing for the inclusion of well-known structures such as lithologic contacts, beta associations also allow for some regions—either because of greater importance to the ultimate management decisions, or because of greater density of data, or both—to have a large number of parameters whereas other regions have homogeneous values. By allowing a large number of parameters only in focused areas of interest, the overall number of parameters can be reduced, thus mitigating some of the concerns about run times.

Line Search The purpose of the line search is similar to the purpose of the Levenberg-Marquardt adjustment used in PEST. Whereas the Levenberg-Marquardt search makes a correction to the search direction when the optimization algorithm might otherwise stray from the optimal direction, the line search adjusts the length along the Quasi-Newton direction to avoid overshooting. The line search, therefore, serves its greatest purpose in its first iteration or two. After that, the value of the line search is limited for mathematical reasons having to do with linearization of the problem (see appendix 3 for more details). As a result, a value of between 2 and 5 for `it_max_linesearch` in the control variables is generally adequate. If the line search algorithm does not converge, a warning will be issued and, although it is good to know that this took place, the line search has served its purpose and little gain will be achieved by increasing `it_max_linesearch`.

Level of Fit "With great power comes great responsibility." In applications where parameters outnumber observations, there lurks a real danger of overfitting. In other words, parameters can be adjusted to achieve of correspondence between simulated and observed equivalents that exceeds a reasonable level. The danger of this is that some of the *lack* of correspondence is often due to random epistemic error and is not representative of actual system behavior. However, if the parameters are adjusted to match observations within this margin of error, they are "fitting the noise." The ramifications of this type of adjustment are mainly diminished predictive power of the model and unrealistic roughness of the parameter fields estimated. There are two means of avoiding this problem. One is the maximum entropy property of BGA. The algorithm is designed to find the *smoothest* solution consistent with the level of fit. If all structural parameters—including σ_R—are estimated, then the algorithm will try to achieve perfect fit with the smoothest solution that can do so. This may still lead to overfitting, however, so in most cases, it is more appropriate to set the level of fit by using `sig_0` in the Epistemic Error Term input block described below to a level of fit chosen by the user to be appropriate given known and suspected uncertainties about both the observation quality and the model. Weights on observations can account for different levels of quality in different observations. In most cases, the user should also set `sig_opt=0` to force the algorithm to use a consistent value for epistemic uncertainty and thus manually control the level of fit. If set this way, the algorithm will adjust the other structural parameters to achieve the smoothest possible solution corresponding to the specified level of fit.

The level of smoothness in the optimal BGA solution is always smoother than conditional realizations (Kitanidis, 1995), which characterize more of the potential variability in each solution. In cases such as transport models where heterogeneity is the most important, conditional realizations (made possible using the optimal solution and the posterior covariance—both provided by bgaPEST) will result in a more precise characterization of system behavior in heterogeneity.

Structural Parameter Optimization A good general guideline for all modeling is to start simple and add complexity as appropriate. In bgaPEST, this goal is achieved by starting with small values of variogram parameters (slope for the linear, variance for the nugget or exponential) such that the solution will be very smooth. By optimizing for structural parameters, roughness will be introduced by the algorithm until convergence at the optimal level of roughness is achieved. At the early, exploratory stages of a project, it might be desirable to set `sig_opt=1` to see what level of fit may be achievable, but the user should be prepared to override this setting in later stages as allowing too much roughness to be introduced. For the prior distribution variogram parameters, however, optimization should always be employed in keeping with the Empirical Bayes perspective designed into the algorithm.

Limitations of bgaPEST Version 1.0

bgaPEST marks the first widely available implementation of BGA for use by practitioners. Limitations, of course, accompany this first implementation. For example, version 1.0 has a limited explicit parallelization facility. This can be overcome by using external programs for derivatives and calling a parallel Jacobian calculation package such a BeoPEST (Schreüder, 2009) or GENIE (Muffels and others, 2012) whenever a Jacobian matrix is required. The impact of this workaround is on the run times required to obtain a solution.

A practical upper limit on the number of parameters estimated is on the order of 100,000. To estimate a larger number of parameters, machines with a large amount of random access memory (RAM) must be used. At some greater limit, methods such as periodic embedding or other decompositions must be incorporated to mitigate the expense of storing and calculating the prior covariance matrix.

The source code is written in Fortran 90 and should be compilable on any platform with a Fortran compiler. Special care was taken to avoid obscure and nonstandard language features. Nonetheless, it is possible that some platform- or compiler-specific problems may be encountered.

It is possible to use bgaPEST with a small number of parameters, but the assumption from the start is that parameters in at least part of the spatio-temporal domain represent a field of correlated instances (for example, model nodes or discrete times) that often outnumber the number of data observations. A combination of homogeneous parameters in zones with a refined area of interest that is distributed is a common application and, as implemented through beta associations, this mix of distributed and zoned parameters is supported and encouraged. Typically, sufficient data to support a distributed parameter set are limited to part of a model domain in space or time.

In considering uncertainty, version 1.0 presents posterior covariance values. For some applications, conditional realizations may be desired to capture candidate roughness of solutions within the ensemble distribution of solutions. Details for conditional realizations are provided by Kitanidis (1995).

Acknowledgments

This work was supported by a variety of projects and programs. General support was provided by the National Research Council Postdoctoral Research Associateship, the U.S. Geological Survey (USGS) Groundwater Resources Program, and the USGS Water Energy and Biogeochemical Budgets (WEBB) Program. In addition, the USGS Exchange Visitor Program and the USGS Wisconsin Water Science Center supported a 6-month exchange in which Dr. Marco D'Oria was hosted in Wisconsin. Dr. D'Oria also received support from the University of Parma, Italy, Department of Civil and Environmental Engineering and Architecture (Università degli Studi di Parma). The authors thank Dr. Michael Cardiff, Boise State University, Stephen Westenbroek, USGS Wisconsin Water Science Center, and Mike Eberle, USGS Science Publishing Network, for valuable technical and editorial reviews of this document.

References Cited

Aster, R.C., Borchers, B., and Thurber, C.H., 2005, Parameter estimation and inverse problems: Amsterdam, Elsevier Academic Press, International Geophysics Series, v. 90, 301 p.

Banta, E.R., Poeter, E.P., Doherty, J.E., and Hill, M.C., 2006, JUPITER: Joint Universal Parameter IdenTification and Evaluation of Reliability—An application programming interface (API) for model analysis: U.S. Geological Survey Techniques and Methods, book 6, chap. E1, 268 p.

Cardiff, M., and Kitanidis, P.K., 2009, Bayesian inversion for facies detection—An extensible level set framework: Water Resources Research, v. 45, W10416, doi:10.1029/2008wr007675.

Cardiff, M., Barrash, W., and Kitanidis, P.K., 2012, A field proof-of-concept of aquifer imaging using 3-D transient hydraulic tomography with modular, temporarily-emplaced equipment: Water Resources Research, v. 48, no. 5, W05531, doi:10.1029/2011WR011704.

Casella, G., 1985, An introduction to empirical Bayes data-analysis: American Statistician, v. 39, no. 2, p. 83–87, doi:10.2307/2682801.

Chamberlin, T.C., 1890, The method of multiple working hypotheses: Science (Old Series), v. 15, no. 92.

Collinson, J.D., 1969, Sedimentology of Grindslow shales and Kinderscout grit—A deltaic complex in Namurian of Northern England: Journal of Sedimentary Petrology, v. 39, no. 1, p. 194–221.

Condor Team, 2012, Condor Version 7.6.6 Manual: Madison, Wisconsin, University of Wisconsin—Madison.

Deutsch, C.V., and Journel, A.G., 1992, GSLIB—Geostatistical software library and users guide: New York, Oxford University Press, 340 p.

Doherty, J., 2010a, PEST, Model-independent parameter estimation—User manual (5th ed., with slight additions): Brisbane, Australia, Watermark Numerical Computing.

Doherty, J., 2010b, PEST, Model-independent parameter estimation—Addendum to user manual (5th ed.): Brisbane, Australia, Watermark Numerical Computing.

Draper, N.R., and Smith, H., 1966, Applied regression analysis: New York, Wiley, 407 p.

Fienen, M., Kitanidis, P., Watson, D., and Jardine, P., 2004, An application of Bayesian inverse methods to vertical deconvolution of hydraulic conductivity in a heterogeneous aquifer at Oak Ridge National Laboratory: Mathematical Geology, v. 36, no. 1, p. 101–126, doi:10.1023/B:MATG.0000016232.71993.bd.

Fienen, M., Luo, J., and Kitanidis, P., 2006, A Bayesian geostatistical transfer function approach to tracer test analysis: Water Resources Research, v. 42, no. 7, W07426, doi:10.1029/2005WR004576.

Fienen, M.N., Clemo, T.M., and Kitanidis, P.K., 2008, An interactive Bayesian geostatistical inverse protocol for hydraulic tomography: Water Resources Research, v. 44, W00B01, doi:10.1029/2007WR006730.

Fienen, M., Hunt, R., Krabbenhoft, D., and Clemo, T., 2009, Obtaining parsimonious hydraulic conductivity fields using head and transport observations—A Bayesian geostatistical parameter estimation approach: Water Resources Research, v. 45, W08405, doi:10.1029/2008wr007431.

Hill, M.C., 2006, The practical use of simplicity in developing ground water models: Ground Water, v. 44, no. 6, p. 775–781, doi:10.1111/j.1745-6584.2006.00227.x.

Hoeksema, R.J., and Kitanidis, P.K., 1984, An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling: Water Resources Research, v. 20, no. 7, p. 1003–1020, doi:10.1029/WR020i007p01003.

Isaaks, E.H., and Srivastava, R.M., 1989, Applied geostatistics: Oxford, UK; New York; Oxford University Press, 561 p.

Jaynes, E.T., and Bretthorst, G.L., 2003, Probability theory—The logic of science: Cambridge, UK; New York; Cambridge University Press, 727 p.

Kitanidis, P.K., 1995, Quasi-linear geostatistical theory for inversing: Water Resources Research, v. 31, no. 10, p. 2411–2419, doi:10.1029/95WR01945.

Kitanidis, P.K., 1997, Introduction to geostatistics—Applications in hydrogeology:

Cambridge, UK; New York; Cambridge University Press, 249 p.

Kitanidis, P.K., and Vomvoris, E.G., 1983, A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations: Water Resources Research, v. 19, no. 3, p. 677–690, doi:10.1029/WR019i003p00677.

Li, W., Nowak, W., and Cirpka, O.A., 2005, Geostatistical inverse modeling of transient pumping tests using temporal moments of drawdown: Water Resources Research, v. 41, no. 8, p. 1–13, doi:10.1029/2004WR003874.

Li, W., Englert, A., Cirpka, O.A., Vanderborght, J., and Vereecken, H., 2007, Two-dimensional characterization of hydraulic heterogeneity by multiple pumping tests: Water Resources Research, v. 43, no. 4, W04433, doi:10.1029/2006WR005333.

Li, W., Englert, A., Cirpka, O.A., and Vereecken, H., 2008, Three-dimensional geostatistical inversion of flowmeter and pumping test data: Ground Water, v. 46, no. 2, p. 193–201, doi:10.1111/j.1745-6584.2007.00419.x

Michalak, A.M., Bruhwiler, L., and Tans, P.P., 2004, A geostatistical approach to surface flux estimation of atmospheric trace gases: Journal of Geophysical Research, v. 109, no. D14, doi:10.1029/2003jd004422.

Michalak, A.M., and Kitanidis, P.K., 2002, Application of Bayesian inference methods to inverse modeling for contaminant source identification at Gloucester Landfill, Canada, *in* Hassanizadeh, S.M., Schotting, R.J., Gray, W.G., and Pinder, G.F., eds., Proceedings of the Fourteenth International Conference on Computational Methods in Water Resources (CMWR XIV): Amsterdam, Elsevier, v. 2, p. 1259–1266.

Michalak, A.M., and Kitanidis, P.K., 2003, A method for enforcing parameter nonnegativity in Bayesian inverse problems with an application to contaminant source identification: Water Resources Research, v. 39, no. 2, 1033, doi:10.1029/2002WR001480.

Mueller, K.L., Gourdji, S.M., and Michalak, A.M., 2008, Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach; 1. Results using atmospheric measurements: Journal of Geophysical Research-Atmospheres, v. 113, no. D21, doi:10.1029/2007jd009734.

Muffels, C., Schreüder, W., Doherty, J., Karanovic, M., Tonkin, M., Hunt, R., and Welter, D., 2012, Approaches in highly parameterized inversion—GENIE, a general model-independent TCP/IP run manager, U.S. Geological Survey Techniques and Methods, book 7, chap. C6, 26 p.

Neupauer, R.M., and Wilson, J.L., 1999, Adjoint method for obtaining backward-in-time location and travel time probabilities of a conservative groundwater contaminant: Water Resources Research, v. 35, no. 11, p. 3389–3398.

Nowak, W., and Cirpka, O.A., 2004, A modified Levenberg-Marquardt algorithm for quasi-linear geostatistical inversing: Advances in Water Resources, v. 27, no. 7, p. 737–750, doi:10.1016/j.advwatres.2004.03.004.

R Development Core Team, 2011, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, ISBN 3-900051-07-0.

RamaRao, B.S., Lavenue, A.M., de Marsily, G., and Marietta, M.G., 1995, Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields; 1. Theory and computational experiments: Water Resources Research, v. 31, no. 3, p. 475–493.

Remy, N., Boucher, A., and Wu, J., 2009, Applied geostatistics with SGeMS: Cambridge, UK; New York; Cambridge University Press, 264 p.

Robbins, H., 1956, An empirical Bayes approach to statistics, *in* Neyman, J., ed., Proceedings of the Third Berkeley Symposium on Mathematical Statistics: University of California Press, v. 1, p. 157–163.

Rubin, Y., 2003, Applied stochastic hydrogeology: Oxford, UK; New York; Oxford University Press, 391 p.

Samper, F.J., and Neuman, S., 1986, Adjoint state equations for advective-dispersive transport, *in* Sixth International Conference on Finite Elements in Water Resources, p. 423–437.

Schreüder, W., 2009, Running BeoPEST, *in* Proceedings, PEST Conference 2009, Potomac, Md., November 1–3, 2009: Bethesda, Md., S.S. Papadopulos and Associates, p. 228–240.

Snodgrass, M.F., and Kitanidis, P.K., 1997, A geostatistical approach to contaminant source identification: Water Resources Research, v. 33, no. 4, p. 537–546.

Snodgrass, M., and Kitanidis, P., 1998, Transmissivity identification through multi-directional aquifer stimulation: Stochastic Hydrology and Hydraulics, v. 12, no. 5, p. 299–316, doi:10.1007/s004770050023.

Swift, D.J.P., Parsons, B.S., Foyle, A., and Oertel, G.F., 2003, Between beds and sequences—Stratigraphic organization at intermediate scales in the Quaternary of the Virginia coast, USA: Sedimentology, v. 50, no. 1, p. 81–111, doi:10.1046/j.1365-3091.2003.00540.x.

Sykes, J.F., Wilson, J.L., and Andrews, R.W., 1985, Sensitivity analysis for steady state groundwater flow using adjoint operators: Water Resources Research, v. 21, no. 3, p. 359–371, doi:10.1029/WR021i003p00359.

Tikhonov, A.N., 1963a, Solution of incorrectly formulated problems and the regularization method [in Russian]: Soviet Mathematics Doklady, v. 4, p. 1035–1038.

Tikhonov, A.N., 1963b, Regularization of incorrectly posed problems [in Russian]: Soviet Mathematics Doklady, v. 4, p. 1624–1637.

Townley, L., and Wilson, J., 1985, Computationally efficient algorithms for parameter estimation and uncertainty propagation in numerical models of groundwater flow: Water Resources Research, v. 21, no. 12, p. 1851–1860.

Walker, R.G., 1984, General introduction—Facies, facies sequences and facies models, *chap. 1 of* Walker, R.G., Facies models (2d ed.): Toronto, Geological Association of Canada, p. 1–9.

Walker, R.G., 1992, Facies, facies models and modern stratigraphic concepts, *chap. 1 of* Walker, R.G., and James, N.P., Facies models—Response to sea level change: St. Johns, Newfoundland, Geological Association of Canada, p. 1–14.

Westenbroek, S., Doherty, J., Walker, J., Kelson, V., Hunt, R., and Cera, T., 2012, Approaches in highly parameterized inversion—TSPROC, a general time-series processor to assist in model calibration and result summarization: U.S. Geological Survey Techniques and Methods, book 7, chap. C7, 79 p.