Use of paired simple and complex models to reduce predictive bias and quantify uncertainty

John Doherty[1,2] and Steen Christensen[3]

[1]   Modern environmental management and decision-making is based on the use of increasingly complex numerical models. Such models have the advantage of allowing representation of complex processes and heterogeneous system property distributions inasmuch as these are understood at any particular study site. The latter are often represented stochastically, this reflecting knowledge of the character of system heterogeneity at the same time as it reflects a lack of knowledge of its spatial details. Unfortunately, however, complex models are often difficult to calibrate because of their long run times and sometimes questionable numerical stability. Analysis of predictive uncertainty is also a difficult undertaking when using models such as these. Such analysis must reflect a lack of knowledge of spatial hydraulic property details. At the same time, it must be subject to constraints on the spatial variability of these details born of the necessity for model outputs to replicate observations of historical system behavior. In contrast, the rapid run times and general numerical reliability of simple models often promulgates good calibration and ready implementation of sophisticated methods of calibration-constrained uncertainty analysis. Unfortunately, however, many system and process details on which uncertainty may depend are, by design, omitted from simple models. This can lead to underestimation of the uncertainty associated with many predictions of management interest. The present paper proposes a methodology that attempts to overcome the problems associated with complex models on the one hand and simple models on the other hand, while allowing access to the benefits each of them offers. It provides a theoretical analysis of the simplification process from a subspace point of view, this yielding insights into the costs of model simplification, and into how some of these costs may be reduced. It then describes a methodology for paired model usage through which predictive bias of a simplified model can be detected and corrected, and postcalibration predictive uncertainty can be quantified. The methodology is demonstrated using a synthetic example based on groundwater modeling environments commonly encountered in northern Europe and North America.

1.   Introduction

[2] Management of the natural environment is often informed by computer simulation of environmental processes. Policy, management, and regulatory decisions rely heavily on model outcomes. Increasingly these models are complex and built at great expense.

[3] Since modeling began, the lure of complexity has had its proponents and its sceptics. In a much cited article, *Lee* [1973] lists the "seven sins of large-scale models" when used in the urban planning context. His views are echoed more recently in the groundwater context by *Clement* [2011] and *Doherty* [2011b]. In another well-known article, *Logan* [1994] defends "big ugly models." *Van Nes and Scheffer* [2005] present arguments for both approaches while finding favor in the use of "medium sized complex models."

[4] It is salient to reflect on what environmental modeling can actually achieve when used in the decision-making context. Unfortunately, "predictive accuracy" is rarely one of them (in spite of the fact that it is often the stated reason for model complexity). *Moore and Doherty* [2005, 2006] point out that the uncertainties associated with many predictions required of (especially groundwater) models is necessarily high. This is an outcome of the so-called "null space" contribution to the uncertainty of these predictions. This contribution is born of an information deficit in most calibration data sets with respect to the hydraulic and other properties of the large, complex, and heterogeneous systems, which large, complex models attempt to simulate. This view is supported by authors such as *Harrar et al.* [2003], *Højberg and Refsgaard* [2005], *Rojas et al.* [2010], and *Troldborg et al.* [2007] who point out that the potential for model predictive error tends to rise with the extent to

[1]National Centre for Groundwater Research and Training, Flinders University, Adelaide, Australia.
[2]Watermark Numerical Computing, Corinda, Australia.
[3]Department of Earth Sciences, University of Aarhus, Denmark.

which the nature of predictions made by a model differs from those comprising the data set against which it is calibrated.

[5] What a complex model can aspire to, however, is, (1) quantification of the uncertainty associated with predictions of management interest through inclusion in the model of all facets of system behavior and of hydraulic property heterogeneity on which these predictions may depend, and (2) reduction of that uncertainty to its theoretical minimum through processing of all available information contained in expert knowledge and in measurements of system state comprising the calibration data set of the model.

[6] Theoretically, posterior (i.e., "postcalibration") predictive uncertainty can be quantified through Bayes equation. When applied to parameters employed by a model, this can be written in its simplest form as,

$$P(\mathbf{k}|\mathbf{h}) \propto P(\mathbf{h}|\mathbf{k}) \, P(\mathbf{k}), \tag{1}$$

where $\mathbf{k}$ is the vector of model parameters; $\mathbf{h}$ is the vector of observations comprising the calibration data set; $P(\mathbf{k})$ is the prior probability density function of parameters, this expressing expert knowledge; $P(\mathbf{h}|\mathbf{k})$ is the likelihood function, this increasing with the level of model-to-measurement fit attained through the calibration process; and $P(\mathbf{k}|\mathbf{h})$ is the posterior parameter probability density function.

[7] The second factor on the right of Bayes equation (i.e., the prior parameter probability density function) depicts what is attractive about a complex model, namely, its ability to encapsulate expert knowledge (which is intrinsically stochastic in nature). This arises from the fact that a complex, physically based model is made of parts whose details can be related to the details of reality. Experts can recognize these details; furthermore, the values of parameters attributed to these details can be informed by field or laboratory measurements of system properties. However, the first term on the right of Bayes equation (i.e., the likelihood term) provides problems for complex models. The adjustment of hundreds (or perhaps thousands) of parameters in order to fit historical observations of system state is a very difficult undertaking because of the long run times, and the risk of numerical instability which often attends the use of such models.

[8] The simple model's greatest asset is its ability to calculate and maximize the likelihood term. Simple models are fast, and they are generally numerically stable. Provided they are endowed with enough parameters, they can fit a calibration data set well; hence, information can flow from that data set to the model. A problem, however, is that the receptacles into which that information flows (i.e., the parameters which this information serves to estimate) may have reduced physical meaning. Hence, the simple model may not be an ideal receptacle for expert knowledge. While expert knowledge may indeed assist in informing some of the parameters of a simple model, the linkage between what can be seen or measured in the field and the values that should be assigned to simple model parameters are not as clear as for a complex model. As will be discussed herein, this may not matter. What is important, however, is that a simple model may not include parameterization and process detail on which important predictions of future system behavior may depend. In contrast, it may be well calibrated, and hence include system detail on which historical system behavior depends. This lack of representation of

prediction-sensitive null space parameter components results in under-estimation of predictive uncertainty. At the same time, as we shall see, predictions made by such models may be associated with a considerable amount of bias.

[9] There is thus a tension between complexity and simplicity, a tension that must be resolved on a case-by-case basis through pursuit of a simplification strategy that is optimal for each case. A number of authors have addressed the issue of optimal model simplification. Strategies range from appropriate abstraction achieved through parameter and process lumping [see, for example, *Pachepsky et al.*, 2005; *Beven*, 2002; and references cited therein], model reduction using orthogonal decomposition [see, for example, *Saide et al.*, 2010; *Lieberman et al.*, 2010; *Willcox and Peraire*, 2002; *Moore*, 1981], and data-driven model emulation [see, for example, *Sivakumar*, 2008; *Young and Ratto*, 2009, 2011; *Conti et al.*, 2009; and references cited therein].

[10] Where complex two- and three-dimensional spatial and temporal processes prevailing in a heterogeneous environment must be simulated, where predictive uncertainty is likely to be high in spite of the availability of a large calibration data set, and where expert knowledge must thus play a significant role in reduction (to the extent that this is possible) and quantification of predictive uncertainty, the tension between the two terms on the right of Bayes equation is strong. This is the context in which most modeling of subsurface fluid movement takes place, for example, in the simulation of regional groundwater systems or of the behavior of petroleum reservoirs. A common response to this tension is to simplify a complex model to the extent that its run times are tractable, and numerical granularity of model outputs is reduced to levels that allow its use with specialist inversion software. Simplification strategies may include an assumption of steady state and/or confined conditions, abstract representation of recharge processes and groundwater/surface water interaction, and/or use of a coarse grid in place of a fine grid. *Zyvoloski and Vesselinov* [2006] point out, however, that the benefits gained through an ability to thereby calibrate such models may be eroded by the need for the calibration process to endow some parameters with erroneous values. These values compensate for a model-to-measurement misfit that would otherwise be induced by model structural defects introduced though the simplification process. If a prediction depends on these parameters in a different way from that in which members of the calibration data set depend on them, predictive bias will be the inevitable result. The magnitude of this bias will be difficult to quantify, so that its inclusion in the predictive uncertainty analysis process becomes difficult or impossible.

[11] To address this problem, the use of paired simple and complex models has been proposed as another means of reconciling the conflicting demands of the two terms on the right of Bayes' equation, and to thereby gain maximum benefit from each of them. Parameterization of the complex model is informed by expert knowledge, typically through the generation of stochastic parameter fields for this model, these encapsulating what is known, and what is unknown, about system properties. For each such realization, the paired simple model receives a complimentary parameterization. This can be achieved either through the implementation of various upscaling techniques (discussed, for example, by *Farmer* [2002]) and/or through calibrating the simple

model against outputs of the complex model. Examples of this approach include the works of *Aanonsen and Eydinov* [2006], *Aanonsen* [2008], and *Scheidt et al.* [2011]. Theoretical underpinnings (based on Bayes' equation), and other examples of this approach, are provided by *Kennedy and O'Hagan* [2000], *Omre and Lødøen* [2004], *Lødøen and Omre* [2008], and *Lødøen and Tjelmeland* [2010]. All of these authors use examples from petroleum reservoir management, where problems associated with history matching are exacerbated by the extremely long run times associated with the simulation of oil extraction.

[12] A paired model approach was also employed by *Cooley* [2004] and *Cooley and Christensen* [2006]. However they restricted their attention to simplification of model parameter fields rather than simplification of model structure. Nevertheless, their methodology for development of a covariance matrix of simplification-induced structural noise, and for use of this matrix in calibration of a model which employs a simplified parameter scheme, is of relevance to the present study.

[13] The present paper adopts a paired model approach. However, in contrast to previous authors, we analyze the model simplification process from a subspace rather than a Bayesian point of view. This allows us to explore the costs associated with model simplification, as well as the nature of the relationship between predictions made by a simple model and those made by a complex model. This, in turn, allows us to suggest a methodology for achieving optimal model usage by employing both of them together.

[14] Our paper is organized as follows: In section 2 we examine model simplification from a theoretical point of view, deriving expressions for model predictive error incurred through the simplification process. Drawing on this theory, we devote section 3 to the explanation of a methodology for conjunctive use of a paired complex/simple model in making unbiased predictions of minimum error variance, and in quantifying the uncertainty associated with those predictions. In sections 4 and 5 we demonstrate the paired model methodology using a synthetic case that is based on a commonly encountered real-world groundwater management problem. Section 6 discusses and draws conclusions from this work.

2. Concepts and Theory

2.1. General

[15] As mentioned in section 1, in contrast to previous investigators, we take a subspace approach to explaining and quantifying the effects of model simplification. Hence, in many places throughout the text we employ the terms "predictive error" and "predictive error variance" rather than "predictive uncertainty," though often we use the terms interchangeably, as is common practice. We view "predictive uncertainty" as an intrinsic property of a prediction that is calculable from the data that is available to us, and from the expert knowledge, which we possess, via Bayes' equation. Furthermore, as discussed above, we have identified the reduction and quantification of predictive uncertainty as the proper goals of the modeling process when implemented as a basis for environmental management.

[16] Unfortunately, because it is a simplification of reality, use of a model to predict the future, and to condition

such predictions on expert knowledge and historical measurements of system state, necessarily involves error. From a Bayesian perspective, use of a model can therefore be viewed as introducing another source of uncertainty to predictions of future system behavior. However, in harmony with the subspace approach that we have taken, we refer to such a model-induced propensity for predictive wrongness as "error."

[17] The concept of "error" is quite in harmony with the notion of model calibration. Indeed, calibration can be viewed as a form of simplification, for it provides a means of securing parameter uniqueness where it would not otherwise exist. The assignment of unique values to parameters in this fashion is not a Bayesian concept. Nevertheless, for reasons which will not be discussed herein, it forms the centerpiece of everyday environmental modeling practice.

[18] The goal of the model/parameter simplification/calibration process must be to attain the ability to make unbiased predictions of minimum error variance. Naturally, that variance must be quantified. If simplification is without cost, the error variance of a prediction is equal to its innate uncertainty. Reduction of predictive error variance to this level then provides the metric for integrity of the simplification process.

[19] Because the mathematical discussion that constitutes section 2 of our paper is formulated in terms of subspace rather than Bayesian concepts, we focus on quantifying predictive error variance, and on reducing it to its theoretical lower limit. The words "error" and "error variance" are employed in their strict mathematical sense here. However, in other parts of the text (including in the title of our paper) we have not been so strict in discriminating between "error" and "uncertainty."

[20] We therefore adopt the following premises:

[21] (1) When the simplification process is such that the propensity for error in a particular model prediction is reduced to its theoretical minimum, the remaining propensity for error is approximately equal to the uncertainty of that prediction.

[22] (2) Quantification of minimized predictive error variance therefore constitutes quantification of predictive uncertainty.

[23] Accordingly, we employ the more general term "uncertainty" in place of "propensity for error" or "error variance" where appropriate in the remainder of our text in order to better convey the repercussions of our work to our readers.

2.2. Null Space

[24] Central to the subspace concepts on which our analysis of simplification is based is the concept of the null space. Before embarking on an explanation of model simplification we take a few lines to explain this concept. For more details we refer the reader to any good book on linear algebra. The salience of the null space to model calibration is well explained in texts such as *Menke* [1984], *Koch* [1987], and *Aster et al.* [2005].

[25] A matrix $\mathbf{Z}$ has a null space if a nonzero vector $\mathbf{k}_n$ exists for which the following relationship holds:

$$\mathbf{0} = \mathbf{Z}\mathbf{k}_n. \qquad (2)$$

[26] Suppose that the vector $\mathbf{Z}$ represents the action of a model on its parameters $\mathbf{k}$ under calibration conditions. Let

the vector **h** represent measurements comprising the calibration data set, and let the vector **ε** characterize (supposedly random) noise associated with these measurements. Then:

$$\mathbf{h} = \mathbf{Zk} + \boldsymbol{\varepsilon}. \tag{3}$$

[27] Suppose for the moment that measurement noise is zero. Suppose also that we have found a parameter set **k** that allows the model to reproduce the calibration data set **h**. Then:

$$\mathbf{h} = \mathbf{Z}\underline{\mathbf{k}}. \tag{4}$$

[28] If (2) is added to (4) it is immediately apparent that

$$\mathbf{h} = \mathbf{Z}(\underline{\mathbf{k}} + \mathbf{k}_n). \tag{5}$$

[29] It follows that the existence of a null space promulgates nonuniqueness of the inverse problem of model calibration. It is easily shown that any matrix that has more columns than rows has a null space. Many matrices that have more rows than columns also have a null space. In practical terms, any calibration process that attempts to estimate many parameters is likely to be nonunique, because the matrix that represents a linearized form of the model is likely to possess a null space. Any complex model which attempts to represent complexity at a scale approaching that of the real world will certainly possess a null space.

[30] Ideally, the calibration process should seek a parameter set **k** that has no null space components. The purposeful exclusion of null space components from the calibrated parameter set has two beneficial effects. First, it provides a means through which uniqueness of the inverse problem of model calibration can be achieved, for only one parameter set **k** exists for which equation (4) is satisfied and null space parameter components are excluded. Second, with appropriate parameter scaling, this represents a minimum error variance solution to the inverse problem of model calibration. This is because, through purposeful exclusion of inestimable parameter combinations (i.e., parameters **k**$_n$ of equation (2)) from the calibrated parameter field, the potential for parameter incorrectness is minimized. This is not to say that this potential is small; minimization is achieved through making this potential symmetrical about the calibration parameter field **k** by staying out of the null space altogether, and thereby refusing to enter the null space in one (possibly wrong) direction or the other.

[31] Estimation of a parameter field **k**, which excludes null space components can be accomplished through singular value decomposition (SVD). The situation is depicted in Figure 1. The vector **k** represents the true, but unknown, parameter set of the model. SVD finds the projection of this set onto the so-called estimable subspace, or solution space, of parameter space defined through the inverse problem of model calibration. The solution space is the orthogonal complement of the null space.

[32] By definition, null space parameter sets (i.e., parameter sets for which equation (2) applies) cannot be inferred through the calibration process. Often (but not always), they represent hydraulic property detail that is beyond the reach of the calibration process. If, for the purposes of linear analysis, parameters are defined as perturbations
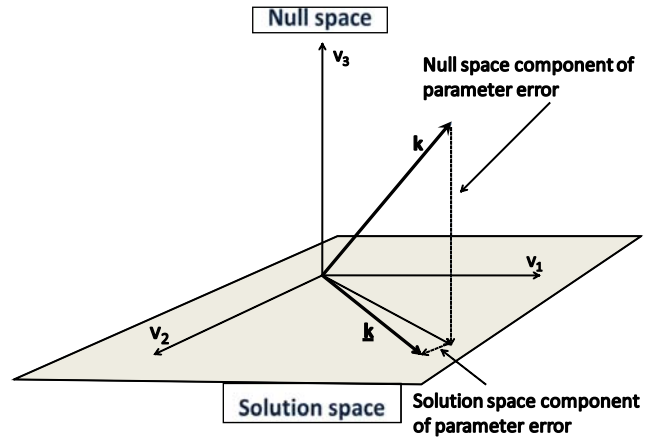

Figure 1. Solution and null parameter spaces. The vectors **v**$_1$, **v**$_2$, and **v**$_3$ are unit orthogonal vectors composed of the columns of the **V** matrix of equation (9) achieved through singular value decomposition of a model matrix **Z**.

from their precalibration preferred values where they are informed solely by expert knowledge, their precalibration minimum error variance values are obviously zero. Perturbations to these values required for model calibration should lie entirely within the solution space. By definition, a nonzero null space projection of any parameter set which purports to calibrate a model is not supported by the calibration data set. An unsupported potential for wrongness (which will be denoted as "bias" herein) is thus built into any prediction which is sensitive to that projection. For a more complete discussion of the use of SVD in model calibration see *Moore and Doherty* [2005, 2006]. See also, the above-cited texts.

2.3. Conceptualizing Model Simplification

[33] In the discussion that follows we represent model simplification as the omission from a model of certain processes, and of parameters which govern the operation of those processes. For simplicity, a model is represented as a linear operator. That is, the model is represented as a matrix.

[34] All models are simplifications of reality. Hence, the concepts discussed herein refer to complex models as well as simple models. For the purposes of the following discussion, however, we equate a complex model with reality. The model's operation is specified by the matrix **Z**. The parameters on which it operates are specified by the vector **k**. As stated above, for the purposes of our linear analysis, values assigned to parameters comprising the vector **k** are considered to represent perturbations from their precalibration, expert-knowledge-based values. Equation (3) thus describes "reality" as it pertains to the calibration process.

[35] Let us subdivide **k** into **k**$_s$ and **k**$_o$, where "s" stands for "simple" and "o" stands for "omitted." The vector **k**$_o$ represents parameters excluded from the complex "reality model" in construction of a simple model which retains only the parameters **k**$_s$. Let **Z**$_s$ and **Z**$_o$ depict the processes which operate on these parameters. Thus **k** of equation (3) is partitioned as

$$\mathbf{k} = \begin{bmatrix} \mathbf{k}_s \\ \mathbf{k}_o \end{bmatrix}, \tag{6}$$

so that equation (3) becomes:

$$\mathbf{h} = \mathbf{Z}\mathbf{k} = \begin{bmatrix} \mathbf{Z}_s & \mathbf{Z}_o \end{bmatrix} \begin{bmatrix} \mathbf{k}_s \\ \mathbf{k}_o \end{bmatrix} + \boldsymbol{\varepsilon} = \mathbf{Z}_s\mathbf{k}_s + \mathbf{Z}_o\mathbf{k}_o + \boldsymbol{\varepsilon}. \quad (7)$$

[36] In the context of this paper, "omission" infers fixing of parameters (at possibly erroneous values), so that they lose their capacity for adjustment during the calibration process of the simple model and in analyzing postcalibration uncertainty of predictions made by the simple model. Thus, for example, where the number of layers in a groundwater model is reduced from many to few, $\mathbf{k}_o$ represents hydraulic properties that are notionally assigned to omitted layers. Alternatively, it can represent differences between the hydraulic properties pertaining to a series of thin layers that exist in reality and that assigned to a single thick layer which replaces it in a simple model. A similar concept applies where a numerical grid is coarsened in the horizontal direction; in this case $\mathbf{k}_o$ represents the differences between properties employed in a dense grid and upscaled properties assigned to elements of the replacement coarse grid. Similarly, where simplicity is based on the assumption of steady state conditions, some elements of $\mathbf{k}_o$ may represent differences between day-by-day recharge rates which prevail in reality and average recharge rates employed by the simplified model. In all of these cases the matrix $\mathbf{Z}_o$ represents the sensitivities of model outputs used in the calibration process to omitted parameters. Alternatively, it can be thought of as encapsulating processes which are omitted from the complex model that then allow omission of redundant adjustable parameters.

2.4. Calibrating the Simple Model

[37] The simple model need not be so simple that its parameters are uniquely estimable. Calibration of the simple model may thus constitute an ill-posed inverse problem. For generality, we will assume that this is the case. We will also (without loss of generality) assume that this inverse problem is solved through SVD. Hence in calibrating the simple model a parameter set $\underline{\mathbf{k}}_s$ is determined such that:

$$\underline{\mathbf{k}}_s = \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{h}, \quad (8)$$

where $\mathbf{V}_1$, $\mathbf{S}_1$, and $\mathbf{U}_1$ are partitioned according to positivity of singular values from $\mathbf{V}$, $\mathbf{S}$, and $\mathbf{U}$ obtained through singular value decomposition of $\mathbf{Z}_s$ as,

$$\mathbf{Z}_s = \mathbf{U}\mathbf{S}\mathbf{V}^t. \quad (9)$$

[38] The columns of $\mathbf{V}_1$ are unit vectors which span the solution space of $\mathbf{Z}_s$. The columns of its orthogonal complement (denoted herein as $\mathbf{V}_2$) span the null space of $\mathbf{Z}_s$ (see *Moore and Doherty* [2005] for full details). Note that, as will be discussed shortly, while the solution space of $\mathbf{Z}_s$ is, by definition, orthogonal to the null space of $\mathbf{Z}_s$, it is not necessarily orthogonal to the null space of the complex model $\mathbf{Z}$ of which $\mathbf{Z}_s$ is a simplification.

[39] On the basis of the above considerations, the calibrated parameter set $\begin{bmatrix} \underline{\mathbf{k}}_s \\ \mathbf{k}_o \end{bmatrix}$ established for the complex model $\mathbf{Z}$ is given by:

$$\underline{\mathbf{k}} = \begin{bmatrix} \underline{\mathbf{k}}_s \\ \mathbf{k}_o \end{bmatrix} = \begin{bmatrix} \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t \\ \mathbf{0} \end{bmatrix} \mathbf{h}. \quad (10)$$

[40] The zero-valued second element of the right side of (10) implies that values for $\mathbf{k}_o$ are not estimated, but remain at their precalibration values. Substituting (7) into (10) we obtain:

$$\underline{\mathbf{k}} = \begin{bmatrix} \underline{\mathbf{k}}_s \\ \mathbf{k}_o \end{bmatrix} = \begin{bmatrix} \mathbf{V}^1\mathbf{S}_1^{-1}\mathbf{U}_1^t \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{Z}_s & \mathbf{Z}_o \end{bmatrix} \begin{bmatrix} \mathbf{k}_s \\ \mathbf{k}_o \end{bmatrix} + \begin{bmatrix} \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t \\ \mathbf{0} \end{bmatrix} \boldsymbol{\varepsilon}. \quad (11)$$

2.5. Compensatory Parameters

[41] For the sake of simplicity (but in no way invalidating the analysis that follows), let us assume for a while that measurement noise $\boldsymbol{\varepsilon}$ is zero. From (11) the calibrated values $\underline{\mathbf{k}}_s$ of simple model parameters $\mathbf{k}_s$ are then computed as,

$$\underline{\mathbf{k}}_s = \mathbf{V}^1\mathbf{S}_1^{-1}\mathbf{U}_1^t[\mathbf{Z}_s\mathbf{k}_s + \mathbf{Z}_o\mathbf{k}_o]. \quad (12)$$

[42] If (9) is substituted into (12), it follows from the orthonormality of $\mathbf{U}$ that

$$\underline{\mathbf{k}}_s = \mathbf{V}_1\mathbf{V}_1^t\mathbf{k}_s + \mathbf{V}^1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{Z}_o\mathbf{k}_o. \quad (13)$$

[43] The first term on the right side of equation (13) is the projection of $\mathbf{k}_s$ onto the solution space of $\mathbf{Z}_s$. This constitutes the traditional solution to the inverse problem of model calibration (where the model is described by $\mathbf{Z}_s$) as obtained using subspace methods. Note that if $\mathbf{Z}_s$ were a matrix of full rank, solution for $\underline{\mathbf{k}}_s$ could be obtained using traditional overdetermined methods. These can be shown to be equivalent to SVD. In this case [see, for example, *Koch*, 1987]:

$$\underline{\mathbf{k}}_s = \mathbf{k}_s + (\mathbf{Z}_s^t\mathbf{Z}_s)^{-1}\mathbf{Z}_s^t\mathbf{Z}_o\mathbf{k}_o. \quad (14)$$

[44] Equations (13) and (14) depict the phenomenon of what we denote herein as "parameter surrogacy." This describes the situation where adjustable parameters become "contaminated" through their capacity to "soak up" a model-to-measurement misfit that would otherwise result from defects of the simple model. Equation (14) in particular, shows that under these circumstances each parameter is partly "what it is supposed to be" and partly something else, this "something else" being an outcome of the existence of omitted parameters $\mathbf{k}_o$. If a good fit between model outputs and field data is pursued in spite of the fact that a model is a simplification of reality (this being common calibration practice), then the likelihood of parameter surrogacy may become large; this likelihood increases with the degree of model simplification.

[45] From (13) it follows that parameter surrogacy is eliminated under either of the following conditions:

$$\mathbf{k}_o = \mathbf{0} \quad (15a)$$

(that is, the model has no defects) or

$$\mathbf{U}_1^t\mathbf{Z}_o = \mathbf{0}. \quad (15b)$$

[46] The second of the above conditions is met if simplification-induced model-to-measurement misfit has zero projection onto the same model output space as that containing information from the calibration data set. Hence the information contained in the calibration data set is separable from the effects on model outputs of model simplification. In order to at least partly achieve this effect, a modeler may restrict the dimensions of the solution space $\mathbf{V}_1$, and hence the complementary model output space $\mathbf{U}_1$, to ensure that (15b) is obeyed to the greatest extent possible. This is done through truncating the SVD process at a smaller number of singular values than that required to fit the calibration data set well. This constitutes an implementation of the traditional means of forestalling parameter surrogacy through the avoidance of over-fitting. While valid in concept, implementation of this methodology in practice is always subjective as $\mathbf{Z}_o$ is not known.

2.6. Null Space Parameter Entrainment

[47] Ideally, model calibration should be based on $\mathbf{Z}$, the true model, rather than on $\mathbf{Z}_s$, the simple model. If $C(\mathbf{k})$, the covariance matrix of innate parameter variability (where $\mathbf{k}$ includes both $\mathbf{k}_s$ and $\mathbf{k}_o$ parameters) is proportional to the identity matrix $\mathbf{I}$ (or is transformed to be such), a minimum error variance solution to the inverse problem of model calibration would thereby be obtained. This would constitute a superior solution to the inverse problem of model calibration compared to that discussed above based on $\mathbf{k}_s$ alone; in this case, all aspects of the model that are not perfectly known are taken into account, and are adjusted in accordance with the information content of the data on the one hand, and their propensity for variability as described by $C(\mathbf{k})$ on the other hand. A minimum error variance status for postcalibration $\mathbf{k}$ would then ensure a minimum error variance status for predictions which depend on $\mathbf{k}$ (this including all model predictions).

[48] Let us subject the "reality model" matrix $\mathbf{Z}$ to singular value decomposition such that

$$\mathbf{Z} = \mathbf{U}_r \mathbf{S}_r \mathbf{V}_r^t, \qquad (16)$$

where the "r" subscript denotes "real." The "real" solution to the inverse problem of model calibration would thus be composed of columns of the orthonormal matrix $\mathbf{V}_{r1}$ after appropriate partitioning of $\mathbf{V}_r$ into the parameter solution and null spaces $\mathbf{V}_{r1}$ and $\mathbf{V}_{r2}$. The number of elements in each unit vector comprising the columns of $\mathbf{V}_{r1}$ and $\mathbf{V}_{r2}$ is equal to the number of simple plus omitted model parameters (the latter being close to infinity if the complexity of a "real" model approaches that of the real world). In contrast, a solution to the inverse problem of model calibration obtained through calibration of the simple model maintains the values of all $\mathbf{k}_o$ parameters at those implied by model simplification, perturbations to these values therefore being zero. As seen in the context of the "real" parameter space, this solution to the inverse problem of model calibration is composed of linear combinations of columns of the matrix $\begin{bmatrix} \mathbf{V}_1 \\ \mathbf{0} \end{bmatrix}$, with $\mathbf{V}_1$ defined through equation (9) on the basis of the limited parameter set employed by the simple model. This matrix, thus defines the calibration solution space of the real model when only simple model parameters are adjusted.

[49] Optimality of model calibration achieved through adjustment of only simple model parameters requires the following:

$$\mathbf{V}_{r2}^t \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{0} \end{bmatrix} = \mathbf{0}, \qquad (17)$$

where $\mathbf{V}_{r2}$ defines the null space of the real model. Equation (17) states that the solution space of the simple model (this being the restricted subspace in which parameter adjustment is actually undertaken) lies within the solution space of the real model. It then follows that calibration of the simple model leads to the introduction of no null space components of the real model. As was previously discussed, the introduction of real model null space parameter components erodes the minimum error variance status of predictions made by the real model.

[50] Equation (17) implies that the solution space of $\mathbf{Z}$ contains no omitted parameters, for by definition, the solution space of $\mathbf{Z}$ is the orthogonal complement of the null space of $\mathbf{Z}$ and, through (17), is thus defined by $\begin{bmatrix} \mathbf{V}_1 \\ \mathbf{0} \end{bmatrix}$. All of the omitted parameters must therefore lie within the null space of the real model operator $\mathbf{Z}$. However this will only occur if model outputs under calibration conditions are insensitive to omitted parameters. Unfortunately, this ideal situation is seldom realized in practice. If any model output under calibration conditions is indeed sensitive to omitted elements of the real model (i.e., real or notional parameters that are hardwired into the construction of the simple model), then equation (17) cannot hold. It therefore follows that when model calibration is effected through adjustment of only simple model parameters (i.e., parameters $\mathbf{k}_s$ pertaining to $\mathbf{Z}_s$), solutions to the inverse problem thus achieved include null space components of the real model $\mathbf{Z}$, and are therefore nonoptimal. In other words, the adjustment of $\mathbf{k}_s$ parameters leads to entrainment of null space components of $\begin{bmatrix} \mathbf{k}_s \\ \mathbf{k}_o \end{bmatrix}$ parameters, and therefore to nonoptimality of the solution of the inverse problem when considered in the broader realm of the real model $\mathbf{Z}$ whose predictions of future environmental behavior we seek.

[51] Figure 2 provides an attempt to depict the situation. Here a real model is represented as having three parameters while a simple model is represented as having two parameters. Suppose that singular value decomposition of the real model matrix $\mathbf{Z}$ leads to a definition of the orthogonal unit vectors $\mathbf{v}_{r1}$, $\mathbf{v}_{r2}$, and $\mathbf{v}_{r3}$ comprising the $\mathbf{V}_r$ matrix of equation (16). Suppose further that the null space has only one dimension, this being spanned by $\mathbf{v}_{r3}$. If the vector $\mathbf{k}$ shown in the figure represents the true parameter set of the real model, then the best solution to the inverse problem of model calibration is represented by $\underline{\mathbf{k}}$, this being the projection of $\mathbf{k}$ onto the solution space of the real model. The latter is spanned by the unit vectors $\mathbf{v}_{r1}$ and $\mathbf{v}_{r2}$. Now let us suppose that $\mathbf{k}_{s1}$ and $\mathbf{k}_{s2}$ represent the two parameters employed by the simple model. Estimation of these two parameters constitutes a well-posed inverse problem. However, the simple model space spanned by these two parameters does not lie wholly within the solution space of the real model because $\mathbf{k}_{s1}$ and $\mathbf{k}_{s2}$ have nonzero
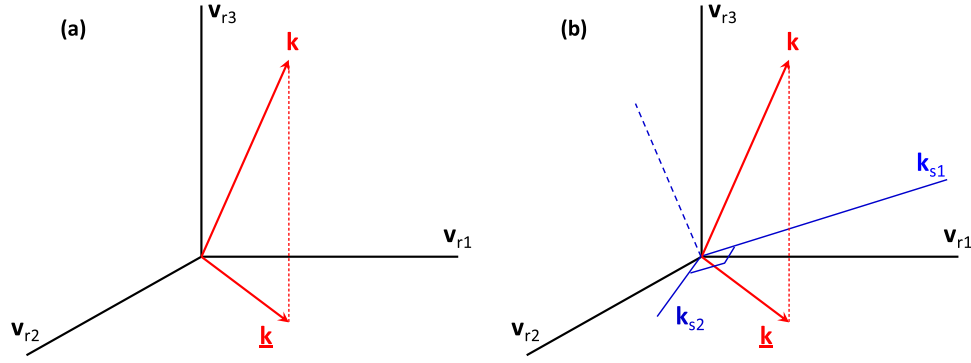
Figure 2. Figures 2a shows three axes in parameter space arising from singular value decomposition of the $\mathbf{Z}$ matrix representing the real (i.e., complex) model. Also shown is a parameter vector $\mathbf{k}$ and its projection onto the solution space of the real model, the latter being spanned by the orthogonal unit vectors $\mathbf{v}_{r1}$ and $\mathbf{v}_{r2}$. Figure 2b depicts the two parameter axes $\mathbf{k}_{s1}$ and $\mathbf{k}_{s2}$ of a simple model. A calibrated parameter vector $\underline{\mathbf{k}}$ for this simple model can be formulated to have the same projection onto the complex model solution space as that of any complex model parameter set $\mathbf{k}$. However, this vector will also have a projection onto the complex model null space.

projections onto the null space of the real model. As these parameters are adjusted through the simple model calibration process such that their vector sum has the same projection onto the complex model solution space as does the real world parameter set $\mathbf{k}$ (this being required to fit the calibration data set), so too are real model null space parameters unavoidably and inadvertently adjusted. null space parameter entrainment is the inevitable consequence.

[52] In summary, if model outputs corresponding to members of the calibration data set are sensitive to any structural aspects of a simple model that are omitted from the simple model adjustable parameter set, those simple model parameters that are adjusted through the calibration process tend to compensate for these simple model inadequacies. However with parameter compensation goes null space entrainment of real model parameters, for the two are inseparable. Parameters achieved through calibration of a simple model thus lose their minimum error variance status. This process can only be avoided if the solution space of the simple model coincides with that of the real model; that is, if the parameter reduction process implied by model simplification produces subspaces that are aligned with those implied by the calibration data set as it interacts with the parameter set of the real model.

2.7. Making a Prediction

[53] Let the scalar s represent a prediction made by the real model. Let the vector $\mathbf{y}$ designate the sensitivity of this prediction to real model parameters $\mathbf{k}$. Thus,

$$s = \mathbf{y}^t \mathbf{k}. \tag{18}$$

[54] The vector $\mathbf{y}$ can be partitioned into $\mathbf{y}_s$ and $\mathbf{y}_o$ components, where its sensitivities to simple model adjustable and omitted parameters are thereby made explicit. Thus,

$$s = \mathbf{y}_s^t \mathbf{k}_s + \mathbf{y}_o^t \mathbf{k}_o. \tag{19}$$

[55] Let $\underline{s}$ designate the prediction made by the calibrated simple model. It is calculated as,

$$\underline{s} = \mathbf{y}^t \underline{\mathbf{k}}_s, \tag{20}$$

so that predictive error is given by

$$\underline{s} - s = \mathbf{y}_s^t \underline{\mathbf{k}}_s - \mathbf{y}_s^t \mathbf{k}_s - \mathbf{y}_o^t \mathbf{k}_o, \tag{21}$$

that is (from (8)):

$$\underline{s} - s = \mathbf{y}_s^t \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \mathbf{h} - \mathbf{y}_s^t \mathbf{k}_s - \mathbf{y}_o^t \mathbf{k}_o. \tag{22}$$

[56] If (7) is substituted into this equation we obtain (after some simplification which recognizes the orthonormality of $\mathbf{V}$):

$$\underline{s} - s = \mathbf{y}_s^t \mathbf{V}_1 \mathbf{V}_1^t \mathbf{k}_s + \mathbf{y}_s^t \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \mathbf{Z}_o \mathbf{k}_o$$
$$+ \mathbf{y}_s^t \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \boldsymbol{\varepsilon} - \mathbf{y}_s^t \mathbf{k}_s - \mathbf{y}_o^t \mathbf{k}_o. \tag{23a}$$

[57] That is

$$\underline{s} - s = -\mathbf{y}_s^t (\mathbf{I} - \mathbf{V}_1 \mathbf{V}_1^t) \mathbf{k}_s + \mathbf{y}_s^t \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \mathbf{Z}_o \mathbf{k}_o$$
$$+ \mathbf{y}_s^t \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \boldsymbol{\varepsilon} - \mathbf{y}_o^t \mathbf{k}_o. \tag{23b}$$

[58] After further manipulation, and employing the relationship:

$$\mathbf{V}_1 \mathbf{V}_1^t + \mathbf{V}_2 \mathbf{V}_2^t = \mathbf{I}, \tag{24}$$

[see *Aster et al.*, 2005], we obtain:

$$\underline{s} - s = -\mathbf{y}_s^t \mathbf{V}_2 \mathbf{V}_2^t \mathbf{k}_s + (\mathbf{y}_s^t \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \mathbf{Z}_o - \mathbf{y}_o^t) \mathbf{k}_o + \mathbf{y}_s^t \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \boldsymbol{\varepsilon}. \tag{25}$$

[59] If $\mathbf{k}_s$, $\mathbf{k}_o$, and $\boldsymbol{\varepsilon}$ are statistically independent, their innate variability being denoted by their individual covariance matrices $C(\mathbf{k}_s)$, $C(\mathbf{k}_o)$, and $C(\boldsymbol{\varepsilon})$, we obtain the following expression for the error variance of the prediction s (inapplicability of this assumption is discussed shortly):

$$\sigma^2_{\underline{s}-s} = \mathbf{y}_s^t \mathbf{V}_2 \mathbf{V}_2^t C(\mathbf{k}_s) \mathbf{V}_2 \mathbf{V}_2^t \mathbf{y}_s + (\mathbf{y}_s^t \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \mathbf{Z}_o - \mathbf{y}_o^t) C(\mathbf{k}_o)$$
$$(\mathbf{y}_s^t \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \mathbf{Z}_o - \mathbf{y}_o^t)^t + \mathbf{y}_s^t \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t C(\boldsymbol{\varepsilon}) \mathbf{U}_1 \mathbf{S}_1^{-1} \mathbf{V}_1^t \mathbf{y}_s \tag{26}$$

[60] This equation can be simplified if the following relationships apply. (In fact, ideally the $\mathbf{Z}_s$ matrix should be formulated specifically through appropriate parameter transformation so that the first two of the following relationships do apply.)

$$C(\mathbf{k}_s) = \sigma_{ks}^2 \mathbf{I}, \qquad (27a)$$

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

$$C(\mathbf{k}_o) = \sigma_{ko}^2 \mathbf{I}. \qquad (27c)$$

[61] Equation (26) then becomes:

$$\sigma^2_{\underline{s}-s} = \sigma^2_{ks}\mathbf{y}_s^t\mathbf{V}_2\mathbf{V}_2^t\mathbf{y}_s + \sigma^2_{ko}(\mathbf{y}_s^t\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{Z}_o-\mathbf{y}_o^t) \\ \cdot (\mathbf{y}_s^t\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{Z}_o-\mathbf{y}_o^t)^t + \sigma^2_{\varepsilon}\mathbf{y}_s^t\mathbf{V}_1\mathbf{S}_1^{-2}\mathbf{V}_1^t\mathbf{y}_s \qquad (28)$$

[62] From (25) the contribution of model simplification to the error of the prediction is given by:

$$(\underline{s} - s)_{\text{simplification}} = (\mathbf{y}_s^t\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{Z}_o-\mathbf{y}_o^t)\mathbf{k}_o. \qquad (29)$$

[63] Even if $\mathbf{y}_o$ is zero (i.e., even if a prediction is insensitive to omitted structural parameters), the contribution of simplification-induced model structural defects to model predictive error may still be nonzero. This is an outcome of the first term in the brackets on the right of equation (29), i.e., the term that describes the compensatory nature that parameters play when a model with defects is calibrated.

2.8. Predictive Immunity to Model Structural Defects

[64] From equation (29) model structural defects will make no contribution to model predictive error if the following condition is met:

$$\mathbf{y}_o^t\mathbf{k}_o = \mathbf{y}_s^t\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{Z}_o\mathbf{k}_o. \qquad (30)$$

[65] This can also be written as,

$$\begin{bmatrix}\mathbf{y}_s & \mathbf{y}_o\end{bmatrix}^t \begin{bmatrix} \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{Z}_o \\ -\mathbf{I} \end{bmatrix} \mathbf{k}_o = \mathbf{0}. \qquad (31)$$

[66] Now consider a case where measurement noise $\boldsymbol{\varepsilon}$ is zero, and where model imperfections are such that they can be "calibrated out" because enough parameters can play compensatory roles for model-to-measurement misfit to be reduced to a very small level through the model calibration process. If this can occur, then it follows that for any data set $\mathbf{h}$ that can be calculated as,

$$\mathbf{h} = \mathbf{Z}_o\mathbf{k}_o, \qquad (32)$$

a simple model parameter set $\underline{\mathbf{k}}_s$ can be found such that:

$$\mathbf{h} = \mathbf{Z}_o\mathbf{k}_o = \mathbf{Z}_s\underline{\mathbf{k}}_s = \mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^t\underline{\mathbf{k}}_s. \qquad (33)$$

[67] That is,

$$\underline{\mathbf{k}}_s = \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{h} = \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{Z}_o\mathbf{k}_o. \qquad (34)$$

[68] Using this together with (33), we obtain:

$$\begin{bmatrix}\mathbf{Z}_s & \mathbf{Z}_o\end{bmatrix}\begin{bmatrix} \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{Z}_o \\ -\mathbf{I} \end{bmatrix}\mathbf{k}_o = \mathbf{Z}\begin{bmatrix} \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{Z}_o \\ -\mathbf{I} \end{bmatrix}\mathbf{k}_o = \mathbf{0}, \qquad (35)$$

from which it can be concluded that the vector $\begin{bmatrix} \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{Z}_o \\ -\mathbf{I} \end{bmatrix}\mathbf{k}_o$ is in the null space of the real model vector $\mathbf{Z}$. Equation (31) therefore states that a prediction will be immune from simplicity-induced model structural defects, even though these defects exist, if that prediction depends only on solution space components of the real model $\mathbf{Z}$. A repercussion of this is that a simple model may indeed possess many imperfections, to the extent that it may be considered to be a "black box." However, if it can be calibrated such that its outputs fit the calibration data set, then predictions made by that model will be immune from structural defects as long as these predictions are dependent only on solution space parameter components of the real model. Such predictions tend to be those which are similar in time and space, and pertain to the same predictive conditions, as measurements comprising the calibration data set. For predictions such as these, the main source of error will be that inherited from measurement noise associated with the calibration data set.

2.9. Comparing Simple and Complex Model Predictions

[69] Equation (25) can be rewritten as,

$$s = \underline{s} + \mathbf{y}_s^t\mathbf{V}_2\mathbf{V}_2^t\mathbf{k}_s - \mathbf{y}_s^t\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\mathbf{Z}_o\mathbf{k}_o - \mathbf{y}_s^t\mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^t\boldsymbol{\varepsilon} + \mathbf{y}_o^t\mathbf{k}_o. \qquad (36)$$

[70] Suppose that many random realizations of a complex model were generated using a pertinent statistical characterization of $\mathbf{k}$ $\left(\text{i.e.,}\begin{bmatrix}\mathbf{k}_s \\ \mathbf{k}_o\end{bmatrix}\right)$. Let us further suppose that in each case a simple model is calibrated against those complex model outputs that pertain to the calibration data set $\mathbf{h}$. A $\mathbf{k}_s$ that complements each stochastic $\mathbf{k}$ is thereby obtained. For each member of the complex/simple model pair thus obtained, we make a prediction of interest, thereby calculating s (for the complex model) and $\underline{s}$ (for the simple model). By doing this enough times, a scatterplot of s versus $\underline{s}$ can be constructed. The features of such a plot will now be analyzed. As will be discussed below, this constitutes the methodology for paired complex/simple model usage which forms the subject matter of this paper.

[71] Let us start our analysis of the s versus $\underline{s}$ scatterplot by assuming that the simple model is not simple at all, but is, in fact, the complex model calibrated against complex-model-generated outputs corresponding to the calibration data set $\mathbf{h}$. The "simple model" thus has no structural defects. Simplification is undertaken only through calibration; the parameter field is thus simplified in order to achieve

uniqueness of the calibration process. Suppose also that no noise is associated with complex model outputs that constitute the calibration data set. Equation (36) thus becomes

$$s = \underline{s} + \mathbf{y}_s^t \mathbf{V}_2 \mathbf{V}_2^t \mathbf{k}_s. \qquad (37)$$

[72] Under these circumstances a plot of s versus $\underline{s}$ shows vertical scatter about a line of slope 1. The extent of this scatter delineates the null space contribution to predictive error; scatter is vertical because this term affects s and not $\underline{s}$.

[73] With the addition of the measurement noise that would afflict a real-world calibration data set, equation (36) becomes:

$$s = \underline{s} + \mathbf{y}_s^t \mathbf{V}_2 \mathbf{V}_2^t \mathbf{k}_s - \mathbf{y}_s^t \mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{U}_1^t \boldsymbol{\varepsilon}. \qquad (38)$$

[74] The additional term adds horizontal scatter about a line with a slope of 1 as the measurement noise term affects $\underline{s}$ and not s. In many instances of real-world modeling, this scatter will be much smaller than the vertical scatter describing the null space contribution to predictive error. This follows from the often heavy dependence of predictions of real world interest on model null space parameter components (i.e., parameter combinations occupying the columns of $\mathbf{V}_2$ whose values cannot be inferred from the calibration data set).

[75] When the effects of model simplification are taken into account we return to equation (36). As has been described, the structural component of predictive error is given by equation (29), which is featured in equation (36). This term can be subdivided into two separate effects. The first of these is that caused by parameter surrogacy. This adds horizontal scatter to the s versus $\underline{s}$ scatterplot as it effects $\underline{s}$ but not s. The second effect arises from prediction sensitivity to structural simplifications that are not compensated for during the model calibration process. If the simple model can fit the calibration data set well, it is apparent from the preceding discussion that these components must lie in the null space of $\mathbf{Z}$, the complex model. Alternatively, where a less-than-perfect fit of the simple model to the calibration data set is purposefully sought in order to reduce parameter surrogacy (see above), some of these components will belong to a null space which is intentionally expanded by the modeler. Such expansion can be implemented through truncation of the simple model SVD-based calibration process at relatively high-singular values; alternatively, it can arise through assignment to the simple model of a purposefully simplistic parameterization scheme. In either case, this term effects s but not $\underline{s}$; it thus contributes to vertical scatter of the s versus $\underline{s}$ scatterplot. In doing so, it allows representation of the effects on simple model predictive error of complex model details not represented in the simple model.

[76] As will be demonstrated shortly, the following features of a practical s versus $\underline{s}$ scatterplot are also worthy of note.

[77] 1. If a high degree of horizontal scatter exists about a line of unit slope, the line of best fit through the scatterplot will possess a slope of less than 1.0. Flattening of the line of best fit will be exacerbated where scatter is prediction-dependent because of model nonlinearity. Under these circumstances, high predictive values will often tend to be more prone to horizontal scatter than low predictive values as there is likely to be a tendency for the effects which induce this scatter to shrink as s and $\underline{s}$ approach zero or a low value. Under these circumstances, the line of best fit through the s versus $\underline{s}$ scatterplot may tend toward a curve rather than a straight line.

[78] 2. Statistical interdependence between $\mathbf{k}_s$ and $\mathbf{k}_o$ (which was ignored in derivation of equation (36)) will produce a similar effect to that just described. In this case, the degree of horizontal scatter in the s versus $\underline{s}$ scatterplot will be related to the degree of vertical scatter. Prediction-dependent horizontal and vertical scatter may then lead to curvature of the line of best fit through the scatterplot.

[79] 3. The above discussion assumes that simplification is random and not consistent. In many cases simplification will, in fact, be consistent. For example, certain elements of $\mathbf{k}_o$ will be awarded fixed values during the model simplification process irrespective of the values awarded to these parameters during the stochastic complex model generation process. This will promulgate either a horizontal or vertical shift of the s versus $\underline{s}$ scatterplot, this depending on whether the affected components of $\mathbf{k}_o$ lie within the null space of $\mathbf{Z}$ (vertical shift) or the solution space of $\mathbf{Z}$ (in which case a horizontal shift will be induced by parameter surrogacy).

2.10. Summary

[80] The main outcomes of the above theoretical analysis will now be summarized.

[81] 1. Where the process of model simplification is such that the calibration solution space of the simple model is not perfectly aligned with that of a "reality model," at least some simple model parameters will take on compensatory roles during the calibration process. This engenders bias in estimated parameters, and in some predictions which depend on them, because of entrainment of real model null space parameter components during the simple model calibration process.

[82] 2. In spite of this, if a model prediction is dependent only on real model solution space parameter components, then that prediction will be immune to null space uncertainty, as well as to parameter bias induced through null space entrainment. Errors in that prediction will arise solely from the presence of measurement noise in the calibration data set. In general, such predictions are those which resemble members of the calibration data set.

[83] 3. An s versus $\underline{s}$ scatterplot can be created through comparing predictions made by a set of stochastically generated complex models with predictions made by a set of paired simple models that are calibrated against those complex model outputs that correspond to the calibration data set. Errors induced by prediction dependency on complex model null space parameter components promotes vertical scatter in this plot. Horizontal scatter arises from the presence of measurement noise in the calibration data set, as well as from the surrogate roles played by some parameters as they compensate for simplicity-induced model structural defects.

3. Paired Model Analysis

3.1. Methodology

[84] We now propose a methodology for making model-based predictions of future environmental behavior, for

analyzing the uncertainty associated with those predictions, and for reducing that uncertainty to as low a level as is commensurate with the current state of expert knowledge and with the information content of currently available measurements of system state.

[85] 1. Build a complex model for a system under study. Ensure that this model is complex enough to constitute a suitable repository for expert knowledge, and that it encapsulates all processes and parameters on which predictions of interest may depend.

[86] 2. Run this model many times using different realizations of parameter sets. Where appropriate, these should be conditioned on available measurements of system properties, as well as on any other data, such as geological and geophysical data, that collectively encapsulate prior knowledge of parameter values. Perhaps include in this stochastic process random realizations of facets of model construction such as layer geometries that are not normally considered as parameters. In each case, calculate model outcomes corresponding to members of the calibration data set **h** for the system under study. (Note that conditioning of random parameter fields on local measurements of system properties is very different from constraining those property fields in such a way that measurements of system state, for example, heads, calculated by the model on the basis of those fields are respected. The former is a relatively easy task. The latter comprises model calibration which, for a complex model endowed with a stochastic parameter field, can be a very difficult task. The difficulty of this task provides one of the motivations for the present work.)

[87] 3. Build a corresponding simple model. Ensure that this model is simple enough to run relatively quickly. If possible, ensure that it is complex enough to allow replication of the calibration data set **h**.

[88] 4. Add realizations of measurement noise to outputs of the complex model that correspond to members of the calibration data set. For each stochastic realization of the complex model, calibrate the simple model against these measurement-noise-enhanced complex model outputs. Obtain

thereby, a calibrated simple model counterpart to each realization of the complex model.

[89] 5. Compute predictions of management interest using complex model realizations. Calculate the same predictions using calibrated simple model counterparts.

[90] 6. For each prediction of interest plot s versus $\underline{s}$, where s is the prediction made by the complex model, and $\underline{s}$ is the same prediction made by its simple model counterpart.

[91] 7. Calibrate the simple model against the actual field data set. Do this using the same objective function and weighting strategy as that employed in step 4 where the simple model was calibrated against different realizations of complex model outputs corresponding to this data set.

[92] 8. Using the line (or curve) of best fit for the plots of s versus $\underline{s}$ obtained in step 6, correct the simple model prediction $\underline{s}$ for bias. At the same time, ascertain the uncertainty of the bias-corrected prediction from scatter around the line (or curve) of s versus $\underline{s}$ best fit.

3.2. Comments on the Methodology

[93] The methodology described in section 3.1 is illustrated in Figure 3. The first part of this figure shows a scatterplot of predictions made with a complex model against predictions made by its paired simple model. Each point in the figure pertains to one particular stochastic realization of a complex model that attempts to simulate processes that are operative at a particular study site. In each case, a paired simple model is parameterized by calibrating it against complex model outputs that correspond to measurements comprising the real-world calibration data set. Suppose that there are one thousand such paired models; thus, one thousand points comprise the scatterplot of Figure 3a.

[94] The simple model is then calibrated one more time, this time against the real-world calibration data set. Conceptually, the real world can be thought of as yet another stochastic realization of the complex model. The simple model has already been calibrated a thousand times against a model of this type. Furthermore, the value of a prediction
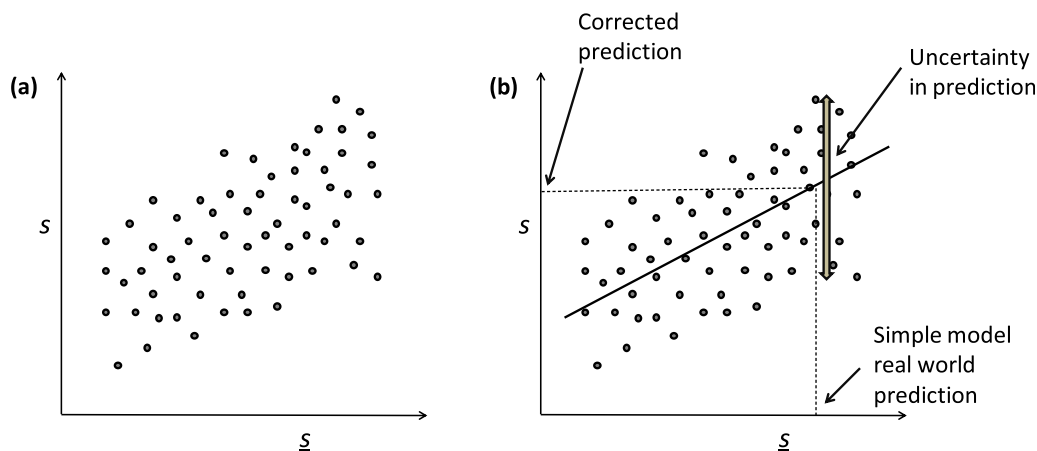

Figure 3. Figure 3a shows a scatterplot of complex model predictions against simple model predictions based on repeated use of stochastically generated complex models and their simple model counterparts. Figure 3b shows how a prediction made with a simple model calibrated against a real world calibration data set is processed using the scatterplot. First, the simple model prediction is corrected for bias. Then the uncertainty associated with that prediction is quantified.

of interest made by a thus-calibrated simple model has been compared with the same prediction as made by its complex model counterpart a thousand times. Hence, when the simple model is calibrated against the real-world data set, the outcomes of the previous calibration and prediction exercises endow us with the knowledge required to correct the simple model prediction for bias, and to quantify the uncertainty of this prediction. The first of these tasks is accomplished by following the dashed line in Figure 3b from the simple model prediction to the line of best fit through the scatterplot, and then to the complex model prediction axis. The second of these tasks is accomplished through measuring the scatter of points along the vertical line rising from the simple model prediction (indicated by the vertical double-arrow in Figure 3b). Meanwhile, the extent to which the line of best fit through the scatterplot deviates from a slope of unity provides the modeler with an indication of the extent to which parameters may have been endowed with surrogate roles during the simple model calibration process.

[95] A scatterplot such as that depicted in Figure 3 can also be used to demonstrate the effectiveness (or otherwise) of the model calibration process in reducing the uncertainty of a prediction of interest. The total vertical scatter of points within the graph (i.e., the scatter of points projected onto the s-axis) represents the precalibration uncertainty of the prediction. Precalibration uncertainty arises from the prior probability term of Bayes' equation, i.e., equation (1). In contrast, the scatter of points along the vertical line which intersects the real-world prediction $\underline{s}$ depicts the posterior uncertainty of the prediction (indicated by the vertical double-arrow in Figure 3b).

[96] The methodology of paired model usage described above has the following significant advantages:

[97] 1. The complex model need not run fast. While its numerical performance should have integrity, strict continuity of model outputs with respect to parameter values is not required, for it does not need to be calibrated. Nor do hydraulic properties employed by the complex model require parametric characterization; they can thus be represented categorically and stochastically.

[98] 2. The paired simple model need only be complex enough to allow a good fit to be achieved with the calibration data set. It need include few, if any, null space parameter components. Hence, calibration of that model could be accomplished with little numerical difficulty. It can thus be built with speed and stability in mind.

[99] 3. Because it is not essential that the simple model be physically based, only that it contain enough parameters to represent the complex model solution space and therefore match the calibration data set reasonably well, it can include devices such as "correction parameters" that can assist it in reducing bias and/or in obtaining a good fit with field data, regardless of whether these parameters possess a strict physical basis or not.

[100] 4. Once a complex model has been constructed, it can work in conjunction with multiple simple models, each focused on a different type of prediction. The design of these models can thus be such as to optimize their performance with respect to different types of environmental decision making.

[101] Before using a synthetic test case to demonstrate the paired complex/simple model methodology, a comment

on the linearity assumption which underpins the theory provided in section 2 is warranted. The linearity assumption facilitates a mathematical analysis of the model simplification process based on subspace concepts. At the same time it explains some of the features of predictions made by a simplified model, and some of the characteristics of a plot of s versus $\underline{s}$. Where the dependence of model outputs on model parameters is nonlinear, the modeling context is more complicated. Solution and null spaces will possess complex, and sometimes discontinuous, geometries. Nevertheless, ramifications of the model simplification process as described herein are likely to be robust. The same applies to the paired complex/simple model methodology described herein for the exploration of model predictive bias and uncertainty. Construction of a plot of s versus $\underline{s}$ requires no linearity assumption. Model nonlinearity may engender curvature of the line of s versus $\underline{s}$ best fit; it may also result in asymmetric and nonstationary scatter of points about this line (or curve). However neither of these detracts from the use of the line (or curve) of best fit in identifying and correcting for model predictive bias, nor in scatter about this line (or curve) in providing a measure of model predictive uncertainty.

4. Test Case Description

[102] The paired model methodology described above is now demonstrated using a synthetic case that is based on two commonly encountered real-world groundwater management problems. These are the prediction of hydraulic head and drawdown caused by pumping of groundwater from a well, and the prediction of recharge areas of that pumping well.

[103] For this particular example, model simplification pertains to spatial aspects of the model's construction and parameterization; furthermore, the calibration data set and predictions that we investigate are composed only of steady state model outputs. As the theory presented herein is general, its application is neither restricted solely to models of this type, nor to simplification strategies of this kind. Application of the concepts and methodology discussed herein to other types of models, and its repercussions for system dynamic response rather than steady state response, will be the subject of another publication.

4.1. Hydrogeological Setting

[104] The model domain is rectangular, being 7 km north-south (N-S) and 5 km east-west (E-W). To the south it is bounded by the ocean while the other external boundaries are closed (no flux).

[105] The geological setting of the area is conceptualized as glacial deposits underlain by impermeable clay. The impermeable system base is horizontal in most of the catchment; however, a 150 m deep valley is eroded into the central part of it. The valley runs north from the coast for a distance of 5 km inland and has sloping sides; see Figures 4 and 5. It is filled with glacial sediments deposited in highly N-S elongated layered structures (representing shifting channel fillings) consisting of either gravel, sand, silt, or clayey till. The entire area is capped by 50 m of glacial sediments deposited as gently N-S elongated layered structures composed of either sand, silt, or clayey till. The setting is thus typical of parts of northern Europe and North America: a glacially formed landscape with a buried tunnel valley.

[106]   The exact stratigraphy is only known at the locations of 35 boreholes of variable depth; see Table 1 and Figure 4. Thirty of the boreholes have been constructed as monitoring wells, each screening the deepest 10 m of sand registered in the borehole. (Borehole stratigraphy is not documented here since this would require a very large table.)

[107]   Groundwater is recharged by the deficit between precipitation and evapotranspiration. The amount of recharge is partly dependent on the type of sediment at the surface; it is larger in sandy areas, intermediate in silty areas, and smaller in clayey areas. Groundwater is pumped from a deep well located in the southern-central part of the buried valley ($x = 2487.5$ m; $y = 1912.5$ m). The well screens the deepest 10 m of the valley, which at this depth contains a laterally extensive body of sand and gravel. The pumping rate is 0.015 $m^3 s^{-1}$.

4.2.   Complex Model

[108]   The complex model is a 3-D groundwater flow and particle tracking model set up in MODFLOW-2000 [*Harbaugh et al.*, 2000] and MODPATH version 5 [*Pollock*, 1994]. All cells of the finite difference grid have horizontal

dimensions of 25 m × 25 m and 10 m thickness, so that the overall dimensions of the grid are $(n_x, n_y, n_z) = (200, 280, 20)$. The total number of cells is thus 1,120,000.

[109]   The categorical depositional geology of the 3-D model grid was simulated using T-PROGS [*Carle*, 1999]; the proportions and mean lengths for the different categories of sediment are provided in Table 2. The bedding is represented as a maximally disordered system using "maximum entropy" transition frequencies [*Carle*, 1999]. Each simulation was conditioned to match the stratigraphy in the 35 boreholes.

[110]   Within each category of sediment the spatial distribution of hydraulic conductivity was simulated as a horizontally correlated random field; Figure 5 shows an example. The same was done for porosity and recharge. The random fields were generated using the sequential Gaussian simulation method [*Deutsch and Journel*, 1998] based on the geostatistical parameters given in Table 3.

[111]   Groundwater flow is confined and at steady state. For each stratigraphic and hydraulic property realization, the simulated hydraulic head field was used to calculate heads in the 30 monitoring wells. Independent Gaussian errors with zero mean and a variance of 0.01 m^2 were added to MODFLOW-calculated well heads to produce head observations used to calibrate the simplified model.

[112]   A total of 1000 complex model simulations were made (each including generation of a categorical depositional fields, a stochastic field of hydraulic conductivity within the categorical depositional field porosity and recharge pertaining to each deposit, and simulation of groundwater flow and particle tracking). Except for the fact that the categorical fields were all conditioned on the same stratigraphy for the 35 boreholes, the 1000 simulations are independent.

4.3.   Simple Model

[113]   The simplified model, which combines MODFLOW-2000 and MODPATH, uses a coarser grid than the complex one. The grid is $(n_x, n_y, n_z) = (50, 70, 4)$, this giving a total of 14,000 cells (80 times less than for the complex model). The two top layers of equal thickness represent the 50 m capping layer. The two deeper layers, of 70 and 80 m thickness, represent the 150 m deep buried tunnel valley. The thicknesses and hydraulic conductivities (i.e., transmissivities) of the four layers mainly control horizontal groundwater flow, while the vertical conductance between the centers of each of these layers controls vertical flow.

[114]   For calibration purposes hydraulic conductivity (K) of the four simple model layers, the vertical conductance (C) between the layer centers, and the recharge (R) supplied to the top model layer are each represented as the product of a spatially uniform value and a spatially variable multiplier field. The multiplier fields for all hydraulic properties, and for all layers, are independent of each other. Both uniform layer hydraulic property values and the spatial distribution of multipliers were estimated through calibration. Each multiplier field was parameterized using pilot points [see *Doherty*, 2003] placed in a uniform 500 m grid. Interpolation from pilot points to multiplier fields was done using simple kriging (with a log mean of zero) based on the geostatistical parameters provided in Table 4.

[115]   Initial calibration experiments showed that the hydraulic conductivity (K) fields of the two upper layers
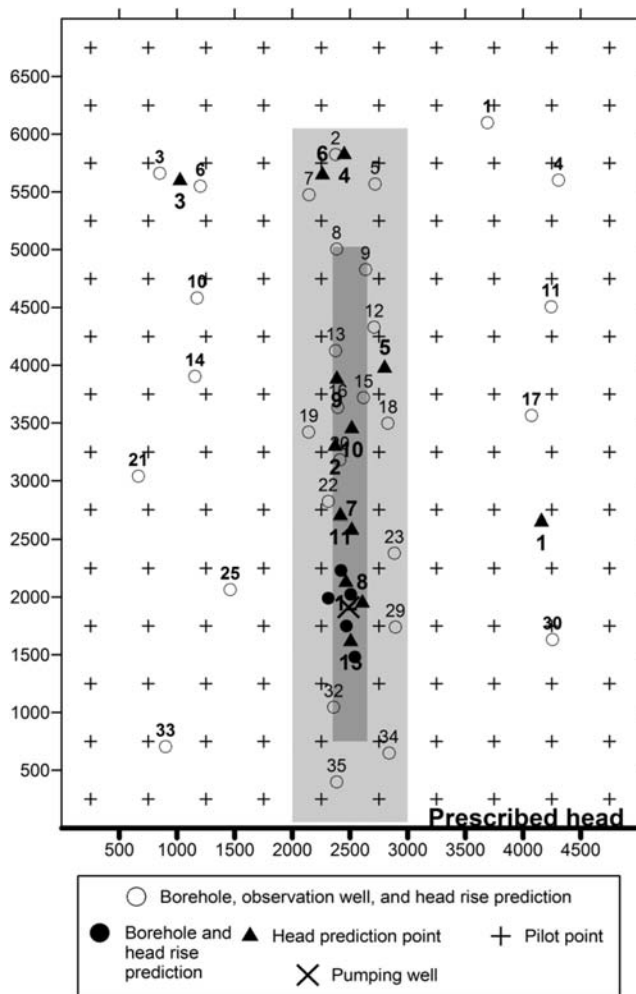

Figure 4.   A map of boreholes, wells, pilot points, and locations at which predictions are made within the model area. The gray-shaded areas indicate top (light) and bottom (dark) portions of the buried valley where that valley is represented by only two layers in the simple model.
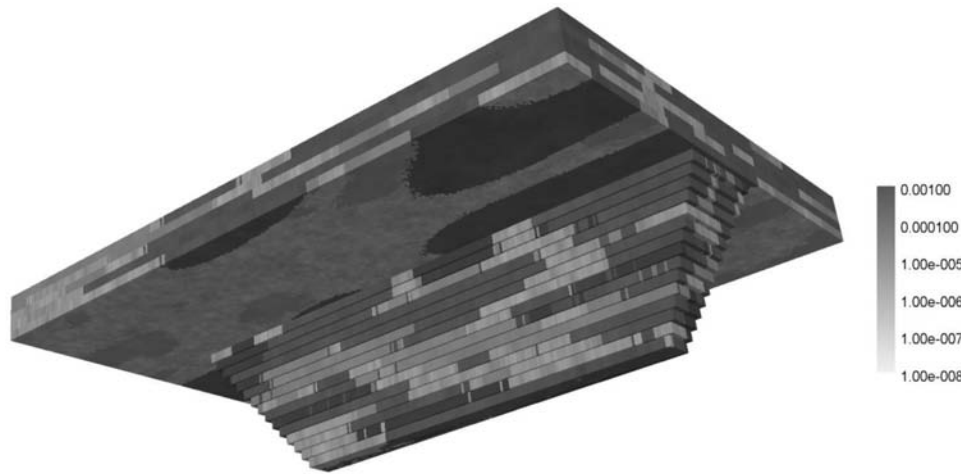
Figure 5. Example of hydraulic conductivity field (m s^{-1}) of complex model. Notice the difference in shape of stratigraphic patterns between the capping layer and the buried valley deposits. The former patterns have large extensions in both horizontal directions while the latter have small extensions transverse to the main direction of the valley, and large extensions in the longitudinal direction.

cannot be estimated independently from the available observations. Identical hydraulic conductivity fields were therefore used for these layers. The total number of adjustable model parameters was thus 507 − seven spatial constants (K_{1+2}, K_3, K_4, C_{1-2}, C_{2-3}, C_{3-4} and R where subscripts specify model layer numbers), and a total of 500 pilot point values for the seven multiplier fields.

[116] Calibration was implemented automatically using PEST [*Doherty*, 2011a] employing a combination of constrained Tikhonov regularization and singular value decomposition (SVD). Use of the Tikhonov component allows the setting of a "target measurement objective function"—this specifying the level of model-to-measurement misfit which will hopefully be achieved, but will not be further reduced (or reduced only slightly) in order to avoid the deleterious effects of over-fitting. This target value was set to 0.35, which is a little larger than $30 \times 0.01 = 0.30$, where 30 is the number of head observations used for calibration and 0.01 is the observation error variance (observation weights were all set to 1.0). This strategy ensured that the resulting calibrated parameter fields are smooth, and deviate as little as possible from spatially constant values, while the model-to-measurement misfit for all observation wells is of the order of observation error. In other words, each calibrated simple model fits the "observations" calculated by the complex model acceptably well (i.e., to a level set by the noise that accompanies these observations), while the calibrated parameter fields are as homogeneous as possible.

[117] Each iteration of the calibration process required computation of the sensitivity matrix of model outputs used in the calibration process with respect to all adjustable parameters. This was somewhat time-consuming because of the number of parameters requiring estimation (507, as stated above); calibration of the simple model therefore required 1–2 hr of computing time. Nevertheless, repeated recalibration of the simplified model was possible; given the size and parameterization scheme of the complex model, calibrating it even once would have been impossible. Using the above strategy, a calibrated simple model

counterpart was thus obtained for each of the 1000 realizations of the complex model.

[118] In summary, simplifications implemented in construction of the simple model from the complex model included the following use of a coarse horizontal grid use of a substantially smaller number of model layers (this resulting in coarser representation of the geometry of the buried valley), representation of the combined vertical conductance of many layers by single interlayer conductance terms, replacement of detailed categorical stochastic parameterization of recharge and conductivity by uniform arrays of pilot points, assumption of vertical homogeneity in the upper two layers of the simplified model, and use of uniform porosity (see below).

4.4. Predictions

[119] A number of predictions were computed using the 1000 complex model realizations; they were also computed using their calibrated simple model counterparts. These predictions were:

[120] (1) Hydraulic head at 13 ungaged positions while pumping from the well,

[121] (2) Rise in the hydraulic head to its new steady state level in the 30 observation wells on cessation of pumping, and

[122] (3) Recharge area of, and minimum travel time to, the pumping well.

[123] The two first types of prediction were made using MODFLOW, while the recharge area and the minimum travel time were calculated from MODFLOW results using MODPATH. The recharge area was computed by placing particles in a horizontally uniform 25 m grid at the surface, and tracking them forward in time until reaching either the pumping well or the southern prescribed-head model boundary. Each particle reaching the well is assumed to represent recharge from a 25×25 m^2 area; the number of particles terminating in the pumping well thus defines the total recharge area. The travel time of the first particle to reach the pumping well is the minimum travel time. The

Table 1. Locations and Depths of Boreholes With Lithological Information, and for Which Head Rise Predictions Were Computed[a]

		Location				Fitting the s Versus $\underline{s}$ Line				95% Uncertainty Interval Testing			
Borehole	X (m)	Y (m)	Bore Hole Depth (m)	Screen in Complex Model Layer	Screen in Simple Model Layer	α	β	r^2	S	Number	Inside	Above	Below
1	3692	6100	50	4	2	−0.09	0.44	0.41	0.18	500	490	3	7
2	2375	5824	90	8	3	−0.26	0.41	0.36	0.19	500	489	3	8
3	850	5662	50	4	2	−0.03	0.45	0.44	0.17	500	486	8	6
4	4308	5602	50	3	2	−0.02	0.44	0.44	0.17	500	487	3	10
5	2717	5570	120	6	3	−0.24	0.41	0.37	0.18	500	488	3	9
6	1201	5550	50	4	2	−0.03	0.45	0.43	0.18	500	487	7	6
7	2144	5477	110	8	3	−0.27	0.41	0.35	0.19	500	489	3	8
8	2384	5006	200	16	4	−0.27	0.41	0.34	0.19	500	488	4	8
9	2634	4830	200	14	4	−0.28	0.41	0.34	0.19	500	488	4	8
10	1174	4583	50	3	2	0.03	0.45	0.43	0.18	500	488	7	5
11	4243	4506	50	4	2	0.00	0.44	0.43	0.17	500	488	4	8
12	2708	4330	170	15	4	−0.32	0.41	0.32	0.21	500	488	3	9
13	2375	4127	200	19	4	−0.22	0.45	0.34	0.22	500	486	5	9
14	1155	3905	50	3	2	0.06	0.45	0.43	0.18	500	482	10	8
15	2616	3720	200	20	4	−0.21	0.46	0.32	0.24	500	485	2	13
16	2394	3637	200	19	4	−0.23	0.46	0.32	0.25	500	483	2	15
17	4073	3565	50	4	2	0.00	0.43	0.42	0.17	500	488	3	9
18	2828	3498	120	12	3	−0.31	0.41	0.32	0.21	500	493	2	5
19	2140	3421	110	10	3	−0.25	0.42	0.35	0.20	500	490	4	6
20	2412	3184	200	20	4	−0.20	0.50	0.27	0.32	500	484	0	16
21	665	3042	50	4	2	0.06	0.44	0.45	0.16	500	484	10	6
22	2311	2823	170	13	3	−0.35	0.40	0.27	0.22	500	492	2	6
23	2884	2379	100	6	3	−0.16	0.38	0.33	0.20	500	489	1	10
24	2421	2231	200	(20)	(4)	0.28	0.79	0.20	1.07	500	485	1	14
25	1460	2064	50	5	2	−0.21	0.34	0.30	0.18	500	481	15	4
26	2506	2024	200	(20)	(4)	−0.90	0.61	0.23	1.50	500	474	3	23
27	2311	1990	180	(18)	(4)	−1.06	0.51	0.17	1.25	500	476	8	16
28	2468	1750	200	(20)	(4)	0.03	0.79	0.15	1.49	500	490	0	10
29	2893	1741	90	9	3	−0.46	0.33	0.22	0.25	500	486	1	13
30	4255	1632	50	4	2	−0.13	0.37	0.46	0.12	500	483	7	10
31	2542	1482	200	(20)	(4)	−0.91	0.52	0.08	0.93	500	488	0	12
32	2357	1047	200	20	4	−0.86	0.33	0.08	0.54	500	482	0	18
33	900	705	50	5	2	−0.05	0.36	0.53	0.06	500	472	10	18
34	2838	649	110	11	3	0.18	0.55	0.74	0.11	500	481	6	13
35	2384	400	120	12	3	0.20	0.60	0.80	0.13	500	476	8	16

[a]For boreholes where the screened layer numbers are in parentheses, there is no head observation used in the calibration process. α and β are regression coefficients used in equation (39), r^2 is the correlation coefficient, and S is the standard deviation. Interval testing provides the number of realizations for which complex model predictions fall inside, above, and below the estimated 95% prediction interval about the line of s versus $\underline{s}$ best fit.

calculation of travel time requires knowledge of porosity. Since porosity cannot be estimated through calibration against steady state head observations, a spatially uniform porosity of 0.2 was attributed to the domain of the simplified model. (Recall that in the complex model, the porosity is represented by a random field with a mean that varies between the four types of categorical deposits.)

4.5. Bias Correction for Simple Model Predictions

[124] As described above, this step involves plotting each prediction made by the complex model (i.e., s) against its simple model counterpart (i.e., $\underline{s}$). For each prediction, a number of (but not all) the pairs were used to estimate the coefficients of a regression (best-fit) line:

$$s = \alpha + \beta \underline{s}, \qquad (39)$$

and its corresponding 95% prediction interval (see, for example, *Draper and Smith* [1998], equation (1.4.12) for details of prediction interval calculations). In equation (39) α and β are regression coefficients (intercept and slope). As explained in section 2 of this paper the regression line corrects the simple model prediction for bias while the prediction interval records the uncertainty associated with the

Table 2. Proportions and Mean Lengths for Sediment Categories

Category	Proportion (−)	$\overline{L}_x$ (m)	$\overline{L}_y$ (m)	$\overline{L}_z$ (m)
Inside Valley				
Gravel	0.10	100.	800.	10.
Sand	0.40	200.	1200.	20.
Silt	0.20	300.	600.	10.
Clay	0.30	400.	500.	20.
Capping Layer				
Sand	0.45	1200.	4800.	20.
Silt	0.20	500.	1000.	10.
Clay	0.35	500.	1000.	10.

Table 3. Geostatistical Parameters for Stochastic Hydraulic Property Fields Employed by the Complex Model[a]

	$\mathrm{Log}_{10}(K)$			$\mathrm{Log}_{10}(R)$			$\mathrm{Log}_{10}(\varphi)$		
Category	μ	a	σ^2	μ	a	σ^2	μ	a	σ^2
Gravel	−3.00	200.	0.0227	−8.20	200.	0.007,752	−0.60	200.	0.000,428
Sand	−4.00	200.	0.0227	−8.20	200.	0.007,752	−0.60	200.	0.000,428
Silt	−6.00	200.	0.0227	−8.60	200.	0.007,752	−0.74	200.	0.000,428
Clay	−7.00	50.0	0.122	−8.82	50.0	0.007,752	−1.00	50.0	0.000,428

[a]K = hydraulic conductivity (m s^{-1}); R = recharge (m s^{-1}); φ = porosity (−); μ = mean; a = range (m); σ^2 = sill. The semivariograms are exponential.

prediction. A regression line and prediction interval was estimated for each of the predictions examined in this study.

4.6. Testing Computed Predictive Uncertainty

[125] We employed predictions from only some model realizations in estimating regression lines and prediction intervals. The remaining realizations were used to test the integrity of the method. The following was done for each remaining realization.

[126] 1. The predicted value of the simple model was employed in conjunction with equation (39) to correct for bias and to determine the upper and lower 95% prediction limits (from the prediction interval).

[127] 2. The value of the prediction made by the corresponding complex model with respect to these prediction limits was noted.

[128] 3. Steps 1 and 2 were repeated for all model realizations not used to estimate the regression coefficients of equation (39).

[129] 4. A count was made of the number of times the complex model prediction fell inside and outside the prediction limits.

[130] If the complex model prediction falls inside the 95% prediction limits for 95% of these verification realizations, this suggests that the method of quantifying predictive uncertainty is accurate for that prediction.

5. Test Case Results

5.1. Simple Model Calibration

[131] The simple model was calibrated to fit head observations generated by the complex model for 1000 realizations of the latter. For 983 realizations, calibration reduced the sum of squared head residuals to less than or equal to the target value of 0.35 m^2. For 15 of the remaining realizations, the sum of squared head residuals was found to be less than 0.8 m^2. For the remaining two realizations, the

Table 4. Geo-Statistical Parameters for Log-Transformed Stochastic Fields of the Simplified Model[a]

	μ	a_x	a_y	σ^2
R	0.0	250.	500.	.122
K, C of capping layer	0.0	250.	1000.	1.56
K, C of buried valley	0.0	250.	500.	1.56

[a]K = multiplier field for hydraulic conductivity of model layers; C = multiplier field for leakance of confining beds (m s^{-1}); R = multiplier field of recharge; μ = mean; a_x = range in x-direction (m); a_y = range in y-direction (m); σ^2 = sill. The semivariograms are exponential.

sum of squared head residuals was 2.09 and 319.8 m^2. These 17 realizations were excluded from further analyses.

[132] While there was a tendency for the calibrated simple model to fit some (but not all) of the observations from within the buried valley more poorly than other observations, few, if any other trends, were apparent in the fitting pattern over the realizations for which the fit was decreed to be acceptable.

5.2. Head Prediction

[133] Table 5 gives s versus $\underline{s}$ regression line coefficients obtained from 483 paired model realizations for the 13 head predictions. In most cases α (the intercept) differs from 0.0, this indicating that the simple model possesses consistent bias in these predictions. As described in section 2 of this paper, this arises from a consistent error in null space parameter components omitted from the simple model. For all head predictions, the estimated slope β of the s versus $\underline{s}$ line is nearly 1.0; hence, parameter surrogacy does not appear to strongly affect the simple model's ability to predict hydraulic heads at the prediction points.

[134] Figure 6 shows plots of four paired model head predictions, two from the capping layer (3 and 4), and two from the buried valley (9 and 12). Prediction 3 is located between two closely spaced observation wells (3 and 6). In this case, the points plot close to the identity line, the correlation coefficient (r^2) is 1.0, and the standard deviation (S) is less than 0.01; the simple model prediction is near perfect. Prediction 4 is located between two distant observation wells (1 and 6). For this prediction there is some bias; see α and β in Table 5. The points are more scattered around the regression line which is also indicated by the higher standard deviation and to some degree by the correlation coefficient in Table 5; this indicates some null space contribution to the uncertainty of this prediction.

[135] The results are similar for predictions 9 and 12 in the buried valley (Figure 6 and Table 5). Prediction 9 is located between observations 13, 15, and 16. The points plot close to the identity line. Prediction 12 is close to the pumping well and distant from the observation wells. Evidently this introduces some bias, together with a large null space component of parameter uncertainty as evinced by a large S and a relatively small r^2. The small r^2 is a result of the large scatter of points around the regression line and a tendency for outliers to fall below the 95% prediction interval for simple model predictions less than approximately 10 m. This nonstationarity of scatter about the s versus $\underline{s}$ line is probably an outcome of model nonlinearity.

[136] As stated above, while 983 realizations of paired model predictions were used in our analysis, only 483 of

Table 5. Head Prediction Results[a]

Head Prediction Point	Location				Fitting the s Versus $\underline{s}$ Line				95% Uncertainty Interval Testing			
	X (m)	Y (m)	Screen in Complex Model Layer	Screen in Simple Model Layer	α	β	r^2	S	Number	Inside	Above	Below
1	4160	2650	4	2	−0.17	0.99	0.92	0.84	500	480	15	5
2	2370	3300	4	2	0.36	0.96	0.97	0.38	500	491	6	3
3	1025	5600	4	2	0.02	1.00	1.00	0.09	500	486	8	6
4	2450	5825	4	2	1.13	0.94	0.91	0.82	500	480	11	9
5	2800	3975	6	3	0.64	1.01	0.96	0.50	500	473	24	3
6	2260	5650	8	3	−0.14	1.00	1.00	0.13	500	480	3	17
7	2515	2580	10	3	0.09	0.98	0.97	0.33	500	478	13	9
8	2606	1950	12	3	0.01	1.00	0.90	0.56	500	469	7	24
9	2385	3880	19	4	0.00	1.00	1.00	0.11	500	481	7	12
10	2514	3452	20	4	0.05	1.00	1.00	0.14	500	478	13	9
11	2416	2700	20	4	0.34	0.95	0.88	0.70	500	471	5	24
12	2464	2128	20	4	0.26	0.96	0.72	1.17	500	478	5	17
13	2505	1615	20	4	−0.43	1.02	0.79	0.87	500	483	3	14

[a]α and β are regression coefficients defined in equation (39), r^2 is the correlation coefficient, and S is the standard deviation. Interval testing provides the number of realizations for which complex model predictions fall inside, above, and below the estimated 95% prediction interval about the line of s versus $\underline{s}$ best fit.
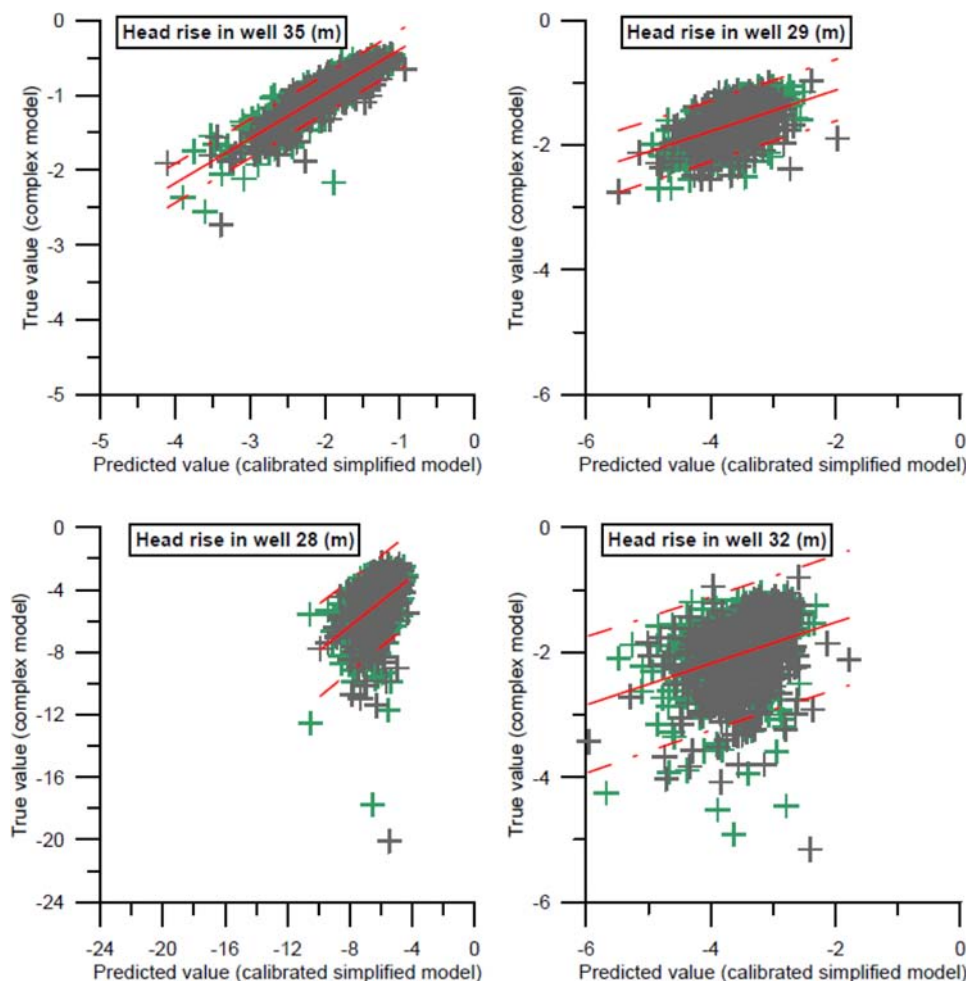

Figure 6. The paired model predictions of hydraulic head. For each prediction 483 of the points (gray) were used to estimate the regression line and the 95% prediction interval; the other 500 points (green) plot similarly and were used for testing.

these were used in computing the statistics presented in Table 5, and the regression lines and prediction intervals presented in Figure 6. The remaining 500 realizations were used to test the integrity of predictive bias correction and quantification of model predictive uncertainty. This was done by comparing each bias-corrected simple model prediction with the corresponding prediction made by the complex model. Table 5 shows that for the 13 head predictions, the number of points falling between the prediction limits varies between 469 (93.8%) and 483 (96.6%). These numbers are close to 475 (this corresponding to 95% of 500 realizations). However Table 5 also shows that for some predictions the number of points falling above the upper prediction limit is noticeably different from the number of points falling below the lower prediction limit. A similar phenomenon is apparent for prediction 12 in Figure 6. This suggests that use of a linearized prediction interval about the line of s versus $\underline{s}$ best fit may not be appropriate for these predictions due to nonstationarity of scatter about this line, and/or that the line of s versus $\underline{s}$ best fit should be allowed to exhibit some curvature.

5.3. Head Rise Prediction

[137] Table 1 gives regression line coefficients and 95% uncertainty outcomes for another model prediction. This is the prediction of head rise following cessation of pumping incurred in all boreholes from which observations were used in the calibration process. It is apparent that in many cases α differs from 0.0, and that β in all cases is less than 0.8 and mostly less than 0.5. The former indicates that simple model predictions are systematically in error. This is probably due to consistent errors in null space parameter components omitted from the simple model as described in the theory. As is also described in the theory, β values of less than 1.0 indicate parameter surrogacy incurred through the calibration of the simple model.

[138] Figure 7 provides examples of s versus $\underline{s}$ plots for predictions of head rise obtained at different observation points. Wells 35 and 29 have their screen in the upper part of the buried valley; the former is further from the pumping well than the latter, and is close to the southern model boundary. The slope of the regression line for well 35 is moderate ($\beta = 0.60$), the scatter around the regression line

Figure 7. The paired model predictions of rise in hydraulic head. For each prediction 483 of the points (gray) were used to estimate the regression line and the 95% prediction interval; the other 500 points (green) plot similarly and were used for testing.

Table 6. Prediction Results for Recharge Area of Pumping Well and Minimum Travel Time to Pumping Well[a]

Prediction	Fitting the s Versus $\underline{s}$ Line				95% Uncertainty Interval Testing			
	α	β	r^2	S	Number	Inside	Above	Below
Recharge area	2.69×10^6	0.16	0.12	7.09×10^5	500	481	9	10
Min. travel time	1.01×10^9	0.02	0.00	2.79×10^8	500	489	7	4

[a]α and β are regression coefficients defined in equation (39), r^2 is the correlation coefficient, and S is the standard deviation. Interval testing provides the number of realizations for which complex model predictions fall inside, above, and below the estimated 95% prediction interval about the line of s versus $\underline{s}$ best fit.

is moderate ($S = 0.13$), and the correlation coefficient is reasonably high ($r^2 = 0.80$). In testing the efficiency of the linear prediction interval, it is found that almost exactly the expected number (476 versus 475) of testing points fall within the prediction limits; however, there is a tendency for too many points to fall below (16) and too few to fall above (8) the lower- and upper-prediction limits, respectively.

[139] For well 29, the slope and the correlation coefficient are much smaller (0.33 and 0.22, respectively), and the standard deviation is larger (0.25) than for well 35. The number of test points falling inside the 95% prediction interval is 486 (97.2%), but of some concern is that 13 points fall below and only one is above the prediction interval. Once again, this raises questions concerning the use of a linearized prediction interval about the line of s versus $\underline{s}$ best fit when quantifying the uncertainty of head rise predictions made by the simple model in well 29. It suggests that some curvature be admitted to the line of s versus $\underline{s}$ best fit, or even that prediction limits be estimated manually.

[140] Figure 7 also shows results for the predicted head rise in wells 28 and 32; both of these have screens in the deepest part of the buried valley, and both are fairly close to the pumping well. For both wells the scatter of points is significant, in particular for well 32; furthermore, several testing points fall below (and some much below) the linearized prediction interval about the line of s versus $\underline{s}$ best fit, while none fall above the prediction interval (Table 1). The dominance of the null space contribution to the uncertainty

of this prediction is evinced by the high degree of scatter about the line of s versus $\underline{s}$ best fit, together with the relatively short length of this line in the horizontal direction.

5.4. Prediction of Recharge Area and Minimum Travel Time

[141] Table 6 and Figure 8 provide regression line coefficients and 95% uncertainty outcomes for particle-tracking-based predictions of the recharge area of the pumping well and of the minimum travel time from a recharge point to the pumping well. For both predictions the intercept α is large, this is probably an outcome of consistent errors in null space parameter components incurred through the simplification process. For the travel time prediction, part of this error is probably attributable to the assumption of a uniform porosity of 0.2 whereas, in reality, the mean porosity differs within each stratigraphic unit. For both types of prediction, the estimated slope β and the correlation coefficient r^2 are small. The small slopes suggest parameter surrogacy. The high degree of scatter about the lines of s versus $\underline{s}$ best fit indicates a large null space contribution to the uncertainties of both of these predictions. For the prediction of minimum travel time $r^2 = 0.00$; this demonstrates that the model calibration process does not constrain the uncertainty of this prediction at all, because of its high dependence on null space components. null space dependency of the recharge area prediction is similarly high. Of interest is the fact that the simple model shows a wider predictive range than the complex model for both of these
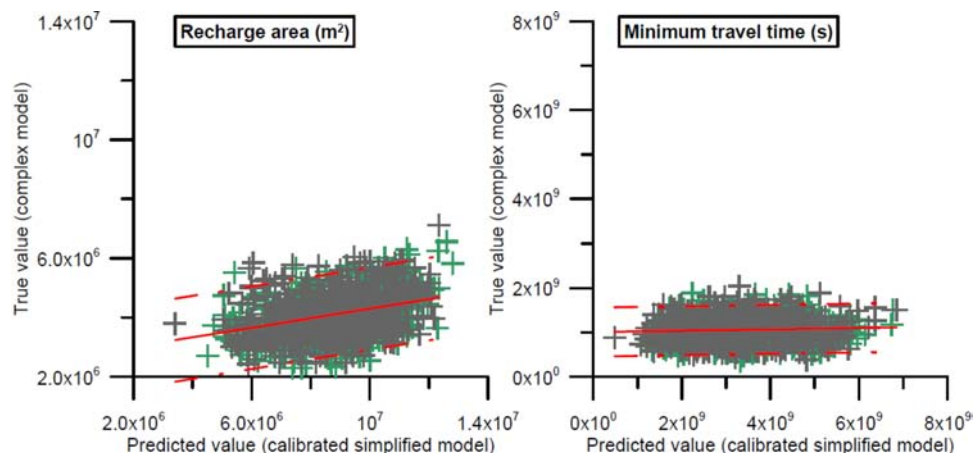

Figure 8. The paired model predictions of the recharge area of the pumping well and minimum travel time to pumping well. For each prediction 483 of the points (gray) were used to estimate the regression line and the 95% prediction interval; the other 500 points (green) plot similarly and were used for testing.

predictions. This is an outcome of parameter surrogacy. The use of paired models allows detection and correction for this effect, at the same time as it allows calculation of the true predictive interval while evincing its heavy null space dependency.

6. Discussion and Conclusions

[142] This paper has presented and demonstrated a methodology for model deployment in contexts where decisions pertaining to environmental management must be based on model outcomes. This requires that the uncertainties associated with model predictions be quantified, at the same time as it requires that these uncertainties be reduced to their theoretical minimum. The methodology is based on the use of paired models. One of these models is complex, and therefore provides appropriate receptacles for information acquired through direct measurements of system properties and/or for geological inference based on expert knowledge. The other is simple, and provides a means of extracting maximum information from historical measurements of system state. When used in concert the models provide a mechanism for making predictions of quantifiable uncertainty whose bias and error variance are reduced as much as possible. While the methodology requires some computational cost for its implementation, it does overcome many of the problems associated with traditional model usage in the environmental management context. In particular, the high run times and numerical granularity that often attend large models do not shackle the processes of model calibration and predictive uncertainty analysis. At the same time, all sources of information are granted access to the model parameterization process.

[143] In designing and implementing the methodology discussed herein, we have assumed that the processes and construction details of the complex model approximate those of reality. It is obvious that this will not always be the case. Indeed, even the most complex model is quite simple compared to reality itself. In spite of this, a modeler can only do his or her best. The methodology proposed above allows the complex model to represent more complexity than it otherwise could, for it does not need to be calibrated. Hence long run times and granularity associated with numerical representation of its processes are not the same issue that they would be if calibration of the complex model itself was required. Nevertheless, the less-than-perfect nature of a complex model, and its consequential failure to represent all nuances of system behavior, may indeed result in some degree of underestimation of predictive uncertainty. This, unfortunately, is unavoidable.

[144] We have also assumed that the complex model represents environmental processes correctly, and that it simulates processes which actually prevail in the real world rather than those which are mistakenly believed to prevail. This assumption too is unavoidable, and is required for any form of model-based data analysis. The methodology presented herein does, however, allow the modeler to include in his or her conceptualization of uncertainty the use of a variety of fundamentally different design concepts in the stochastic representation of the real world that comprises the complex model domain and processes that operate within it. In fact, lack of knowledge of some of the details of reality as it prevails at a particular study site demands such representation in the stochastic complex model generation process if predictive uncertainty analysis is to have integrity. The methodology discussed herein imposes no limitations on the nature of complex model stochasticity. In doing so it presents a mechanism for inquiry into the effects that the lack of such knowledge has on predictions of management interest. For some predictions the effects of large knowledge gaps on null space uncertainty and parameter surrogacy may be profound. For other predictions, particularly those that have a high solution space dependency, the repercussions may not be as great.

[145] As well as explaining and demonstrating the use of paired models, the theory and practical example presented in this paper expose many important features of modeling in general. These include the following.

[146] 1. In spite of the surrogate roles that some parameters are condemned to assume as they are adjusted through the calibration process, some model predictions can be made with a remarkably high level of precision. These are predictions that depend on parameter combinations that are estimable through the calibration process. Any surrogate role that these parameter combinations play affects these predictions in a similar manner to that in which it affects model outputs employed in the calibration process. For these predictions the calibration process is therefore "self-correcting."

[147] 2. Predictions that are significantly different in character from model outputs used in the calibration process may possess significant uncertainty, this arising from their dependence on null space parameter components (i.e., parameter combinations that are uninformed by the calibration process). Unless the null space is awarded explicit representation in the modeling process (either in a stand-alone simple model or in a complementary complex model), the uncertainty associated with such predictions will be underestimated.

[148] 3. Calibration of a simplified model can induce bias of unknown size in certain model predictions, this arising from spurious null space components that are entrained with solution space components, and then unwittingly adjusted to nonzero values during calibration of the simple model. If model simplification is undertaken carefully, it may be possible to reduce the propensity for such bias. However it is unlikely that any model will be completely immune from this problem. To the extent that predictions of future system behavior are different from model outputs employed in the calibration process, the propensity for such bias will be higher. Fortunately, this bias can be recognized and corrected through implementation of the paired model methodology demonstrated herein.

[149] 4. Ideally, model simplification should take place in the context of the data set against which the model must be calibrated. If possible, the simplification process should respect subspaces determined by the calibration data set, thereby reducing the propensity for predictive bias.

[150] While it may be possible (and desirable) to tune model simplification to the particular modeling and data context with which a modeler is faced, there will be many occasions where a modeler's options are limited. Unfortunately, environmental management often demands that complex, heterogeneous systems be simulated; simultaneously, it often demands that model parameterization be based on a combination of sporadic and irregular (in space

and time) measurements of system properties, supplemented by equally sporadic and irregular measurements of system states comprising a calibration data set. In such circumstances the case for complexity is strong. So too, is the case for simplicity. The methodology described herein provides a means whereby the benefits of both are accessible to the modeling process. The outcome of this process then becomes a prediction of minimized (though not necessarily small) error variance whose uncertainty is quantified. Furthermore, this is achieved at manageable computational cost. These outcomes are essential requirements for model-based decision-making.

[151] We conclude this paper with a few remarks on some broader modeling issues exposed by our study. The deleterious effects of parameter surrogacy on predictions which are of different types, and/or are made at different locations, from those which comprise the calibration data set have been discussed. The fact that such predictions may be associated with a high degree of uncertainty due to their dependence on null space parameter components is well-documented. Such uncertainty can be explored using, for example, methods described by *Tonkin and Doherty* [2009]. However plots such as those depicted in Figure 8 are particularly disturbing, for they exhibit a larger degree of horizontal scatter than vertical scatter. Hence the range of predictive variability exhibited by the sequence of calibrated simple models is greater than that exhibited by the sequence of uncalibrated complex models. Parameter surrogacy, and consequential predictive bias, thus profoundly affects the ability of the simple model used in our study to make predictions of this type, notwithstanding the relatively complex nature of that model. The propensity for predictive error has therefore actually been increased, rather than decreased, as an outcome of model calibration. An obvious question then follows. Given that all models are simplifications of reality, on how many occasions of real-world model usage, where predictions have been made by relatively complex models, has the calibration of such models eroded rather than enhanced their predictive ability?

[152] Further philosophical questions follow. It was shown in section 2 that if a prediction is highly solution space dependent, the calibration process cannot help but reduce the propensity for error in this prediction, regardless of model defects. In this case parameter surrogacy is not an issue, even though it may affect some or many model parameters a great deal. In fact, for this kind of prediction, parameter surrogacy may underpin predictive accuracy, for the adjustment of parameters through the calibration process allows them to assimilate the information content of the calibration data set. This then allows better definition of the parameter solution space, even though parameters may take on values that are unrealistic when considered in terms of their names and user-assigned roles. This questions the use of expert-knowledge-based prior parameter information in the simple model calibration process. For predictions such as these, the imposition of constraints imposed by prior information may actually hamper the ability of the calibration process to reduce predictive error. In contrast, the use of prior information in the simple model calibration process would probably reduce the propensity for bias in predictions such as those depicted in Figure 8 which exhibit lower solution space dependency. It is thus apparent that when recognition is given to the imperfect nature of any

model, the role of the calibration process, and the extent to which calibration and reality constraints should be imposed on model parameters, becomes very unclear. When a model is imperfect, the link between parameters of minimum error variance and predictions of minimum error variance is broken. Recognition of this strongly suggests that the same (necessarily imperfect) model may need to be calibrated in different ways, according to different calibration philosophies, in order to optimize its ability to make predictions of different types.

References

Aanonsen, S. I. (2008), Efficient history matching using a multiscale technique, *SPE Reservoir Eval. Eng.*, *11*(1), 154–164.

Aanonsen, S. I., and D. Eydinov (2006), A multiscale method for distributed parameter estimation with application to reservoir history matching, *Comput. Geosci.*, *10*, 97–117, doi:10.1007/s10596-005-9012-4.

Aster, R. C., B. Borchers, and C. H. Thurber (2005), *Parameter Estimation and Inverse Problems*, 301301 pp., Elsevier, Amsterdam, Netherlands.

Beven, K. J. (2002), Uncertainty and the detection of structural change in models of environmental systems, in *Environmental Foresight and Models: A Manifesto*, edited by M. B. Beck, pp. 227–250, Elsevier, Oxford, U. K.

Carle, S. F. (1999), *T-PROGS: Transition Probability Geostatistical Software—Version 2.1.*, 7777 pp., Univ. California, Davis.

Clement, T. P. (2011), Complexities in hindcasting models: When should we say enough is enough, *Ground Water*, *49*(5), 620–629, doi:10.1111/j.1745-6584.2010.00765.x.

Conti, S., J. P. Gosling, J. E. Oakley, and A. O'Hagan (2009), Gaussian process emulation of dynamic computer codes, *Biometrika*, *96*, 663–676.

Cooley, R. L. (2004), A theory for modeling ground-water flow in heterogeneous media, *U.S. Geol. Surv. Prof. Pap. 1679*, 220220 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*(5), 639–656.

Deutsch, C., and A. Journel (1998), GSLIB Geostatistical Software Library and User's Guide, 2nd ed, 368368 pp., Oxford Univ. Press, N. Y.

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

Doherty, J. (2011a), PEST: Model-independent parameter estimation, user manual, Watermark Numerical Computing, available at http://www.pesthomepage.org/Downloads.php.

Doherty, J. (2011b), Modeling: Picture perfect or abstract art, *Ground Water*, *49*(4), 455–456.

Draper, N. R., and H. Smith (1998), *Applied Regression Analysis*, 3rd ed., 706706 pp., John Wiley, N. Y.

Farmer, C. L. (2002), Upscaling: A review, *Int. J. Numer. Meth. Fluids*, *40*, 63–78, doi:10.1002/fld.267.

Harbaugh, A. W., E. R. Banta, M. C. Hill, and M. G. McDonald (2000), MODFLOW-2000, the U.S. Geological Survey modular ground-water model: User guide to modularization concepts and the Ground-Water Flow Process, *U.S. Geol. Surv. Open-File Rep. 00-92*, 121 p.

Harrar, W. G., T. O. Sonnenborg, and H. J. Henriksen (2003), Capture zone, travel time, and solute transport model predictions using inverse modeling and different geological models, *Hydrogeol. J.*, *11*(5), 536–547.

Hojberg A. L., and J. C. Refsgaard (2005). Model uncertainty-parameter uncertainty versus conceptual models, *Water Sci. Technol.*, *52*(6), 177–186.

Kennedy, M. C., and A. O'Hagan (2000), Predicting the output from a complex computer code when fast approximations are available, *Biometrika*, *87*(1), 1–13.

Koch, K. R. (1987), *Parameter Estimation and Hypothesis Testing in Linear Models*, 378 pp., Springer, Berlin, Germany.

Lee, D. B. (1973), Requiem for large-scale models, *J. Am. I. Planners 39*, 163–178.

Lieberman, C., K. Willcox, and O. Ghattas (2010), Parameter and state model reduction for large-scale statistical inverse problems, *SIAM J. Sci. Comput.*, *32*(5), 2523–2542.

Lødøen, O. P., and H. Omre (2008), Scale-corrected ensemble Kalman filtering applied to production-history conditioning in reservoir evaluation, *SPE J.*, *13*(2), 177–194.