WATER RESOURCES RESEARCH, VOL. 46, W05525, doi:10.1029/2009WR008377, 2010

A short exploration of structural noise

John Doherty[1,2] and David Welter[3]

Received 9 July 2009; revised 27 November 2009; accepted 6 January 2010; published 27 May 2010.

[1] "Structural noise" is a term often used to describe model-to-measurement misfit that cannot be ascribed to measurement noise and therefore must be ascribed to the imperfect nature of a numerical model as a simulator of reality. As such, it is often the dominant contributor to model-to-measurement misfit. As the name "structural noise" implies, this type of misfit is often treated as an additive term to measurement noise when assessing model parameter and predictive uncertainty. This paper inquires into the nature of defect-induced model-to-measurement misfit and provides a conceptual basis for accommodating it. It is shown that inasmuch as defect-induced model-to-measurement misfit can be characterized as "noise," this noise is likely to show a high degree of spatial and temporal correlation; furthermore, its covariance matrix may approach singularity. However, the deleterious impact of structural noise on the model calibration process may be mitigated in a variety of ways. These include adoption of a highly parameterized approach to model construction and calibration (including the strategic use of compensatory parameters where appropriate), processing of observations and their model-generated counterparts in ways that are able to filter out structural noise prior to fitting one to the other, and/or through implementation of a weighting strategy that gives prominence to observations that most resemble predictions required of a model.

Citation: Doherty, J., and D. Welter (2010), A short exploration of structural noise, *Water Resour. Res.*, *46*, W05525, doi:10.1029/2009WR008377.

1. Introduction

[2] This article explores some of the properties of structural noise, particularly as they relate to model calibration and to predictions made by a calibrated model. Structural noise can be loosely defined as model-to-measurement misfit that cannot be explained by measurement noise. From this definition, its strong relationship to the calibration process becomes apparent, for it is through calibration that its presence is often exposed. Structural noise arises from the fact that a model is not a perfect simulator of environmental reality. A model's imperfections are consequences of many facets of its design. These include (but are not limited to) coarseness of spatial and temporal discretizations used in numerical solution schemes, simplifications used in mathematical representation of environmental processes, and simplified representations of hydraulic property heterogeneity in model parameterization schemes.

[3] The nature of structural noise needs to be understood before it can be accommodated (to the extent that this is possible) when calibrating a model and when using a model to make a prediction of future environmental behavior. The following issues are of particular relevance.

[1]Watermark Numerical Computing, Brisbane, Queensland, Australia.
[2]Also at National Centre for Groundwater Research and Training, Flinders University, Adelaide, South Australia, Australia.
[3]South Florida Water Management District, West Palm Beach, Florida, USA.

0043-1397/10/2009WR008377

[4] 1. To what extent are values estimated for model parameters compromised by the presence of structural noise? Can any deleterious effects be mitigated through adoption of certain calibration strategies?

[5] 2. Can the propensity for error incurred by the presence of structural noise in values assigned to model parameters be quantified?

[6] 3. If, through the calibration process, it has been demonstrated that a model is inadequate in some respects, is it necessarily inadequate in all respects?

[7] Although model inadequacies are made apparent to most modelers on a daily basis, there is far from universal agreement on what to do about it. Despite the fact that model-to-measurement misfits are normally far greater, and far less random, than would be expected on the basis of measurement noise alone, the structural component of this misfit is often ignored. Alternatively, it is treated as simply an additive term to measurement noise, both in assigning observation weights prior to calibration and in exploring parameter uncertainty after calibration. Sometimes its existence is acknowledged through use of heuristic calibration methodologies that attempt to tune the calibration process to a model's strengths, while limiting the damage inflicted on parameter estimates by its weaknesses; justification for the adoption of such strategies often relies on intuition and sometimes meets with opposition from critical reviewers who demand greater mathematical rigor.

[8] As reported in the literature, approaches to the issue of structural noise differ widely. Much to the consternation of its Bayesian critics [e.g., *Mantovan and Todini*, 2006], the popular generalized likelihood uncertainty estimation (GLUE) methodology allows use of a subjective likelihood

function to characterize model-to-measurement misfit, and parameter uncertainty incurred as an outcome of this misfit. This is justified by *Beven* [2005] and *Beven et al.* [2008] by noting that the uncertainty associated with estimation of identifiable parameters does not decrease with the addition of more data to the calibration data set (as it should according to Bayes theorem), if those data are "contaminated" by noise of structural origin (or, equivalently, if the data must inform parameters of a defective model). Parameters estimated on the basis of a noise-contaminated data set are, of course, subject to error. Theoretically, the potential for error in estimated parameters can be calculated from the statistical character of noise associated with that data set. This is not the case with structural noise, however, for if the nature of model structural defects is unknown, so too is the nature of model-to-measurement misfit induced by these defects. Subjective decisions must then be made on weighting schemes to apply to the calibration data set when fitting model outputs to that data set in order to estimate parameters. As estimated parameter values depend on these decisions, these estimates, as well as assessments made of the uncertainty associated with these estimates, inherit a subjective component.

[9] An approach that bears some similarity to that taken by Beven and his coworkers is that adopted by (among others) *Gupta et al.* [1998], *Madsen* [2000, 2003], *Vrugt et al.* [2003] and *Wöhling and Vrugt* [2008]. These authors explore the concept of Pareto optimality as model parameters are changed along a Pareto front whereby one objective function component is traded off against another; each component specifies the ability of the model to replicate one particular aspect of a system's historical behavior. A Pareto front defines the locus of points in parameter space for which it is impossible to improve all components of the objective function simultaneously. Extremes of the front define locations in parameter space for which a single objective function component is minimized. It is recommended that many points along the Pareto front be used to parameterize a model when making predictions of future system behavior, thereby allowing at least partial exploration of the uncertainty associated with this prediction. In formulating a calibration strategy that seeks to specifically inform parameters pertaining to distinct system modes, *Wagener et al.* [2003] adopt a related strategy through a procedure which they refer to as "dynamic identifiability analysis."

[10] Other authors have pursued the issue of structural noise from a parameter, rather than a model output, point of view (this, they argue, being closer to the root of the problem). *Kennedy and O'Hagan* [2001] present a formalism wherein a model's propensity for structural error is explicitly included in the relationship between its inputs and outputs. Statistical attributes are associated with this propensity, for which it may be possible to estimate so-called hyperparameters as part of a Bayesian-based parameter inference process. *Vrugt et al.* [2005] include an ensemble of stochastic forcing terms which alter the values of individual model states in their conceptual formulation of model structural inadequacies. They then attempt stochastic characterization of these forcing terms from an analysis of model-to-measurement misfit and use an extended Kalman filter methodology to generate one-step-ahead estimates of streamflow predictive uncertainty based on this characterization. *Kavetski et al.* [2006a, 2006b] and *Kuczera et al.* [2006] introduce the notion of storm-dependent parameters which are estimated as part of the calibration process of hydrologic rainfall-runoff models. Use of these parameters reduces model-to-measurement misfit considerably; at best they supply missing system features that explain that misfit, while at worst they function as effective surrogates for those that do. A not dissimilar approach is followed by *Young* [2002], *Thiemann et al.* [2001], and *Goegebeur and Pauwels* [2007], who use recursive updating of hydrologic model parameters to accommodate the fact that the roles of these parameters change over time in order to compensate for inadequacies in a model's ability to respond to a succession of rainfall events. *Lin and Beck* [2007] demonstrate the use of an innovative recursive prediction error algorithm in inferring time variability of parameter estimates. The exposed nature of this variability then provides a means to reduce or eliminate model defects that give rise to it.

[11] Another approach is to characterize structural noise as being dependent on time-variant, "explanatory" variables such as previous model outputs, past model error, current model inputs, and seasonality. The parameters through which structural error is linked to these explanatory variables can be obtained through regression fitting as part of the calibration process [*Montanari and Grossi*, 2008] and/or through so-called error-correcting, data-driven statistical models such as artificial neural networks, decision trees, instance-based weighting, and support vector machines [*Demissie et al.*, 2009].

[12] Closely related to the concept of defect-induced model-to-measurement misfit is that of ranking different conceptual models on the basis of comparative levels of fit attained through calibrating those models against the same data set. Uncertainty incurred by model structural inadequacy is then partly quantified through making predictions with more than one model when one cannot be declared as definitely superior to another; model ranking is based on statistics such as Akaike information criterion (AIC), second-order bias-corrected AIC (AICc), Bayesian information criterion (BIC), and Kashyap information criterion (KIC) which include a misfit term in their definition. For definitions and examples of use of these statistics, see, for example, *Neuman* [2003], *Ye at al.* [2004], and *Poeter and Anderson* [2005]. *Gaganis and Smith* [2001, 2006, 2008] present related methodologies that seek to address the affects of different parameterization schemes and/or of different model structures on model outcomes, either as a whole or on subsets of these outcomes that pertain to different aspects of the system which they simulate. Predictions are then made with all parameter sets and/or model structures that cannot be eliminated; prediction-specific ranking of alternative parameterizations and/or models may also be undertaken. The fact that model-to-measurement misfit is largely attributable to model structural defects rather than to measurement noise is an explicit part of the analysis, although the relationship between the two is not explored.

[13] *Cooley* [2004], *Cooley and Christensen* [2006], and *Gallagher and Doherty* [2007a] analyze the effects of simplification-induced structural noise on model outputs that are matched to field measurements through the model calibration process, and on model outputs that form the basis for predictions of practical interest. Although their considerations are restricted to parameter simplification rather than model simplification, their work provides some

useful insights into the nature of simplification-induced model-to-measurement misfit; in particular, its spatially and temporally correlated nature are made explicit. These authors show that if these of its characteristics are ignored, the potential for error in estimated model parameters may be inflated, at the same time as the magnitude of this potential may be underestimated.

[14] This section has provided a brief overview of attempts that have been made to accommodate model structural defects when calibrating models, and when using these models to make predictions. If nothing else, it illustrates the range of approaches that have been taken on this issue. While some approaches appear to work well in some modeling contexts, none of them is universally applicable, and all of them involve a certain degree of subjectivity. It is an unavoidable outcome of the nature of the problem that quantification of the effects of model structural defects on model predictions can only be approximate, and that methodologies which attempt to achieve this will always be somewhat heuristic.

[15] The present study attempts to further our understanding of structural noise by providing some theoretical insights into its nature. On the basis of these insights, some methodologies are suggested for accommodating its presence in everyday modeling practice (particularly as it pertains to estimating values for model parameters). Some of these are similar to those that have been provided by previous authors, while others resemble methodologies that have been implemented by modelers for many years with justification based on intuition rather than on theory. It is shown that although the propensity for structure-induced model parameter and predictive error can rarely if ever be quantified, some simple steps can often be taken to mitigate this propensity.

[16] The remainder of this paper is arranged as follows. First, a number of examples of structural noise as it is encountered in everyday modeling practice are presented, together with a synthetic example that illustrates some important features of this phenomenon. Following this, structural noise is conceptualized mathematically, and some of its salient features are explained. On the basis of this conceptualization, a number of methodologies for its accommodation are suggested. Though all of these are, of necessity, approximate in nature, they can be expected to yield beneficial outcomes when applied in at least some practical modeling contexts. The paper finishes with a summary of findings from this work.

2. Examples of Structural Noise

2.1. Examples From Everyday Modeling

[17] The inevitability of defect-induced structural error on surface water model outcomes is described by many authors, including *Moussa and Bocquillon* [1996] who discuss the impact of different mathematical representations of hydraulic processes, and of numerical schemes for solving the resulting equations, on model output error. *Booij* [2003] examines the effect of scale on the integrity of hydraulic process representation in environmental models.

[18] It is widely recognized that a watershed hydrologic model is rarely able to replicate all aspects of a historical streamflow time series equally well. Depending on the manner in which weights are assigned to observations during the calibration process, it may simulate low flows better than high flows, or vice versa. Sometimes its imperfections may be expressed in more subtle ways. For example, a model may repeatedly over-predict or under-predict the magnitude of the first flow peak of a new wet season. From these examples it is apparent that if structural noise as it pertains to watershed models is to be described from a statistical point of view, the existence of temporal correlation must be an integral part of that description. Moreover, this description must recognize a strong statistical linkage between noise as it pertains to flow events of the same type, irrespective of their temporal separation. In doing so it must therefore recognize that the same model inadequacies are expressed in the same way whenever rainfall sequences are similar.

[19] Historical streamflow data sets are often very long, sometimes comprising daily (or even hourly) flows spanning many years. Bayesian analysis suggests that the uncertainty associated with estimates of identifiable model parameters should diminish (eventually to almost zero) as the length of the calibration data set increases. Intuitively this is not the case, for it is self-evident that there is some lower limit below which the uncertainty of a given parameter cannot be reduced, irrespective of the length of the time series on which its estimation is based. This conclusion is supported by experimental evidence forthcoming from studies such as those undertaken by *Perrin et al.* [2007]. This implies that the "noise" associated with the streamflow calibration data set has certain statistical properties that are not normally associated with random variables. In particular, a singular covariance matrix is implied.

[20] When calibrating a groundwater model, rarely if ever are model-to-measurement fits achieved which are close to the accuracy with which heads can be measured in wells. Depending on the size of a groundwater model domain, model-to-measurement discrepancies of several meters may be construed as demonstrating a "good fit." Thus structural error is readily accepted as being the dominant component of model-to-measurement misfit in most groundwater modeling contexts. A common feature of this misfit is that it shows a high degree of spatial and temporal correlation. The former is often evinced by residuals of the same sign over large contiguous portions of the model domain. Temporal correlation is often demonstrated by consistent under- or over-prediction of water levels in a particular well by the model.

[21] While groundwater model-to-measurement fit may sometimes be considered unsatisfactory when judged on the basis of a single indicator such as an objective function, certain outcomes of the calibration process can nevertheless be very pleasing (or can be readily made so). For example, despite the fact that it may be impossible to overlay modeled heads on measured heads for a single well when these are plotted against time, it may be possible to calibrate the model in such a way that the amplitude and timing of observed seasonal water level variations are almost perfectly replicated by the model. Furthermore, it is often easy to find a set of parameters that allows a model to reproduce observed head differences between aquifers separated by a leaky aquitard, even if heads on either side of the aquitard cannot be exactly matched by the model. Similarly, observed lateral head gradients can often be well matched by a model, even if individual heads cannot. In recognition of this, ground-

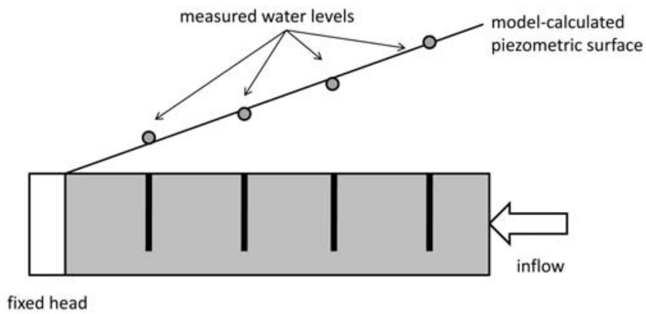

Figure 1a. Four water level measurements in a confined aquifer.

water modelers have often pursued a calibration strategy that endows a model with the ability to replicate well those aspects of historical system behavior that it is capable of replicating, while accepting the fact that other aspects of system behavior cannot be as well replicated. Model-based management is then based on those types of model outputs for which credibility has thus been demonstrated.

2.2. A Synthetic Example

[22] Figure 1a shows a confined aquifer. Water enters the aquifer at its eastern end and exits through a fixed head boundary at its western end. Hydraulic conductivity within the aquifer is uniform and recharge is zero; hence the hydraulic gradient is uniform. Water level measurements in four wells (each contaminated by a small amount of measurement noise) are represented in Figure 1a.

[23] If, in simulating groundwater movement in this aquifer, the modeler assigns an incorrect head value to the western fixed head boundary, the situation depicted in Figure 1b will arise. A least squares fit to the four measured heads would then result in erroneous estimation of aquifer transmissivity. Furthermore, the propensity for error in this estimate would not be reduced with the addition of head measurements in extra wells emplaced between the current wells.

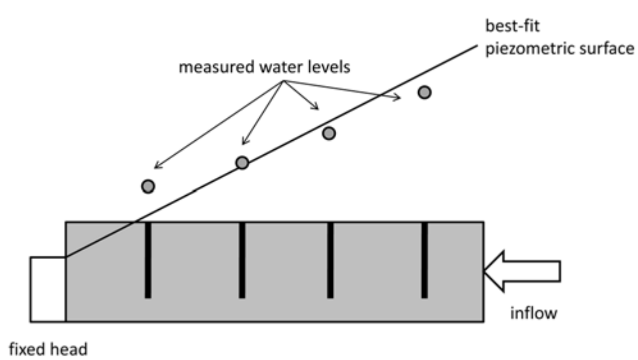

Figure 1b. Simulation of piezometric heads where an incorrect value has been assigned to the western model boundary.

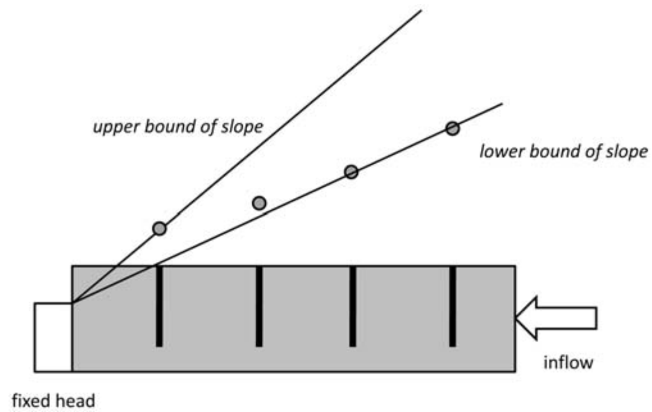

Figure 1c. Possible choices for piezometric surfaces used in obtaining upper and lower transmissivity estimates.

[24] A modeler may notice that the model-to-measurement misfit exhibited in Figure 1b is not suggestive of random and independent measurement noise, either in magnitude or in disposition, and that it is therefore probably of structural origin. In recognition of the potential for error in estimation of transmissivity that the presence of this noise may incur, he/she may attempt a heuristic quantification of this potential using the methodology demonstrated in Figure 1c in which transmissivity is altered to allow variation of the slope of the line between the two depicted extremes. This heuristic strategy may be justified by noting that because the covariance matrix of structural noise is unknown, credibility of fit between the model-generated and measured heads must be judged subjectively (as in the GLUE methodology). Unfortunately, however, the true value of transmissivity does not lie between those calculated on the basis of these two extremes.

[25] The true value of transmissivity is calculated from the line shown in Figure 1d. This does not appear to provide a good fit between model outputs and field measurements as all model-to-measurement residuals are of the same sign. The fact that this fit does, indeed, lead to the correct value of transmissivity (and is therefore the "line of best fit") is further evidence that to the extent that structural

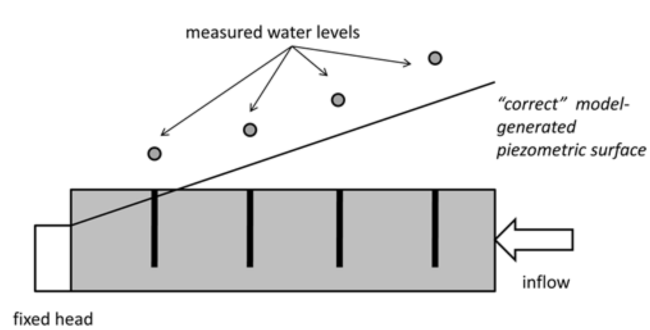

Figure 1d. The "correct" fit between model outputs and field measurements.

noise can be assigned a covariance matrix, it is not a typical covariance matrix.

3. Mathematical Conceptualization of Structural Noise

3.1. Model Simplification

[26] For ease of analysis, system and model linearity are assumed herein; hence both can be represented by matrices. Let the matrix $\mathbf{Z}$ describe a natural system as it acts on system properties encapsulated in a vector $\mathbf{k}$. Natural complexity dictates that $\mathbf{k}$ has (infinitely) many elements. Let $\mathbf{h}$ be a vector of observations comprising a calibration data set, and let ε represent the noise associated with those measurements. Then

$$\mathbf{h} = \mathbf{Zk} + \varepsilon. \tag{1}$$

Let the matrix $\mathbf{X}$, acting on a set of parameters $\mathbf{p}$, represent a model of the system described by $\mathbf{Z}$. As $\mathbf{p}$ has far fewer elements than $\mathbf{k}$, $\mathbf{X}$ has far fewer columns than $\mathbf{Z}$ and hence a lower dimensional range space (which means that it has a reduced capacity for representing system behavioral nuances). Conceptually $\mathbf{X}$ is related to $\mathbf{Z}$ through the equation

$$\mathbf{X} = \mathbf{ZL}, \tag{2}$$

where $\mathbf{L}$ represents a "simplification matrix" which reduces the range space of $\mathbf{Z}$ to that of $\mathbf{X}$. In the linear context in which the current analysis takes place, this represents the means by which the simple model $\mathbf{X}$ is computed from the complex model $\mathbf{Z}$. This simplification process is represented as a reduction in the model range space because a simplified model cannot (by definition) represent as many facets of system behavior as can be represented by a complex model. However, the complex model's parentage of the simple model is represented by equation (2). From (1),

$$\mathbf{h} = \mathbf{Xp} + \boldsymbol{\eta} + \varepsilon, \tag{3}$$

where structural noise $\boldsymbol{\eta}$ is recognized as

$$\boldsymbol{\eta} = \mathbf{Zk} - \mathbf{Xp}. \tag{4}$$

That is, structural noise is the difference (in system output space) between the system $\mathbf{Z}$ acting on real-world system properties $\mathbf{k}$, and the simulator $\mathbf{X}$ acting on a limited set of parameters $\mathbf{p}$ which represent those properties. Suppose that the relationship between system properties $\mathbf{k}$ and parameters used by the model can be expressed by a matrix $\mathbf{N}$ such that

$$\mathbf{p} = \mathbf{Nk}. \tag{5}$$

If a model possesses any degree of physical basis, such a matrix at least notionally exists. In most cases it represents some kind of averaging process. For example, it may describe the process through which soil moisture characteristics throughout a watershed are represented by a single storage element of appropriate size. Or it may describe the process through which a complex distribution of hydraulic conductivity within a mapped geological unit is represented by a zone whose uniform conductivity purports to represent the average hydraulic conductivity in that unit. Equation (4) then becomes

$$\boldsymbol{\eta} = \mathbf{Z}(\mathbf{I} - \mathbf{LN})\mathbf{k}, \tag{6}$$

where $\mathbf{I}$ is the identity matrix. The matrix $\mathbf{LN}$ has some interesting properties. To show this, let us first look at equation (2) more closely. For a given set of model parameters $\mathbf{p}$, let $\mathbf{j}$ represent the set of system properties that these describe. Thus

$$\mathbf{j} = \mathbf{Lp}, \tag{7}$$

where $\mathbf{j}$ has the same number of elements as $\mathbf{k}$. For example, if $\mathbf{X}$ is a groundwater model which uses zones of piecewise constancy to represent spatial variability of hydraulic conductivity, then $\mathbf{j}$ represents a map of piecewise-uniform hydraulic conductivity; the elements of $\mathbf{j}$ represent hydraulic conductivities assigned to individual pixels of the map. Meanwhile $\mathbf{L}$ is a matrix that selects a value for each pixel on the basis of the zone in which that pixel lies. For the matrix $\mathbf{N}$ to fulfill its role as a device for mapping real-world system properties to model parameters (including "real-world" system properties calculated from a set of model parameters $\mathbf{p}$ using $\mathbf{L}$), the following equation must apply:

$$\mathbf{p} = \mathbf{Nj}. \tag{8}$$

From (8) it follows that

$$\mathbf{p} = \mathbf{Nj} = \mathbf{NLp} = \mathbf{NLNj}, \tag{9}$$

from which

$$\mathbf{N} = \mathbf{NLN} \tag{10}$$

so that $\mathbf{L}$ is a generalized inverse of $\mathbf{N}$. Also,

$$\mathbf{LN} = \mathbf{LNLN} \tag{11}$$

so that $\mathbf{LN}$ is idempotent, and hence rank-deficient [*Koch*, 1988]. So too then is $(\mathbf{I} - \mathbf{LN})$ of equation (6). From (6) the covariance matrix of structural noise $\mathrm{C}(\boldsymbol{\eta})$ can be written as

$$\mathrm{C}(\boldsymbol{\eta}) = \mathbf{Z}(\mathbf{I} - \mathbf{LN})\mathrm{C}(\mathbf{k})(\mathbf{I} - \mathbf{LN})^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}, \tag{12}$$

where the $\mathrm{C}(\mathbf{k})$ covariance matrix depicts the innate variability of system properties (including any spatial correlation which they may possess). From (12) the rank of $\mathrm{C}(\boldsymbol{\eta})$ can be no higher than that of $(\mathbf{I} - \mathbf{LN})$, which is k-p where k is the number of elements of $\mathbf{k}$ and p is the number of elements of $\mathbf{p}$. However, this does not necessarily endow $\mathrm{C}(\boldsymbol{\eta})$ with singular status, as the number of elements h of $\mathbf{h}$ will mostly be far fewer than k-p.

[27] What will normally engender $\mathrm{C}(\boldsymbol{\eta})$ with singular or near-singular status is the fact that good model design must seek to endow $\mathbf{Z}(\mathbf{I} - \mathbf{LN})$ with a very low rank. Ideally, $(\mathbf{I} - \mathbf{LN})$ should fall entirely within the null-space of $\mathbf{Z}$ so that structural noise is zero; conceptually, its rank deficiency can be made large enough to guarantee this if the number of elements p of $\mathbf{p}$ used by the model $\mathbf{X}$ is made large enough. The fact that the null-space of $\mathbf{Z}$ is very large, having dimensions of at least k-h (and probably many more), provides ample room for this. However, this is unlikely to be achieved in practice because of restrictions on

the size of p and of the composition of $\mathbf{p}$, for these are set by the design specifications of the model and the simplification strategies that are implicit in these specifications. Thus the range of $(\mathbf{I} - \mathbf{LN})$ is likely to possess a nonzero projection onto the orthogonal complement of the null-space of $\mathbf{Z}$. The dimensionality of this projection can be a maximum of h but will mostly be considerably less than this (as must be the case if the model is to be considered as an acceptable system simulator). To the extent that it is less than h, this endows $C(\boldsymbol{\eta})$ with singular status. (The rank of $C(\boldsymbol{\eta})$ is just 1 for the case illustrated in Figures 1a–1d.)

[28] To explore this concept in another way, let us subject $\mathbf{Z}$ to singular value decomposition such that

$$\mathbf{Z} = \mathbf{USV}^{\mathrm{T}}, \tag{13}$$

where $\mathbf{U}$ is an orthonormal matrix whose columns are unit vectors spanning the range space of $\mathbf{Z}$, $\mathbf{V}$ is an orthonormal matrix whose columns are unit vectors spanning the domain of $\mathbf{Z}$, and $\mathbf{S}$ is a diagonal matrix of singular values arranged from highest to lowest. The rate at which singular values decay to zero as we move down the diagonal of $\mathbf{S}$ depends on the nature of the model (i.e., $\mathbf{Z}$), and on the data comprising the calibration data set $\mathbf{h}$. For a surface water model calibrated on the basis of a single time series, about eight (but possibly fewer) singular values will be significantly nonzero [*Jakeman and Hornberger*, 1993; *Doherty and Hunt*, 2009]. For a groundwater model calibrated against a comprehensive transient data set, there may be a few hundred significantly nonzero singular values, though a few of these are likely to dominate the others [*Gallagher and Doherty*, 2007b]. On most occasions, however, there will be far fewer significantly nonzero singular values than this.

[29] From (13) (and acknowledging the orthonormality of $\mathbf{U}$),

$$\mathbf{U}^{\mathrm{T}}\mathbf{o} = \mathbf{SV}^{\mathrm{T}}\mathbf{k}, \tag{14}$$

where $\mathbf{o}$ is a vector of model-generated equivalents to observations $\mathbf{h}$. Equation (14) states that in calculating system response to parameters, projections of parameters onto columns of $\mathbf{V}$ are multiplied by singular values corresponding to these columns in order to evaluate projections of model outputs onto columns of $\mathbf{U}$; the latter are loosely equivalent to system response modes. Where a singular value is low, the corresponding response mode is of low amplitude. Meanwhile orthogonal unit vectors comprising columns of $\mathbf{V}$ corresponding to zero-valued singular values span the null-space of $\mathbf{Z}$. A primary model design criterion must be to ensure that $(\mathbf{I} - \mathbf{LN})$ projects as little as possible onto columns of $\mathbf{V}$ that correspond to nonzero singular values. To the extent that this fails to occur, the projection is likely to be limited to only a few of these unit vectors (which hopefully correspond to relatively low singular values). A natural outcome of model design is therefore for $C(\boldsymbol{\eta})$ to be dominated by only a few eigencomponents, irrespective of the number of elements h of $\mathbf{h}$ (and therefore of $\boldsymbol{\eta}$), and to thus approach singularity (or to actually be singular). As such, $C(\boldsymbol{\eta})$ expresses variability of fewer random variables than it has dimensions. Hence a complex pattern of model-to-measurement misfit (for example, for a surface water model calibrated over many years of data) may be the expression of only a handful of random variables.

[30] Regardless of the rank of $C(\boldsymbol{\eta})$, one stochastic property that structural noise is most unlikely to possess is that of statistical independence. In other words, $C(\boldsymbol{\eta})$ is most unlikely to be diagonal. Even if $C(\mathbf{k})$ is diagonal and $(\mathbf{I} - \mathbf{LN})$ is symmetric (meaning that it would be an orthogonal projection operator and thus possess eigenvalues of either one or zero), the pattern of diminishing values of diagonal elements of $\mathbf{S}$ would ensure nondiagonality of $C(\boldsymbol{\eta})$. Thus structural noise will show spatial and temporal correlation. Furthermore, the nature of this correlation will be such that its removal through strategies such as autoregressive moving average (ARMA) transformation of flow time series [see, e.g., *Kuczera*, 1983; *Campbell et al.*, 1999; *Campbell and Bates*, 2001] is unlikely to be successful, for temporal correlation of structural noise will not be stationary, but will rather be event-based.

[31] From (3) and (6),

$$\mathbf{h} = \mathbf{Xp} + \mathbf{Z}(\mathbf{I} - \mathbf{LN})\mathbf{k} + \varepsilon, \tag{15}$$

which is now rewritten as

$$\mathbf{h} = \mathbf{X}_1\mathbf{p}_1 + \mathbf{X}_2\mathbf{p}_2 + \varepsilon, \tag{16}$$

where $\mathbf{X}_1$ and $\mathbf{p}_1$ of equation (16) are the same as $\mathbf{X}$ and $\mathbf{p}$ of equation (15) and $\mathbf{X}_2$ and $\mathbf{p}_2$ of equation (16) are chosen such that their product is equal to $\mathbf{Z}(\mathbf{I} - \mathbf{LN})\mathbf{k}$ of equation (15). For reasons already discussed, $\mathbf{p}_2$ need not have many elements. This can be seen by rewriting (15) as

$$\mathbf{h} = \mathbf{Xp} + \mathbf{US}_1\mathbf{V}_1^{\mathrm{T}}(\mathbf{I} - \mathbf{LN})\mathbf{k} + \varepsilon, \tag{17}$$

where $\mathbf{V}_1$ is a matrix containing columns of $\mathbf{V}$ that are associated with nonzero singular values. $\mathbf{X}_2$ and $\mathbf{p}_2$ of equation (16) can then be defined as

$$\mathbf{X}_2 = \mathbf{US}_1 \tag{18a}$$

$$\mathbf{p}_2 = \mathbf{V}_1^{\mathrm{T}}(\mathbf{I} - \mathbf{LN})\mathbf{k}, \tag{18b}$$

from which it is apparent that $\mathbf{p}_2$ has at most as many elements as there are nonzero singular values associated with $\mathbf{Z}$ (maximum of h but mostly less than this). In practice, if model simplification is optimal, it will have fewer elements than this, as parameters comprising $\mathbf{p}_1$ will not be repeated in $\mathbf{p}_2$.

3.2. Some Repercussions

[32] In light of the above analysis, the process of model simplification can be conceptualized as being roughly equivalent to the process of omitting parameters (or combinations of parameters) from a model. These are the $\mathbf{p}_2$ parameters of equation (16). Reality is then represented by the model using the equation

$$\mathbf{h} = \mathbf{X}_1\mathbf{p}_1 + \boldsymbol{\tau}, \tag{19}$$

where $\boldsymbol{\tau}$ is combined measurement noise ε and structural noise $\mathbf{X}_2\mathbf{p}_2$. Meanwhile the omitted $\mathbf{p}_2$ parameters are assigned fixed values. It is in the nature of model simplification that some of these values will be wrong. For example, if multiple hydrostratigraphic layers comprising reality are replaced by a single layer in a groundwater model, the properties assigned to missing layers are notionally the same as those assigned to the single layer represented in the model. If boundary conditions are assigned incorrect values (as in the example of Figures 1a–1d), the missing $\mathbf{p}_2$ parameters are corrections to these misassigned values. In both cases, correct values for the misassigned, fixed (and therefore missing) $\mathbf{p}_2$ parameters are unknown. The propensity for error when misassigning these values for the sake of simplifying model construction can, however, be calculated from the covariance matrix of reality C($\mathbf{k}$) using the relationship

$$\mathrm{C}(\mathbf{p}_2) = \mathbf{V}_1^{\mathrm{T}}(\mathbf{I} - \mathbf{LN})\mathrm{C}(\mathbf{k})(\mathbf{I} - \mathbf{LN})^{\mathrm{T}}\mathbf{V}_1, \tag{20}$$

from which the covariance matrix of structural error can be calculated as

$$\mathrm{C}(\boldsymbol{\eta}) = \mathbf{X}_2\mathrm{C}(\mathbf{p}_2)\mathbf{X}_2^{\mathrm{T}} = \mathbf{X}_2\mathbf{V}_1^{\mathrm{T}}(\mathbf{I} - \mathbf{LN})\mathrm{C}(\mathbf{k})(\mathbf{I} - \mathbf{LN})^{\mathrm{T}}\mathbf{V}_1\mathbf{X}_2^{\mathrm{T}}. \tag{21}$$

A fruitful way of looking at structural noise as it appears in the model calibration process is thus as "information with nowhere to go." The high level of temporal and spatial covariance that is often visible in minimized model-to-measurement misfit should alert the modeler that such information is present, for misfit of this nature is not normally ascribable to measurement noise. The information that this misfit expresses belongs to parameters that are not being given the opportunity for estimation through a calibration process based on the model as presently designed.

[33] From the singular or near-singular nature of the covariance matrix of measurement noise, it follows that defect-induced model-to-measurement misfit may express variability of only a few random variables. Furthermore, as equation (21) indicates, these variables are better conceived of as parameters than as noise associated with field measurements. Unfortunately, structural noise is often thought of in the latter terms. Problems then arise in inferring the propensity for error in parameters that are estimated through model calibration (i.e., the $\mathbf{p}_1$ parameters of equations (16) and (19)) and in predictions which depend on those parameters. This arises from the unknown nature of the covariance matrix of structural noise, and the need to therefore apply heuristic methods to its (necessary) inclusion in uncertainty analysis. A partially effective heuristic method is exemplified in Figure 1c and in the use of subjective likelihood functions in GLUE analysis.

[34] The above analysis suggests that one way in which structural noise can be reduced is to include as many parameters as possible in the model, and in the parameter estimation process through which the model is calibrated. In the surface water modeling context, the use of storm-dependent parameters with/without Kalman filter upgrading is an example of this. In the groundwater modeling context this philosophy can be implemented through the use of dense spatial parameterization, and inclusion of uncertain boundary conditions and system stresses in the parameter estimation process. Estimation of many parameters leads, of course, to an ill-posed inverse problem. Methods based on mathematical regularization must be used to solve this problem. If properly formulated, use of these methods can lead to the minimum error variance solution of that problem [see, e.g., *Moore and Doherty*, 2005, 2006; *Hunt et al.*, 2007; *Menke*, 1984; *Aster et al.*, 2005]. Furthermore, because aspects of the modeled system that are not constrained by model calibration can be described probabilistically (for example, by a covariance matrix such as C($\mathbf{k}$) discussed above) as a result of expert knowledge and site characterization studies, predictive uncertainty analysis can be undertaken with far more integrity than would be the case if reliance were placed on a C($\boldsymbol{\eta}$) matrix whose properties are almost completely unknown. For a description of linear and nonlinear methods of parameter and predictive uncertainty analysis that are available for use in highly parameterized contexts, see *Gallagher and Doherty* [2007b], *Tonkin et al.* [2007], *Tonkin and Doherty* [2009], and *Doherty* [2010].

[35] Even in complex, physically based models endowed with many parameters, structural noise can never be eliminated, for some of the notional or actual erroneously fixed parameters encapsulated in $\mathbf{p}_2$ of equation (16) include facets of model construction such as grid design, spatial and temporal discretization, assignment of elevations to geological layering, representation of processes that affect surface water quality, etc., that constitute simplifications of reality that are necessary elements of the simulation process. These aspects of structural noise can be responsible for locally high levels of model-to-measurement misfit, misfit that must be accommodated in other ways, including some that are discussed in section 4. However, before leaving the present section it will be demonstrated that even if a model is a perfect simulator of reality, structural noise cannot be eradicated as a component of model-to-measurement misfit.

3.3. Lower Limit of Structural Noise

[36] Suppose that we have a model that is a perfect simulator or reality, i.e., the model $\mathbf{Z}$ of equation (1). Suppose also that variability of parameters $\mathbf{k}$ is described by the covariance matrix

$$\mathrm{C}(\mathbf{k}) = \sigma_{\mathrm{k}}^2\mathbf{I}, \tag{22}$$

where $\mathbf{I}$, once again, is the identity matrix. (Normally native model parameters would need to undergo transformation before (22) would apply; however, we will skip this step for the sake of brevity.) For $\mathbf{k}$ satisfying (22) a minimum error variance solution to the inverse problem of model calibration is approximately achieved through truncated singular value decomposition such that estimated parameters $\widehat{\mathbf{k}}$ are calculated from observations $\mathbf{h}$ using the formula

$$\widehat{\mathbf{k}} = \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{h}, \tag{23}$$

where the subscript 1 indicates pretruncation singular values and corresponding orthonormal columns of the $\mathbf{V}$ matrix of equation (13). Substitution of (1) and (13) into (23), and recalling that $\mathbf{U}$ is an orthonormal matrix, leads to

$$\widehat{\mathbf{k}} = \mathbf{V}_1\mathbf{V}_1^{\mathrm{T}}\mathbf{k} + \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}^{\mathrm{T}}\varepsilon, \tag{24}$$

from which parameter error is calculated as

$$\widehat{\mathbf{k}} - \mathbf{k} = -\mathbf{V}_2\mathbf{V}_2^{\mathrm{T}}\mathbf{k} + \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}^{\mathrm{T}}\varepsilon, \tag{25}$$

where the subscript 2 refers to posttruncation columns of $\mathbf{V}$, and the relationship

$$\mathbf{V}_1\mathbf{V}_1^{\mathrm{T}} + \mathbf{V}_2\mathbf{V}_2^{\mathrm{T}} = \mathbf{I} \tag{26}$$

has been used. If measurement noise is homoscedastic and independent so that

$$\mathrm{C}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2\mathbf{I}, \tag{27}$$

then from (22) and (25) the covariance matrix of parameter error is calculated as

$$\mathrm{C}\left(\widehat{\mathbf{k}} - \mathbf{k}\right) = \sigma_{\mathrm{k}}^2\mathbf{V}_2\mathbf{V}_2^{\mathrm{T}} + \sigma_\varepsilon^2\mathbf{V}_1\mathbf{S}_1^{-2}\mathbf{V}_1^{\mathrm{T}}. \tag{28}$$

Moore and Doherty [2005] show that parameter error variance falls and then rises as the truncation point shifts to singular values of higher index (and hence lower value), this being an outcome of the $\mathbf{S}_1^{-2}$ component of the second term of equation (28). *Doherty and Hunt* [2009] and *Doherty* [2010] suggest that a suitable truncation point can be chosen as that at which the error variance associated with a parameter eigencomponent (i.e., column of $\mathbf{V}$) rises rather than falls as the truncation point is advanced. This occurs where error variance lost through exclusion of an eigencomponent from the $\mathbf{V}_2$ subspace is approximately equal to that gained through including it in the $\mathbf{V}_1$ subspace. From (28), the first is equal to σ_{k}^2 while the second is equal to $\sigma_\varepsilon^2/s_i^2$ where i is the truncation point index and s_i is the singular value at that index. Thus truncation should occur where

$$s_i^2 = \sigma_\varepsilon^2/\sigma_{\mathrm{k}}^2. \tag{29}$$

The structural noise associated with this near-optimal method of model simplification (i.e., parameter simplification achieved through singular value truncation) can be identified through expansion of equation (1) in the following manner:

$$\mathbf{h} = \mathbf{Z}\mathbf{k} + \boldsymbol{\varepsilon} = \mathbf{Z}\mathbf{V}_1\mathbf{V}_1^{\mathrm{T}}\mathbf{k} + \mathbf{Z}\mathbf{V}_2\mathbf{V}_2^{\mathrm{T}}\mathbf{k} + \boldsymbol{\varepsilon}. \tag{30}$$

From (24), the model calibration process seeks to estimate $\mathbf{V}_1\mathbf{V}_1^{\mathrm{T}}\mathbf{k}$ (for this is $\widehat{\mathbf{k}}$ in the absence of measurement noise). This leads to the immediate conclusion that structural noise is given by

$$\boldsymbol{\eta} = \mathbf{Z}\mathbf{V}_2\mathbf{V}_2^{\mathrm{T}}\mathbf{k}, \tag{31a}$$

which from (13) can be rewritten as

$$\boldsymbol{\eta} = \mathbf{U}\mathbf{S}_2\mathbf{V}_2^{\mathrm{T}}\mathbf{k}, \tag{31b}$$

from which the covariance matrix of structural noise can be calculated as

$$\mathrm{C}(\boldsymbol{\eta}) = \sigma_{\mathrm{k}}^2\mathbf{U}\mathbf{S}_2^2\mathbf{U}^{\mathrm{T}}. \tag{32}$$

This is normally a singular matrix because in most cases all but the leading singular values of $\mathbf{S}_2$ will be zero. The variance of $\mathbf{u}_1^{\mathrm{T}}\boldsymbol{\eta}$, this being the projection of $\boldsymbol{\eta}$ onto the first observation eigencomponent (i.e., the first column of $\mathbf{U}$), is easily calculated as

$$\sigma_{\mathbf{ut1}\boldsymbol{\eta}}^2 = \sigma_{\mathrm{k}}^2 s_i^2, \tag{33}$$

where the index i, once again, refers to the index at truncation (which we assume to be the first singular value featured in $\mathbf{S}_2$). From (29),

$$\sigma_{\mathbf{ut1}\boldsymbol{\eta}}^2 = \sigma_\varepsilon^2. \tag{34}$$

From this it can be concluded that even where model simplification (which we have loosely equated to parameter simplification) is optimal, the level of structural noise cannot be reduced below the level of measurement noise. Where simplification is less than optimal (as it always is), the level of structural noise can be presumed to be higher than this. Some repercussions of this are as follows:

[37] 1. Given that the level of structural noise will always be at least as high as that of measurement noise (and probably much higher), model-to-measurement residuals cannot be expected to lack spatial and temporal correlation. Advice that is sometimes provided to modelers suggesting that spatial/temporal correlation of residuals is indicative of less-than-effective model calibration must therefore be questioned.

[38] 2. If computation of postcalibration statistics such as AIC, AICc, BIC, and KIC ignores the unavoidable presence of structural noise of unknown covariance matrix that accompanies the use of any model (even a perfect model), then recommendations made on the basis of these statistics that favor one model over another, or one parameterization scheme over another, should be taken as suggestive rather than definitive.

4. Accommodation of Structural Noise

4.1. General

[39] Theory presented in section 3 is now developed further with the aim of suggesting three means by which structural noise can be accommodated in the model calibration process, and when using a model to make predictions. It is freely acknowledged that no method of accommodating structural noise can be exact. However, on the basis of understandings gained through the previous analysis, it is possible to suggest means through which the potential damage incurred by ignoring its presence or, failing to understand its nature, can be ameliorated to some extent in some modeling contexts.

4.2. Filtering Out Structural Noise

[40] Let the operation of a natural system be described by the equation

$$\mathbf{h} = \mathbf{X}\mathbf{p} + \boldsymbol{\varepsilon}. \tag{35}$$

This equation is identical to equation (1) except for the fact that $\mathbf{Z}$ in equation (1) has been replaced by $\mathbf{X}$ to designate well-posedness of the inverse problem of estimating $\mathbf{p}$. This assumption is adopted for convenience; the following analysis is readily extended to under-determined systems. Let the covariance matrix of measurement noise be designated as $\mathrm{C}(\boldsymbol{\varepsilon})$. Under these conditions, the minimum error variance solution $\widehat{\mathbf{p}}$ for estimation of parameters $\mathbf{p}$ is obtained through

$$\widehat{\mathbf{p}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{h}, \tag{36}$$

where $\mathbf{Q}$ is chosen as

$$\mathbf{Q} = \sigma_r^2 C^{-1}(\varepsilon) \tag{37}$$

and σ_r^2 is a constant of proportionality [see, e.g., *Koch*, 1988]. The covariance matrix of parameter error associated with this estimate is then readily calculated as

$$C(\widehat{\mathbf{p}} - \mathbf{p}) = \sigma_r^2 \left(\mathbf{X}^T\mathbf{Q}\mathbf{X}\right)^{-1}. \tag{38}$$

Equation (35) becomes equivalent to equation (16) if the vector $\mathbf{p}$ is partitioned into two subvectors $\mathbf{p}_1$ and $\mathbf{p}_2$. As above, a minimum error variance estimate $\widehat{\mathbf{p}}_1$ of $\mathbf{p}_1$ can be obtained through conjunctive estimation of $\mathbf{p}_1$ and $\mathbf{p}_2$ as $\mathbf{p}$ through equation (36); $\mathbf{X}$ in that equation is simply replaced by the equivalent matrix $[\mathbf{X}_1\ \mathbf{X}_2]$. The error variance of estimation of $\mathbf{p}_1$ by $\widehat{\mathbf{p}}_1$ is available through appropriate partitioning of $C(\widehat{\mathbf{p}} - \mathbf{p})$ obtained through (38). After some laborious matrix algebra (principally focused on deriving expressions for the partitions of the inverse of the $\mathbf{X}^T\mathbf{Q}\mathbf{X}$ matrix), the following formulas are obtained for $\widehat{\mathbf{p}}_1$ and $C(\widehat{\mathbf{p}}_1 - \mathbf{p})$. For illustrative purposes, measurement noise is assumed to be homoscedastic and independent, so that the weight matrix $\mathbf{Q}$ is replaced by the identity matrix $\mathbf{I}$,

$$\widehat{\mathbf{p}}_1 = \left(\mathbf{X}_1^T\mathbf{W}^T\mathbf{W}\mathbf{X}_1\right)^{-1}\mathbf{X}_1^T\mathbf{W}^T\mathbf{W}\mathbf{h} \tag{39}$$

$$C(\widehat{\mathbf{p}}_1 - \mathbf{p}) = \sigma_\varepsilon^2\left(\mathbf{X}_1^T\mathbf{W}^T\mathbf{W}\mathbf{X}_1\right)^{-1}, \tag{40}$$

where

$$\mathbf{W}^T\mathbf{W} = \mathbf{I} - \mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-1}\mathbf{X}_2^T. \tag{41}$$

Equations (39) to (41) allow computation of $\widehat{\mathbf{p}}_1$ and $C(\widehat{\mathbf{p}}_1 - \mathbf{p})$ from $\mathbf{h}$ without explicit citation of values estimated for, or assigned to, $\mathbf{p}_2$. Thus estimates of $\mathbf{p}_1$ are granted immunity from actual or notional misassignment of values to $\mathbf{p}_2$ during the model simplification process. Meanwhile, the error variance of estimation of $\mathbf{p}_1$ by $\widehat{\mathbf{p}}_1$ is as low as can be theoretically attained. Therefore, in spite of its deficiencies, the simplified model of equation (16) where only $\mathbf{p}_1$ is available for estimation is as capable of making predictions that are sensitive only to the $\mathbf{p}_1$ parameter set as is the presimplified model of equation (35). While this may be comforting, unfortunately the above parameter estimation procedure cannot be implemented in real-world model calibration contexts because $\mathbf{X}_2$ is unknown. Nevertheless, if an approximate $\mathbf{W}$ can be found for use in the above equations, some progress can be made.

[41] To explore one mechanism through which an approximate $\mathbf{W}$ can be found, we note that $\mathbf{W}$ satisfies the following equations:

$$\mathbf{W}\mathbf{X}_2 = \mathbf{0} \tag{42a}$$

$$\mathbf{W}\mathbf{W}^T = \mathbf{I}_n, \tag{42b}$$

where n is the number of rows in $\mathbf{W}$. Given that $\mathbf{X}_2\mathbf{p}_2$ of equation (16) represents structural noise, equation (42a) states that $\mathbf{W}$ effectively filters out structural noise. If both sides of equation (16) are multiplied by $\mathbf{W}$, we therefore obtain

$$\mathbf{W}\mathbf{h} = \mathbf{W}\mathbf{X}_1\mathbf{p}_1 + \mathbf{W}\varepsilon. \tag{43}$$

This is of the same form as (35); estimation of $\mathbf{p}_1$ can then take place using equation (36) with $\mathbf{W}\mathbf{X}_1$ substituted for $\mathbf{X}$ in that equation, and $\mathbf{Q}$ set equal to $\mathbf{I}$ because of the fact that if measurement noise is homoscedastic and independent, then so too is $\mathbf{W}$-filtered measurement noise. Explicitly, $\mathbf{W}$-filtered measurement noise $\boldsymbol{\omega}$ is given by

$$\boldsymbol{\omega} = \mathbf{W}\varepsilon, \tag{44a}$$

the covariance matrix of which is

$$C(\boldsymbol{\omega}) = \mathbf{W}C(\varepsilon)\mathbf{W}^T. \tag{44b}$$

Because of (42b), $C(\boldsymbol{\omega})$ is proportional to $\mathbf{I}$ if $C(\varepsilon)$ is proportional to $\mathbf{I}$.

[42] $\mathbf{W}^T\mathbf{W}$ calculated according to (41) is unique; in this case n, the number of rows of $\mathbf{W}$ and the number of observations in the reformulated inverse problem expressed by equation (43), is equal to the number of elements of $\mathbf{h}$ minus the number of elements of $\mathbf{p}_2$. In contrast, many different $\mathbf{W}$ which satisfy both of equations (42) can be found with a smaller number of rows than this. These can all filter out structural noise; however, none can be guaranteed to provide a minimum error variance estimate of $\mathbf{p}_1$ as does $\mathbf{W}$ computed through (41).

[43] If the presence of structural noise is ignored, estimates $\widehat{\mathbf{p}}_1$ of $\mathbf{p}_1$ would be obtained using the following equation (recalling that $C(\varepsilon)$ is assumed to be $\sigma_\varepsilon^2\mathbf{I}$):

$$\widehat{\mathbf{p}}_1 = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{h}. \tag{45}$$

Through substitution of (16) into (45) and propagating covariance we obtain for the covariance matrix of error of $\widehat{\mathbf{p}}_1$ estimated through (45):

$$C(\widehat{\mathbf{p}}_1 - \mathbf{p}) = \left(\mathbf{X}_1^T\mathbf{X}_1\right)^{-1}\mathbf{X}_1^T\mathbf{X}_2C(\mathbf{p}_2)\mathbf{X}_2^T\mathbf{X}_1\left(\mathbf{X}_1^T\mathbf{X}_1\right)^{-1} + \sigma_\varepsilon^2\left(\mathbf{X}_1^T\mathbf{X}_1\right)^{-1}. \tag{46}$$

A comparison with (40) reveals that if $C(\mathbf{p}_2)$ is small, the potential for error in parameters $\mathbf{p}_1$ estimated through (45) can actually be smaller than if estimated through (39). However, a small $C(\mathbf{p}_2)$ infers near-perfect model simplification, or simplification with no compromises, which is not in fact simplification at all. In that case, "simplification" has actually been a process of hardwiring correct values for $\mathbf{p}_2$ into the model. This in turn reduces error variance of estimation of $\mathbf{p}_1$ (unless $\mathbf{X}_2$ and $\mathbf{X}_1$ happen to be orthogonal, in which case (40) and (46) become equivalent). In more general circumstances, the first term of (46) increases as the extent of compromises involved in the model construction/simplification process increases. Its actual magnitude will always be indeterminate, as the difference between reality and a model can never be known. For this reason it is advisable to avoid this source of error; filtering out of structural noise presents a possible avoidance mechanism.

[44] In real-world modeling, implementation of a strategy for filtering out structural noise can only ever be heuristic. In many circumstances a suitable strategy is to formulate a

multicomponent objective function and to ensure that each component is made approximately equally visible in the overall objective function; thus no component either dominates the objective function, or is dominated by other components, this ensuring that the information content of all components is able to inform parameters estimated through the calibration process. Different components may be formed simply through appropriate partitioning of the calibration data set into different groups, some of which may be considered to be relatively uncontaminated by structural noise. Alternatively, or as well, all or part of the observation data set **h** may be processed in a number of different ways (see below) and matched to model outputs which are processed in the same way, each such processing mechanism leading to a different objective function component. Ideally, these different objective function components should be designed in such a way as to be informative of different aspects of the system under study. That is, they should be informative of different subgroups of parameters, or of different (ideally orthogonal) combinations of parameters. Furthermore, processing should be such as to "orthogonalize out" structural noise, thereby emulating the action of the matrix **W** in the above equations. The efficacy (or otherwise) of structural noise removal can then be judged through the absence (or otherwise) of a high degree of spatial/temporal correlation in model-to-measurement misfit as it pertains to observations and processed observations comprising each objective function component.

[45] Designation of relative weighting between objective function components cannot help but be subjective. The extent of measurement-derived noise associated with each component as made apparent through calibration-minimized misfit may serve as a guide. However, in appropriate circumstances, another guide may be the pragmatic consideration that no component of the overall objective function should either dominate the total objective function, or be dominated by other components (at least at the commencement of the calibration process). In this way, information on all aspects of the system under study is "heard" through the calibration process. Structural noise that may be associated with some objective function components does not therefore allow these components to drown out subtle but valuable information that may be present in other objective function components. More complex and flexible weighting strategies may also be implemented, including exploration of the Pareto front as described above.

[46] Practical examples of formulation of objective function components in accordance with this strategy include the following.

[47] 1. Strategic use of vertical, lateral and/or temporal head differences in groundwater model calibration. Collected head differences of each type should be assigned to different observation groups and weighted to ensure visibility of each with respect to other observation groups. In this way, a model's inability to replicate absolute heads need not compromise its ability to replicate their gradients, thus allowing better estimation of vertical conductance, horizontal transmissivity and/or storage coefficient than would otherwise be the case. (Note that use of lateral head gradient would allow correct estimation of transmissivity for the case depicted in Figure 1b.)

[48] 2. Assignment of observations of different types to different observation groups with weighting designed to ensure approximately equal visibility of each group in the overall objective function.

[49] 3. Processing of an observed streamflow time series and its model-generated counterpart in order to separately match flows, monthly volumes, base flow approximations, and/or other salient (orthogonalized) flow components [see, e.g., *Doherty and Johnston*, 2003; *Gutiérrez-Magness and McCuen*, 2005; *Fenicia et al.*, 2007; *Marcé et al.*, 2008].

[50] 4. Deweighting of spatially and/or temporally clustered observations which would otherwise dominate an objective function through sheer weight of numbers.

[51] The similarity between this strategy and signature based model evaluation as described by *Gupta et al.* [2008] is noteworthy. Through formulation of a multicomponent objective function based on processing of measurements in different ways, these authors state that they are pursuing a more rigorous basis for model evaluation than that provided simply by matching model outputs to field data through use of a single fit criterion. Though based on different arguments from those presented herein, the approach described by these authors has as its outcome a similar calibration methodology to that described in the present paper. See also *Gupta et al.* [2009].

4.3. Prediction-Specific Weights Assignment

[52] Equations (39) and (40) are equivalent to (36) and (38) if **Q** in the latter equations is written as

$$\mathbf{Q} = \mathbf{W}^{\mathrm{T}}\mathbf{W} = \left(\mathbf{I} - \mathbf{X}_2\left(\mathbf{X}_2^{\mathrm{T}}\mathbf{X}_2\right)^{-1}\mathbf{X}_2^{\mathrm{T}}\right). \qquad (47)$$

Q of equation (47) defines a projection operator onto the orthogonal complement of the range space of $\mathbf{X}_2$. The range space of $\mathbf{X}_2$ is, of course, the space in which structural noise exists. This raises an important point. Where the system under study is as simple as the model, and hence equation (35) with $\mathbf{X}_1$ substituted for **X** constitutes a complete description of that system, a minimum error variance estimate $\widehat{\mathbf{p}}_1$ of $\mathbf{p}_1$ is obtained when the weight matrix **Q** satisfies (37). However, where the system is more complex than the model (as it always is), and where model-to-measurement misfit is dominated by structural noise (as it mostly is), observations, or observation combinations, should be given higher weight to the extent that they are protected from contamination by structural noise (because they are orthogonal to its range space). If a model is well constructed for its intended purpose, the range space of structural noise will be approximately orthogonal to that of model outputs, the latter being specified by the range space of $\mathbf{X}_1$.

[53] Implementation of this type of weighting strategy is, once again, difficult in real-world modeling contexts as $\mathbf{X}_2$ is unknown. Where the existence of structural noise is obvious, then observations that apparently express it should be provided with low weights; sometimes these observations (or groups of observations) are made apparent through the calibration process.

[54] Other situations may require implementation of a proactive rather than reactive weighting strategy. Consider models used in flood prediction. Most hydrologic models of this type are lumped and simplified, as they must simulate surface water movement over large watersheds; hence structural noise is to be expected. In spite of the fact that

simplifications and approximations are an inherent aspect of the structure of such models, it is to be hoped that these models have been designed in such a way that their structural defects will not detract from their ability to simulate flow peaks. Thus a modeler is justified in believing that flood peaks are combinations of observations that lie within the range space of the simplified model $\mathbf{X}_1$ and are hopefully thereby approximately orthogonal to the range space of $\mathbf{X}_2$.

[55] When calibrating a hydrologic streamflow model, it is often difficult to fit both low flows and high flows simultaneously. Flow transformations such as Box-Cox [*Box and Cox*, 1964; *Box and Jenkins*, 1976], including logarithmic [*Lumb et al.*, 1994] are often suggested as a means of distributing misfit evenly over the response modes of the system under the assumption that measurement noise transformed in this way is approximately homoscedastic. However, the above analysis (and work of *Gallagher and Doherty* [2006]) suggests that this is not necessarily the best way to optimize the performance of a model whose purpose is to predict peak flows. A better means of ensuring its performance in this regard may be to start with a traditional transformation strategy such as those cited above, but to then vary weights as necessary so that if flow peaks and troughs cannot both be simultaneously matched, the calibration process is provided with some incentive for fitting the former in preference to the latter.

[56] This philosophy raises the specter of prediction-specific model calibration. This, in the authors' opinion, may indeed have validity in some circumstances, particularly in environments where structural noise is potentially high, and models are necessarily simple. Hence a hydrologic model calibrated for optimal performance in matching flow peaks may then be recalibrated to optimize its performance in simulating low flows if it were then to be used for predicting some aspect of aquatic health related to low-flow regimes.

4.4. Compensatory Parameters

[57] The use of explanatory parameters as a means of accommodating structural noise was briefly mentioned in section 1. The idea of "soaking up" the information with nowhere to go that is structural noise has an appeal. Better fits can be obtained between model outputs and field measurements than would be possible without their use, thus enhancing estimation of nonexplanatory model parameters. In the meantime, the fact that certain model parameters may be assuming undefined roles so that they can compensate for model defects and so provide a repository for wayward information emanating from those defects, may not be a bad thing. This is especially so if such parameters can be readily identified as assuming this compensatory role, and if predictions required of the model are either independent of compensatory parameters, or depend on them in the same way as do model outputs employed in its calibration (see below).

[58] Consider the example of Figures 1a–1d. A modeler may consider it inappropriate to adjust the height assigned to the western fixed head boundary. Instead, he/she may insert a thin zone between this boundary and the first measurement well, declaring the transmissivity of this zone to be an adjustable parameter. It can be shown that if this is done, not only will a correct estimate for transmissivity in the remainder of the model domain be obtained, but also that this estimate is of minimum error variance (the same variance that would be attained if the head associated with the fixed boundary were estimated instead of the transmissivity in the artificial zone juxtaposing that boundary).

[59] The adjustment of transmissivity in a thin zone adjacent to the fixed head boundary as compensation for failure to adjust the height of the fixed head boundary itself may or may not compromise the model's ability to make a prediction, for this depends on the prediction. For example, the ability of the model depicted in Figures 1a–1d to predict heads up the hill from the most easterly well under the same inflow regime is not impaired through use of this compensatory parameter. The model will therefore be far superior in this regard than if it were configured as in Figure 1b. On the other hand, the model's ability to predict heads in observation wells following an increase in uphill inflow will be impaired if this parameterization strategy is used (but probably not as much as the model configured as in Figure 1b).

[60] Similar situations are often encountered when using devices such as pilot points for spatial parameterization of groundwater model domains. Sometimes local "bulls eyes" appear in parameter fields that emerge from pilot-point-based calibration. If proper precautions in model construction and pilot point emplacement have been taken (including the use of as many pilot points as possible together with appropriate mathematical regularization), the estimation of locally unreasonable parameter values may not necessarily be a bad thing. To the extent that they soak up information present in structural noise that would otherwise have no place to go, and that they then allow better estimates to be made of other parameters through allowing information pertinent to those parameters unfettered access to the calibration process, their compensatory role may serve the parameter estimation process well. If possible, of course, any model imperfections that they evince should be rectified. However, if this is not possible, the model calibrated in this way may be rendered capable of making many model predictions better than would have otherwise been the case.

[61] The argument for judicious use of compensatory parameters can be formalized. Once again, equation (16) is assumed to represent the modeling context in which parameter estimation takes place, with $\mathbf{p}_1$ parameters being visible (and adjustable) within the model, and $\mathbf{p}_2$ parameters being notionally or actually fixed through the model construction process. Suppose, however, that parameter estimation is now based on the following equation:

$$\mathbf{h} = \mathbf{X}_1\mathbf{p}_1 + \mathbf{X}_c\mathbf{p}_c + \varepsilon, \tag{48}$$

where the $\mathbf{p}_c$ vector holds compensatory parameters and the $\mathbf{X}_c$ matrix represents sensitivities of model outputs used in the calibration process to these compensatory parameters. For compensatory parameters to be capable of absorbing information that is intended for the missing $\mathbf{p}_2$ parameter set, the following condition must hold:

$$\mathbf{X}_c\widehat{\mathbf{p}}_c \approx \mathbf{X}_2\mathbf{p}_2, \tag{49}$$

where $\widehat{\mathbf{p}}_c$ holds estimated values for $\mathbf{p}_c$ and $\mathbf{p}_2$ holds the "correct" values for the fixed $\mathbf{p}_2$ parameters. Let s be a prediction of interest, and let its sensitivities to $\mathbf{p}_1$, $\mathbf{p}_2$ and $\mathbf{p}_c$

be described by the vectors $\mathbf{y}_1$, $\mathbf{y}_2$ and $\mathbf{y}_c$. The value of the prediction is given by the equation

$$s = \mathbf{y}_1^T \mathbf{p}_1 + \mathbf{y}_2^T \mathbf{p}_2, \tag{50}$$

whereas the prediction made by the model is calculated using the equation

$$\widehat{s} = \mathbf{y}_1^T \widehat{\mathbf{p}}_1 + \mathbf{y}_c^T \widehat{\mathbf{p}}_c. \tag{51}$$

Assuming that $\widehat{\mathbf{p}}_1$ is a good approximation to $\mathbf{p}_1$ (the approximation hopefully being enhanced by the use of compensatory parameters), $\widehat{s}$ will approach s if

$$\mathbf{y}_2^T \mathbf{p}_2 = \mathbf{y}_c^T \widehat{\mathbf{p}}_c. \tag{52}$$

From (49), this will occur if

$$\mathbf{y}_2^T \approx \mathbf{c}^T \mathbf{X}_2. \tag{53a}$$

$$\mathbf{y}_c^T \approx \mathbf{c}^T \mathbf{X}_c, \tag{53b}$$

where $\mathbf{c}$ is an arbitrary vector. What this means is that the prediction should either depend only on parameters $\mathbf{p}_1$, or that compensatory parameters perform a similar role in relation to the prediction as they do in relation to aspects of system state that are employed in the calibration process. This is a condition that is often met in practice.

5. Conclusions

[62] Accommodation of model imperfections is an essential component of model usage. It is also the most problematical aspect of model usage and one for which no universally applicable quantitative set of procedures will ever be available. The existence, and some characteristics, of model defects are often identified through the calibration process; indeed part of the role of the calibration process should be to accomplish exactly this. Sometimes the exposure of these defects allows a modeler to rectify them. On other occasions a modeler's task must be to mitigate the deleterious effects that these defects have on extracting information from an environmental data set, and in using that information as a basis for scientifically based environmental management.

[63] This paper has sought to explore some of the ways in which model imperfections express themselves and some of the means by which model usage can progress in a meaningful and useful manner despite the less-than-perfect capacity of models to simulate real-world systems. The analysis was based on an assumption of model linearity. While most models are nonlinear, the powerful tools and concepts that are available through linear analysis have been used to demonstrate some important outcomes of the imperfect nature of models as they express themselves in everyday model usage. A number of suggestions have emerged from this study of ways in which model imperfections may be better accommodated in various real-world modeling contexts. However, the only conclusion that is universally applicable in this regard is that ignoring their existence, or treating their manifestations as an additive term to measurement noise under the assumption that the statistical properties of these manifestations are similar to those of measurement noise, will almost certainly constitute a suboptimal means of accommodating them.

[64] In light of this, perhaps the term "structural noise" is best avoided, for it is suggestive of properties that model-to-measurement misfit induced by model inadequacies simply does not possess. Nevertheless, the term has become such an accepted part of the modeling vernacular that it would be pointless to suggest that its use be banned. However, the above study (in common with some previous studies cited above) suggests that a better understanding of the nature of this phenomenon, and a better formulation of means through which its potential for inflicting damage on the model calibration and predictive processes can be mitigated, is to be gained by focusing on model parameters rather than on model outputs.

[65] Strategies suggested herein for accommodation of model structural defects are necessarily heuristic in nature. Some of them contradict advice that has been given elsewhere, advice that is based on an assumption that structural noise can either be ignored or treated as a form of measurement noise. For example, suggestions made herein of a flexible approach to observation weighting as it is applied to separate objective function components, and of tuning a weighting strategy to the type of prediction required of a model, may be seen by some as defective. Nevertheless, both this strategy and others suggested herein have often been adopted by modelers in the past in recognition of the fact that models are imperfect simulators of reality, and that the rules that should be applied in textbook cases to models that simulate systems that are as simple as models themselves simply do not apply in the real world. Unfortunately, replacements for these rules have not been available; nor are they available now. However, it is hoped that the present study provides some mathematical justification for pursuing a suite of alternative approaches whose primary justification until now has simply been that of common sense.

[66] It is hoped that this paper provides a framework for further research in the important area of accommodation of model structural inadequacies. An obvious means by which the ideas and theory presented herein can be given practical application is through the use of synthetic models based on real-world modeling problems to investigate the nature of structurally induced model-to-measurement misfit as it occurs in common modeling contexts. Through this means, strategies that are maximally effective in those contexts can be developed.

[67] This paper concludes with a summary of findings drawn from this study, presented in point form.

[68] 1. The term "structural noise" is applied herein to describe model-to-measurement misfit associated with a calibrated model beyond that which can be attributed to measurement noise; it can be conceptualized as "information with nowhere to go." It arises from the imperfect nature of a model as a simulator of reality, and constitutes an announcement of at least some of its imperfections. The natural repositories for the information contained within structural noise are parameters which have explicitly or implicitly been assigned incorrect values in the necessarily simplified representation of reality that is a model.

[69] 2. To the extent that structural noise can be ascribed a stochastic characterization, its covariance matrix is likely to possess significant off-diagonal terms which will bestow

on it a high degree of spatial and/or temporal correlation. Furthermore, this covariance matrix will often approach singularity. This erodes the credibility of parameter and predictive uncertainty analysis based on rigorous statistical approaches that attempt to relate parameter confidence to levels of model-to-measurement misfit. Such approaches are approximate at best.

[70] 3. Formulation of a multicomponent objective function, with each component containing observations, or processed observations, that are high in information pertaining to one particular aspect of a simulated system, can grant that information visibility in a composite objective function that may otherwise be dominated by structurally induced misfit. Through ensuring that each such component is granted sufficient weight to make it visible in the overall objective function, the risk of denying access to the model of important information that is resident within the calibration data set is diminished, at the same time that structural noise is at least partially filtered out.

[71] 4. If a model's construction is such that it cannot fit all aspects of the calibration data set equally well, then its ability to make a prediction of a certain type may be strengthened if higher weights are provided to members of the calibration data set that bear greatest similarity to that prediction.

[72] 5. Structural noise is reduced where a model is endowed with a large number of estimable parameters. Large numbers of parameters are readily accommodated in calibration procedures that involve the use of mathematical regularization. As *Moore and Doherty* [2005] show, concomitant parameter and predictive uncertainty analysis can then utilize the statistics of innate parameter variability (which are often an outcome of site characterization studies) rather than relying on the unknown statistics of structural noise. Ideally, model parameters should be physically based. However, if predictions required of a model are similar in nature to model outputs matched through the calibration process, the presence of compensatory parameters may not erode model predictive credibility.

[73] **Acknowledgments.** The authors wish to thank the South Florida Water Management District for funding and supporting the work documented in this paper.

References

Aster, R. C., B. Borchers, and C. H. Thurber (2005), *Parameter Estimation and Inverse Problems*, 301 pp., Elsevier, Amsterdam.

Beven, K. (2005), On the concept of model structural error, *Water Sci. Technol.*, *52*(6), 167–175.

Beven, K. J., P. J. Smith, and J. E. Freer (2008), So why would a modeller choose to be incoherent?, *J. Hydrol.*, *354*, 15–32, doi:10.1016/j.jhydrol.2008.02.007.

Booij, M. J. (2003), Determination and integration of appropriate spatial scales for river basin modeling, *Hydrol. Processes*, *17*, 2581–2598, doi:10.1002/hyp.1268.

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

Box, G. E. P., and G. M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, Calif.

Campbell, E. P., and B. C. Bates (2001), Regionalization of rainfall-runoff model parameters using Markov chain Monte Carlo samples, *Water Resour. Res.*, *37*(3), 731–739, doi:10.1029/2000WR900349.

Campbell, E. P., D. R. Fox, and B. C. Bates (1999), A Bayesian approach to parameter estimation and pooling in nonlinear flood event models, *Water Resour. Res.*, *35*(1), 211–220, doi:10.1029/1998WR900043.

Cooley, R. L. (2004), A theory for modeling ground-water flow in heterogeneous media, *U.S. Geol. Surv. Prof. Pap.*, *1679*, 220 pp.

Cooley, R. L., and S. Christensen (2006), Bias and uncertainty in regression-calibrated models of groundwater flow in heterogeneous media, *Adv. Water Resour.*, *29*, 639–656, doi:10.1016/j.advwatres.2005.07.012.

Demissie, Y. K., A. J. Valocchi, B. S. Minsker, and B. A. Bailey (2009), Integrating a calibrated groundwater flow model with error-correcting data-driven models to improve predictions, *J. Hydrol.*, *364*, 257–271, doi:10.1016/j.jhydrol.2008.11.007.

Doherty, J. (2010), PEST: Model independent parameter estimation: User documentation, Watermark Numer. Comput., Brisbane, Queensland, Australia. (Available at http://www.pesthomepage.org)

Doherty, J., and R. J. Hunt (2009), Two statistics for evaluating parameter identifiability and error reduction, *J. Hydrol.*, *366*, 119–127, doi:10.1016/j.jhydrol.2008.12.018.

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, doi:10.1111/j.1752-1688.2003.tb04381.x.

Fenicia, F., H. H. G. Savenije, P. Matgen, and L. Pfister (2007), A comparison of alternative multiobjective calibration strategies for hydrologic modeling, *Water Resour. Res.*, *43*, W03434, doi:10.1029/2006WR005098.

Gaganis, P., and L. Smith (2001), A Bayesian approach to the quantification of the effect of model error on the predictions of groundwater models, *Water Resour. Res.*, *37*(9), 2309–2322, doi:10.1029/2000WR000001.

Gaganis, P., and L. Smith (2006), Evaluation of the uncertainty of groundwater model predictions associated with conceptual errors: A per-datum approach to model calibration, *Adv. Water Resour.*, *29*, 503–514, doi:10.1016/j.advwatres.2005.06.006.

Gaganis, P., and L. Smith (2008), Accounting for model error in risk assessments: Alternatives to adopting a bias towards conservative risk estimates in decision models, *Adv. Water Resour.*, *31*, 1074–1086, doi:10.1016/j.advwatres.2008.04.010.

Gallagher, M. R., and J. Doherty (2006), Parameter estimation and uncertainty analysis for a watershed model, *Environ. Model. Software*, *22*, 1000–1020, doi:10.1016/j.envsoft.2006.06.007.

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

Gallagher, M. R., and J. Doherty (2007b), Predictive error analysis for a water resource management model, *J. Hydrol.*, *334*, 513–533, doi:10.1016/j.jhydrol.2006.10.037.

Goegebeur, M., and V. R. N. Pauwels (2007), Improvement of the PEST parameter estimation algorithm through extended Kalman filtering, *J. Hydrol.*, *337*, 436–451, doi:10.1016/j.jhydrol.2007.02.006.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple noncommensurable measures of information, *Water Resour. Res.*, *34*, 751–763, doi:10.1029/97WR03495.

Gupta, H. V., T. Wagener, and Y. Liu (2008), Reconciling theory with observations: Towards a diagnostic approach to model evaluation, *Hydrol. Processes*, *22*(18), 3802–3813, doi:10.1002/hyp.6989.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. G. Martinez-Bacquero (2009), Decomposition of the mean squared error & NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, *377*, 80–91, doi:10.1016/j.jhydrol.2009.08.003.

Gutiérrez-Magness, A., and R. H. McCuen (2005), Effect of flow proportions on HSPF model calibration accuracy, *J. Hydrol. Eng.*, *10*(5), 343–352, doi:10.1061/(ASCE)1084-0699(2005)10:5(343).

Hunt, R. J., J. Doherty, and M. J. Tonkin (2007), Are models too simple? Arguments for increased parameterisation, *Ground Water*, *45*(3), 254–262, doi:10.1111/j.1745-6584.2007.00316.x.

Jakeman, A. J., and G. M. Hornberger (1993), How much complexity is warranted in a rainfall-runoff model?, *Water Resour. Res.*, *29*(8), 2637–2649, doi:10.1029/93WR00877.

Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resour. Res.*, *42*, W03407, doi:10.1029/2005WR004368.

Kavetski, D., G. Kuczera, and S. W. Franks (2006b), Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resour. Res.*, *42*, W03408, doi:10.1029/2005WR004376.

Kennedy, M. C., and A. O'Hagan (2001), Bayesian calibration of computer models, *J. R. Stat. Soc., Ser. B*, *63*, 425–464, doi:10.1111/1467-9868.00294.

Koch, K. R. (1988), *Parameter Estimation and Hypothesis Testing in Linear Models*, xvi + 378 pp., Springer, Berlin.

Kuczera, G. (1983), Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty, *Water Resour. Res.*, *19*(5), 1151–1172, doi:10.1029/WR019i005p01151.

Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *J. Hydrol.*, *331*, 161–177, doi:10.1016/j.jhydrol.2006.05.010.

Lin, Z., and M. B. Beck (2007), On the identification of model structure in hydrological and environmental systems, *Water Resour. Res.*, *43*, W02402, doi:10.1029/2005WR004796.

Lumb, A. M., R. B. McCammon, and J. L. Kittle Jr. (1994), Users manual for an expert system (HSPexp) for calibration of the hydrologic simulation program—Fortran, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, *94-4168*, 102 pp.

Madsen, H. (2000), Automatic calibration of a conceptual rainfall-runoff model using multiple objectives, *J. Hydrol.*, *235*, 276–288, doi:10.1016/S0022-1694(00)00279-1.

Madsen, H. (2003), Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, *Adv. Water Resour.*, *26*, 205–216, doi:10.1016/S0309-1708(02)00092-1.

Mantovan, P., and E. Todini (2006), Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *J. Hydrol.*, *330*, 368–381, doi:10.1016/j.jhydrol.2006.04.046.

Marcé, R., C. E. Ruiz, and J. Armengol (2008), Using spatially distributed parameters and multiresponse objective functions to solve parameterization of complex applications of semidistributed hydrologic models, *Water Resour. Res.*, *44*, W02436, doi:10.1029/2006WR005785.

Menke, W. (1984), *Geophysical Data Analysis: Discrete Inverse Theory*, Academic, San Diego, Calif.

Montanari, A., and G. Grossi (2008), Estimating the uncertainty of hydrological forecasts: A statistical approach, *Water Resour. Res.*, *44*, W00B08, doi:10.1029/2008WR006897.

Moore, C., and J. Doherty (2005), The role of the calibration process in reducing model predictive error, *Water Resour. Res.*, *41*, 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, doi:10.1016/j.advwatres.2005.07.003.

Moussa, R., and C. Bocquillon (1996), Criteria for the choice of flood-routing methods in natural channels, *J. Hydrol.*, *186*, 1–30, doi:10.1016/S0022-1694(96)03045-4.

Neuman, S. P. (2003), Maximum likelihood Bayesian averaging of alternative conceptual-mathematical models, *Stochastic Environ. Res. Risk Assess.*, *17*(5), 291–305, doi:10.1007/s00477-003-0151-7.

Perrin, C., L. Oudin, V. Andreassian, C. Rojas-Serna, C. Michel, and T. Mathevet (2007), Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models, *Hydrol. Sci. J.*, *52*(1), 131–151, doi:10.1623/hysj.52.1.131.

Poeter, E., and D. Anderson (2005), Multimodel ranking and inference in ground water modeling, *Ground Water*, *43*, 597–605, doi:10.1111/j.1745-6584.2005.0061.x.

Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, *37*(10), 2521–2535, doi:10.1029/2000WR900405.

Tonkin, M., and J. Doherty (2009), Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques, *Water Resour. Res.*, *45*, W00B10, doi:10.1029/2007WR006678.

Tonkin, M., J. Doherty, and C. Moore (2007), Efficient nonlinear predictive error variance for highly parameterized models, *Water Resour. Res.*, *43*, W07429, doi:10.1029/2006WR005348.

Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian (2003), Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, *39*(8), 1214, doi:10.1029/2002WR001746.

Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, *41*, W01017, doi:10.1029/2004WR003059.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modeling: Dynamic identifiability analysis, *Hydrol. Processes*, *17*, 455–476, doi:10.1002/hyp.1135.

Wöhling, T., and J. A. Vrugt (2008), Combining multiobjective optimization and Bayesian model averaging to calibrate forecast ensembles of soil hydraulic models, *Water Resour. Res.*, *44*, W12432, doi:10.1029/2008WR007154.

Ye, M., S. P. Neuman, and P. D. Meyer (2004), Maximum likelihood Bayesian averaging of spatial variability in unsaturated fractured tuff, *Water Resour. Res.*, *40*, W05113, doi:10.1029/2003WR002557.

Young, P. C. (2002), Advances in real-time flood forecasting, *Philos. Trans. R. Soc. London, Ser. A*, *360*, 1433–1450, doi:10.1098/rsta.2002.1008.

J. Doherty, Watermark Numerical Computing, 336 Cliveden Ave., Brisbane, Qld 4075, Australia. (johndoherty@ozemail.com.au)

D. Welter, South Florida Water Management District, 3301 Gun Club Rd., West Palm Beach, FL 33406, USA.