Bayesian analysis of data-worth considering model and parameter uncertainties

Shlomo P. Neuman a,⁎, Liang Xue b, Ming Ye b, Dan Lu b

a Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ 85721, United States
b Department of Scientific Computing, Florida State University, Tallahassee, FL 32306, United States

Abstract

The rational management of water resource systems requires an understanding of their response to existing and planned schemes of exploitation, pollution prevention and/or remediation. Such understanding requires the collection of data to help characterize the system and monitor its response to existing and future stresses. It also requires incorporating such data in models of system makeup, water flow and contaminant transport. As the collection of subsurface characterization and monitoring data is costly, it is imperative that the design of corresponding data collection schemes be cost-effective, i.e., that the expected benefit of new information exceed its cost. A major benefit of new data is its potential to help improve one’s understanding of the system, in large part through a reduction in model predictive uncertainty and corresponding risk of failure. Traditionally, value-of-information or data-worth analyses have relied on a single conceptual-mathematical model of site hydrology with prescribed parameters. Yet there is a growing recognition that ignoring model and parameter uncertainties render model predictions prone to statistical bias and underestimation of uncertainty. This has led to a recent emphasis on conducting hydrologic analyses and rendering corresponding predictions by means of multiple models. We describe a corresponding approach to data-worth analyses within a Bayesian model averaging (BMA) framework. We focus on a maximum likelihood version (MLBMA) of BMA which (a) is compatible with both deterministic and stochastic models, (b) admits but does not require prior information about the parameters, (c) is consistent with modern statistical methods of hydrologic model calibration, (d) allows approximating lead predictive moments of any model by linearization, and (e) updates model posterior probabilities as well as parameter estimates on the basis of potential new data both before and after such data become actually available. We describe both the BMA and MLBMA versions theoretically and implement MLBMA computationally on a synthetic example with and without linearization.

1. Introduction

The world’s water supply is threatened by overexploitation and contamination. To manage this supply in an optimal and sustainable manner, it is necessary to understand the response of water resource systems to existing and planned schemes of exploitation and pollution prevention and/or remediation. Such understanding requires the collection of suitable data to help characterize the system and monitor its response to existing and future stresses. It also requires incorporating such data in suitable models of water flow and contaminant transport.

As noted by Back [3], three strategies have traditionally been used to determine the magnitude of a data collection effort: minimizing cost for a specific level of accuracy or precision, minimizing uncertainty for a given budget, or responding to regulatory demands on data quantity and quality. Various combinations of these strategies have also been described such as a fitness-for-purpose approach [55]. Many today prefer a fourth approach based on value-of-information or data-worth analysis. Here the decision to collect additional data, or the design of a data collection program, is based on cost-effectiveness. A program is considered cost-effective if the expected benefit from the new information exceeds the cost. A major benefit of new data is its potential to help improve one’s understanding of the system, in large part through a reduction in model predictive uncertainty. This benefit, however, is worth the cost only if it has the potential to impact decisions concerning management of the water resource system.

Value-of-information or data-worth analyses incorporating statistical decision theory have been applied to various water-related problems in the 1970s [12,13,23,36] and to groundwater problems in the 1980s [4,22,33,47]. More recent applications to groundwater resource and contamination issues have been reported in [1,11,21,27–30,37,45,50,51]. James and Freeze [28] proposed a Bayesian decision-making framework to evaluate the worth of data in the context of contaminated groundwater that has been widely cited in the subsequent literature. A comprehensive review focusing on health risk assessment can be found in [61]. Additional recent publications of relevance include [18,41].
A major limitation of many existing approaches is that they rely on a single conceptual-mathematical model of geologic or watershed makeup and of hydrologic processes therein. Yet hydrologic environments are open and complex, rendering them prone to multiple interpretations and mathematical descriptions, including parameterizations. This is true regardless of the quantity and quality of available data. Predictions and analyses of uncertainty based on a single hydrologic concept are prone to statistical bias (caused by reliance on an inadequate model) and underestimation of uncertainty (caused by under-sampling of the relevant model space). Analyses of environmental data-worth which explore how different sets of conditioning data impact the predictive uncertainty of multiple models in a Bayesian context include [19,49,54]; whereas Frer et al. [19] employ Generalized Likelihood Uncertainty Estimation (GLUE; see [5,6]), Rojas et al. [49] combine GLUE with Bayesian model averaging (BMA; [7,17,25,32,34]). Diggle and Lophaven [16] describe a Bayesian approach to geostatistical design. Nowak et al. [43] introduce a Bayesian approach to data worth analysis when flow and transport take place in a random log hydraulic conductivity field. Whereas flow and transport are described by a single (linearized) model each having known parameters, other than those describing spatial variations in log hydraulic conductivity, the latter is characterized by a single drift model and a continuous family of variogram models having uncertain parameters.

In a similar spirit, we propose in this paper a multimodel approach to optimum value-of-information or data-worth analyses that is based on model averaging within a Bayesian framework. Our approach is general in that it considers multiple models of any kind, all having uncertain parameters; whereas parameterizing models in the manner of [43] is elegant and computationally efficient, it is unfortunately limited to a narrow range of variogram models and does not, generally, apply to other models such as those of flow and transport. We prefer BMA over GLUE because it (a) rests on rigorous statistical theory, (b) is compatible with deterministic as well as stochastic models and, (c) in its maximum likelihood (ML) version (MLBMA), is consistent with current ML methods of hydrologic model calibration [35,38–40,57–60]. Whereas BMA (like the closely related approach in [43]) relies heavily on prior parameter statistics, MLBMA can do without such statistics or otherwise update them on the basis of potential new data both before and after they are collected. We describe the proposed BMA and MLBMA approaches theoretically, outline ways to implement the MLBMA version computationally, and illustrate the latter on a synthetic example. Our proposed methodology should be of help in designing the collection of hydrologic characterization and monitoring data in a cost-effective manner by maximizing their benefit under given cost constraints. The benefit would accrue from optimum gain in information, or reduction in predictive uncertainty, upon considering jointly not only traditional sources of uncertainty such as those affecting model parameters and the reliability of data but also, most importantly, lack of certainty about the conceptual-mathematical models that underlie the analysis and the scenarios under which the system would operate in the future. The methodology should apply to a broad range of models representing natural processes in ubiquitously open and complex earth and environmental systems.

2. Background

2.1. Bayesian decision analysis framework

One way to cast the data-worth issue would be within a Bayesian risk-cost-benefit decision framework such as that of Freeze et al. [20,21]. Suppose without loss of generality that the data are intended to help one decide whether or not a contaminated site should be remediated. This decision problem is illustrated in Fig. 1 [3] by a decision tree in which $\Phi$ is the decision objective; $\Phi_i$ is an objective function associated with each decision alternative ($i = 1, 2$) defined as

$$\Phi_i = B_i - C_i - \gamma P_i C_f,$$

where $B_i$ is the benefit and $C_i$ the investment cost, risk being expressed as the product $\gamma P_i C_f$ of a risk aversion factor $\gamma$, the probability of failure $P_i$ and the cost of failure $C_f$. $C$ designates a contaminated and $C$ an uncontaminated state of the site; costs and benefits occurring at the triangular terminal nodes (only the cost of failure is indicated in the figure). Collecting additional information generally causes the risk term to decrease due to a decrease in the probability of failure. The corresponding increase in $\Phi_i$ is the expected value (worth) of the new data. The final outcome of the analysis depends on the choice of decision rule one adopts [42]; for example, maximizing $\Phi_i$ would result in the largest benefit and lowest cost.

According to Back [3] the Bayesian approach to data-worth analysis entails five steps as illustrated in Fig. 2. The steps include (1) defining one or more data collection (sampling) programs, (2) postulating a prior probability for the state of the site (e.g., contaminated or uncontaminated), (3) using Bayes’ theorem to update the prior probability to a posterior probability conditional on the new data, corresponding to each data collection program, (4) estimating corresponding costs and benefits, and (5) computing the worth of data or value of information using a given decision model and using the results to optimize the data collection scheme. This work also considers using linearized estimation of uncertainty to update the prediction variance in step 3. We add that data-worth analyses actually include a pre-posterior and a posterior mode. In the posterior mode, a given set of data is evaluated in hindsight, after spending the money to collect it. In the pre-posterior mode, possible sampling schemes are analyzed for the worth of data not yet collected. This mode requires averaging over all such data, weighted by their pre-posterior probabilities, as we do below.
As noted, collecting additional information generally reduces risk due to a decrease in the probability of failure. A reduction in the probability of failure comes about through a reduction in uncertainty about the expected system state, present or future. The impact of hydrologic data on this expectation and the associated uncertainty are often evaluated by means of a hydrologic model. Commonly, the model is considered to be certain while its parameters (and in some cases its forcing terms such as sources and boundary conditions) are treated as being uncertain due to insufficient and error-prone data [18]. As already noted, we know of only one work that considers the impact of data on model predictive uncertainty within a Bayesian framework by considering the model itself to be uncertain [19] and another work that parameterizes this uncertainty [43]. Below we provide background about Bayesian model averaging and its maximum likelihood version which we propose to employ for this same purpose.

2.2. Bayesian model averaging (BMA)

Consider a random vector, A, the multivariate statistics of which are to be predicted with a set M of K mutually independent models, Mk, each characterized by a vector of parameters θk, conditional on a discrete set of data, D (the case of correlated models has recently been considered in [52]). In analogy to the case of a scalar A [25] we write the joint posterior (conditional) distribution of A as

\[ p(A|D) = \sum_{k=1}^{K} p(A|M_k)p(M_k|D). \]

(2)

i.e., as the average over all models of the joint posterior distributions p(A|M_k) associated with individual models, weighted by the model posterior probabilities p(M_k|D). These weights are given by Bayes’ rule in the form

\[ p(M_k|M) = \frac{p(D|M_k)p(M_k)}{\sum_{l=1}^{K} p(D|M_l)p(M_l)}. \]

(3)

where

\[ p(D|M_k) = \int p(D|M_k, \theta_k)p(\theta_k|M_k)d\theta_k \]

is the integrated likelihood of model M_k, p(D|M_k, \theta_k) being the joint likelihood of this model and its parameters, p(\theta_k|M_k) the prior density of \theta_k under model M_k, and p(M_k) the prior probability of M_k. The likelihood p(D|M_k, \theta_k) contains a model of D measurement errors, and the prior density p(\theta_k|M_k) may contain a model of parameter measurement errors [8]. All probabilities are implicitly conditional on the choice of models entering into the set M.

The posterior mean and covariance of A are given, through extension of Draper’s [17] analysis to our multivariate case, by

\[ E(A|D) = \sum_{k=1}^{K} E(A|M_k)p(M_k|D), \]

(5)

\[ \text{Corr}(A|D) = \sum_{k=1}^{K} \text{Corr}(A|M_k)p(M_k|D) + \sum_{k=1}^{K} [E(A|D,M_k) - E(A|D)]^T p(M_k|D), \]

(6)

where the superscript T denotes transpose. Eq. (6) is a discrete expression of the law of total covariance,

\[ \text{Corr}(A|D) = E_{M_k\theta_k} \text{Corr}(A|M_k, \theta_k) + \text{Corr}_{M_k\theta_k} E(A|M_k, \theta_k), \]

(7)

where E_{M_k\theta_k} \text{Corr}(A|M_k, \theta_k) is the within-model component of \text{Corr}(A|D) and Corr_{M_k\theta_k} E(A|M_k, \theta_k) is its between-model component. We will also be interested in the trace

\[ \text{Tr}[\text{Corr}(A|D)] = \text{Tr}[E_{M_k\theta_k} \text{Corr}(A|M_k, \theta_k)] + \text{Tr}[\text{Corr}_{M_k\theta_k} E(A|M_k, \theta_k)] \]

(8)

which provides a scalar measure of the posterior variance of A. The latter is of interest because, for K ≥ 1, one generally has \text{Tr}[\text{Corr}_{M_k\theta_k} E(A|M_k, \theta_k)] > 0 so that \text{Tr}[\text{Corr}(A|D)] > \text{Tr}[E_{M_k\theta_k} \text{Corr}(A|M_k, \theta_k)]. Hence the consideration of multiple models generally results in greater predictive uncertainty, as measured by \text{Tr}[\text{Corr}(A|D)], than the uncertainty associated with a single model, as measured by \text{Tr}[E_{M_k\theta_k} \text{Corr}(A|M_k, \theta_k)].

Note that, for any consistent norm ||.||, ||\text{Corr}(A|D)|| ≤ ||E_{M_k\theta_k} \text{Corr}(A|M_k, \theta_k)|| + ||\text{Corr}_{M_k\theta_k} E(A|M_k, \theta_k)|| where the equality holds if and only if the two right hand side arguments are linearly dependent, which is generally not the case. It follows that ||\text{Corr}_{M_k\theta_k} E(A|M_k, \theta_k)|| > 0 does not generally imply ||\text{Corr}(A|D)|| > ||E_{M_k\theta_k} \text{Corr}(A|M_k, \theta_k)||.

2.3. Maximum likelihood Bayesian model averaging (MLBMA)

BMA defines the integrated likelihood p(D|M_k) of model M_k entirely in terms of the prior parameter density p(\theta_k|M_k) of model parameters, having thus no provision for the conditioning of model parameters on measurements D (i.e., for the estimation of optimum model parameters on the basis of D using inverse methods). Instead, it requires computing the integral in (4) through exhaustive sampling of the parameter space \theta_k for each model followed by numerical integration. One way to resolve both issues is to replace \theta_k by an estimate, \hat{\theta}_k, which maximizes the likelihood p(D|M_k, \theta_k). Obtaining such maximum likelihood (ML) estimates entails calibrating each model against (conditioning on) the data D using well-established statistical inverse methods [8–10,24,44,48]. Approximating p(D|M_k, \theta_k) in (2) by p(D|M_k, \hat{\theta}_k), where the sub-script indicates approximation based on ML estimation of \theta_k, was shown to be useful (for the case of a scalar A) in [17,46,56]. Neuman [40] proposed evaluating the weights p(M_k|D) in (2), (5), and (6) based on a result due to Kashyap [31]. The corresponding expression is [57]

\[ p(M_k|D) \approx p(M_k|D)_\text{ML} = \frac{\exp \left( -\frac{1}{2} \delta KIC_k^D \right) p(M_k)}{\sum_{k=1}^{K} \exp \left( -\frac{1}{2} \delta KIC_k^D \right) p(M_k)}, \]

(9)

where

\[ \delta KIC_k^D = KIC_k^D - KIC_k^D\text{min}. \]

(10)

\[ KIC_k^D = -2 \ln p(D|M_k, \hat{\theta}_k)_\text{ML} - 2 \ln p(\hat{\theta}_k|M_k)_\text{ML} \]

\[ + N_k \ln \left( \frac{N^D}{2\pi} \right) + \ln |F_k(D|M_k)|_\text{ML}. \]

(11)

KIC_k^D being the so-called Kashyap model selection (or information) criterion for model M_k, KIC_k^D\text{min} its minimum value over all candidate models, and -2lnp(D|M_k, \hat{\theta}_k)_\text{ML} - 2lnp(\hat{\theta}_k|M_k)_\text{ML} a negative log-likelihood incorporating prior measurements of the parameters (if available; see [8]), evaluated at \hat{\theta}_k. Here N_k is the dimension of \theta_k (number of adjustable parameters associated with model M_k), N^D is the dimension of D (number of discrete data points, which may include measured parameter values), and F_k is the normalized (by N^D observed (as opposed to ensemble mean) Fisher information matrix having components

\[ F_{k,jj} = -\frac{1}{N^D} \left[ \frac{\partial^2 \ln p(D|M_k, \theta_j)}{\partial \theta_j \partial \theta_j} \right]_{\theta_j = \hat{\theta}_j}. \]

(12)

In the limit of large N^D/N_k, KIC_k^D reduces asymptotically to the so-called Bayesian selection (or information) criterion (e.g. [59])

\[ BIC_k^D = -2 \ln p(D|M_k, \hat{\theta}_k)_\text{ML} + N_k \ln N^D. \]

(13)
3. Effect of data augmentation on uncertainty

3.1. BMA framework

Suppose that the original data set $\mathbf{D}$ is augmented by another hypothetical data set, which has not yet been collected and is therefore uncertain. Assume that the multivariate statistics of $\mathbf{D}$, predicted with the model set $\mathbf{M}$, can be conditioned on the augmented data set $\{\mathbf{D}, \mathbf{C}\}$. Assume further that the multivariate statistics of $\mathbf{C}$, conditional on $\mathbf{D}$, can be predicted either via BMA or via MLBMA with a set $\mathbf{P}$ of $I$ mutually independent statistical models, $P_i$, having parameters $\pi_i$. The models $P_i$ may be independent of $M_k$, may form extensions of $M_k$ or may coincide with the latter as in the computational example given later in this paper. Then analogy to (2) and (5)–(8) implies

$$p(\mathbf{A} | \mathbf{D}) = E_{\mathbf{C} \in \mathbf{P}} p(\mathbf{A} | \mathbf{D}, \mathbf{C}),$$

$$p(\mathbf{A} | \mathbf{D}, \mathbf{C}) = \sum_{k=1}^{K} p(\mathbf{A} | \mathbf{D}, \mathbf{C}, M_k) p(M_k | \mathbf{D}, \mathbf{C}),$$

$$E(\mathbf{A} | \mathbf{D}) = E_{\mathbf{C} \in \mathbf{P}} E(\mathbf{A} | \mathbf{D}, \mathbf{C}),$$

$$E(\mathbf{A} | \mathbf{D}, \mathbf{C}) = \sum_{k=1}^{K} E(\mathbf{A} | \mathbf{D}, \mathbf{C}, M_k) p(M_k | \mathbf{D}, \mathbf{C}),$$

$$Cov(\mathbf{A} | \mathbf{D}, \mathbf{C}) = E_{\mathbf{C} \in \mathbf{P}} Cov(\mathbf{A} | \mathbf{D}, \mathbf{C}) + Cov_{\mathbf{C} \in \mathbf{P}} E(\mathbf{A} | \mathbf{D}, \mathbf{C}).$$

$$Tr[Cov(\mathbf{A} | \mathbf{D})] = Tr[E_{\mathbf{C} \in \mathbf{P}} Cov(\mathbf{A} | \mathbf{D}, \mathbf{C})] + Tr[Cov_{\mathbf{C} \in \mathbf{P}} E(\mathbf{A} | \mathbf{D}, \mathbf{C})],$$

where $\mathbf{C}$ is implicitly conditional on the choice of models $\mathbf{P}$ and on $\mathbf{D}$. The term $Tr[Cov_{\mathbf{C} \in \mathbf{P}} E(\mathbf{A} | \mathbf{D}, \mathbf{C})] = Tr[Cov(\mathbf{A} | \mathbf{D})] - Tr[E_{\mathbf{C} \in \mathbf{P}} Cov(\mathbf{A} | \mathbf{D}, \mathbf{C})]$ represents the difference between the total trace conditional on $\mathbf{D}$ and the expected trace conditional jointly on $\mathbf{D}$ and $\mathbf{C}$. As this difference is positive, conditioning on $\mathbf{D}$ and $\mathbf{C}$ jointly results in a lower trace than conditioning on $\mathbf{D}$ alone. The difference could be viewed as an extended version of the A-criterion in optimal design.

The joint posterior distribution (weight) of each model is given, analogously to (3), by Bayes’ rule

$$p(M_k | \mathbf{D}, \mathbf{C}) = \frac{p(\mathbf{D}, \mathbf{C} | M_k) p(M_k)}{\sum_{l=1}^{K} p(\mathbf{D}, \mathbf{C} | M_l) p(M_l)},$$

where, in analogy to (4), the joint integrated likelihood of $\mathbf{D}$ and $\mathbf{C}$ is

$$p(\mathbf{D}, \mathbf{C} | M_k) = \int p(\mathbf{D}, \mathbf{C} | M_k, \theta_k) p(\theta_k | M_k) d\theta_k.$$  

The joint likelihood of $\mathbf{D}$ and $\mathbf{C}$ under the integral lends itself to Bayesian updating according to

$$p(\mathbf{D}, \mathbf{C} | \theta_k) = p(\mathbf{D} | \mathbf{C}, \theta_k) p(\mathbf{C} | \theta_k).$$

3.2. MLBMA framework

Let $\theta^*_{\text{ML}}$ be an ML estimate of $\theta_k$ obtained by maximizing the joint likelihood $p(\mathbf{D}, \mathbf{C} | \theta_k)$ under the integral in (22) with respect to these parameters. Then MLBMA would entail approximating Eqs. (15), (17), (19)–(21) via

$$p(M_k | \mathbf{D}, \mathbf{C}) = \frac{p(\mathbf{D}, \mathbf{C} | M_k) p(M_k)}{\sum_{l=1}^{K} p(\mathbf{D}, \mathbf{C} | M_l) p(M_l)},$$

$$E(\mathbf{A} | \mathbf{D}, \mathbf{C}) = \sum_{k=1}^{K} E(\mathbf{A} | \mathbf{D}, \mathbf{C}, M_k) p(M_k | \mathbf{D}, \mathbf{C}).$$

$$Cov(\mathbf{A} | \mathbf{D}, \mathbf{C}) = Cov(\mathbf{A} | \mathbf{D}, \mathbf{C})_{\text{ML}} = \sum_{k=1}^{K} Cov(\mathbf{A} | \mathbf{D}, \mathbf{C}, M_k)_{\text{ML}} p(M_k | \mathbf{D}, \mathbf{C})_{\text{ML}},$$

$$E(\mathbf{A} | \mathbf{D}) = E(\mathbf{A} | \mathbf{D})_{\text{ML}} = \sum_{k=1}^{K} E(\mathbf{A} | \mathbf{D}, \mathbf{C}, M_k)_{\text{ML}} p(M_k | \mathbf{D}, \mathbf{C})_{\text{ML}}.$$
where $s = |x_1 - x_2|$ is separation distance (lag) between values at any two points $x_1$ and $x_2$, $\lambda_0$ is an upper cutoff scale proportional to domain size, $A$ is a coefficient, $H$ is a Hurst scaling exponent, $\sigma^2(\lambda_0) = A \lambda_0^{-2H}/2H$ is variance (sill) and $I(\cdot)$ is the incomplete gamma function. The corresponding integral scale is $l(\lambda_0) = 2H \lambda_0 (1 + 2H)$. We set the parameters of the TpvG model equal to $\theta = (A, H, \lambda_0)^T = (0.1, 0.25, 5)^T$ which correspond to $\sigma^2 = 0.45$ and $I = 1.67$. We then generate a “true” sample of 2500 $Z$ values at $50 \times 50$ nodes of a square grid, spaced a unit distance apart, as shown in Fig. 3. After verifying that a sample variogram based on all the generated values reproduces the original TpvG very closely we select 100 $Z$ values at randomly located nodes to comprise a vector $D$ of “available” data, 20 values to form a vector $C$ ($C$ representing true but unknown data values, $C$ their estimates) at other randomly located “potential new” sampling nodes, and those at the remaining 2380 nodes to make up a vector $A$ of “unknown” values that we wish to predict via MLMA. The latter is based either on $D$ or on $[D, C]$ where $C$ are values simulated randomly at the “potential new” sampling nodes conditional on $D$.

To predict $A$ we consider a set $M$ of $K = 3$ alternative variogram models, $M_k$, having parameters $\theta_k$ (purposely excluding the generating, or “true,” TpvG model): exponential (Exp), Gaussian (Gau) and spherical (Sph). Each model, $M_k$, is assigned an equal prior probability, $p(M_k) = 1/3$, and is calibrated against $D$ to yield ML estimates $\hat{\theta}_k$ of $\theta_k$ by minimizing the joint negative log likelihood $-2 \ln p(D|M_k, \theta_k) = 2 \ln p(\theta_k|M_k)$. The process, denoted by ML$^2$ as explained in Appendix A, also yields corresponding parameter estimation covariance matrices $\mathbf{T}^{\hat{\theta}_k}$, Khyashp criteria $KIC^2_k$, according to (11) and Bayesian criteria $BIC^2_k$ according to (13). For comparison we also compute information theoretic criteria

$$AIC^0_k = -2 \ln p(D|M_k)_{ML} + 2N_k,$$

$$AIC^P_k = -2 \ln p(D|M_k)_{ML} + 2N_k + \frac{2N_k(N_k + 1)}{N_k^2 - N_k - 1}$$

introduced, respectively, by Akaike [2] and Hurvich and Tsai [26]. $KIC^2_k$ and $BIC^2_k$ are used to compute posterior model probabilities (or, in the case of $AIC^2_k$ and $AIC^P_k$, model averaging weights), $p(M_k|D)_{ML^2}$, for each model via (9). The results are listed in Table 1 and the corresponding fits illustrated in Fig. 4. It is evident that the sample $D$ of available data is not large enough to reproduce correctly the TpvG model used to generate it; the corresponding sample and fitted variograms underestimate the true sill and overestimate the true integral scale. All criteria favor the spherical model, assigning a very low posterior probability (or weight) to the Gaussian model. Our model averaged results are based on $KIC^2_k$.

Augmenting the sample to include $[D, C]$ is still not enough to reproduce correctly the TpvG model; the sample and fitted variograms underestimate the true sill to a greater extent than was the case with $D$ alone but overestimate the true integral scale to a lesser extent (compare Figs. 4 and 5, and Tables 1 and 2). The preference of all criteria for the spherical model is now more pronounced (less ambiguous) than it was in the case of $D$. These results indicate a need to account for the effect of potential new data on parameter estimation and model weighting, as we do next by following the procedure in Appendix A.

Predictions of $A$ are uncertain due to random spatial fluctuations in $Z$ as well as uncertainty about the variogram model, $M_k$ and its parameters, $\theta_k$. As $A$ is generally nonlinear in $\theta_k$, one can estimate its lead moments either through linearization, as described in Appendix B, or via Monte Carlo simulation. We start with the latter option by (a) drawing $R_0 = 2000$ random realizations, $\theta_k^*$, of $\theta_k$ from a multivariate normal distribution $\theta_k \sim N(\theta_k^0, \mathbf{T}^{\hat{\theta}_k})$ where $r_0 = 1, 2, \ldots, R_0$ for each $M_k$; (b) obtaining kriging estimates $E(\Delta_p|D, M_k, \theta_k^*)_{ML^2}$ and kriging variances $\text{Var}(\Delta_p|D, M_k, \theta_k^*)_{ML^2}$ and covariances $\text{Cov}(\Delta_p, \Delta_q|D, M_k, \theta_k^*)_{ML^2}$ for all components $\Delta_p$ and $\Delta_q$ of $A$; and (c) averaging these over all $R_0$ realizations to obtain

$$E(\Delta_p|D, M_k)_{ML^2} = E_{\theta_k} E(\Delta_p|D, M_k, \theta_k)_{ML^2}$$

$$\approx \frac{1}{R_0} \sum_{r=1}^{R_0} E(\Delta_p|D, M_k, \theta_k^*)_{ML^2} \quad \text{covariances}$$

$$\text{Cov}(\Delta_p, \Delta_q|D, M_k)_{ML^2} = E_{\theta_k} \left[ \text{Cov}(\Delta_p, \Delta_q|D, M_k, \theta_k)_{ML^2} \right]$$

$$+ \text{Cov}_{\theta_k} \left[ E(\Delta_p, \Delta_q|D, M_k, \theta_k)_{ML^2} \right].$$

Table 1

<table>
<thead>
<tr>
<th>Model</th>
<th>Exp</th>
<th>Gau</th>
<th>Sph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sill estimate</td>
<td>0.404</td>
<td>0.233</td>
<td>0.283</td>
</tr>
<tr>
<td>Std of sill</td>
<td>2.26</td>
<td>2.26</td>
<td>2.26</td>
</tr>
<tr>
<td>Integral scale</td>
<td>3.686</td>
<td>1.979</td>
<td>2.681</td>
</tr>
<tr>
<td>Std of integral scale</td>
<td>0.574</td>
<td>0.187</td>
<td>0.283</td>
</tr>
<tr>
<td>NLL</td>
<td>−145.14</td>
<td>−136.13</td>
<td>−147.23</td>
</tr>
<tr>
<td>NLL rank</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>$p(M_k)$</td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
</tr>
<tr>
<td>$AIC$</td>
<td>−139.14</td>
<td>−130.13</td>
<td>−141.23</td>
</tr>
<tr>
<td>$AIC^P$</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>$p(M_k</td>
<td>D)_{ML^2}$</td>
<td>25.98%</td>
<td>25.98%</td>
</tr>
<tr>
<td>$AIC_c$</td>
<td>−138.91</td>
<td>−129.90</td>
<td>−141.00</td>
</tr>
<tr>
<td>$AIC^P_c$</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>$p(M_k</td>
<td>D</td>
<td>_{ML^2}$</td>
<td>25.98%</td>
</tr>
<tr>
<td>$KIC$</td>
<td>−146.89</td>
<td>−135.94</td>
<td>−147.29</td>
</tr>
<tr>
<td>$KIC^P$</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>$p(M_k</td>
<td>D</td>
<td>_{ML^2}$</td>
<td>44.92%</td>
</tr>
</tbody>
</table>
where

\[ \text{E}_n \left[ \text{Cov} \left( \Delta y | \mathbf{D}, M_k, \theta_k^c \right) \right] \approx \frac{1}{R_y} \sum_{j=1}^{R_y} \text{Cov} \left( \Delta y | \mathbf{D}, M_k, \theta_k^c \right) \]  

and [53]

\[ \text{Cov} \left( \Delta y | \mathbf{D}, M_k, \theta_k^c \right) = \text{Cov} \left( \mathbf{X}_y - \mathbf{X}_y | \mathbf{M}_k, \theta_k^c \right) \]

\[ + \sum_{m=m_1}^{m_2} \sum_{n=n_1}^{n_2} \lambda_{mn} \lambda_{nm} \text{Cov} \left( \mathbf{X}_m - \mathbf{X}_n | \mathbf{M}_k, \theta_k^c \right) \]

\[ - \sum_{m=m_p} \lambda_{mp} \text{Cov} \left( \mathbf{X}_m - \mathbf{X}_p | \mathbf{M}_k, \theta_k^c \right) \]

\[ - \sum_{n=n_p} \lambda_{np} \text{Cov} \left( \mathbf{X}_n - \mathbf{X}_p | \mathbf{M}_k, \theta_k^c \right). \]  

Fig. 5. TpG, sample and fitted variograms based on \( \mathbf{D} \); see statistics in Table 1.

Fig. 5. TpG, sample and fitted variograms based on \( \mathbf{D}, \mathbf{C} \); see statistics in Table 2.

Table 2
Parameter estimates; negative log likelihoods NLL; model selection criteria AIC, AICc, BIC and KIC; prior and posterior model probabilities; and rankings of variogram models based on \( \mathbf{D}, \mathbf{C} \).

<table>
<thead>
<tr>
<th>Model</th>
<th>Exp</th>
<th>Gau</th>
<th>Sph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sill estimate</td>
<td>0.371</td>
<td>0.372</td>
<td>0.373</td>
</tr>
<tr>
<td>Std of sill</td>
<td>0.214</td>
<td>0.220</td>
<td>0.210</td>
</tr>
<tr>
<td>Integral scale estimate</td>
<td>2.409</td>
<td>1.907</td>
<td>2.672</td>
</tr>
<tr>
<td>Std of integral scale</td>
<td>0.403</td>
<td>0.151</td>
<td>0.311</td>
</tr>
<tr>
<td>NLL</td>
<td>-183.86</td>
<td>-175.59</td>
<td>-188.36</td>
</tr>
<tr>
<td>NLL rank</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>( p(M_k) )</td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
</tr>
<tr>
<td>AIC</td>
<td>-177.86</td>
<td>-169.59</td>
<td>-182.36</td>
</tr>
<tr>
<td>AIC rank</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>( p(M_k</td>
<td>\mathbf{D}, \mathbf{C} )</td>
<td>9.52%</td>
<td>0.15%</td>
</tr>
<tr>
<td>( p(M_k</td>
<td>\mathbf{D}, \mathbf{C} )</td>
<td>-177.67</td>
<td>-169.40</td>
</tr>
<tr>
<td>( p(M_k</td>
<td>\mathbf{D}, \mathbf{C} )</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>( p(M_k</td>
<td>\mathbf{D}, \mathbf{C} )</td>
<td>9.52%</td>
<td>0.15%</td>
</tr>
<tr>
<td>BIC</td>
<td>-169.25</td>
<td>-160.99</td>
<td>-173.76</td>
</tr>
<tr>
<td>BIC rank</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>( p(M_k</td>
<td>\mathbf{D}, \mathbf{C} )</td>
<td>9.52%</td>
<td>0.15%</td>
</tr>
<tr>
<td>KIC</td>
<td>-184.80</td>
<td>-174.87</td>
<td>-188.51</td>
</tr>
<tr>
<td>KIC rank</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>( p(M_k</td>
<td>\mathbf{D}, \mathbf{C} )</td>
<td>13.51%</td>
<td>0.09%</td>
</tr>
</tbody>
</table>

* Std represents standard deviation.

without neglecting the effect of $C$ on parameter and model uncertainties. Fig. 9 verifies that $E(\mathbf{C|D})_{\text{MLBMA}}$ represents a smoothed version of $\mathbf{C}$, which in turn is well within the range of simulated $\mathbf{C}^c$ realizations.

For every realization $\mathbf{C}^c$ we predict $\mathbf{A}$ the same way as before but now conditional on an expanded data base $\{\mathbf{D, C}^c\}$. This yields $E(\mathbf{A|D, C}^c)_{\text{MLBMA}}$ and $\text{Cov}(\mathbf{A|D, C}^c)_{\text{MLBMA}}$, where the subscript $\text{MLBMA}$ denotes ML estimation based on $\{\mathbf{D, C}^c\}$. The latter are then averaged over all realizations of $\mathbf{C}^c$ to obtain

$$E(\mathbf{A|D})_{\text{MLBMA}} = E_{\text{MC}}E(\mathbf{A|D, C}^c)_{\text{MLBMA}} \approx \frac{1}{R_c} \sum_{r = 1}^{R_c} E(\mathbf{A|D, C}^c)_{\text{MLBMA}},$$

(44)

and $\text{Tr}[\text{Cov}(\mathbf{A|D})_{\text{MLBMA}}]$. The final step entails computing $\text{Tr}[\text{Cov}(\mathbf{A|D})_{\text{MC}}]$ which, we recall, represents the variance reduction measure $\text{Tr}[\text{Cov}(\mathbf{A|D})_{\text{MLBMA}}]$. We verify that sample estimates of all these three terms stabilize after 200 realizations. Variations of $\text{Var}(\mathbf{A|D})_{\text{MLBMA}}$, $E_{\text{MC}}\text{Var}(\mathbf{A|D, C}^c)_{\text{MLBMA}}$ and $\text{Var}_{\text{MC}}(\mathbf{A|D, C}^c)_{\text{MLBMA}}$ across the grid are shown, respectively. In Figs. 10–12. A comparison of Figs. 6 and 10 reveals that even though $\text{Var}(\mathbf{A|D})_{\text{MLBMA}}$ tends to exceed $\text{Var}(\mathbf{A|D})_{\text{MC}}$, the ratio of $\text{Tr}[\text{Cov}(\mathbf{A|D})_{\text{MLBMA}}] = 712.95$, while $\text{Tr}[\text{Cov}(\mathbf{A|D})_{\text{MC}}] = 682.44$ due, most likely, to sampling errors stemming from size and accuracy limitations on $\mathbf{D}$. The two quantities have near identical spatial patterns. Likewise, though $E_{\text{MC}}\text{Var}(\mathbf{A|D, C}^c)_{\text{MLBMA}}$, in Fig. 11 tends to exceed $\text{Var}(\mathbf{A|D, C}^c)_{\text{MLBMA}}$, in Fig. 7 these too have near identical spatial patterns. One therefore expects the estimated variance reduction $E_{\text{MC}}\text{Var}(\mathbf{A|D, C}^c)_{\text{MLBMA}}$ to exhibit a pattern similar to that of the true variance reduction $\text{Var}(\mathbf{A|D})_{\text{MLBMA}} - \text{Var}(\mathbf{A|D, C}^c)_{\text{MLBMA}}$, a fact verified through a comparison of Figs. 12 and 8. Correspondingly $\text{Tr}[\text{Cov}_{\text{MC}}(\mathbf{A|D, C}^c)_{\text{MLBMA}}] = 54.61$ approximately closely the true trace reduction $\text{Tr}[(\mathbf{A|D})_{\text{MLBMA}}] - \text{Tr}[\text{Cov}(\mathbf{A|D, C}^c)_{\text{MLBMA}}] = 58.95$.

The above results are based on Monte Carlo evaluation of all moments. Estimating the lead moments of $\mathbf{A}$ through linearization as described in Appendix B brings about a 10-fold reduction in central processor time without any serious effect on accuracy: the value of $\text{Tr}[\text{Cov}_{\text{MC}}(\mathbf{A|D, C}^c)_{\text{MLBMA}}]$ drops from 54.61 to 54.54 and that of $\text{Tr}[\text{Cov}(\mathbf{A|D})_{\text{MLBMA}}] - \text{Tr}[\text{Cov}(\mathbf{A|D, C}^c)_{\text{MLBMA}}]$ from 58.95 to 55.3, all spatial patterns remaining virtually unchanged. A visual comparison of $\text{Tr}[\text{Cov}(\mathbf{A|D})_{\text{MLBMA}}], \text{Tr}[\text{Cov}(\mathbf{A|D, C}^c)_{\text{MLBMA}}], \text{Tr}[\text{Cov}(\mathbf{A|D})_{\text{MC}}]$, and $\text{Tr}[\text{Cov}(\mathbf{A|D, C}^c)_{\text{MC}}]$ values obtained by the two methods is provided in Fig. 13.

5. Conclusions

Our paper leads to the following major conclusions:

1. A multimodel approach to optimum value-of-information or data-worth analyses has been proposed based on a Bayesian model averaging (BMA) framework. We have focused on a maximum likelihood (MLBMA) variant of BMA that (a) is compatible with both deterministic and stochastic models, (b) admits but does not require prior information about the parameters, (c) is consistent with modern statistical methods of hydrologic model calibration, (d) allows approximating lead predictive moments of any model by linearization, and (e) updates model posterior probabilities as well as parameter estimates on the basis of potential new data both before and after such data become actually available.

2. The proposed approach should be of help in designing the collection of hydrologic characterization and monitoring data in a cost-effective manner by maximizing their benefit under given cost constraints. Benefits would accrue from optimum gain in information, or reduction in predictive uncertainty, upon considering jointly not only traditional sources of uncertainty such as those affecting model parameters and the reliability of data but also lack of certainty about the underlying models.

3. Implementation of the proposed approach on a synthetic geostatistical problem in two space dimensions demonstrates a need to account for the impact of potential new data on model and parameter uncertainties. Though neither existing nor a potentially augmented set of data are sufficient to identify correctly the underlying geostatistical model (variogram) and its parameters, they nevertheless yield self-consistent results and allow identifying quite accurately the impacts of potential new data on the spatial distribution and magnitude of corresponding reductions in predictive variance.

4. Approximating lead predictive moments associated with each model by linearization, as described in Appendix B, has yielded results comparable to those obtained via Monte Carlo simulation with a much lesser expenditure of computational effort. The extent to which such linearization would work in strongly nonlinear situations remains an open question.
Acknowledgements

This research was supported in part through a contract between the University of Arizona and Vanderbilt University under the Consortium for Risk Evaluation with Stakeholder Participation (CRESPP III, funded by the US Department of Energy. The third and fourth authors were supported in part by NSF-EAR Grant 0911074 and DOE-ERSP Grant DE-SC0002687.

Appendix A. Computational implementation of MLBMA framework

To assess the impact of data augmentation within the above MLBMA framework computationally we propose the following approach:

1. Postulate a set $M$ of $K$ mutually independent geostatistical, statistical or stochastic models $M_k$, with parameters $\theta_k$ for the desired output vector, $\Delta$.

2. Obtain $ML$ estimates $\hat{\theta}_k^M$ of $\theta_k$ by calibrating each $M_k$ against available data $D$ through minimization of the log likelihood $-2\ln p(D|M_k, \theta_k) - 2\ln p(\theta_k|M_k)$, then compute the corresponding estimation covariance $C_k^D$ and $KICD_k$.

3. Compute $p(M_k|D)_{ML} = \frac{\exp(-\frac{1}{2}C_k^D(M_k|D))}{\sum_{k=1}^{K} \exp(-\frac{1}{2}C_k^D(M_k|D))}$ where the subscript $ML$ designates the ML estimation process in step 2.

4. For each model $M_k$ estimate $E(\Lambda|D, M_k)_{ML}$ and $Cov(\Lambda|D, M_k)_{ML}$ either through second-order approximations, as described in Appendix B, or via Monte Carlo simulation (both options are explored in our synthetic example):

   a. Draw random samples (realizations) of $\theta_k$ from a multivariate Gaussian distribution with mean $\hat{\theta}_k^M$ and covariance $C_k^M$.

   b. Estimate $E(\Lambda|D, M_k, \theta_k)_{ML}$ and $Cov(\Lambda|D, M_k, \theta_k)_{ML}$ for each realization of $\theta_k$.

   c. Average over all realizations of $\theta_k$ to obtain sample estimates of $E(\Lambda|D, M_k)_{ML} = E_k E(\Lambda|D, M_k, \theta_k)_{ML}$ and $Cov(\Lambda|D, M_k)_{ML} = E_k Cov(\Lambda|D, M_k, \theta_k)_{ML}$.

5. Compute $E(\Lambda|D)_{ML} = \sum_{k=1}^{K} E(\Lambda|D, M_k)_{ML} p(M_k|D)_{ML}$ and $Cov(\Lambda|D)_{ML} = \sum_{k=1}^{K} [E(\Lambda|D, M_k)_{ML} - E(\Lambda|D)_{ML}] E(\Lambda|D, M_k)_{ML} - E(\Lambda|D)_{ML}^2 p(M_k|D)_{ML}$ and/or $Tr(Cov(\Lambda|D)_{ML})$.

6. Postulate a set $P$ of $L$ alternative geostatistical, statistical or stochastic models $P_l$, with parameters $\pi_l$ for a potential data set $C$, the models $P_l$ may be independent of $M_k$ or form extensions of $M_k$ or may coincide with the latter as in the computational examples given in this paper.

7. Predict multivariate statistics of $C$ conditional on $D$, either via BMA or via MLBMA by means of the model set $P$, in the case of MLBMA the procedure would parallel that described for $\Lambda$ in steps (2)–(6).

8. Estimate $E(\Lambda|D)_{ML}C_{ML}$ and $Cov(\Lambda|D)_{ML}C_{ML}$, where the subscript $MLC$ designates the ML estimation process in step 2 but now with respect to an augmented data set $[D,C]$, either through second-order approximations, as described in Appendix B, followed by step 10 or via Monte Carlo simulation by using the statistics of $C$ from step 7 to generate random realizations of $C$ (both options are explored in our synthetic example); for each realization and for each model $M_k$:

   a. Optionally linearize the residuals entering into the negative log likelihood $-2\ln p(D|C, M_k, \theta_k) - 2\ln p(\theta_k|M_k)$ about $\theta_k^C$ (this option is not explored in this paper).

   b. Obtain $ML$ estimates $\hat{\theta}_k^C$ of $\theta_k$ by minimizing this negative log likelihood with respect to $\theta_k$, then compute the corresponding estimation covariance $C_k^C$ and $KICD_k^C$.

   c. Compute $p(M_k|D, C)_{MLC} = \frac{\exp(-\frac{1}{2}C_k^C(M_k|D,C))}{\sum_{k=1}^{K} \exp(-\frac{1}{2}C_k^C(M_k|D,C))}$ where the subscript $MLC$ designates the ML estimation process in step 2.

   d. For each model $M_k$ estimate $E(\Lambda|D, C, M_k)_{MLC}$ and $Cov(\Lambda|D, C, M_k)_{MLC}$ via Monte Carlo simulation:

   i. Draw random samples (realizations) of $\theta_k$ from a multivariate Gaussian distribution with mean $\hat{\theta}_k^C$ and covariance $C_k^C$.

   ii. Estimate $E(\Lambda|D, C, M_k, \theta_k)_{MLC}$ and $Cov(\Lambda|D, C, M_k, \theta_k)_{MLC}$ for each realization of $\theta_k$.

   iii. Average over all realizations of $\theta_k$ to obtain sample estimates of $E(\Lambda|D, C, M_k)_{MLC} = E_k E(\Lambda|D, C, M_k, \theta_k)_{MLC}$ and $Cov(\Lambda|D, C, M_k)_{MLC} = E_k Cov(\Lambda|D, C, M_k, \theta_k)_{MLC}$.

   e. Compute $E(\Lambda|D, C)_{MLC} = \sum_{k=1}^{K} E(\Lambda|D, C, M_k)_{MLC} p(M_k|D, C)_{MLC}$ and $Cov(\Lambda|D, C)_{MLC} = \sum_{k=1}^{K} Cov(\Lambda|D, C, M_k)_{MLC} p(M_k|D, C)_{MLC}$

   $+ \sum_{k=1}^{K} E(\Lambda|D, M_k)_{ML} - E(\Lambda|D)_{ML} \right] E(\Lambda|D, M_k)_{ML} - E(\Lambda|D)_{ML}^2 p(M_k|D, C)_{MLC}$

   and/or $Tr(Cov(\Lambda|D)_{MLC})$.

9. Average over all realizations of $C$ to obtain sample estimates of $E(\Lambda|D)_{MLC} = E_k E(\Lambda|D, C)_{MLC}$.

$E(\Lambda|D)_{MLC} = E_k E(\Lambda|D, C)_{MLC} + Cov(\Lambda|D, C)_{MLC}$

$+ Tr(E_k Cov(\Lambda|D, C)_{MLC})$; note that due to inevitable sampling errors and approximations associated with the ML estimation process $E(\Lambda|D)_{MLC}$, $Cov(\Lambda|D)_{MLC}$ and $Tr(Cov(\Lambda|D)_{MLC})$ obtained at this step would generally differ, though ideally not by much, from $E(\Lambda|D)_{MLC}$, $Cov(\Lambda|D)_{MLC}$ and $Tr(Cov(\Lambda|D)_{MLC})$ obtained at step 5.

10. Repeat steps 6–9 for different sets $C_1, C_2, C_3, \ldots$ of potential data and select that set which maximizes the difference $Tr(Cov(\Lambda|D)_{MLC}) - Tr(E_k Cov(\Lambda|D, C)_{MLC})$ between the trace conditional on $D$ and the expected trace conditional on $D$ and $C$ (this step is not explored in the present paper).

Appendix B. Second-order moment approximations

Step 4 of the MLBMA procedure in Appendix A entails an optional approximation of $E(\Lambda|D, M_k)_{ML}$ and $Cov(\Lambda|D, M_k)_{ML}$ via lead order expansions of $E(\Lambda|D, M_k, \theta_k)_{ML}$ and $Cov(\Lambda|D, M_k, \theta_k)_{ML}$ in $\theta_k$.
about $\theta_c^*$; step 8 entails a similar approximation of $E(\Delta D, C, M_s)_{\lambda_{0,c}}$ and $\text{CoV}(\Delta D, C, M_s)_{\lambda_{0,c}}$ via lead order expansion of $E(\Delta D, C, M_s, \theta_1)_{\lambda_{0,c}}$ and $\text{CoV}(\Delta D, C, M_s, \theta_1)_{\lambda_{0,c}}$ in $\theta_1$ about $\theta_1^*$. In the context of step 4 one would start with second-order expansions

$$E(\Delta D, M_s, \theta_1)_{\lambda_{0,c}} \approx E(\Delta D, M_s, \theta_1^*) + \sum_{n=1}^{N_0} \frac{\partial E(\Delta D, M_s, \theta_1)_{\lambda_{0,c}}}{\partial \theta_1} \Delta \theta_1,$$

$$1 + \frac{1}{2} \sum_{n=1}^{N_0} \sum_{m=1}^{N_1} \frac{\partial^2 E(\Delta D, M_s, \theta_1)_{\lambda_{0,c}}}{\partial \theta_1 \partial \theta_2} \Delta \theta_1 \Delta \theta_2,$$

where $\Delta D$ is the $p$th component of $\Delta$ and $\Delta \theta_1 = \theta_1 - \theta_1^*$. Since $\Delta D_{\theta_1} = 0$ one has, to second order,

$$E_n E(\Delta D, M_s, \theta_1)_{\lambda_{0,c}} \approx E(\Delta D, M_s, \theta_1^*) + \sum_{n=1}^{N_0} \frac{\partial E(\Delta D, M_s, \theta_1)_{\lambda_{0,c}}}{\partial \theta_1} \Delta \theta_1,$$

$$1 + \frac{1}{2} \sum_{n=1}^{N_0} \sum_{m=1}^{N_1} \frac{\partial^2 E(\Delta D, M_s, \theta_1)_{\lambda_{0,c}}}{\partial \theta_1 \partial \theta_2} \Delta \theta_1 \Delta \theta_2,$$

and

$$\text{CoV}(\Delta D, M_s, \theta_1)_{\lambda_{0,c}} \approx \text{CoV}(\Delta D, M_s, \theta_1^*) + \sum_{n=1}^{N_0} \frac{\partial \text{CoV}(\Delta D, M_s, \theta_1)_{\lambda_{0,c}}}{\partial \theta_1} \Delta \theta_1,$$

$$1 + \frac{1}{2} \sum_{n=1}^{N_0} \sum_{m=1}^{N_1} \frac{\partial^2 \text{CoV}(\Delta D, M_s, \theta_1)_{\lambda_{0,c}}}{\partial \theta_1 \partial \theta_2} \Delta \theta_1 \Delta \theta_2,$$

where $\Gamma_{\text{mkm}} = \Delta E(\Delta D, M_s, \theta_1)_{\lambda_{0,c}}$ is the $(n,m)$th component of $\Gamma_q$. An analogous approach would apply to step 8 of Appendix A.

References


