Parameter interdependence and uncertainty induced by lumping in a hydrologic model

Mark R. Gallagher[1,2] and John Doherty[2,3]

Received 16 July 2006; revised 22 November 2006; accepted 22 December 2006; published 16 May 2007.

[1] Throughout the world, watershed modeling is undertaken using lumped parameter hydrologic models that represent real-world processes in a manner that is at once abstract, but nevertheless relies on algorithms that reflect real-world processes and parameters that reflect real-world hydraulic properties. In most cases, values are assigned to the parameters of such models through calibration against flows at watershed outlets. One criterion by which the utility of the model and the success of the calibration process are judged is that realistic values are assigned to parameters through this process. This study employs regularization theory to examine the relationship between lumped parameters and corresponding real-world hydraulic properties. It demonstrates that any kind of parameter lumping or averaging can induce a substantial amount of ''structural noise,'' which devices such as Box-Cox transformation of flows and autoregressive moving average (ARMA) modeling of residuals are unlikely to render homoscedastic and uncorrelated. Furthermore, values estimated for lumped parameters are unlikely to represent average values of the hydraulic properties after which they are named and are often contaminated to a greater or lesser degree by the values of hydraulic properties which they do not purport to represent at all. As a result, the question of how rigidly they should be bounded during the parameter estimation process is still an open one.

Citation: Gallagher, M. R., and J. Doherty (2007), Parameter interdependence and uncertainty induced by lumping in a hydrologic model, *Water Resour. Res.*, *43*, W05421, doi:10.1029/2006WR005347.

1. Introduction

[2] While distributed, physically based watershed models are finding growing use among hydrologists, their use is by no means ubiquitous. The advantages of physically based simulation of hydrologic processes have been demonstrated by models such as MIKE SHE [*Refsgaard and Storm*, 1995], based on SHE, the European Hydrological System [*Abbott et al.*, 1986a, 1986b], MODHMS [*Panday and Huyakorn*, 2004], and Gridded Surface Subsurface Hydrologic Analysis (GSSHA) [*Downer and Ogden*, 2002], which is derived from CASC2D [*Ogden and Julien*, 2002]. Part of their attraction rests on their ability to provide detailed information on flow conditions at all locations within the model domain, rather than simply at the watershed pour point. They also have the advantage that more of the processes operating within an environmental system can be represented than is possible using simpler, lumped-parameter models, and that some of the parameters pertaining to some of the processes (particularly geometrical parameters) can be assigned on the basis of easily measured properties of the system.

[3] Notwithstanding the benefits gained from using physically based models, most hydrologic simulation is still conducted using lumped parameter models, especially where simulation of large watersheds over moderate-to-large periods of time is undertaken, as is commonly required for flood prediction and for water quality and other studies. One reason for this is that lumped parameter models often perform better in such contexts [*Reed et al.*, 2004]. Another is that in modeling contexts such as this, the data requirements of physically based models, as well as their lengthy run times, rule them out of contention. The fact that such models require calibration in order to enhance their predictive ability in any particular watershed in which they are deployed, and that calibration, whether undertaken by hand or automatically, generally requires that many model runs be undertaken, further erodes the extent to which physically based models are presently deployed in engineering and regulatory applications. Added to this is the fact that, unless a calibration data set includes, in addition to outlet flows, a significant amount of historical measurements of system state internal to the system, only a handful of parameters are identifiable through the calibration process [*Sorooshian and Gupta*, 1983; *Hornberger et al.*, 1985; *Young et al.*, 1996; *Beven*, 1989; *Jakeman and Horberger*, 1993; *Wagener et al.*, 2003; *Beven*, 2006]. Hence only averaged parameters can be estimated (and a few of these at best), so that in spite of all their elegance, physically based models deployed in the simulation of real-world systems may, in fact, assume many of the characteristics of lumped parameter models, some of which will be explored in this paper.

[1]Natural Resource Sciences, Department of Natural Resources and Water, Indooroopilly, Queensland, Australia.
[2]Department of Civil Engineering, University of Queensland, St. Lucia, Queensland, Australia.
[3]Watermark Numerical Computing, Brisbane, Queensland, Australia.

0043-1397/07/2006WR005347

[4] Despite their ''lumped'' nature, commonly used lumped-parameter models, such as the Sacramento model [*Burnash et al.*, 1973], Hydrologic Simulation Program-FORTRAN (HSPF) [*Bicknell et al.*, 1997], Water Bounded Network Model (WBNM) [*Boyd et al.*, 1979], and Hydrologiska Byråns Vattenbalansavdelning (HBV) [*Bergstrom*, 1976], attempt to simulate hydrologic processes realistically, employing parameters that are designed to bear some relation to hydraulic properties that have a physical basis and for which realistic values can thus be supplied from outside of the calibration process, even if they are not directly measurable. Guidance documents that have been written to support the use of these models [see, e.g., *U.S. Environmental Protection Agency (EPA)*, 2000] provide appropriate ranges for many of their parameters and advise users not to transgress the limits of these ranges when calibrating a model, for transgression of these limits may lead to unrealistic predictions of system behavior. These ranges are often linked to measurable or inferable watershed attributes such as land use and soil type, this providing further illustration of their hydraulic origins.

[5] Much has been written about the ''regionalization'' of parameters employed by lumped-parameter watershed models. This concept is based on the premise that while the parameter values assigned to watershed models may not be equivalent to quantities that are directly measurable in the field or laboratory, there is nevertheless a high degree of correlation between them and observable watershed characteristics. On this basis, it is then hoped that knowledge gained from calibrating models for gauged watersheds can be transferred to ungauged watersheds, where parameters inferred on the basis of watershed characteristics alone can then provide reasonable estimates of streamflow from that watershed. However, while some attempts at regionalization appear to have been relatively successful [*Seibert*, 1999; *Campbell and Bates*, 2001; *Kokkonen et al.*, 2003], success in this endeavor has been far from widespread. For example, *Merz and Bloschl* [2004] demonstrated that parameter transfer to ungauged basins based on spatial proximity performed better than that based on physiographic catchment attributes. *Post and Jakeman* [1999] also experienced mixed results for daily streamflow predictions as a result of poor correlation between model parameters and landscape attributes.

[6] More recently, the model parameter estimation experiment (MOPEX) was established to develop techniques for the a priori estimation of parameters used in land surface parameterization schemes of hydrologic models; see *Hogue et al.* [2004] for further details. As part of this initiative, the issue of transferability of optimized hydrologic model parameters on a geographical basis is being explored. *Wagener and Wheater* [2006] provide a detailed discussion of the manifold challenges facing this research. In summarizing MOPEX work to date, *Duan et al.* [2006] suggest that further research is required into the transfer of parameters from calibrated systems to ungauged catchments, in spite of some encouraging results involving MOPEX catchments reported by *Ao et al.* [2006] and *Young* [2006].

[7] Similarity relationships between parameters of the same type representing similar land uses in neighboring watersheds have been employed in composite multiwatershed calibration based on regularized inversion by *Doherty and Johnston* [2003] and *Doherty and Skahill* [2005]. This strategy was indeed successful in reducing the propensity for parameter nonuniqueness. The extent to which inferred parameter values were more ''realistic'' as a result of this strategy is, however, unknown.

[8] The philosophical basis for linking the values of lumped parameters to measurable properties of real-world systems (and of naming parameters after these properties) is that lumped parameters are often assumed to be good approximations to the average values of these properties over a watershed. However, the extent to which such average properties, or anything even related to these average properties, are in fact estimated on the basis of data sets normally employed in the calibration of watershed models has only been investigated to a limited extent. The same applies to the extent to which such ''averaging'' actually transgresses hydraulic property boundaries, so that the value assigned to one parameter is in fact partly attributable to a hydraulic property which it is not supposed to represent at all.

[9] The notion of effective (or average) parameters employed by models has been investigated quite extensively in relation to both unsaturated flow [e.g., *Yeh et al.*, 1985; *Montoglou and Gelhar*, 1987; *Russo*, 1992; *Ferrante and Yeh*, 1999; *Zhu et al.*, 2006] and saturated flow [e.g., *Warren and Price*, 1961; *Dagan*, 1979; *El-Kadi and Brutsaert*, 1985; *Gomez-Hernandez and Gorelick*, 1989; *Ababou and Wood*, 1990; *Neuman and Orr*, 1993]. Such research typically focuses on the translation of parameters representing the hydraulic response of a heterogeneous medium measured at the point scale to an equivalent homogeneous medium as represented by one or many cells of a model grid or mesh (a process known as ''upscaling''). A detailed review of the various methods of saturated hydraulic conductivity upscaling is given by *Wen and Gomez-Hernandez* [1996] and more recently by *Sanchez-Vila et al.* [2006]. Unlike the present study, however, these studies focus on the ''forward problem'' of assigning appropriate parameter values to all or part of the model domain so that it simulates real-world response as well as possible; in contrast, our study investigates regional or effective parameter inference as it takes place through solution of the inverse problem of model calibration.

[10] In an examination of the uncertainty in effective surface roughness parameters for a flood inundation model estimated through the calibration process, *Pappenberger et al.* [2005] noted that as model geometries are only an approximation to real system geometry, estimates of effective Manning's roughness parameters may be somewhat compromised as they compensate for model geometrical imperfections, a point also noted by *Aronica et al.* [1998] and *Marks and Bates* [2000]. While these studies are indeed more related to the present one in that they examine parameter inference based on fitting model outputs to historical field observations, no detailed theoretical examination of the problem of ''parameter contamination'' either by other parameters or by model inadequacies was undertaken as it is below.

[11] A related (as we shall show) question is that of measurement noise. It is commonly assumed in watershed

model calibration that model-to-measurement residuals will approach homoscedascity if a fit is sought between appropriately transformed modeled and measured flows rather than directly between the flows themselves. A Box-Cox transformation [*Box and Cox*, 1964], which includes log transformation as a special case, is often employed for this purpose [*Chander et al.*, 1978; *Hirsch*, 1979; *Sorooshian and Dracup*, 1980]. Reduction of temporal correlation between residuals is often sought through fitting of an autoregressive moving average (ARMA) [*Box and Jenkins*, 1976] model to them [e.g., *Kuczera*, 1983; *Bates and Campbell*, 2001]. The values estimated for model parameters through the calibration process will depend on which, if any, of these transformations are undertaken, and on the weights assigned to different transformed flows in formulation of the objective function (and, indeed, on how the objective function is defined) that is minimized through this process. Where a weighted least-squares objective function is employed, and where weights are inversely proportional to measurement noise (or if the inverse of the measurement noise covariance matrix is employed as a weighting matrix in formulation of a quadratic form objective function where measurement noise is assumed to be correlated), estimated values for parameters are those of greatest likelihood if noise is Gaussian [*Bard*, 1974]. However, because a ''good fit'' between model outputs and field measurements can be obtained with many different weighting strategies, and because each of these strategies normally leads to the estimation of different parameter values, the question must be asked: which of these parameter values (if any) represents the true ''average'' of the corresponding hydraulic property over the watershed. The urgency of this question is heightened by the fact that (in the authors' experience at least) temporal correlation of residuals is not easily removed by appropriate ARMA transformation. Furthermore, even if the stochastic structure of measurement noise were available, use of the inverse covariance matrix of this structure in definition of the objective function is numerically intractable (unless this matrix is highly band limited) because of the sheer number of observations employed in most flow time series used for calibration purposes and hence the large size of the covariance matrix.

[12] The purpose of the present paper is to address some of these issues through undertaking a brief theoretical study, supported by a synthetic example, that seeks to improve our understanding of them. We do this by generating a flow time series using a simple, synthetic, physically based model of a hillside drained by a channel. Parameters of each of a number of types employed by the model are assumed to be spatially variable over the hillside (as are hydraulic properties in practice) and characterized by a known spatial covariance function. We then ''lump'' each parameter type by estimating a single value for that type through the calibration process. We examine the nature of ''structural noise'' generated through this lumping process and assess the applicability of different commonly employed flow transformations in promoting homoscedascity and reducing temporal correlation. We borrow concepts employed in solution of ill-posed inversion problems to build a ''resolution matrix'' linking estimated, lumped parameters to their ''real-world'' counterparts, showing that the relationship between the two is actually quite complex. In doing this, we demonstrate that it is not always correct to expect that the value assigned to a lumped parameter through the calibration process will fall within the realistic numerical range of its real-world counterpart.

[13] Our study was limited to some extent by the numerical complexity and run time requirements of physically based simulators of linked surface and groundwater flow. Nevertheless, we have attempted to employ a synthetic case that is complex enough to be realistic (and for conclusions drawn from our study to thereby have validity) but simple enough to be numerically tractable. As shown below, the necessity to obtain derivatives of model outputs with respect to spatially distributed model parameters required that the model be run many times in succession. It also required that the model's numerical behavior be such that small differences between large flow outputs, calculated on the basis of incrementally varied parameter values, had sufficient integrity for calculation of finite-difference derivatives with respect to those parameters. Where computed surface flows are high and vary rapidly with time, this requirement can be difficult to meet because of the use of various numerical strategies such as ''adaptive time stepping'' that are required to meet the demands of this type of simulation [*Kavetski et al.*, 2006]; hence the change in model outputs resulting from a small change to one parameter may reflect in part to certain alterations to the numerical solution path in computation of the two respective outputs. Indeed, this problem was encountered in the present study. However, by restricting parameter types to those discussed below, and by carefully monitoring model outputs for signs of numerical contamination of derivatives in this manner, the problem was overcome to a sufficient extent for our study to achieve its aims.

[14] The analysis presented below assumes model linearity. Though this is indeed an approximation, it is not expected to invalidate the conclusions drawn herein concerning the effect of lumping on model-to-measurement misfit and on the relationships between estimated parameters and real-world hydraulic properties. Nonlinear analysis would be mathematically intractable, and the outcomes of numerical experiments would be far more difficult to interpret.

[15] The following section introduces the model; the section after that discusses the theory. This order of presentation was chosen so that the theoretical discussion could take place in a context that would readily reveal its relevance. The outcomes of our analyses are presented in sections following the theory on which these analyses are based.

2. The Model

2.1. The Model Domain

[16] Figure 1 shows a hillside sloping to the southwest. The elevation of the northeast corner of the hillside is 57 m relative to the elevation of the southwest corner. To the south and west, the hill is bounded by trapezoidal channels of 3-m depth. Overland flow takes place across the hillside to both channels. However, some flow infiltrates to the groundwater system where it enters the

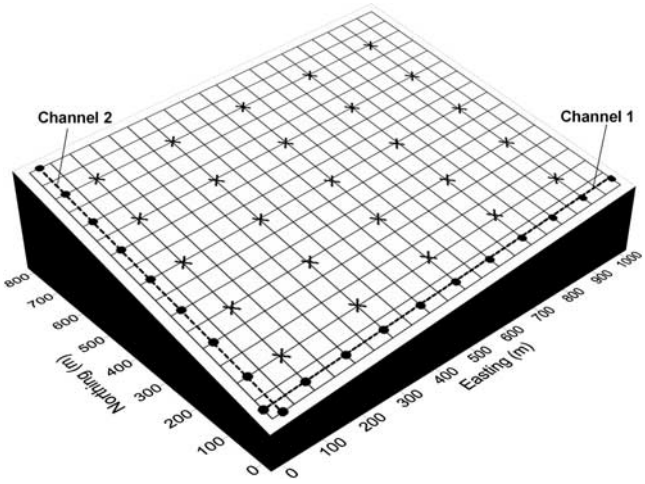

Figure 1. MODHMS model grid for groundwater and overland flow, with channel flow segments shown as dashed lines. Pilot points used for parameterization of the groundwater and overland flow domains (crosses) and channels (dots) are also shown.

channels as base flow. The underlying aquifer is unconfined with a horizontal planar basement at an elevation of −20 m.

2.2. Flow Simulator

[17] Overland, channel, and subsurface flow of water was simulated using MODHMS [*Hydrogeologic Inc.*, 2003]. MODHMS simulates both overland flow and streamflow by solving the diffusion wave approximation of the Saint Venant equation [*Viessman and Lewis*, 1996] for shallow, unsteady flow in one and two dimensions, respectively. Interaction between the overland and channel flow regimes is governed by weir-type equations. Overland and channel fluxes are implicitly coupled to the subsurface regime via soil surface and streambed leakage terms, respectively; the partial differential equations for two- and three-dimensional saturated and unsaturated flow are solved using a cell-centered, finite difference approach. The reader is referred to *Hydrogeologic*, Inc. [2003] and to the work of *Panday and Huyakorn* [2004] for derivation of the flow equations and details of the numerical solution techniques employed by this model.

2.3. Simulation Details

[18] Overland and groundwater flow in our present study were simulated using a 20 × 16 grid of square 50 × 50 m cells; channel flow was simulated using a 1 × 20 array of cells comprising two reaches, the confluence of these reaches being at the southwestern corner of the model domain. The groundwater system was simulated as a single-layer unconfined aquifer.

[19] A simulation period of 24 hours was chosen for this study. A longer period would have been desirable; however, long run times prevented this. The length of the run time was exacerbated by the need to employ short time steps and tight convergence criteria for the MODHMS numerical solver in order to maintain integrity of finite difference derivatives calculation. Figure 2a shows rainfall during the simulation time, while Figure 2b shows the summed flow at the confluence of the two channels (90% of which was contributed by channel 1, the southern channel) over this period. Samples of this flow at 18-s intervals constituted the "calibration data set" for the notional calibration exercise discussed herein through which lumped system parameters were inferred; sampling from 2 hours into the simulation was undertaken (prior to this, no appreciable change in outflow above initial conditions was simulated). The calibration data set comprised 4381 samples in all. During the calibration period, 12% of channel flow was contributed by overland flow from the hillside, 79% was contributed by groundwater flow, while 9% was due to rainfall falling within the channel itself. A warm-up period comprising an identical event, followed by 2 days decay, was used to set initial conditions for all system state variables simulated by the model.

[20] As will become clear in the analysis to follow, the actual flow values comprising the synthetic time series of Figure 2b were not employed in the analysis; rather, the sensitivities of these flows to different hydraulic properties form the basis for the equations that we derive below and the discussion that follows from them. To simplify derivation of these equations and the discussion that follows, it is assumed that flows are not contaminated by measurement error.

2.4. Surface and Subsurface Properties

[21] Table 1 provides a description of pertinent MODHMS parameter types and the respective values of these parameters that were employed for generation of the flow time series of Figure 2b; in generating this flow series, all of these parameter types were uniform over their respective domains. Six of these parameter types (those highlighted in Table 1) were selected for our investigation of lumping. Two of these are groundwater parameters (SF1 and HY), two are overland flow parameters (X_FRICTN and BOT_LKG), and two are channel flow parameters (RGHN_CH1 and BEDCOM_CH1); Y_FRICTN was linked to X_FRICTN so that the two together effectively constituted one parameter type. Selection of these parameter types was based partly on the fact that the "observed" channel flow time series is more sensitive to these parameters than to others, and partly on the fact that derivatives of flows with respect to these parameters were calculable using finite parameter differences with integrity using MODHMS; although some "numerical granularity" was apparent as these parameters were varied slightly for the purpose of finite difference derivatives calculation, this granularity was slight compared with the (approximated) derivatives themselves. It is freely acknowledged that the study presented herein would also have benefited from the inclusion of other parameter types (and other factors such as spatial variability of rainfall and/or initial conditions). However, computational demands limited the study to just these six parameter types. It is expected that the inclusion of other parameter types in the analysis would only have strengthened the outcomes of this study and not altered them substantively.

2.5. Parameter Heterogeneity

[22] For the purpose of investigating the effects of lumping on parameter values estimated through the calibration process, the six selected parameters were assumed to be

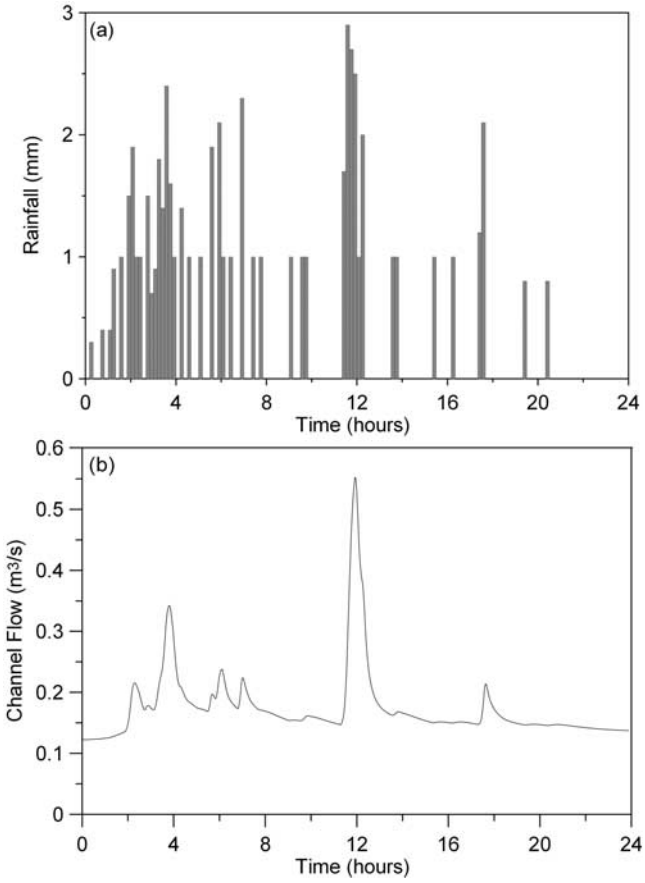

Figure 2. (a) Rainfall hyetograph (10 min) for single day event and (b) MODHMS-calculated stream discharge hydrograph at system outlet.

variable over the model domain. Note that, in the study described below, the model was not actually run using other than the uniform parameter values shown in Table 1; however, the stochastic nature of their spatial variability was incorporated into the equations developed below. As will be further described, part of the strategy of the current study is to inquire into just what is actually being estimated through a calibration process that seeks to compute the values of only six "effective parameters," i.e., one value for each type of parameter which, for the purpose of formulating a well-posed inverse problem, is assumed to be uniform over the model domain.

[23] In assessing the effect of spatial variability of these properties on the estimation of their lumped equivalents, each of the six parameters selected for examination in this study was represented using a set of "pilot points" spread over the model domain. Pilot points were originally used in groundwater model parameterization by *Certes and De Marsily* [1991], and have been used by many others since then in the groundwater modeling context [see, e.g., *RamRao et al.*, 1995; *LaVenue et al.*, 1995; *Doherty*, 2003; *Moore and Doherty*, 2005; *Tonkin and Doherty*, 2005]. For the overland and groundwater flow parameters, a two-dimensional array of these points was arranged in a 5×5 grid, comprising 25 pilot points for each such parameter. For the channel parameters, only those in the main channel were represented using pilot points (those in the tributary segment were relatively insensitive because of the low flows occurring within this channel and were not used in this analysis). Eleven pilot points were placed within this main channel. Figure 1 shows all sets of pilot points. Hydraulic property values assigned to pilot points were interpolated to model grid cells using kriging based on the same spatial covariance matrix as that used to ascribe notional spatial variability to these properties over the model domain (see below).

3. Theory

3.1. General

[24] In the present study, "lumping" is considered to be equivalent to "averaging." Equations are now developed based on the premise that the six highlighted parameters of Table 1 are, in fact, spatially variable over the model domain.

Table 1. Values of MODHMS Model Parameters Used in Generating the Synthetic Flow Time Series[a]

Parameter	Parameter Definition	Value
X_FRICTN	Manning's roughness coefficient of land surface in x-direction	1.5e-2 $sm^{1/3}$
Y_FRICTN	Manning's roughness coefficient of land surface in y-direction	1.5e-2 $sm^{1/3}$
BOT_LKG	Vertical leakance of land surface to the subsurface, typically the product of subsurface hydraulic conductivity and the area of interaction divided by the distance between the overland flow node and the groundwater flow node	5.0e-7 s^{-1}
RILLSH	Height of rill storage of land surface	5.0e-4 m
OBSTRH	Height of obstruction storage of land surface	1.0e-8 m
RGHN_CH1	Manning's roughness coefficient of main channel reach	4.0e-2 $sm^{1/3}$
RGHN_CH2	Manning's roughness coefficient of tributary channel reach	4.5e-2 $sm^{1/3}$
BEDCOM_CH1	Vertical leakance of main channel reach, expressed as a ratio of stream bed sediment hydraulic conductivity to material thickness	8.5e-5 s^{-1}
BEDCOM_CH2	Vertical leakance of tributary channel reach, expressed as a ratio of stream bed sediment hydraulic conductivity to material thickness	9.5e-5 s^{-1}
RILLSHCH–CH1	Height of rill storage of main channel reach	3.5e-4 m
RILLSHCH_CH2	Height of rill storage of tributary channel reach	8.0e-4 m
OBSTRH_CH1	Height of obstruction storage of main channel reach	1.0e-8 m
OBSTRH_CH2	Height of obstruction storage of tributary channel reach	1.0e-8 m
SF1	Storage coefficient of aquifer	9.0e-2
HY	Horizontal hydraulic conductivity of aquifer	1.5e-4 ms^{-1}

[a]Those selected for investigation of lumping are shaded.

However, it is the average value of each that is estimated through a notional calibration exercise based on the synthetic flow time series of Figure 2b. The characteristics of a parameter estimation problem thus formulated are examined.

3.2. Spatial Variability of Parameters

[25] Let "true" system parameters be designated using the symbol **k**. These are parameter types that are highlighted in Table 1 as assigned to the respective pilot points of Figure 1. Thus the total parameter vector of "real parameters" for the present study is composed of 122 elements, these being 25 elements each of the X_FRICTN, BOT_LKG, SF1, and HY parameter types, and 11 for each of the RGHN_CH1 and BEDCOM_CH1 parameter types. It is recognized that it is impossible to individually estimate each of these parameters; rather, one parameter representing each of the six different parameter types is actually estimated through a notional "lumped parameter" calibration process. However, as we will demonstrate, the effect of these 122 parameters (representing system detail which is beyond the reach of that which can be inferred through any calibration process because of limitations in the information content of the calibration data set) nevertheless have an effect on that process through the "structural noise" which they generate and the "parameter contamination" that their presence incurs.

[26] For simplicity, the values taken by parameters of different types are assumed to be statistically independent. Thus the matrix C(**k**) representing the spatial covariance of "true system properties" is assumed to be block diagonal. However, spatial correlation is assumed to exist between elements of the same parameter type. For our model, correlation between the logs of parameter values is assumed to decay exponentially with distance with a decay constant of 150 m; different variances (diagonal elements of the spatial covariance matrix) are employed for different aspects of the following analysis. Note that because these parameters are characterized by a spatial variability that pertains to their logs rather than to their native values, the elements of **k** also pertain to the logs of these parameter values.

3.3. Parameter Lumping

[27] Let $\mathbf{k}_i$ represent the elements of **k** pertaining to parameter type *i*. Let the average (over all pilot points) value of this parameter be symbolized as p_i; note that this is a scalar. Let the vector of the six average parameters be designated as **p**. The relationship between **p** and **k** is of the type:

$$\mathbf{p} = \mathbf{N}\mathbf{k} \tag{1}$$

where **N** is a kind of "averaging matrix." If averaging weights are all equal, **N** is a 6×122 matrix, each row of which has either 25 consecutive elements of value 1/25 or 11 consecutive elements of 1/11, with all other elements being zero.

[28] Let **j** be a vector with dimensions the same as **k**, but in which values of each parameter type are all equal to the average value of the respective parameter. **j** can be calculated from **p** using the equation:

$$\mathbf{j} = \mathbf{L}\mathbf{p} \tag{2}$$

in which **L** is a 122×6 "selection matrix." Each element of **j** pertains to a pilot point; each row of **L** has zero-valued elements except for that corresponding to the element of **p** which characterizes its parameter type; that element has a value of 1.

3.4. Representation of the Model

[29] Let the flow time series from which parameters are estimated be represented by the vector $\mathbf{q}_0$; let **q** denote a flow series generated by the model for an arbitrary parameter set. Assuming that the model is linear, and ignoring parameter and flow offsets for simplicity (without loss of generality in the following discussion where the focus is on error terms rather than absolute values), let $\mathbf{q}_0$ be calculable from $\mathbf{k}_0$ using the relationship:

$$\mathbf{q}_0 = \mathbf{Z}\mathbf{k}_0 \tag{3a}$$

In this equation, we are assuming that exact calculation of $\mathbf{q}_0$ is possible using the "true" set of parameters $\mathbf{k}_0$; thus the model is assumed to be a perfect simulator of reality, and the calibration data set is assumed to be uncontaminated by measurement noise. Elements of the **Z** matrix represent the sensitivity of each element of the flow time series to each parameter. Where a more general **k** is used, i.e., a parameter set that is not necessarily the true parameter set, then equation (3a) becomes

$$\mathbf{q} = \mathbf{Z}\mathbf{k} \tag{3b}$$

Let us now employ the matrix **X** to represent the lumped form of the model, i.e., the model in which all elements of the same type are assumed to possess the same value, and thus only 6, rather than 122, parameters are required to generate a flow time series **q**. The equation by which this lumped parameter model generates a flow time series is given by:

$$\mathbf{q} = \mathbf{X}\mathbf{p} \tag{4a}$$

Because a parameter set **j** calculated from **p** using equation (2) produces identical model outputs, equation (4a) can also be written as:

$$\mathbf{q} = \mathbf{Z}\mathbf{L}\mathbf{p} \tag{4b}$$

from which it is evident that:

$$\mathbf{X} = \mathbf{Z}\mathbf{L} \tag{5}$$

(Note that in the example discussed below, both **X** and **Z** were estimated using finite parameter differences as replacements for the derivatives of model outputs with respect to parameters which actually constitute the elements of these matrices.)

3.5. Model Calibration

[30] In general, $\mathbf{q}_0$, while being in the range space of **Z**, is not in the range space of **X**. Thus a perfect fit between lumped model outputs and field measurement cannot be achieved with any particular **p**. This simulates what occurs in general model usage where, because a model is a simplification of reality, its outputs cannot replicate every detail of a measurement data set, the difference between the two being attributed to "structural noise." In many model-

ing contexts, this contribution to model-to-measurement misfit is much greater than that of measurement error.

[31] In the present case, because we are trying to estimate average system properties using a lumped model, we define structural noise, for any "true" parameter field **k**, as the difference between model outputs calculated on the basis of **k** and those calculated on the basis of the spatially averaged field as encapsulated in **p**. Thus symbolizing this noise as ε, we have:

$$\varepsilon = \mathbf{Zk} - \mathbf{Xp} = \mathbf{Z}(\mathbf{I} - \mathbf{LN})\mathbf{k} \tag{6}$$

[32] As a basis for formulation of the inverse problem for estimation of the lumped parameter set **p** of average values of respective elements of **k**, we can write:

$$\mathbf{q} = \mathbf{Zk} = \mathbf{Zk} - \mathbf{Xp} + \mathbf{Xp} = \mathbf{Xp} + \varepsilon \tag{7a}$$

that is:

$$\mathbf{q} = \mathbf{Xp} + \varepsilon \tag{7b}$$

[33] The aim of lumped-parameter model calibration is to calculate a $\underline{\mathbf{p}}$ which minimizes the discrepancies between model (as represented by **X**) outputs and field measurements (as represented by $\mathbf{q}_0$). Where model-to-measurement misfit is measured using a weighted least-squares approach, with the measurement weight matrix being designated as **Q**, the optimum **p** (i.e., $\underline{\mathbf{p}}$) can be calculated [see, e.g., *Koch*, 1987] as:

$$\underline{\mathbf{p}} = (\mathbf{X}^t\mathbf{QX})^{-1}\mathbf{X}^t\mathbf{Qq} \tag{8}$$

In general parameter estimation practice, equation (8) is employed with both linear and nonlinear models. In the latter case, the Jacobian (or sensitivity) matrix **J** replaces **X**, and the $\mathbf{J}^t\mathbf{QJ}$ matrix is supplemented with an additive diagonal term that helps to regularize the inverse problem and improves estimation efficiency in the early stages of the nonlinear parameter estimation process; equation (8) is then solved iteratively, with sensitivities updated on each iteration of the inverse process as **J** is recalculated based on current estimates of **p**.

[34] In equation (8), **q** can be considered either as an "unprocessed" flow time series or as a transformed (for example, using a Box-Cox transformation as will be further discussed below) time series if desired. If flow transformation is linear, the **X** and **Z** matrices of equations (4a) and (4b) are adjusted accordingly though premultiplication by the transformation matrix. For a nonlinear transformation, the transformation sensitivity matrix is employed for premultiplication.

3.6. Covariance Matrix of Estimated Parameters

[35] In the traditional least-squares formulation of the inverse problem, if the model is linear and if the weight matrix **Q** is the inverse of the covariance matrix of measurement noise, $\underline{\mathbf{p}}$ can be shown to be the best unbiased estimator of **p**. Where **p** is the average value of a spatially variable parameter **k**, *Cooley* [2004] shows that the same still holds if the true nature of "measurement noise," which in the present instance is actually structural noise, is taken into account. Thus $\underline{\mathbf{p}}$ becomes the estimate of highest likelihood of the average value of **k** over the model domain (with averaging defined by the **N** matrix of equation (1)) if a **Q** matrix is supplied which is the inverse of the covariance matrix of structural noise $C(\varepsilon)$. Under these circumstances, the covariance matrix of estimated parameters $C(\boldsymbol{p})$ can be formulated as:

$$C\left(\underline{\mathbf{p}}\right) = (\mathbf{X}^t\mathbf{QX})^{-1} \tag{9a}$$

However, where **Q** is not the inverse of $C(\varepsilon)$, the more general expression of $C(\mathbf{p})$ is

$$C\left(\underline{\mathbf{p}}\right) = (\mathbf{X}^t\mathbf{QX})^{-1}\mathbf{X}^t\mathbf{Q}C(\varepsilon)\mathbf{QX}(\mathbf{X}^t\mathbf{QX})^{-1} \tag{9b}$$

Bard [1974] shows that the variance of estimated parameters is minimized if **Q** is, in fact, equal to a scalar multiple of $C^{-1}(\varepsilon)$.

3.7. Covariance Matrix of Structural Noise

[36] For a linear model, $C(\varepsilon)$ can be calculated from equation (6) using the formula:

$$C(\varepsilon) = \mathbf{Z}(\mathbf{I} - \mathbf{LN})C(\mathbf{k})(\mathbf{I} - \mathbf{LN})\mathbf{Z}^t \tag{10}$$

In the nonlinear case, *Cooley* [2004] shows that $C(\varepsilon)$ can be calculated empirically using paired model runs based on stochastic realizations of **k** on the one hand, and corresponding **p** vectors derived from **k** through equation (1) on the other hand.

[37] Note that $C(\varepsilon)$ of equation (10) (which is a square matrix with dimensions equal to the number of flow observations) is singular as **L** and **N** are each of rank 6 and $(\mathbf{I} - \mathbf{LN})$ is idempotent. It would thus appear that a **Q** matrix calculated as $C^{-1}(\varepsilon)$ for use in equation (9a) cannot be found. It is not difficult to show that under these circumstances, the Moore-Penrose inverse of $C(\varepsilon)$ can be used in its place.

[38] Equations (9a) and (9b) need to be properly understood. One way in which their meaning can be visualized is as follows. Suppose that 1000 realizations of **k** are generated on the basis of $C(\mathbf{k})$ using a stochastic field generator. Suppose that, for each of these **k** fields, $\underline{\mathbf{p}}$ is then inferred (using equation 8) from a flow series generated on the basis of that **k** realization. The variability of $\underline{\mathbf{p}}$ between these 1000 realizations will then be expressed by equations (9a) and (9b). This has important repercussions when assessing the values of lumped parameters calculated through the parameter estimation process against values supplied for the "ranges of credibility" of these parameters as provided in guidance documents. Implicit in the notion that a lumped parameter's variability should be the same as that of a real physical property is that each diagonal element of $C(\underline{\mathbf{p}})$ is no larger than the diagonal elements of $C(\mathbf{k})$ corresponding to the same parameter type. We will examine the plausibility of this assumption shortly.

3.8. The Relationship Between Lumped and Real Parameter Values

[39] Through combining equations (3b) and (8), we have:

$$\underline{\mathbf{p}} = (\mathbf{X}^t\mathbf{QX})^{-1}\mathbf{X}^t\mathbf{QZk} = \mathbf{R}'\mathbf{k} \tag{11}$$

We use the symbol $\mathbf{R}'$ rather than $\mathbf{R}$ because the relationship expressed by equation (11) is not unlike that which defines the resolution matrix in regularization theory, for which the symbol $\mathbf{R}$ is normally provided (see, e.g., *Moore and Doherty* [2005] and the many references cited therein for a discussion of the resolution matrix). However, the resolution matrix is normally square, representing the relationship between estimated distributed (nonlumped) parameters and their "true" counterparts, rather than between lumped parameters and the "true" hydraulic properties which they represent on a spatially averaged basis. The resolution matrix itself is easily derived from equation (11) using equation (2). That is, with

$$\underline{\mathbf{k}} = \underline{\mathbf{j}} = \mathbf{L}\underline{\mathbf{p}} \tag{12a}$$

we have

$$\underline{\mathbf{k}} = \mathbf{L}(\mathbf{X}^t\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Q}\mathbf{Z}\mathbf{k} = \mathbf{R}\mathbf{k} \tag{12b}$$

Note that because $\mathbf{R}$ and $\mathbf{R}'$ are functions of $\mathbf{X}$ and $\mathbf{Z}$, the nature of the relationship between estimated parameters $\underline{\mathbf{k}}$ and $\mathbf{p}$, and real hydraulic properties $\mathbf{k}$ is a function of the amount and type of data employed in the calibration process.

[40] Another important relationship is that between the estimates of average parameter values $\underline{\mathbf{p}}$ obtained through model calibration and that obtained through averaging the real-world parameter field $\mathbf{k}$, this being $\mathbf{p}$. From equation (11):

$$\mathbf{p} - \underline{\mathbf{p}} = \mathbf{p} - \mathbf{R}'\mathbf{k} = (\mathbf{N} - \mathbf{R}')\mathbf{k} \tag{13a}$$

Alternatively, from equation (12b)

$$\mathbf{p} - \underline{\mathbf{p}} = \mathbf{N}(\mathbf{I} - \mathbf{R})\mathbf{k} \tag{13b}$$

[41] The covariance matrix of model parameter error (where "error" is defined here as the difference between estimated lumped parameter values, and the average values over the model domain of the "real" hydraulic properties which they represent) is then easily evaluated as:

$$\mathrm{C}\left(\mathbf{p} - \underline{\mathbf{p}}\right) = (\mathbf{N} - \mathbf{R}')\mathrm{C}(\mathbf{k})(\mathbf{N} - \mathbf{R}')^t \tag{14}$$

4. Application of Theory to Present Model

4.1. The Structure of Structural Noise

[42] Let $[\sigma_\varepsilon]_i$ denote the square root of the ith diagonal element of the $\mathrm{C}(\varepsilon)$ matrix. Recall that $\mathrm{C}(\varepsilon)$ is a 4381 × 4381 matrix, 4381 being the number of observations comprising the flow series on which calibration of the model is based. Thus $[\sigma_\varepsilon]_i$ is the standard deviation of structural noise associated with the ith observed flow; as stated above, this is thus the standard deviation of error induced by representation of a heterogeneous system by a lumped (in this case averaged) system. σ_ε is plotted against time in Figure 3 together with flow itself for the case where the standard deviation of log hydraulic property variability is 1.24E−01, corresponding to a factor change of about 33%. In general, σ_ε is between 6% and 10% of flow for our example. Note that it could have been made proportionally higher or lower simply through increasing or decreasing the standard deviation of spatial hydraulic property variability. It is apparent, however, that at this level of heterogeneity, the magnitude of lumping-induced noise is of the same order as that often associated with flow measurement noise.

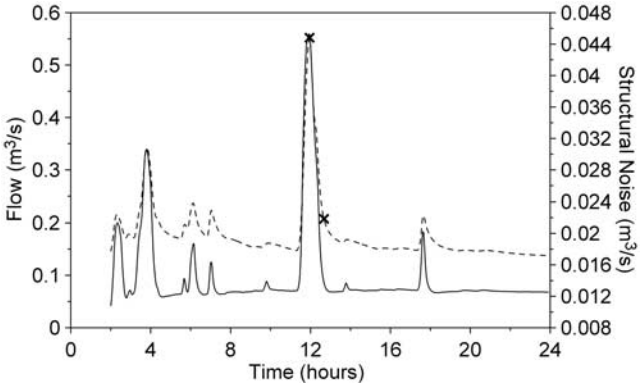

Figure 3. Model-predicted flows (dashed) and lumping-induced structural noise (solid). Crosses indicate peak flow and a representative point on the falling flow limb.

[43] Temporal correlation of structural noise can be examined through inspection of correlation coefficients of flows at different times. The correlation coefficient $\rho_{i,j}$ between noise pertaining to flow at time i and that pertaining to flow at time j is defined as:

$$\rho_{ij} = \frac{\sigma_{ij}^2}{\sqrt{\sigma_{ii}^2\sigma_{jj}^2}} \tag{15}$$

where σ_{ij} is the same as the square root of $[\mathrm{C}(\varepsilon)]_{ij}$. ρ_{ij} ranges between −1 and 1; the closer it is to zero, the less correlation exists between corresponding flow terms. Figure 4a shows ρ_{ij} plotted against time, where i for this plot corresponds to the time of peak flow. Figure 4b shows a similar plot where, on this occasion, i corresponds to a point on the falling limb of the hydrograph following the flow peak; see Figure 3 for the locations of these flows in the overall flow time series.

[44] It is immediately apparent from Figures 4a and 4b that the degree of correlation between structural error at different times during the flow is high. What is also obvious is that correlation between flow terms tends to be highest not between flows that are temporally close, but between flows that occupy similar locations on the hydrographs of neighboring events. This is not a surprising outcome as intuition would lead us to expect that a high degree of commonality would exist between a lumped model's failure to reproduce the same aspects of different events.

[45] As was discussed above, flow time series are often subject to various types of transformation in an attempt to create a time series for which noise is both homoscedastic and uncorrelated. For our present example, although we tried a number of different Box-Cox transformations and ARMA models, we were able to find none that even came close to achieving these goals. A "preferred ARMA order" for our model-generated flow time series was estimated as the minimum of the asymptotical selection criterion described by *Broersen* [2000]. This criterion is a function of

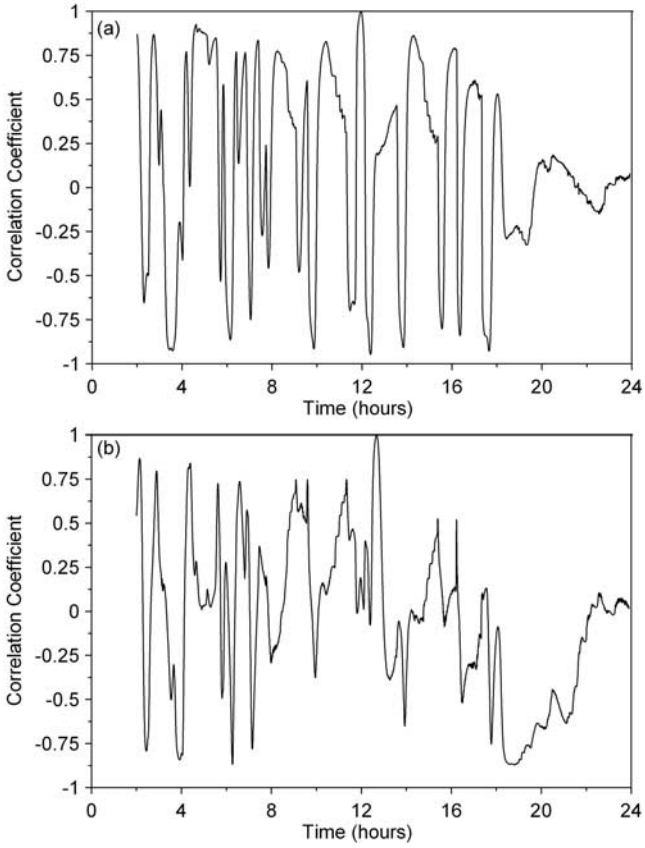

Figure 4. Temporal correlation coefficients for (a) peak flow and (b) a representative point on the recession limb (refer to Figure 3) of the hydrograph.

the variance of the residuals for an ARMA model of given order and the number of time series observations. ARMA model coefficients were estimated using an algorithm described by *Broersen* [2002] which employs a computationally efficient approach to determine ARMA coefficients using the method of *Durbin* [1960]. Figures 5a and 5b show the same correlations as are depicted in Figures 4a and 4b, but for natural log-transformed flows subject to an ARMA (5,4) transformation.

[46] The inability to fully "transform out" temporal correlation is hardly surprising in view of the fact that correlation of structural noise is based on similarity of events rather than on contemporaneity of flow terms as discussed above. It would thus appear that where structural noise is a large contributor to the total noise of a measurement data set, the likelihood of finding an ARMA transformation that can remove temporal correlation is very small indeed. (We have noted that the successful deployment of an ARMA model to eliminate serial residual correlation in hydrologic calibration data sets most often involves an observation time increment much larger than the 18 s employed herein, in fact often of the order of a day or even a month [see, e.g., *Kuczera*, 1983; *Bates and Campbell*, 2001]). This interval may be large enough to significantly reduce the influence of discrete events, and the correlation of structural noise between them, prior to fitting of the ARMA model.

[47] Through consideration of equation (10), the notion that an ARMA model can be fitted to the noise associated with a flow time series can be shown to be theoretically impossible where that noise is entirely structural. Under these circumstances, the rank of $C(\varepsilon)$ can be no greater than that of $(\mathbf{I} - \mathbf{NL})$. As stated above, $\mathbf{I} - \mathbf{NL}$ is idempotent and is therefore rank deficient [*Koch*, 1987]. In the present case, its rank is at most $n - p$, where n is the rank of $\mathbf{I}$ (in this case 122) and p is the rank of $\mathbf{NL}$ (in this case 6). Because $C(\varepsilon)$ is thus of diminished rank, no linear transformation of it can create a diagonal matrix.

[48] A final remark is worth making before leaving the subject of lumping-induced structural noise. *Moore and Doherty* [2006], and papers cited therein, discuss extensively the price to be paid for seeking a unique solution to the inverse problem of calibrating an environmental model which simulates a complex world. The "ideal model" should be capable of reproducing the flow time series depicted in Figure 3 exactly. The matrix $\mathbf{Z}$ (with a rank of 122, and therefore with a range space composed of 122 dimensions in observation space) represents such a model. However, it is not possible to estimate anything like this number of parameters, especially if the observation time series is contaminated by even a tiny amount of measurement noise. Hence a small, identifiable subset of these 122 parameters (with the others held fixed) must be estimated, this subset being composed of either discrete members of the original 122 parameters or of linear combinations of

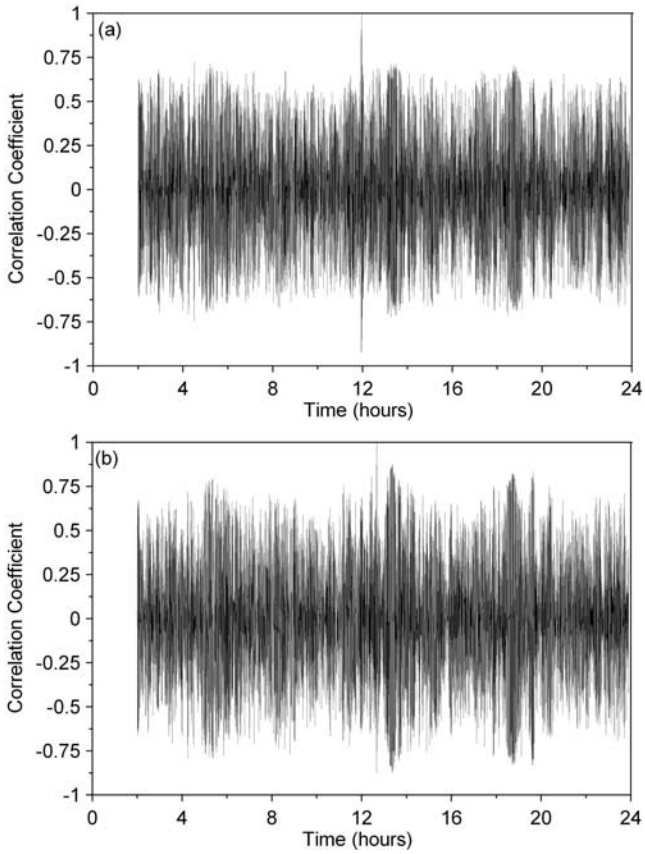

Figure 5. Temporal correlation coefficients for (a) peak flow and (b) a representative point on the recession limb (refer to Figure 3) of the hydrograph following peak flow. Flows are log-transformed and subject to an ARMA(5,4) transformation.

these parameters, as in the present case. This requires that we work with a model matrix **X** in place of the true model matrix **Z**. While this allows us to formulate a well-posed inverse problem that guarantees unique estimation of the reduced number of model parameters, the reduced rank of **X** impedes the ability of the model to reproduce all nuances of the flow time series under all conditions, for this can only be calculated on the basis of the true model matrix **Z**. The result, as equation (6) demonstrates, is nonzero structural noise; this noise cannot be eliminated as **NL** cannot equal **I** because of the rank-deficient nature of the former matrix.

4.2. The Relationship Between Estimated and True Parameters

[49] Equation (11) allows us to calculate the relationship between the lumped parameters estimated through model calibration using equation (8) and "real-world" hydraulic properties. For a particular lumped parameter, the plotting of elements comprising the pertinent row of $\mathbf{R}'$ against the location of the pilot point corresponding to each element allows the contribution made by different areas within the model domain to the estimated value of that lumped parameter to be assessed. Through this mechanism, it is also possible to characterize the contribution made to the estimated value of that parameter by other parameter types; thus the degree to which system properties of certain types contribute to the value of an "effective parameter" of another nominal type can be assessed (this being sometimes referred to as "parameter contamination" [see, e.g., *Lines and Treitel*, 1984]).

[50] Figures 6a–6f show a series of such plots for parameter log(X_FRICTN) where parameter estimation using equation (8) takes place on the basis of log-transformed flows, and where the standard deviation of log spatial variability is the same value (i.e., 1.24E−01) for all hydraulic properties. (Logs are used because, for the parameter types employed in our study, the relationship between model outputs and parameter values is generally more linear than that between model outputs and native parameter values.) Note, however, that these plots are independent of the absolute variability of different hydraulic properties; they are dependent only on the variability of different hydraulic properties relative to each other. Note also that **Q** in equation (8) was set to **I**. In the plots of Figure 6, each element of the $\mathbf{R}'$ matrix row is normalized by the value of log(X_FRICTN) so that relative contributions made by pilot points within different areas of the model domain, and by pilot points pertaining to different parameter types, are readily visible. It is immediately apparent that the estimated value for the lumped parameter log(X_FRICTN) can hardly be considered to be an average of the log of values pertaining to pilot points for that same parameter over the model domain; in fact, the averaging function is heterogeneous and is significantly nonzero only over the lower parts of the model domain. This is not surprising, for it is here that overland flow is greatest and where X_FRICTN is thus of most importance. For the present model configuration, it is zero in the upland parts of the watershed where the water table does not reach the surface even following the heavy rainfall event at time 11.75 hours. Figure 7 shows the maximum spatial extent of complete saturation of the subsurface during this rainfall event; this constitutes approximately 4% of the total model area available for overland flow. As studies such as those of *Dunne and Black* [1970], *Moore et al.* [1988], *Anderson and Burt* [1990], and *Guntner et al.* [2004] show, runoff from these locations mostly dominates the overall response of a watershed to rainfall.

[51] What is of greater interest, however, are the significant positive and negative contributions to the value of X_FRICTN made by local values of the BOT_LKG hydraulic property, and to a lesser extent by HY. Thus X_FRICTN, as represented in our calibrated lumped-parameter model, is partly X_FRICTN, partly BOT_LKG, and partly HY. What it is definitely not, is the spatial average of X_FRICTN over the model domain.

[52] Though not shown herein for lack of space, similar conclusions apply to other parameters.

[53] The situation is somewhat improved where **Q** of equation (8) is assigned a value equal to the Moore-Penrose inverse of $\mathbf{C}(\varepsilon)$ (with $\mathbf{C}(\varepsilon)$ computed for log flows using equation (10); see Figure 8). However, while averaging tends to be more even over the model domain, and while the contributions made by parameters of certain types to effective parameters of another type are lower, unevenness of averaging and contamination of the value estimated for one parameter by that of another property type have not been eradicated. (In both of Figures 6 and 8, complete eradication of both of these phenomena would be indicated by a uniform X_FRICTN contribution plot with a value of 0.04 (this being equal to one divided by the number of pilot points (25) used in its characterization) and a contribution of zero by all other hydraulic property types.)

4.3. Parameter Error

[54] Let us denote $\sigma_{p_i - \underline{p}_i}$ by the standard deviation of the potential error of the ith parameter value $\underline{p}_i$ as an estimate of p_i. Recall that p_i in the present case is the average of the log of those elements of **k** which represent parameter type i over the model domain. $\sigma_{p_i - \underline{p}_i}$ is calculated by taking the square root of the ith diagonal element of $\mathbf{C}(\mathbf{p} - \underline{\mathbf{p}})$ calculated using equation (14).

[55] Where parameter estimation is carried out on the basis of log-transformed flows, $\sigma_{p_i - \underline{p}_i}$ for the six different parameter types are plotted in the middle row of Figure 9a. The back row of this plot shows σ_{k_i} for comparison. σ_{k_i} is the standard deviation of spatial variability of each log-transformed hydraulic property type, these being the diagonal elements of $\mathbf{C}(\mathbf{k})$; in the present case, all of these have the same value of 1.24E−01.

[56] Figure 9b shows a similar plot, but in this case based on a weighting matrix **Q** equal to the Moore-Penrose inverse of $\mathbf{C}(\varepsilon)$. It is apparent that with this choice of **Q**, estimated parameter error is considerably reduced. Thus the values of parameters estimated for the lumped parameter model much more closely approximate the respective average values of the pertinent hydraulic property type over the model domain.

4.4. Estimated Parameter Ranges

[57] Let σ_{p_i} denote the square root of the ith diagonal element of $\mathbf{C}(\underline{\mathbf{p}})$ as calculated using equation (9b). This thus quantifies the variability that can be expected in estimates of each average parameter value that arises as a result of the fact that true hydraulic properties themselves are heterogeneous over the model domain. These are shown for our

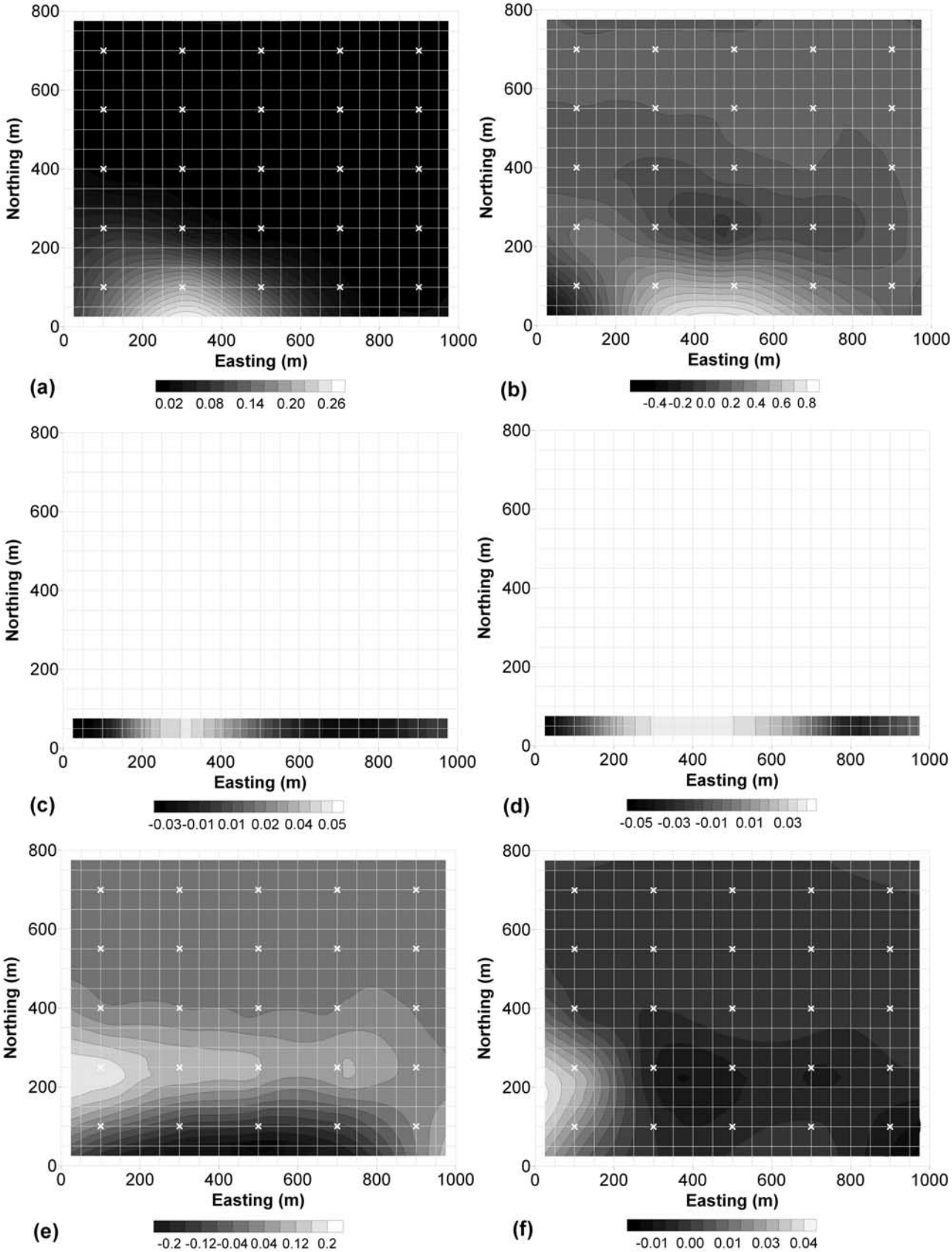

Figure 6. Contribution to the value of lumped parameter log(X_FRICTN) by the logs of (a) X_FRICTN, (b) BOT_LKG, (c) RGHN_CH1, (d) BEDCOM_CH1, (e) HY, and (f) SF1. Parameter estimation takes place on the basis of log flows and a weight matrix **Q** equal to **I**. Scale represents fraction of the value of log(X_FRICTN).

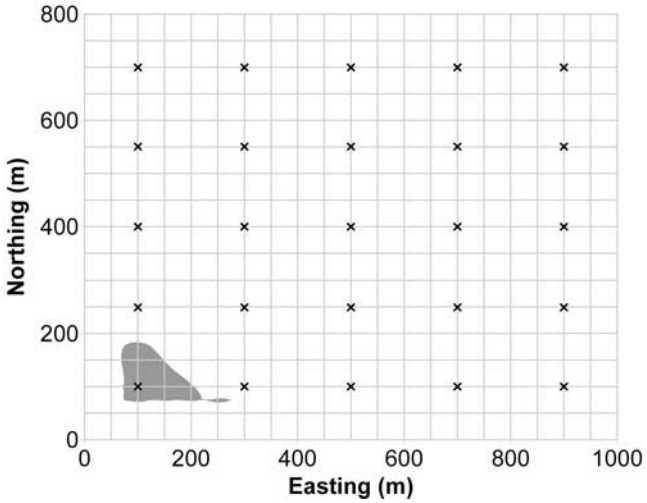

Figure 7. Maximum spatial extent of complete subsurface saturation.

six parameter types in the front row of Figure 9a where parameter estimation is undertaken on the basis of log-transformed flows with **Q** equal to **I**. Where **Q** is equal to the Moore-Penrose inverse of C(ε), the parameter ranges depicted in the front row of Figure 9b result.

[58] Up until now, it has been assumed that the log of each real-world parameter type represented by **k** has an innate variability that is characterized by a standard deviation of 1.24E−01 and hence a variance (diagonal elements of C(**k**)) of 1.55E−02. As stated above, this corresponds to a variability of about 33% for each native hydraulic property. The more realistic assumption will now be made that the amplitude of spatial variability is different for different parameter types; however, the degree of spatial correlation between parameters of the same type is assumed to be the same as in our previous analysis (i.e., exponential with a decay constant of 150 m^{-1}). Log parameter standard deviations employed for this numerical experiment are listed in the first column of Table 2. These, together with σ_{pi-pi} and σ_{pi} calculated for different lumped parameter types are plotted in Figure 10a. For **Q** set to the Moore-Penrose inverse of C(ε), estimated lumped parameter error is once again noticeably diminished (see Figure 10b).

[59] σ_{pi} and σ_{pi-pi} as calculated using equations (9b) and (14), respectively, are depicted in Figure 11a for the following conditions:

[60] • Inversion is based on log flows

[61] • All such flows are assigned a weight of 1.0 (i.e., with **Q** equated to **I**)

[62] • Spatial parameter variability is characterized directly in terms of native parameters, and native parameters are estimated through the parameter estimation process (as is often done in practice)

[63] • The standard deviation of spatial variability of each hydraulic property is assigned the respective value shown in the second column of Table 2

[64] As for the two previous figures, the back row of Figure 11a depicts σ_{ki} (these standard deviations of variability corresponding to the second column of Table 2). For **Q** equal to the Moore-Penrose inverse of C(ε), estimated parameter error is again markedly reduced (see Figure 11b).

5. Discussion

[65] Though our numerical experiments were restricted to the lumping of only six parameters on a hillside of small spatial extent, and though "lumping" in our study consisted only of averaging, a number of conclusions can be drawn from this work that have important repercussions when considering the calibration of lumped parameter models representing much larger systems. For such systems, the "lumping" required for representation of environmental processes by a simplified model is far more pervasive and far more complex than that discussed above, as processes, rather that just parameters which describe the properties pertaining to these processes, undergo vast simplification so that they may be represented numerically. While we have not specifically studied the effects of these kinds of lumping, it is hoped that an examination of the effects of lumping-through-averaging as was undertaken herein will at least throw some light on the effects of the much more pervasive and complex lumping that is the "bread and butter" of hydrologic modeling.

[66] Some initial conclusions are that:

[67] 1. Lumping induces structural noise.

[68] 2. The level of this structural noise is commensurate with that commonly encountered in everyday modeling practice (at least in the case of our synthetic example).

[69] 3. This noise shows a high degree of correlation which is more a function of event similarity than temporal juxtaposition.

[70] An immediate repercussion of the above conclusions is that strategies such as Box-Cox transformation of flows and ARMA modeling of flow residuals which attempt to promulgate homoscedascity and reduce temporal correlation are unlikely to meet with a great deal of success. On the other hand, a homoscedastic error time series for which structural noise shows no correlation can indeed be formulated through premultiplication of (transformed) flows by $\mathbf{Q}^{1/2}$ (this being equivalent to employing the original (transformed) flow time series in the inversion process in conjunction with a weighting matrix of **Q**), where **Q** is calculated in a manner that explicitly accounts for noise that is incurred as a result of system and parameter simplification (including averaging). In all of the cases demonstrated herein, indicators of parameter estimation malperformance, such as the standard deviation of parameter error, are greatly reduced through this mechanism. Unfortunately, however, such a strategy is not feasible in real-world hydrologic modeling because:

[71] 1. It is impossible to compute the covariance matrix of structural noise for the types of lumping employed by real-world lumped parameter models.

[72] 2. Even if this were possible, it would be impractical to incorporate such large nondiagonal matrices in the inversion process.

[73] Another inevitable consequence of parameter lumping is that it cannot be presumed that any particular parameter employed by a lumped parameter model represents only the hydraulic property after which it is named. Figure 6 demonstrates that even for comparatively mild forms of lumping such as spatial averaging, a considerable

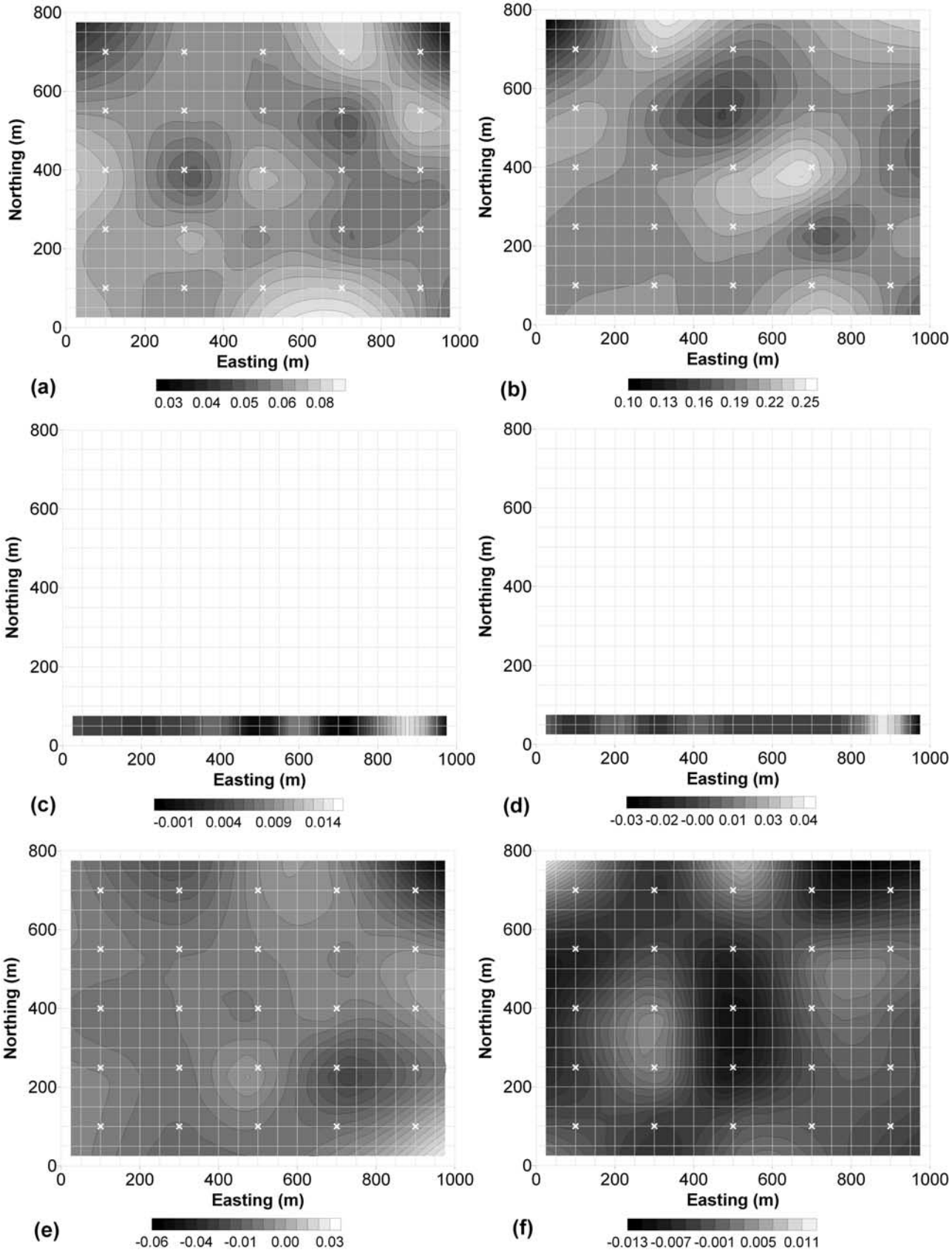

Figure 8. Contribution to the value of lumped parameter log(X_FRICTN) by the logs of (a) X_FRICTN, (b) BOT_LKG, (c) RGHN_CH1, (d) BEDCOM_CH1, (e) HY, and (f) SF1. Parameter estimation takes place on the basis of log flows and the weight matrix **Q** equal to the Moore-Penrose inverse of $C(\epsilon)$. Scale represents fraction of the value of log(X_FRICTN).

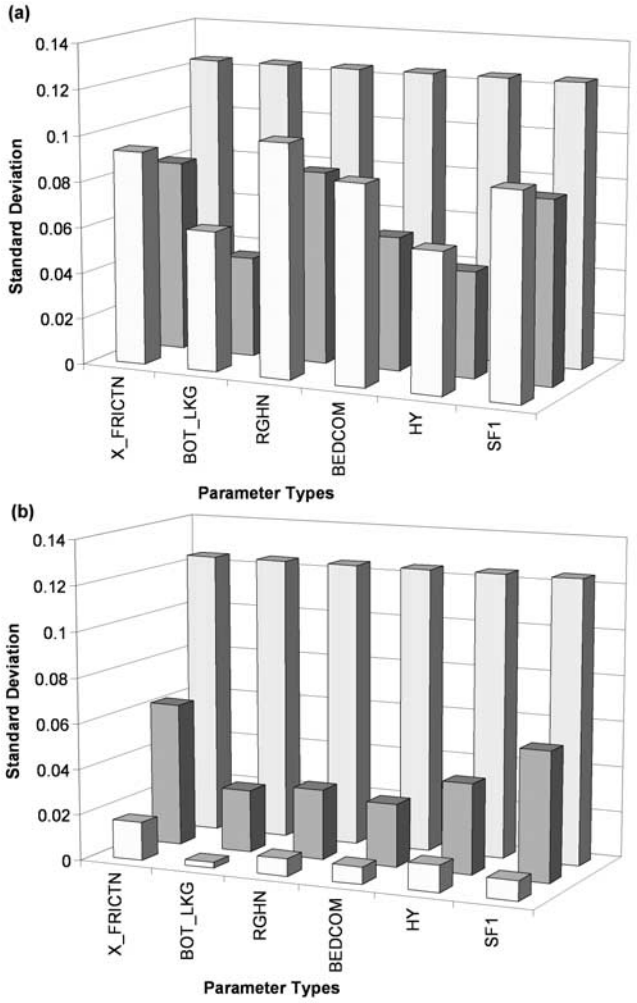

Figure 9. Error standard deviations of estimated lumped parameters $\underline{\mathbf{p}}$ (front row) and of parameter error $\mathbf{p} - \underline{\mathbf{p}}$ (middle row). Standard deviations of spatial variability of each hydraulic property type (i.e., $\mathbf{k}$) are shown in the back row. All parameters are $\log_{10}$ transformed with a standard deviation of log hydraulic property variability of 1.24E−01. In Figure 9a, $\mathbf{Q}$ is equal to $\mathbf{I}$, while in Figure 9b, it is equal to the Moore-Penrose inverse of $\mathbf{C}(\varepsilon)$.

contribution of one hydraulic property type to effective values of a parameter type representing another hydraulic property is likely to exist. Hence the values assigned to some parameters may be more of a reflection of hydraulic properties which they are not supposed to represent than of those which they are. Furthermore, a lumped parameter cannot be the spatial average of a particular hydraulic property simply because it is named such, is designed to have a similar role to that property, or plays the same role as the spatial average of that property in forward simulations. On the other hand, if it is desired that lumped parameters employed by a calibrated model represent spatial averages of hydraulic properties to the greatest degree possible, then as *Cooley* [2004] shows, this must be specifically sought through selection of an observation weight matrix that reflects the stochastic character of lumping-induced measurement noise. Unfortunately, such a matrix will not normally be available, for hydrologic model lumping involves far more than the averaging of spatially variable parameters over the domain of a complex physically based model; instead it involves the use of a completely different model with spatial hydraulic property averaging undertaken as an implicit aspect of the much more pervasive

Table 2. Spatial Variability Adopted for Hydraulic Properties Represented in MODHMS

Parameter	Standard Deviation (Approximate Percent Variation)	
	Case 1: Log Space	Case 2: Native
X_FRICTN	9.20E−02 (24%)	3.71E−03 (25%)
BOT_LKG	1.80E−01 (52%)	2.83E−07 (57%)
RGHN_CH1	8.06E−02 (20%)	8.48E−03 (21%)
BEDCOM_CH1	6.86E−02 (17%)	1.79E−05 (21%)
SF1	1.80E−01 (52%)	8.49E−05 (57%)
HY	6.86E−02 (17%)	1.26E−02 (14%)

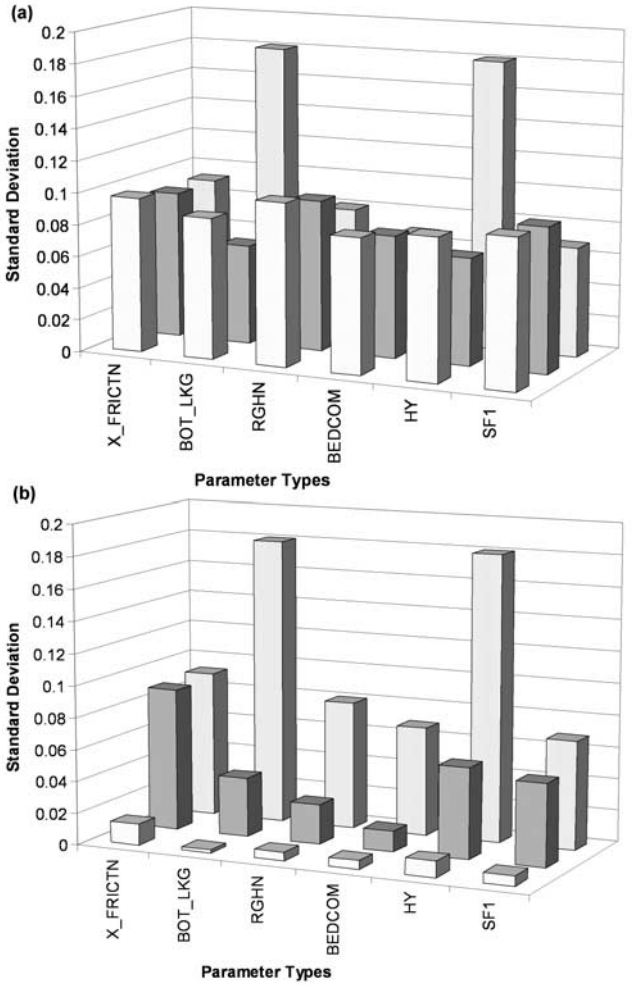

Figure 10. Error standard deviations of estimated lumped parameters $\underline{\mathbf{p}}$ (front row) and of parameter error $\mathbf{p} - \underline{\mathbf{p}}$ (middle row). Standard deviations of spatial variability of each hydraulic property type (i.e., $\mathbf{k}$) are shown in the back row. All parameters are $\log_{10}$ transformed; variability is assumed to be different for each hydraulic property type (back row). In Figure 10a, $\mathbf{Q}$ is equal to $\mathbf{I}$, while in Figure 10b, it is equal to the Moore-Penrose inverse of $\mathbf{C}(\varepsilon)$.

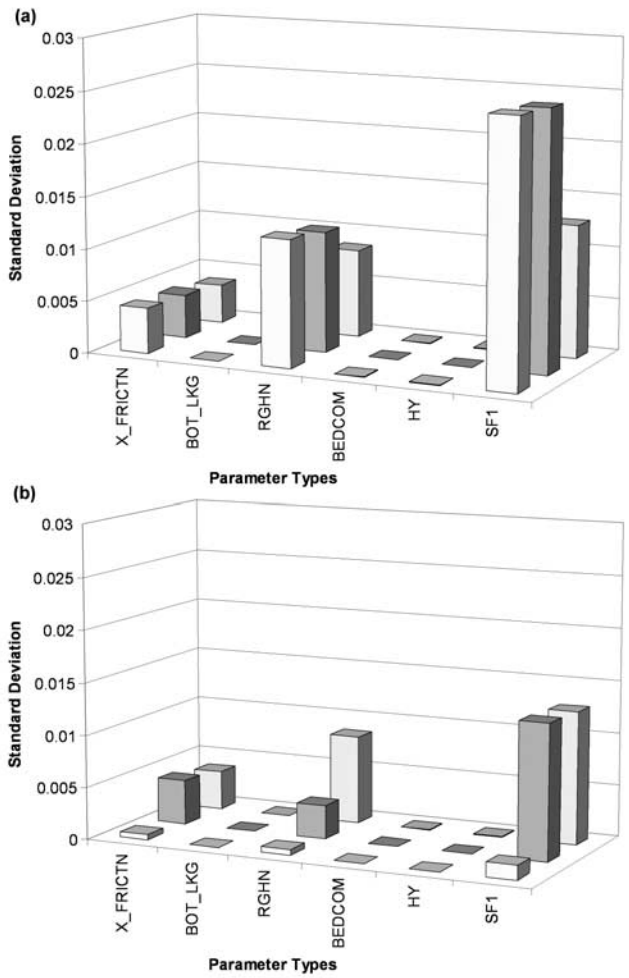

Figure 11. Error standard deviation of estimated lumped parameters $\underline{\mathbf{p}}$ (front row) and of model parameter error $\mathbf{p} - \underline{\mathbf{p}}$ (middle row). Standard deviations of spatial variability of each hydraulic property type (i.e., $\mathbf{k}$) are shown in the back row. Hydraulic property variation is characterized as native, as is estimation of parameters. Parameter estimation is based on the log of flows with $\mathbf{Q}$ being equal to $\mathbf{I}$ in Figure 11a and to the Moore-Penrose inverse of $C(\varepsilon)$ in Figure 11b.

simplifications required for design of the model. The inevitable result of this process is an erosion of the credibility of the estimated value of any lumped parameter as a representor of the average value of the corresponding hydraulic property over the model domain. Thus in short, every parameter employed by a lumped parameter model will be partly what it says it is and partly something else (possibly many other things).

[74] If parameter error ($\mathbf{p} - \underline{\mathbf{p}}$) is large through failure to employ a $\mathbf{Q}$ matrix that reflects lumping-induced measurement noise, then predictive error may also be large (because predictions are, of course, functions of parameters, and thus predictive error is a function of parameter error). Under certain circumstances of model usage, this problem can be exacerbated. This can occur when the model calibration process is employed to estimate one or a number of parameters, which are then assumed to change under a posited land use for which the model is then employed to investigate hydrological repercussions. For example, watershed impervious area may have been estimated through the model calibration process. But if its estimated value is "contaminated" by the values of other hydraulic properties to an unknown extent, then the alteration of that value by a certain percentage to reflect posited alterations to real-world hydraulic properties by the same amount may result in predictions which are just as "contaminated".

[75] Figures 10a and 11a have immediate repercussions for the manner in which "reality checks" are imposed on parameters estimated through the calibration process. Conventional wisdom is categorical in its demands that parameter values "stay within range," and most modeling reports make reference to the fact that estimated parameters have been constrained to lie within ranges that are considered acceptable for the hydraulic properties that they represent, notwithstanding the lumped nature of these parameters. However, Figures 10a and 11a demonstrate that the variability of estimated parameters may be greater than those of the system properties that they represent as an outcome of the fact that their values are partially determined by other hydraulic properties. Figures 10a and 11a suggest that the enforcement of bounds on estimated lumped parameter values, where the values of these bounds are set in accordance with the innate variability of real-world hydraulic properties, may therefore lead to a deterioration in the level of fit between modeled and observed flows achieved through the calibration process. Whether a prediction will be better or worse because of this is a subject that requires further investigation, using methodologies such as those employed by *Moore and Doherty* [2005] which allow calculation of predictive error variance using a simple extension of the regularization concepts discussed above.

6. Conclusions

[76] Through undertaking a series of numerical experiments, some repercussions of the necessity to "lump" or "average" parameters as a necessary component of real-world (especially large scale) modeling have been examined. It has been demonstrated that, when estimated through the calibration process, lumped parameters are partly what they are intended to be, but take on other roles as well. Where lumping is achieved purely through averaging of heterogeneous property fields as in the present study, it has been demonstrated that the values assigned to some parameters can be heavily influenced by the values pertaining to hydraulic properties that they do not purport to represent. This will almost certainly also be the case in more complex modeling contexts where lumping of both processes and properties occurs. In such cases, despite the best of intentions in the design, construction, and calibration of such models, lumped parameters cannot help but become, to a greater or lesser degree, unavoidable surrogates for model inadequacies and implied averaging processes of complex and unknown nature. Furthermore, to the extent that a parameter does indeed represent the hydraulic property after which it is named, its value may be the outcome of an averaging process that takes place over only a small part of the domain of the study area. To the extent that a prediction depends on processes that may operate over other parts of the domain, the calibration process may do little to increase the precision with which that prediction can be made.

[77] Over many years, the authors have witnessed heated philosophical debates between modelers over whether a good fit should hold more sway than parameter reasonableness as an outcome of the model calibration process. The answer to this question must await a continuation of the present investigation in which predictive error variance analysis is incorporated into the study. (Using a simple extension of the mathematics presented herein, *Moore and Doherty* [2006] demonstrate that choice of a weighting/transformation strategy that is "tuned" to a prediction may result in lower predictive error variance for that prediction.) Nevertheless, the study has effectively demonstrated insights that have been known for some time through application of regularization theory in other fields. These are as follows:

[78] 1. Solution of an ill-posed inverse problem requires some form of regularization. In other fields, this often takes the form of mathematical regularization; in many instances of hydrological modeling, it takes the form of parameter lumping. The ill-posedness of the inverse problem is a fundamental outcome of the inherent complexity of the real world.

[79] 2. Regularization implies the existence of an implicit or explicit resolution matrix, in which the relationship between the "effective parameters" employed by a model and real-world hydraulic properties is recorded.

[80] 3. This relationship will involve some kind of averaging of real-world properties, with the averaging function often crossing parameter boundaries.

[81] In the mean time, perhaps an interim conclusion can be drawn from an inspection of Figures 8b, 9b, and 10b. Recall that these figures depict calibrated lumped parameter error and ranges of variability for an idealized weighting strategy inferred from a structural noise covariance matrix that we will never know in any real-world modeling context involving a lumped parameter model. The level of model-to-measurement fit in terms of flows, natural logs of flows, Box-Cox-transformed flows, etc., will never be as good through use of the optimized **Q** matrix on which these figures are based than if the differences between these quantities are minimized directly. This is because use of the former weighting strategy aims to minimize $\mathbf{r}^t\mathbf{Q}\mathbf{r}$, where **r** represents (transformed) model-to-measurement flow residuals, rather than $\mathbf{r}^t\mathbf{r}$ directly. Hence the model will not appear to be as well "calibrated" in the sense that model-to-measurement fit will not appear to be as good. However, this figure shows a low range of variability for the estimated values of lumped parameters: lower than the innate variability of the hydraulic properties that these parameters represent because the integrity of the lumping-through-averaging process is preserved through use of this **Q** weighting matrix which respects the type of structural noise induced by this particular kind of lumping. In other words, parameter values employed by the model deviate little from values that are presumed to be reasonable on the basis of expert knowledge of system properties. These figures also show that parameter error is likely to be smaller for this weighting strategy than it is for all of the other weighting strategies that we have used. This leads to the suggestion that because we will never know the optimum **Q**, or the degree of variability of hydraulic properties C(**k**) over a study area for that matter, we can achieve a similar outcome to that achieved through use of an optimized **Q** through maintaining plausibility of optimized parameter values (in terms of the hydraulic properties that they purport to represent), even if we sacrifice fit in achieving this. This is a powerful argument in favor of the approach recommended by those such as *Lumb and Kittle* [1993] who eschew use of sophisticated parameter estimation tools in favor of expert knowledge as the soundest basis for model calibration.

[82] **Acknowledgments.** We wish to thank Hydrogeologic Inc. for providing us with a copy of MODHMS free of charge for use in our investigation.

References

Ababou, R., and E. F. Wood (1990), Comment on 'Effective groundwater model parameter values: Influence of spatial variability of hydraulic conductivity, leakage, and recharge' by J. J. Gomez-Hernandez and S. M. Gorelick, *Water Resour. Res.*, *26*(8), 1843–1846.

Abbott, M. B., J. C. Bathurst, J. A. Cunge, P. E. O'Connell, and J. Rasmussen (1986a), An introduction to the European Hydrological System-Systeme Hydrologique Europeen, 'SHE', 1: History and philosophy of a physically based distributed modelling system, *J. Hydrol.*, *87*, 45–59.

Abbott, M. B., J. C. Bathurst, J. A. Cunge, P. E. O'Connell, and J. Rasmussen (1986b), An introduction to the European Hydrological System-Systeme Hydrologique Europeen, 'SHE', 2: Structure of a physically based distributed modelling system, *J. Hydrol.*, *87*, 61–77.

Anderson, M. G., and T. P. Burt (1990), *Process Studies in Hillslope Hydrology*, John Wiley, Hoboken, N. J.

Ao, T., H. Ishidaira, K. Takeuchi, A. Kiem, J. Yoshitar, K. Fukami, and J. Magome (2006), Relating BTOPMC model parameters to physical features of MOPEX basins., *J. Hydrol.*, *320*, 84–102.

Aronica, G., B. Hankin, and K. J. Beven (1998), Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data, *Adv. Water Resour.*, *22*(4), 349–365.

Bard, Y. (1974), *Nonlinear Parameter Estimation*, Elsevier, New York.

Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, *37*(4), 937–948.

Bergstrom, S. (1976), Development and application of a conceptual runoff model for Scandinavian countries, *SMHI Rep. No. 7*, Norrkoping, Sweden.

Beven, K. J. (1989), Changing ideas in hydrology: the case of physically-based models, *J. Hydrol.*, *105*, 157–172.

Beven, K. J. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, *320*, 18–36.

Bicknell, B. R., J. C. Imhoff, J. L. Kittle Jr., A. S. Donigian Jr., and R. C. Johanson (1997), Hydrological Simulation Program—FORTRAN, User's Manual for Version 11, *EPA/600/R-97/080*, U. S. EPA, National Exposure Research Laboratory, Athens, GA.

Box, G. E. P., and D. R. Cox (1964), An analysis of transformations, *J. R. Stat. Soc., Ser. B*, *26*, 211–243.

Box, G. E. P., and G. M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, Holden-Day, Boca Raton, Fla.

Boyd, M. J., D. H. Pilgrim, and I. Cordery (1979), A watershed bounded network model for flood estimation—Computer programs and user guide, *Water Research Laboratory Rep. No. 154*, Univ. New South Wales, Sydney, Australia.

Broersen, P. M. T. (2000), Facts and fiction in spectral analysis, *IEEE Trans. Instrum. Meas.*, *49*, 766–772.

Broersen, P. M. T. (2002), Automatic spectral analysis with time series models, *IEEE Trans. Instrum. Meas.*, *51*(2), 211–216.

Burnash, R. J. C., R. L. Farral, and R. A. McGuire (1973), A generalised streamflow simulation system—Conceptual modelling for digital computers, pp. 204, Joint Federal–State River Forecast Center, U. S. A.

Campbell, E. P., and B. C. Bates (2001), Regionalisation of rainfall-runoff model parameters using Markov Chain Monte Carlo samples, *Water Resour. Res.*, *37*(3), 731–739.

Certes, C., and G. de Marsily (1991), Application of the pilot points method to the identification of aquifer transmissivities, *Adv. Water Resour.*, *14*(5), 284–300.

Chander, S., A. Kumar, and S. K. Spolia (1978), Flood frequency analysis by power transformation, *J. Hydraul. Div. Am. Soc. Civ. Eng.*, *104*(11), 1495–1504.

Cooley, R. L. (2004), A theory for modeling ground-water flow in heterogeneous media, *Professional Paper 1679*, U. S. Geological Survey, Reston, Va.

Dagan, G. (1979), Models of groundwater flow in statistically homogeneous porous formations, *Water Resour. Res.*, *15*(l), 47–63.

Doherty, J. (2003), Groundwater model calibration using pilot points and regularization, *Ground Water*, *41*(2), 170–177.

Doherty, J., and J. M. Johnston (2003), Methodologies for calibration and predictive analysis of a watershed model, *J. Am. Water Resour. Assoc.*, *39*(2), 251–265.

Doherty, J., and B. E. Skahill (2005), An advanced regularization methodology for use in watershed model calibration, *J. Hydrol*, in press, doi:10.1016/j.jhydrol.2005.11.058.

Downer, C. W., and F. L. Ogden (2002), GSSHA user's manual, gridded surface subsurface hydrologic analysis version 1.43 for WMS 6.1, *ERDC Technical Report*, Engineering Research and Development Center, Vicksburg, MS.

Duan, Q., et al. (2006), Model Parameter Estimation Experiment (MOPEX): Overview and summary of the second and third workshop results, *J. Hydrol.*, *320*(1–2), 3–17.

Durbin, J. (1960), The fitting of time series models, *Rev. Inst. Int. Stat.*, *28*, 233–243.

Dunne, T., and R. D. Black (1970), An experimental investigation of runoff production in permeable soils, *Water Resour. Res.*, *6*, 478–490.

El-Kadi, A. I., and W. Brutsaert (1985), Applicability of effective parameters for unsteady flow in nonuniform aquifers, *Water Resour. Res.*, *21*(2), 183–198.

Ferrante, M., and T.-C. J. Yeh (1999), Head and flux variability in heterogeneous unsaturated soils under transient flow conditions, *Water Resour. Res.*, *35*(5), 1471–1479.

Gomez-Hernandez, J. J., and S. M. Gorelick (1989), Effective groundwater model parameter values: Influence of spatial variability of hydraulic conductivity, leakance, and recharge, *Water Resour. Res.*, *25*(3), 405–419.

Guntner, A., J. Seibert, and S. Uhlenbrook (2004), Modeling spatial patterns of saturated areas: an evaluation of different terrain indices, *Water Resour. Res.*, *40*(5), W05114, doi:10.1029/2003WR002864.

Hirsch, R. M. (1979), Synthetic hydrology and water supply reliability, *Water Resour. Res.*, *15*(6), 1603–1615.

Hogue, T., T. Wagener, J. Schaake, Q. Duan, A. Hall, H. Gupta, G. Leavesley, and V. Andreassian (2004), Model parameter estimation experiment begins new phase, *EOS*, *85*(22), 217–218.

Hornberger, G. M., K. J. Beven, B. J. Cosby, and D. E. Sappington (1985), Shenandoah watershed study—Calibration of a topography-based, variable contributing area hydrological model to small forested catchment, *Water Resour. Res.*, *21*, 1841–1850.

Hydrogeologic, Inc. (2003), MODHMS software (Version 2.0) documentation. Volume I: Groundwater flow modules, Volume II: Transport modules, Volume III: Surface water flow modules., Herndon, Va.

Jakeman, A., and G. Hornberger (1993), How much complexity is warranted in a rainfall-runoff model?, *Water Resour. Res.*, *29*(8), 2637–2649.

Kavetski, D., G. Kuczera, and S. W. Franks (2006), Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artifacts, *J. Hydrol.*, *320*, 173–186.

Koch, K. R. (1987), *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer, New York.

Kokkonen, T. S., A. J. Jakeman, P. C. Young, and H. J. Koivusalo (2003), Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina, *Hydrol. Processes*, *17*, 2219–2238.

Kuczera, G. (1983), Improved parameter inference in catchment models 1. Evaluating parameter uncertainty, *Water Resour. Res.*, *19*(5), 1151–1162.

La Venue, A. M., B. S. RamRao, G. de Marsily, and M. G. Marietta (1995), Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 2. Application, *Water Resour. Res.*, *31*, 495–516.

Lines, L. R., and S. Treitel (1984), Tutorial: a review of least-squares inversion and its application to geophysical problems, *Geophys. Prospect.*, *32*(2), 159–186.

Lumb, A. M., and J. L. Kittle Jr., (1993), Expert system for calibration and application of watershed models, in *Proceedings of the Federal Interagency Workshop on Hydrologic Modeling Demands for the 90's*, edited by J. S. Burton, U. S. Geological Survey Water Resources Investigation Report 93-4018.

Marks, K., and P. Bates (2000), Integration of high-resolution topographic data with floodplain flow models, *Hydrol. Processes*, *14*(11–12), 2109–2122.

Merz, R., and G. Bloschl (2004), Regionalisation of catchment model parameters, *J. Hydrol.*, *287*(1–4), 95–123.

Montoglou, A., and L. W. Gelhar (1987), Effective hydraulic conductivity of transient unsaturated flow in stratified soils, *Water Resour. Res.*, *23*(l), 57–67.

Moore, C., and J. Doherty (2005), The role of the calibration process in reducing model predictive error, *Water Resour. Res.*, *41*(5), W05020, doi:10.1029/2004WR003501.

Moore, C., and J. Doherty (2006), The cost of uniqueness in groundwater model calibration, *Adv. Water Resour.*, *29*(4), 605–623.

Moore, I. D., G. J. Burch, and D. H. Mackenzie (1988), Topographic effects on the distribution of surface soil water and the location of ephemeral gullies, *Trans. ASAE*, *31*, 1098–1107.

Neuman, S. P., and S. Orr (1993), Prediction of steady state flow in nonuniform geologic media by conditional moments—Exact formalism, effective conductivities, and weak approximations, *Water Resour. Res.*, *29*(2), 341–364.

Ogden, F. L., and P. Y. Julien (2002), CASC2D: A two-dimensional, physically based, Hortonian hydrologic model, in *Mathematical Models of Small Watershed Hydrology and Applications*, edited by V. P. Singh and D. K. Frevert, pp. 69–112, Water Resources Publications, Highlands Ranch, Colo.

Panday, S., and P. S. Huyakorn (2004), A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow, *Adv. Water Resour.*, *27*, 361–382.

Pappenberger, F., K. J. Beven, M. Horritt, and S. Blazkova (2005), Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations, *J. Hydrol.*, *302*(1–4), 46–69.

Post, D. A., and A. J. Jakeman (1999), Predicting the daily streamflow of ungauged catchments in S. E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model, *Ecol. Modell.*, *123*, 91–104.

RamRao, B. S., A. M. La Venue, G. de Marsily, and M. G. Marietta (1995), Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments, *Water Resour. Res.*, *31*, 475–493.

Reed, S., V. Koren, M. Smith, Z. Zhang, F. Moreda, D.-J. Seo, and DMIP Participants (2004), Overall distributed model intercomparison project results, *J. Hydrol.*, *298*, 27–60.

Refsgaard, J. C., and B. Storm (1995), MIKE SHE in *Computer Models of Watershed Hydrology*, edited by V. P. Singh, pp. 809–846, Water Resources Publications, Highlands Ranch, Colo.

Russo, D. (1992), Upscaling of hydraulic conductivity in partially saturated heterogeneous porous formation, *Water Resour. Res.*, *28*(2), 397–409.

Sanchez-Vila, X., A. Guadagnini, and J. Carrera (2006), Representative hydraulic conductivities in saturated groundwater flow, *Rev. Geophys.*, *44*, RG3002, doi:10.1029/2005RG000169.

Seibert, J. (1999), Regionalization of parameters for a conceptual rainfall-runoff model, *Agric. For. Meteorol.*, *98–99*, 279–293.

Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, *16*(2), 430–442.

Sorooshian, S., and V. K. Gupta (1983), Automatic calibration of conceptual rainfall-runoff models: The question of parameter observability and uniqueness, *Water Resour. Res.*, *19*(1), 251–259.

Tonkin, M., and J. Doherty (2005), A hybrid regularised inversion methodology for highly parameterised models, *Water Resour. Res.*, *41*(10), W10412, doi:10.1029/2005WR003995.

U. S. Environmental Protection Agency (EPA) (2000), BASINS Technical Note 6: Estimating Hydrology and Hydraulic Parameters for HSPF, *EPA-823-R00-012*, July.

Viessman, W., and G. L. Lewis (1996), *Introduction to Hydrology*, HarperCollins, New York.

Wagener, T., H. S. Wheater, and H. V. Gupta (2003), Identification and evaluation of watershed models, in *Advances in Calibration of Watershed Models*, edited by Q. Duan, S. Sorooshian, H. V. Gupta, A. Rouseau and R. Turcotte, pp. 29–47, AGU: Monograph.

Wagener, T., and H. S. Wheater (2006), Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, *J. Hydrol.*, *320*(1–2), 132–154.

Warren, J. E., and H. H. Price (1961), Flow in heterogeneous porous media, *Soc. Pet. Eng. J.*, *1*, 153–169.

Wen, X.-H., and J. J. Gomez-Hernandez (1996), Upscaling hydraulic conductivities in heterogeneous media: An overview, *J. Hydrol.*, *183*, 9–32.

Yeh, T. C., L. W. Gelhar, and A. L. Gutjahr (1985), Stochastic analysis of unsaturated flow in heterogeneous soils, 1: Statistically isotropic media, *Water Resour. Res.*, *21*(4), 447–456.

Young, P., S. Parkinson, and M. Lees (1996), Simplicity out of complexity in environmental modelling: Occam's razor revisited, *J. Appl. Stat.*, *23*(2–3), 165–210.

Young, A. R. (2006), flow simulation within UK ungauged catchments using a daily rainfall-runoff model, *J. Hydrol.*, *320*(1–2), 155–172.

Zhu, J., B. P. Mohanty, and N. Das Narenda (2006), On the effective averaging schemes of hydraulic properties at the landscape scale, *Vadose Zone J.*, *5*, 308–316.

J. Doherty, Department of Civil Engineering, University of Queensland, St Lucia, Qld 4072, Australia.

M. R. Gallagher, Natural Resource Sciences, Department of Natural Resources and Water, Indooroopilly, Qld 4068, Australia. (mark.gallagher@nrw.qld.gov.au)