Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques

Matthew Tonkin[1,2] and John Doherty[3]

Received 16 November 2007; revised 31 May 2008; accepted 5 November 2008; published 15 January 2009.

[1] We describe a subspace Monte Carlo (SSMC) technique that reduces the burden of calibration-constrained Monte Carlo when undertaken with highly parameterized models. When Monte Carlo methods are used to evaluate the uncertainty in model outputs, ensuring that parameter realizations reproduce the calibration data requires many model runs to condition each realization. In the new SSMC approach, the model is first calibrated using a subspace regularization method, ideally the hybrid Tikhonov-TSVD ''superparameter'' approach described by Tonkin and Doherty (2005). Sensitivities calculated with the calibrated model are used to define the calibration null-space, which is spanned by parameter combinations that have no effect on simulated equivalents to available observations. Next, a stochastic parameter generator is used to produce parameter realizations, and for each a difference is formed between the stochastic parameters and the calibrated parameters. This difference is projected onto the calibration null-space and added to the calibrated parameters. If the model is no longer calibrated, parameter combinations that span the calibration solution space are reestimated while retaining the null-space projected parameter differences as additive values. The recalibration can often be undertaken using existing sensitivities, so that conditioning requires only a small number of model runs. Using synthetic and real-world model applications we demonstrate that the SSMC approach is general (it is not limited to any particular model or any particular parameterization scheme) and that it can rapidly produce a large number of conditioned parameter sets.

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

1. Introduction

[2] If a model accurately represents processes relevant to the simulated system, errors in simulated predictions depend on parameter detail that is not represented in the model, and/or is not accurately inferred through calibration. In recognition of this, research has been undertaken to develop approaches for evaluating the potential error associated with model outputs. *Tonkin et al.* [2007] describe two broad methodological approaches as predictive error variance analysis (PEVA), and predictive uncertainty analysis (PUA). Some benefits and shortfalls of each approach are now briefly summarized.

[3] Linear and nonlinear PEVA evaluate the potential error in predictions made by a calibrated model using methods based upon variance propagation [e.g., *Bard*, 1974; *Seber and Wild*, 1989]. Since PEVA evaluates the error in predictions made by a calibrated model, some form of regularization is used to obtain a unique solution to the inverse problem. This is usually a form of parameter parsimony. Therefore, PEVA is usually undertaken using a small number of parameters. Nonlinear PEVA is formulated as a constrained optimization problem in which a single prediction is maximized or minimized subject to the constraints of maintaining the model in a calibrated state at a certain level of confidence [*Vecchia and Cooley*, 1987]. *Tonkin et al.* [2007] present a method for undertaking nonlinear PEVA, based on the method described by *Vecchia and Cooley* [1987] but extended to highly parameterized models in which mathematical regularized inversion methods are used to estimate parameters. In that approach, there is (notionally) no limit to the number of parameters that can be included in the analysis.

[4] Nonlinear PEVA possesses efficiencies that stem from its basis in Gauss-Newton techniques [e.g., *Cooley and Naff*, 1990], and can be effective when calculating confidence intervals for a single prediction. However, nonlinear PEVA must be undertaken independently to investigate different model predictions, so that efficiencies achieved from the analysis of a single prediction are eliminated by the necessity of undertaking as many constrained minimizations/maximizations as there are predictions.

[5] PUA is a more intrinsic concept that explicitly acknowledges that many parameter sets enable the model to reproduce available observations within confidence limits determined by the error associated with measurements of system state and the innate variability of system properties. In principle, no single calibrated model is identified. In-

[1]S. S. Papadopulos and Associates, Bethesda, Maryland, USA.
[2]Also at School of Engineering, University of Queensland, Brisbane, Queensland, Australia.
[3]Watermark Numerical Computing, Corinda, Queensland, Australia.

0043-1397/09/2007WR006678$09.00

stead, a suite of parameter realizations is generated. PUA techniques include calibration-constrained Monte Carlo, Markov Chain Monte Carlo (MCMC), and other methods that propagate prior stochastic parameter descriptions through a model [e.g., *Kitanidis*, 1996; *Yeh et al.*, 1996; *Oliver et al.*, 1997; *Kuczera and Parent*, 1998; *Woodbury and Ulrych*, 2000]; the method of stochastic equations [*Rubin and Dagan*, 1987; *Guadagnini and Neuman*, 1999; *Hernandez et al.*, 2006]; Generalized Likelihood Uncertainty Estimation (GLUE) [*Beven and Binley*, 1992]; and deformation techniques [e.g., *Lavenue and de Marsily*, 2001; *Gómez-Hernandez et al.*, 1997, 2003].

[6] MC-based PUA is appealing since it does not typically rest upon assumptions of model (quasi-) linearity; and, once a suite of parameter realizations has been obtained, this ensemble can be used to evaluate the uncertainty associated with *any* model output. However, the application of MC-based PUA can be complex and computationally intensive when honoring calibration constraints on parameters. In particular, calibration-constrained MC-based PUA can be onerous when forward model run times are long and/or when a large number of parameters are included in the analysis.

[7] In one comparative analysis, *Gallagher and Doherty* [2007] determined that while nonlinear, calibration-constrained PEVA was more efficient than MCMC PUA when examining the range of uncertainty associated with a single model prediction, this relative efficiency diminished when the uncertainty of many predictions was evaluated, and/or where there was a need to examine the statistical relationships between multiple model predictions. This suggests that the desirable benefits of MC-based PUA to evaluate multiple predictions and their statistical relationships could be capitalized upon if the computational efficiency could be improved upon.

[8] In this paper we describe a new subspace Monte Carlo (SSMC) technique that enables efficient evaluation of the range of error associated with outputs from highly parameterized models. The SSMC technique is founded in error variance analysis theory: however, it incorporates several developments that render it akin to deformational MC techniques. In the new SSMC technique, the model is calibrated using a subspace regularization technique such as Truncated Singular Value Decomposition (TSVD) or the hybrid Tikhonov-TSVD superparameter approach. Using sensitivities calculated with the calibrated model, the calibration null-space is defined, which is spanned by parameter combinations that (if the model were linear) have negligible effect on simulated equivalents to available observations [*Tonkin and Doherty*, 2005]. Next, a stochastic parameter generator is used to produce multiple realizations of the model parameters. For each realization, a difference is formed between the stochastic parameters and the calibrated parameters. This difference is then projected onto the calibration null-space, added to the calibrated parameters, and the model executed.

[9] If the model is no longer calibrated, parameter combinations that comprise the calibration solution space, which is orthogonal to the calibration null-space, are reestimated with the null-space-projected parameter differences retained as additive values. This recalibration can often be undertaken using superparameters constructed on the basis of existing sensitivities, so that conditioning may require only a small number of model runs. It is also demonstrated that the SSMC technique enables the inclusion of fine (e.g., model grid) scale parameterization even when the calibration is undertaken using parameterization devices such as zones or pilot points.

[10] This paper is structured as follows. First, the theory of regularized inversion using subspace techniques is presented. Equations are presented that describe (1) how calibration can be formulated as subspace-based regularized inversion, (2) how postcalibration parameter error variance is calculated, (3) the projection of stochastic parameter values onto the calibration null-space, and finally, (4) Broyden's rank-one update [*Broyden*, 1965] and how this can benefit the SSMC technique. Following this, the SSMC analysis procedure is described fully, together with an introduction to methods for incorporating fine (e.g., model grid scale) detail that was not included in the calibration but that honors spatiotemporal parameter statistics. Assumptions that underlie the SSMC technique, and their possible repercussions, are then discussed. The SSMC technique is then demonstrated by application to a synthetic groundwater model, and to a real-world watershed model. These applications demonstrate that the SSMC approach is general (that is, it is not limited to any particular model or to any particular parameterization device) and that it can rapidly produce a large number of conditioned parameter sets for use in assessing the uncertainty in a variety of model outputs. Results of these analyses are discussed together with concluding remarks.

[11] To our knowledge this is the first presentation of a general method for efficiently evaluating the range of model outputs for any model employing any parameterization technique. Fundamental to the method is the use of a large number of parameters, and the use of subspace techniques. Use of a large number of parameters may allow property detail to be represented in a model at a scale approaching true variability of these properties (or at least a scale that approaches that to which predictions of interest are sensitive); at the same time, the use of subspace techniques allows the analysis to be undertaken efficiently as an adjunct to calibration undertaken using regularized inversion.

2. Theory

2.1. Parameter Estimation

[12] For linear models, the relationship between the n model parameters, $\mathbf{p}$, and the m model outputs, $\mathbf{y}$, for which there are corresponding measurements can be represented in matrix form as:

$$\mathbf{y} = \mathbf{X}\mathbf{p} \tag{1}$$

where the m-row by n-column (m x n) matrix $\mathbf{X}$ contains the sensitivity of the simulated equivalent of each measurement with respect to each parameter. Equation (1) describes the forward action of the model. If the model is a perfect simulator of reality, differences between measurement data and their simulated equivalents are solely a function of the error associated with each measurement, i.e.,

$$\mathbf{y} = \mathbf{X}\mathbf{p} + \varepsilon \tag{2}$$

where ε characterizes measurement error. A weighted sum-of-squared residuals objective function, Φ, can be defined that is a measure of the fit between simulated values and the measured data that are being fit through the calibration process:

$$\Phi = \left(\mathbf{y} - \mathbf{X}\underline{\mathbf{p}}\right)^{t}\mathbf{Q}\left(\mathbf{y} - \mathbf{X}\underline{\mathbf{p}}\right) \quad (3)$$

where the vector $\underline{\mathbf{p}}$ lists current parameter values, the superscript t indicates the matrix transpose, and the m x m weight matrix $\mathbf{Q}$ is ideally proportional to the inverse of the covariance matrix of measurement errors [*Bard*, 1974]. Other objective functions exist; however the L_2 norm has desirable qualities including the ability to assess parameter and predictive error [*Bard*, 1974].

[13] For a linear model, the parameter values that minimize Φ are obtained from:

$$\underline{\mathbf{p}} = \left(\mathbf{X}^{t}\mathbf{Q}\mathbf{X}\right)^{-1}\mathbf{X}^{t}\mathbf{Q}\mathbf{y} \quad (4)$$

where the superscript -1 indicates the matrix inverse. For nonlinear models $\underline{\mathbf{p}}$ cannot be calculated directly from $\mathbf{y}$ using (4). Instead, a parameter upgrade or change vector, $\Delta\underline{\mathbf{p}}$, is calculated from the residual vector, $\mathbf{r}$ as:

$$\Delta\underline{\mathbf{p}} = \left(\mathbf{X}^{t}\mathbf{Q}\mathbf{X}\right)^{-1}\mathbf{X}^{t}\mathbf{Q}\mathbf{r} \quad (5)$$

Equations (1)–(5) underpin the Gauss-Newton method of nonlinear parameter estimation [*Bard*, 1974]. Its use is predicated on the assumption that an inverse of the n x n matrix $\mathbf{X}^{t}\mathbf{Q}\mathbf{X}$ exists, and hence that the inverse problem includes a small enough number of parameters to be uniquely estimable. Regression performance is commonly improved by supplementing equation (5) with devices such as the (Levenberg-) Marquardt parameter [*Seber and Wild*, 1989].

2.2. Mathematical Regularization

[14] Where there are more parameters than can be estimated uniquely, $\mathbf{X}^{t}\mathbf{Q}\mathbf{X}$ is not directly invertible, and many different parameter sets may lead to an equally low objective function. The goal of calibration is then to obtain a solution to the inverse problem for which the parameters possess the most desirable properties. Suppose that this set of parameters is calculated from observations using matrix $\mathbf{G}$, such that:

$$\underline{\mathbf{p}} = \mathbf{G}\mathbf{y} \quad (6)$$

A variety of methods have been developed to obtain a suitable $\mathbf{G}$. *Tonkin and Doherty* [2005] contrast methods based on Tikhonov regularization [*Tikhonov and Arsenin*, 1977] with those that employ TSVD [e.g., *Aster et al.*, 2005]. In these cases the determination of $\mathbf{G}$ for use in equation (6) requires that $\mathbf{X}^{t}\mathbf{Q}\mathbf{X}$ of equation (5) be replaced, respectively, by the following:

$$\mathbf{X}^{t}\mathbf{Q}\mathbf{X} \cong \left(\mathbf{X}^{t}\mathbf{Q}\mathbf{X} + \beta^{2}\mathbf{T}^{t}\mathbf{S}\mathbf{T}\right) \quad (7)$$

$$\mathbf{X}^{t}\mathbf{Q}\mathbf{X} \cong \left(\mathbf{V}_1\mathbf{E}_1\mathbf{V}_1^{t}\right) \quad (8)$$

where: $\mathbf{T}$ is a matrix of Tikhonov regularization constraints; $\mathbf{S}$ is the regularization weight matrix; β^2 is the regularization weight factor; and $\mathbf{V}_1$ and $\mathbf{E}_1$ are obtained through singular value decomposition of $\mathbf{X}^{t}\mathbf{Q}\mathbf{X}$ such that:

$$\mathbf{X}^{t}\mathbf{Q}\mathbf{X} = [\mathbf{V}_1\mathbf{V}_2]\begin{bmatrix}\mathbf{E}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{E}_2\end{bmatrix}[\mathbf{V}_1\mathbf{V}_2]^{t} = \mathbf{V}_1\mathbf{E}_1\mathbf{V}_1^{t} + \mathbf{V}_2\mathbf{E}_2\mathbf{V}_2^{t} \quad (9)$$

where: ns is the number of pretruncation eigenvalues; $\mathbf{E}_1$ is a diagonal matrix listing the ns pretruncation eigenvalues; $\mathbf{E}_2$ is a diagonal matrix listing the $n - ns$ posttruncation eigenvalues; $\mathbf{V}_1$ contains the ns orthogonal unit eigenvectors corresponding to $\mathbf{E}_1$; and, $\mathbf{V}_2$ contains the $n - ns$ orthogonal unit eigenvectors corresponding to $\mathbf{E}_2$.

[15] *Tonkin and Doherty* [2005] also describe a hybrid Tikhonov-TSVD inversion technique developed to combine the benefits of each of these methods. That technique is based upon constructing a "base" parameterization comprising a large number of parameters, calculating base parameter sensitivities, and using TSVD to decompose the base parameter matrix $\mathbf{X}^{t}\mathbf{Q}\mathbf{X}$ according to equation (9). This decomposition is used to construct superparameters, where superparameters are factors by which principal eigenvectors of the base parameter matrix $\mathbf{X}^{t}\mathbf{Q}\mathbf{X}$, i.e., the ns eigenvectors $\mathbf{V}_1$, are scaled to minimize a least squares objective function. Linear combinations of base parameters that correspond with singular values falling below the TSVD truncation threshold, i.e., the $n - ns$ eigenvectors $\mathbf{V}_2$, are not estimated. Because the number of superparameters may be far less than the number of base parameters, the computational burden for solution of the reformulated inverse problem is greatly reduced.

[16] To implement Tikhonov regularization in the hybrid technique, Tikhonov equations are developed for the base parameters and encapsulated in matrix $\mathbf{T}$, as described above. For the Tikhonov equations to be augmented in the solution to the reformulated problem, matrix $\mathbf{T}$ must first be postmultiplied by $\mathbf{V}^{1}$ and equation (5) can be reformulated as follows:

$$\mathbf{D} = \mathbf{T}\mathbf{V}^{1} \quad (10a)$$

$$\mathbf{X}^{t}\mathbf{Q}\mathbf{X} \cong \mathbf{V}_1\left(\mathbf{Z}^{t}\mathbf{Q}\mathbf{Z} + \beta^{2}\mathbf{D}^{t}\mathbf{S}\mathbf{D}\right) \quad (10b)$$

where $\mathbf{V}_1$ is obtained through TSVD of $\mathbf{X}^{t}\mathbf{Q}\mathbf{X}$ as described above, and the m x ns matrix $\mathbf{Z}$ contains the calculated sensitivities of model outputs with respect to superparameters [see *Tonkin and Doherty*, 2005, equation 12]. Since the appropriate value of β^2 is unknown at the outset, equation (10) is formulated as a constrained minimization, in which the reciprocal of β^2 performs a role to similar to that of the Lagrange multiplier [*Haber*, 1997; *Tonkin and Doherty*, 2005].

[17] As noted by *Tonkin et al.* [2007], the division of parameter space into solution and null-spaces is approximate, since parameter values (and, more importantly, sensitivities upon which the TSVD is based) change throughout the calibration. However, as described later, the recalibration step of the SSMC technique produces numerous parameter combinations that are calibrated to a level com-

patible with measurement noise, even if the relationship between solution and null-spaces is compromised by nonlinearity.

[18] For pure TSVD and the hybrid Tiknonov-TSVD technique respectively, $\mathbf{G}$ is described by:

$$\mathbf{G} = \left(\mathbf{V}_1\mathbf{E}_1\mathbf{V}_1^t\right)\mathbf{X}^t\mathbf{Q} \tag{11}$$

$$\mathbf{G} = \mathbf{V}_1\left(\mathbf{Z}^t\mathbf{Q}\mathbf{Z} + \beta^2\mathbf{D}^t\mathbf{S}\mathbf{D}\right)^{-1}\mathbf{Z}^t\mathbf{Q} \tag{12}$$

Postmultiplication of $\mathbf{G}$ by $\mathbf{X}$ gives the n x n model resolution matrix, $\mathbf{R}$ [e.g., *Menke*, 1989]. Combining equations (2) and (6) gives the following relationship:

$$\underline{\mathbf{p}} = \mathbf{G}\mathbf{X}\mathbf{p} + \mathbf{G}\boldsymbol{\varepsilon} \tag{13}$$

which since $\mathbf{R} = \mathbf{G}\mathbf{X}$ is equivalent to:

$$\underline{\mathbf{p}} = \mathbf{R}\mathbf{p} + \mathbf{G}\boldsymbol{\varepsilon} \tag{14}$$

The resolution matrix illustrates that the estimated value of a single parameter is in fact a weighted average of many true parameters. For overdetermined parameter estimation, each resolving kernel (i.e., each row of $\mathbf{R}$) contains a single nonzero element, of value 1, so that $\mathbf{R}$ is equal to the identity matrix, $\mathbf{I}$. However, in the underdetermined contexts described in this study, the resolving kernel comprising each row of $\mathbf{R}$ possesses a spread of nonzero elements that expresses the averaging relationship through which an element of $\underline{\mathbf{p}}$ is related to the elements of $\mathbf{p}$ [*Menke*, 1989].

2.3. Parameter Error

[19] Equation (14) expresses, in a linear sense, how well parameters $\mathbf{p}$ can be resolved by the estimated parameters, $\underline{\mathbf{p}}$. From equation (14), parameter error, $\mathbf{p} - \underline{\mathbf{p}}$, is given by:

$$\mathbf{p} - \underline{\mathbf{p}} = (\mathbf{I} - \mathbf{R})\mathbf{p} - \mathbf{G}\boldsymbol{\varepsilon} \tag{15}$$

Since $\mathbf{p}$ is unknown, equation (15) cannot be used directly. However, the statistical properties of parameter error can be evaluated on the basis of knowledge of the statistical properties of real-world hydraulic properties, expressed by the covariance matrix $\mathrm{C}(\mathbf{p})$, together with the statistical properties of measurement noise, expressed by the covariance matrix $\mathrm{C}(\boldsymbol{\varepsilon})$. The covariance matrix of parameter error, $\mathrm{C}(\underline{\mathbf{p}}-\mathbf{p})$, can then be derived from equation (15) as:

$$\mathrm{C}\left(\mathbf{p} - \underline{\mathbf{p}}\right) = (\mathbf{I} - \mathbf{R})\mathrm{C}(\mathbf{p})(\mathbf{I} - \mathbf{R})^t + \mathbf{G}\mathrm{C}(\boldsymbol{\varepsilon})\mathbf{G}^t \tag{16}$$

Detailed discussion of equation (16) is given by *Moore and Doherty* [2005]. Recalling the derivation of the terms of equations (14) and (15), it is clear that equation (16) expresses contributions to parameter error arising from (1) the inability to estimate parameter combinations that are insensitive and therefore fall below the TSVD truncation limit and (2) contamination by measurement noise, $\boldsymbol{\varepsilon}$, of the estimates of parameter combinations that lie above the TSVD truncation limit. That is, the first term of equation (16) expresses the contribution to parameter error from the calibration null-space, spanned by the columns of $\mathbf{V}_2$ of equation (9); while the second term of equation (16) expresses the contribution to parameter error due to errors in the estimates of parameter combinations occupying the calibration solution space, spanned by the columns of $\mathbf{V}_1$ of equation (9), stemming from measurement noise. In the absence of Tikhonov regularization, the division between calibration solution and null-spaces is exact, so that the two terms on the right side of (16) are orthogonal; in the presence of Tikhonov regularization, this division is approximate.

[20] *Tonkin et al.* [2007] demonstrate how equation (16) can be used to develop a calibration-constrained PEVA method for use in underdetermined problems. Although this approach can be efficient, it can only investigate the error variance associated with one prediction at a time. If the intent is to evaluate the uncertainty associated with numerous predictions, an efficient Monte Carlo approach that produces an ensemble of calibration-constrained parameter sets would be more suitable for analysis of calibration-constrained variability. The new SSMC approach is now described.

2.4. Calibration-Constrained Subspace Monte Carlo

[21] If $\mathrm{C}(\mathbf{p})$ of equation (16) is constructed to characterize hydrologic properties and their spatiotemporal variation, random realizations of parameter values can be generated from this matrix. However, when the model is executed with these realizations, it is generally the case that very few parameter sets result in an acceptable value for Φ since parameters generated directly from $\mathrm{C}(\mathbf{p})$ are unconstrained by measurements of system state, such as groundwater elevations.

[22] As described above, the first term of equation (16) arises because parameter combinations that lie within the calibration null-space are inaccessible to the calibration since they do not affect simulated equivalents to measured values. In the case that TSVD is used to solve the inverse problem:

$$\mathbf{I} - \mathbf{R} = \mathbf{V}_2\mathbf{V}_2^t \tag{17}$$

where $\mathbf{V}_2$ contains the posttruncation eigenvectors obtained from TSVD of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$. Using (17), differences between stochastic parameter realizations generated from $\mathrm{C}(\mathbf{p})$ and the calibrated parameter values, which are constrained by available measurements, can be projected onto the calibration null-space using:

$$\left(\underline{\mathbf{p}} - \underline{\mathbf{p}}_{st}\right)' = \mathbf{V}_2\mathbf{V}_2^t\left(\underline{\mathbf{p}} - \underline{\mathbf{p}}_{st}\right) \tag{18}$$

where $\underline{\mathbf{p}}_{st}$ represents the stochastic parameter values; $(\underline{\mathbf{p}} - \underline{\mathbf{p}}_{st})'$ is the projected vector of parameter differences; and, $(\underline{\mathbf{p}} - \underline{\mathbf{p}}_{st})$ represents the vector of parameter differences prior to projection. Once calculated, $(\underline{\mathbf{p}} - \underline{\mathbf{p}}_{st})'$ can be added to the calibrated parameters to produce new parameter sets. Hence, the projection operation described by equation (18) removes components of the calculated parameter differences $(\underline{\mathbf{p}} - \underline{\mathbf{p}}_{st})$ that possess nonzero projections onto the calibration solution space, so that $(\underline{\mathbf{p}} - \underline{\mathbf{p}}_{st})'$ contains only those components of the calculated parameter differences that possess a zero projection onto the calibration solution space

(see *Golub and Van Loan* [1989] for discussion of matrix projections).

[23] If the model is linear and the TSVD truncation level coincides exactly with the occurrence of zero singular values, then every new parameter set derived by adding $(\underline{\mathbf{p}} - \underline{\mathbf{p}}_{st})'$ to the calibrated parameters would, by definition, produce the same calibrated objective function, since $(\underline{\mathbf{p}} - \underline{\mathbf{p}}_{st})'$ would have no effect on simulated equivalents to observation data. These conditions are rarely met since the model is typically at least weakly nonlinear, and some diminishingly small but nonzero singular values lie below the TSVD truncation level. As a consequence, each new set of parameters typically results in a (slightly) uncalibrated model. This misfit must be remedied through recalibration. In the context of groundwater modeling, recalibration could be based upon a deformational technique such as that described by *Lavenue and de Marsily* [2001]. However, as described by *LaVenue et al.* [1995] this, and related techniques, can be computationally intensive.

[24] Conditioning of each parameter set can, however, be undertaken efficiently by employing existing sensitivities calculated using the calibrated parameters. Further efficiencies can be achieved if the superparameter technique is used to estimate a small number of linear combinations of the base parameters, rather than estimating each base parameter individually. Redefining each parameter set in terms of superparameters, constructed on the basis of matrix $\mathbf{X}$ corresponding to the calibrated parameter values, possesses distinct advantages over conditioning on the basis of the individual base parameters, and obtaining parameter sensitivities independently for each stochastic parameter set.

[25] First, these superparameters span the calibration solution space, and therefore comprise the orthogonal complement of the null-space projected parameter differences that were generated as described above. Therefore, recalibration focuses on parameter combinations that lie within the calibration solution space, leading to efficient conditioning. Second, no model runs are required to obtain superparameter sensitivities during the first iteration of the conditioning process: these sensitivities are calculated from $\mathbf{X}$ and can be used directly within a modified Gauss-Newton optimization of superparameters. Third, the use of superparameters greatly reduces the dimensionality of the inverse problem that must be undertaken to condition each stochastic parameter set. Therefore, fewer model runs are required to calculate sensitivities if the first inverse iteration does not achieve an acceptable objective function so that additional conditioning iterations are required.

2.5. Incorporating Fine Detail

[26] The foregoing describes situations where the scale or detail level of the model parameterization corresponds with the true scale of parameter variability, that is, if no spatiotemporal averaging of model parameters takes place. This, however, is rarely the case. For example, in the context of groundwater models, a value for hydraulic conductivity is required for each model cell: however, calibration is usually undertaken on the basis of parameterization devices such as zones or pilot points.

[27] If calibration is undertaken using a parameterization scheme that combines or otherwise groups model inputs in any manner, the theory described above can still be applied. Under these circumstances, the following process is undertaken: (1) A stochastic parameter generator is used to generate parameter sets on the model grid or element scale, rather than using the same parameterization mechanism employed in the calibration process. (2) Interpolation is used to form the closest approximation to these stochastic parameters that is achievable using the same parameterization device that is used during the calibration. (3) These parameters are projected onto the calibration null-space using equation (18). (4) A difference is calculated between the stochastic parameters and the interpolated parameters calculated under step 2. (5) The values obtained from steps 3 and 4 are added to the calibrated parameters. (6) Recalibration is undertaken using superparameters as described above.

[28] This approach removes components of fine-scale stochastic variability generated under (1) that are incompatible with calibration data. This allows the calibration process to employ a lumped parameterization scheme, while leading to conditioned parameter sets that contain fine-scale variability that is characteristic of real-world property variability. The synthetic and real-world applications described later employ fine-scale stochastic parameter generation in combination with coarser-scale parameterization schemes.

[29] If the model were linear (or weakly nonlinear), recalibration of each parameter set generated using the techniques described above would require a single iteration to minimize the objective function. If solution-space sensitivities calculated using the calibrated parameters were employed, this would require only one model run. However, since nonlinearity causes parameter sensitivities to change as parameter values change, the required nonlinear parameter estimation is iterative in nature: at each iteration, $\Delta\underline{\mathbf{p}}$ is computed using (5) and parameter sensitivities are recalculated using currently estimated parameter values $\underline{\mathbf{p}}$. The process is then repeated. Note however that, as described above, superparameters are estimated, so that the burden of recalibration is comparatively light.

2.6. Use of the Broyden Rank-One Update

[30] Since the SSMC approach attempts to obtain the greatest reduction in the objective function, Φ, that can be achieved in one or perhaps two iterations, it is desirable to maximize the progress that can be made using existing sensitivities contained in the Jacobian matrix obtained during original calibration of the model, obviating the need to recalculate these sensitivities. Testing to date indicates that inclusion of the Broyden rank one update [*Broyden*, 1965; *Madsen and Nielsen*, 2004; *Skahill and Baggett*, 2006] within the SSMC procedure improves objective function minimization within the first, and, if necessary, any subsequent recalibration iterations.

[31] The Broyden rank one update is a technique for updating matrix $\mathbf{X}$ using information gained on any occasion that a model run is undertaken, for example, when testing the reductions in Φ computed using alternate values of the Marquardt parameter. It is particularly beneficial when it is expensive to compute $\mathbf{X}$ and/or when it is desirable to obtain as great a reduction in Φ as possible given one or perhaps two iterations. Indeed, *Dennis and Marwil* [1982] indicate that use of such an update may be most fruitful when the cost of obtaining a full new Jacobian is sufficiently high: as a result, they recommend that the use

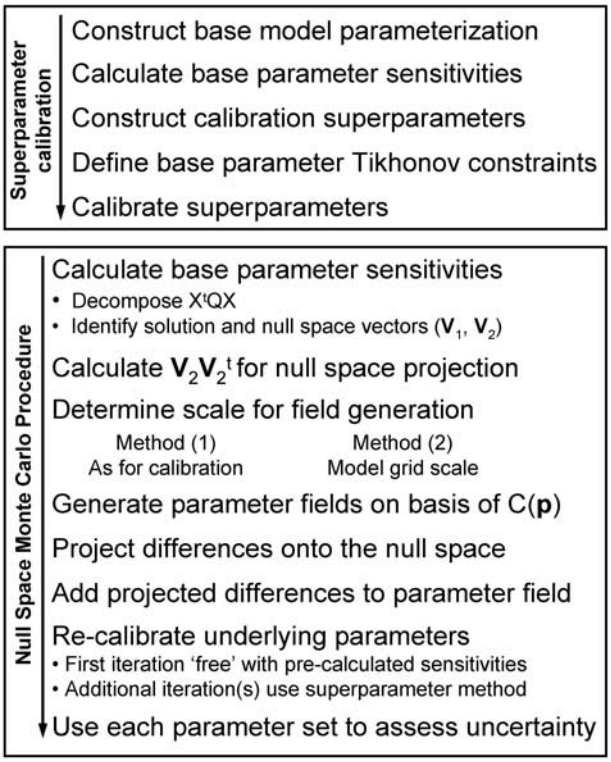

Figure 1. Schematic of the subspace Monte Carlo (SSMC) analysis procedure.

of such an update is most effective between reevaluations of $\mathbf{X}$.

[32] It is in this sense that the Broyden update is employed in the SSMC procedure: that is, the update is employed only when calculating the parameter upgrade vector $\Delta\underline{\mathbf{p}}$ that results from alternate values of the Marquardt parameter. In the case of the first iteration of the SSMC recalibration procedure there is (essentially) nothing to be lost, and everything to be gained in the use of the update since this iteration is undertaken for "free" using an existing Jacobian (Figure 1). If subsequent iterations are required, the trade-off between a rapid update versus full reevaluation of the Jacobian comes into play, and the potential benefits are liable to be problem-specific. Therefore, in the development of the SSMC approach, the ability to undertake the Broyden rank-one update was included with particular focus on the first recalibration iteration, and with the possibility that it may prove beneficial during later iterations, if required.

[33] In general application, where $\mathbf{X}$ is calculated using parameters $\underline{\mathbf{p}}$, and $\Delta\underline{\mathbf{p}}$ is calculated using equation (5) in conjunction with a value for the Marquardt parameter, then:

$$\Delta\underline{\mathbf{p}} = \underline{\mathbf{p}}_u - \underline{\mathbf{p}} \tag{19}$$

where $\underline{\mathbf{p}}_u$ is the trial parameter set. If $\mathbf{y}_u$ and $\mathbf{y}$ represent simulated equivalents of the measured values, calculated using $\underline{\mathbf{p}}_u$ and $\underline{\mathbf{p}}$ respectively, Broyden's rank one update is used to calculate an updated sensitivity matrix $\mathbf{X}_u$ using:

$$\mathbf{X}u = \mathbf{X} + \mathbf{w}\left(\Delta\underline{\mathbf{p}}\right)^t \tag{20}$$

where:

$$\mathbf{w} = \frac{1}{\Delta\underline{\mathbf{p}}^t\Delta\underline{\mathbf{p}}}\left(\mathbf{y}_u - \mathbf{y} - \mathbf{X}\Delta\underline{\mathbf{p}}\right) \tag{21}$$

$\mathbf{X}_u$ is then used to calculate subsequent parameter upgrades using alternative values of the Marquardt parameter. In the context of the SSMC procedure, the Jacobian matrix and the parameter upgrade vector pertain to superparameters: nonetheless, application of the Broyden rank one update to the estimation of superparameters directly parallels that described by equations (20) and (21) for classical Gauss-Newton parameter estimation.

2.7. Full SSMC Procedure

[34] The following narrative describes the stepwise application of the SSMC technique, illustrated in Figure 1, assuming that the superparameter technique is used for the initial calibration, and subsequent conditioning of stochastic parameters sets:

[35] 1. Construct the base model parameterization, calculate base parameter sensitivities, and construct and calibrate superparameters.

[36] 2. Calculate the base parameter sensitivity matrix, $\mathbf{X}$, using the calibrated parameters.

[37] 3. Undertake SVD of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ or $\mathbf{Q}^{1/2}\mathbf{X}$, to obtain $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{E}_1$, and $\mathbf{E}_2$.

[38] 4. Determine the dimensionality of the calibration solution space: that is, the number of columns in $\mathbf{V}_1$ and $\mathbf{V}_2$, respectively.

[39] 5. Determine the scale for stochastic parameter generation. Generate multiple sets of random parameters on the basis of C($\mathbf{p}$).

[40] 6. Calculate $\mathbf{V}_2\mathbf{V}_2^t$, and project differences between stochastic parameters and the calibrated parameters onto the calibration null-space.

[41] 7. Add these projected differences to calibrated parameter values.

[42] 8. Recalibrate the underlying superparameters. The first iteration is achieved using existing sensitivities calculated under step 2. Additional iterations require recalculation of superparameter sensitivities. Since there is usually a small number of superparameters, the calculation of superparameter sensitivities is efficient.

[43] In summary, the new SSMC technique is founded upon generating stochastic parameter sets that reflect the errors in estimated parameters $\underline{\mathbf{p}}$ that are described by the first term of equation (16). To the extent that addition of these values to the calibrated parameters decalibrates the model, solution space parameter projections are then adjusted to achieve an objective function that is commensurate with the characteristics of measurement noise as described by the second term of equation (16). Hence both terms of this equation are respected in the stochastic generation and recalibration processes. Recalibration is efficient because (1) the addition of null-space projected stochastic parameter differences to the calibrated parameters does not lead to substantial decalibration; (2) only the small number of superparameters requires estimation to recalibrate the model; and (3) for each stochastic parameter set, the first recalibration iteration is undertaken using existing solution space sensitivities calculated with the calibrated model.

2.8. Assumptions and Approximations

[44] Some assumptions and approximations involved in the new SSMC procedure are now summarized.

[45] First, when inversion is undertaken using subspace methods, the definition of the solution and null subspaces is approximate, since the parameters generated throughout the SSMC process deviate from the calibrated values. Hence, the null-space projection described by equation (18) becomes inexact. However, errors incurred through this approximation are mitigated since the recalibration that is undertaken to condition each stochastic parameter set is nonlinear. Furthermore in each case the SSMC technique results in numerous parameter sets that are calibrated in accordance with $C(\varepsilon)$.

[46] Second, the results of the analysis may differ depending on the definition of the dimensions of the calibration and null subspaces. If the TSVD truncation level is set too high, each stochastic parameter set may require many model executions to condition; whereas, if the truncation level is set too low, each stochastic parameter set may not require conditioning, but may lead to unrealistically narrow predictive variability. Testing using the synthetic model described later suggests that this trade-off can be refined using a sequentially increasing number of superparameters.

[47] Third, in the examples discussed later, recalibration is considered successful if a level of misfit is achieved that is commensurate with that achieved in the initial calibration. Rigorous theory for calculating objective function thresholds corresponding to various levels of calibration confidence is in fact available [*Vecchia and Cooley*, 1987; *Tonkin et al.*, 2007]. Although the theory exists, rigorous implementation of it can be difficult in practice since the stochastic characteristics of measurement noise and of structural noise induced through the parameterization scheme (and indeed by the model itself) are never known precisely and are computationally intensive to evaluate [see *Cooley*, 2004]. Hence, in the examples presented later this theory is not employed since formulation of a stochastic descriptor for errors associated with the various components of the multicomponent objective functions employed in them is beyond the scope of this article.

[48] Finally, in this study it is assumed that perturbation sensitivities are used throughout the SSMC analysis. However, the sensitivity equation method, adjoint techniques, and/or the method of automatic differentiation (AD) could be used. AD is a broad methodology for obtaining forward or reverse (adjoint) derivatives. AD exploits the fact that computer programs execute a sequence of arithmetic operations and/or functions, regardless of the complexity of the computer model, and that application of the chain rule of derivative calculus to these operations can be used to automatically compute derivatives [*Sambridge et al.*, 2007]. It should be noted, however, that the use of perturbation sensitivities does enable the SSMC method to be used with any model or sequence of models, without any specialized programming requirements.

3. Applications

[49] The methodology is first applied to a synthetic groundwater flow-and-transport model that represents the major features of a real-world site. It is then applied to a real-world lumped parameter watershed model, developed to simulate parts of the Dawson River catchment of Central Queensland, Australia.

[50] The purpose of describing the synthetic application is threefold. The first objective is to illustrate the steps undertaken in an SSMC analysis. The second objective is to contrast the results obtained using alternate Monte Carlo techniques with the SSMC procedure. The third objective is to contrast the efficiency of the SSMC recalibration process when conditioning parameters (1) to groundwater heads only and (2) to groundwater heads and contaminant concentrations. The purpose of describing the real-world application is twofold. The first objective is to illustrate that the SSMC approach is suitable for real world applications. The second objective is to demonstrate that the SSMC approach is not limited to any particular model or to any particular parameterization device.

3.1. Synthetic Groundwater Model

[51] The synthetic groundwater flow-and-transport model is based upon that described by *Tonkin et al.* [2007], and developed to evaluate the possible threat posed to a deep production well by a shallow source of contaminants. The question posed is "Could contaminants released to a shallow aquifer contaminate water recovered from a deep aquifer?"

[52] The flow system is characterized by two aquifers separated by a semiconfining unit. The system is simulated using MODFLOW-2000 [*Harbaugh et al.*, 2000], using two explicit layers separated by a confining bed. The logarithms (to base 10) of hydraulic conductivity in the aquifers and aquitard are generated using the sequential field generator SGSIM [*Deutsch and Journel*, 1992] using exponential variograms with a range of 150 m and sill of 0.15 for the aquifers, and a range of 200 m and sill of 0.3 for the aquitard. The vertical anisotropy within the aquifers is set at 1:10. Steady state constant transmissivity ("confined") flow is simulated, using general head boundaries up and down gradient, and no-flow lateral boundaries. A production well is located in the deep aquifer (Layer 2); see Figure 2.

[53] Using this hypothetical scenario, two model calibrations and subsequent SSMC analyses were undertaken: the first, based only on observations of groundwater head; the second, based on observations of groundwater head and contaminant concentration. The objective of the first SSMC analysis is to illustrate the steps undertaken in the SSMC analysis process, and contrast the results obtained using alternate Monte Carlo techniques with the new SSMC approach when only head data are available to constrain parameters. The objective of the second SSMC analysis is to contrast the results obtained using only head data with those obtained when head and concentration data are available to constrain parameters.

[54] Figure 2 depicts the layout of the synthetic model domain. For calibration purposes, hydraulic conductivity is parameterized using 792 pilot points, 264 for each aquifer and for the aquitard, distributed evenly throughout the model domain. In the pilot point approach, parameter values are estimated at a number of discrete locations (the pilot point locations), and assignment of parameter values throughout the model domain takes place through spatial interpolation from the pilot points to the model nodes [*Certes and de Marsily*, 1991; *RamaRoa et al.*, 1995; *Hunt*

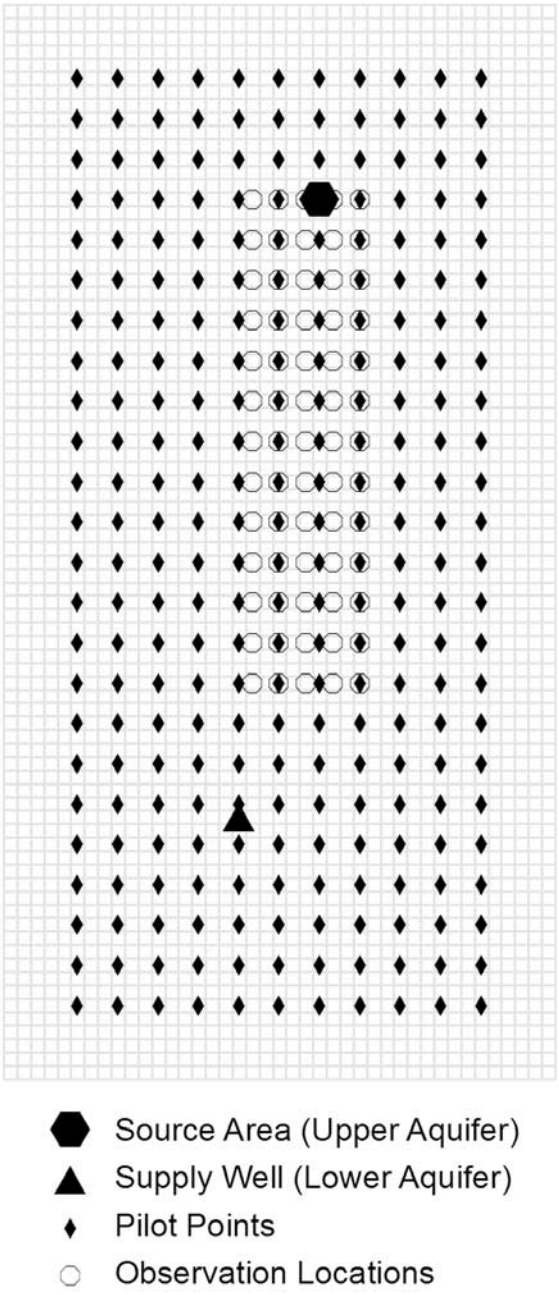

Figure 2. Synthetic model domain illustrating the source area, supply well, hypothetical monitoring wells, and pilot points used to parameterize hydraulic conductivity.

et al., 2007]. In the present study, initial values for pilot points were provided as the (known) mean hydraulic conductivity of each layer. These pilot points constitute the base parameters of the superparameter scheme used to calibrate the model.

[55] *RamaRoa et al.* [1995] and *LaVenue et al.* [1995] describe the use of pilot points for automated calibration of an ensemble of transmissivity fields, to develop frequency distributions for aquifer transmissivities and for model outputs constituting predictions of future system state. *LaVenue et al.* [1995] state that one drawback of their method is the computational burden. It is now demonstrated that the SSMC technique greatly reduces the computational burden of this kind of analysis.

3.1.1. Calibration: Heads Only

[56] Heads calculated by the model using the true parameters were used to generate data for 65 hypothetical monitoring wells in each model layer (Figure 2). Paired screens in the shallow and deep aquifers provide colocated water level data. Gaussian noise with a standard deviation of 0.05 m was added to simulated heads. The calibration data therefore comprised 130 head observations.

[57] To obtain the base parameter sensitivity matrix, $\mathbf{X}$, and undertake the model calibration and SSMC analysis, only the flow model was executed, this requiring less than a second for each model run. One hundred and twenty (120) superparameters were constructed through TSVD of $\mathbf{Q}^{1/2}\mathbf{X}$. Calibration of the superparameters was completed on a single PC. Execution of the model with the best fit parameters results in a value for Φ of 35.

3.1.2. SSMC Analysis: Heads Only

[58] Several adaptations of stochastic parameter generation were employed to generate multiple realizations of hydraulic conductivity parameter sets. In all cases the true (known) variogram was used to generate the stochastic parameters. The parameter sets were then evaluated by contrasting the Φ's calculated using each of these parameter sets with that achieved using the true parameters. The techniques employed are as follows:

[59] 1. Method H1 is stochastic parameter generation and direct assignment of values to pilot point parameters.

[60] 2. Method H2 is extension of H1 through projection of parameter value differences onto the calibration null-space using (18).

[61] 3. Method H3 is unconditioned stochastic parameter generation on the model node scale.

[62] 4. Method H4 is the SSMC method, incorporating node-scale stochastic parameterization as described in section 2.5, with one recalibration iteration completed using superparameters and the existing superparameter sensitivities.

[63] 5. Method H5 is the SSMC method, incorporating node-scale stochastic parameterization as described in section 2.5, with two recalibration iterations completed using superparameters. Existing superparameter sensitivities are used for the first iteration, and recalculated during the second iteration.

[64] Methods H1 and H2 employ stochastic parameter generation at the level of detail used in the calibration (in this case, pilot points), while methods H3, H4, and H5 employ stochastic parameter generation at the node scale. Methods H4 and H5 both employ the Broyden rank-one update to upgrade the terms of the Jacobian matrix during testing of alternate values of the Marquardt parameter. The range of objective functions obtained through each of the 5 methods can be assessed by computing the mean and variance of the objective function; or, by collecting results into a frequency histogram and/or a cumulative density function.

[65] Figure 3 illustrates the values for Φ for the ensemble of models obtained using each of the 5 methods. The increasingly skewed pattern of the histograms illustrates the improvements gained through the projection of differ-

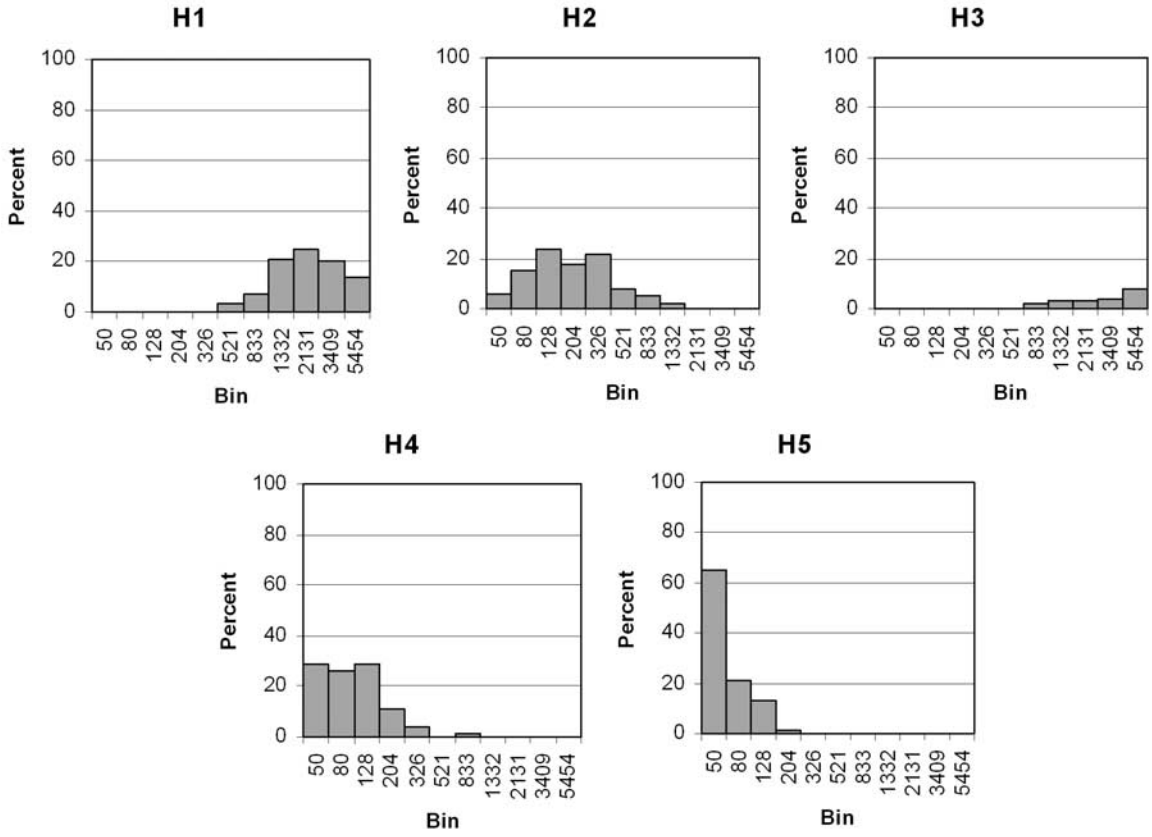

Figure 3. Comparison of objective functions obtained by parameter ensembles using five Monte Carlo techniques.

ences between stochastic pilot point parameter values and their calibrated counterparts onto the calibration null-space (contrasting H2 with H1); and, the improvements gained through the use of the SSMC technique versus unconditioned node-based stochastic parameter generation when undertaken with one recalibration iteration (contrasting H4 with H3), or, with two recalibration iterations (contrasting H5 with H4).

[66] Perhaps most strikingly, the histogram illustrating the ensemble values for Φ achieved by the single-iteration application of the SSMC technique (H4) is located entirely to the left of the histogram illustrating the ensemble values for Φ achieved by the unconditioned multi-Gaussian field generator (H3). Results of the two-iteration application of the new SSMC technique (H5) are also striking. The improvements in the objective functions between H4 and H2 were obtained with an average of 6 model runs for each parameter set. Subsequent improvements in the values for Φ between H5 and H4 were obtained with an average of 126 model runs for each parameter set. At the conclusion of the two-iteration SSMC procedure 90% of the stochastic parameter sets led to an objective function of less than 1.5 times the calibrated value of 35.

[67] The lowest value of Φ achieved using each method was 363 (H1), 39.5 (H2), 636 (H3), 33.9 (H4) and 31.6 (H5), respectively: note that the SSMC recalibrations actually led to minimum values for Φ lower than that obtained with the original calibration. This is unsurprising since the SSMC recalibrations combine the pilot point parameterization scheme with stochastic parameter fields generated on the same scale and using the same variogram as the "true" model, and hence more effectively represent the variability in the "true" model than does the pilot point parameterization scheme when used alone in the original calibration.

[68] As mentioned earlier, the results of the SSMC analysis may differ depending on the user-defined dimensions of the solution and null subspaces. Testing on the synthetic flow model using a sequentially increasing number of superparameters (30, 60, and 120) revealed some differences, in terms of the histograms of the ensemble values for Φ, between results achieved with 30 and 60 superparameters, but no substantial differences between results achieved with 60 and 120 superparameters. This reinforces the concept of the trade-off between computational efficiency and the rigorous exploration of predictive error described by *Tonkin et al.* [2007].

[69] This example demonstrates that the SSMC technique can produce many sets of hydraulic conductivity parameters that can be rapidly recalibrated when using only heads for conditioning. However, the benefit in using concentration data together with groundwater heads for conditioning groundwater model parameters is well established [e.g., *Hendricks Franssen et al.*, 2003]. The objective of the second SSMC analysis is to contrast the performance of the SSMC technique when using only head data with its performance when both head and concentration data are available to constrain parameters.

3.1.3. Calibration: Heads and Concentrations

[70] For this analysis, it is assumed that measurements of groundwater head and concentration are available. Contam-

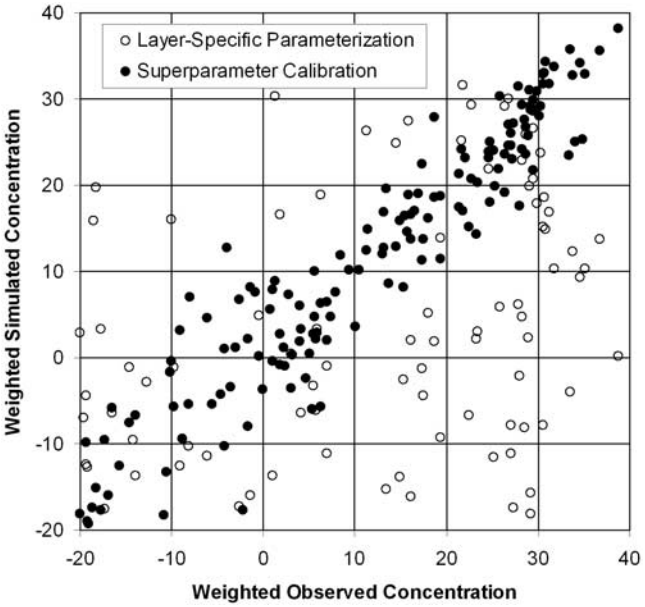

Figure 4. Simulated versus observed concentrations for the layer-specific calibration and superparameter calibration.

inant transport is simulated using MT3DMS [*Zheng and Wang*, 1999]. In the true model, a known time-varying contaminant source is simulated in the shallow aquifer upgradient and lateral to the production well. The true model calculates a peak concentration at the production well of 221 micrograms per liter (μg/L), 1289 days following the contaminant release (t = 1289). Of the many parameter sets generated, this case was selected as reality because the concentration calculated at the production well by the model when calibrated using layer-specific parameters is benign, although in the true case the well is contaminated.

[71] The 130 head observations discussed earlier (Figure 2) were supplemented by concentrations simulated at the wells by the model using the true parameters, with paired screens in the shallow and deep aquifers providing colocated concentration data. Gaussian noise was added to the logs of simulated concentrations to generate concentration observation data; the standard deviation of this noise is equal to 15% of the mean of the natural logs of the concentrations. A hypothetical analytical method-reporting limit of 0.01 μg/L is mimicked in the regression. The calibration data set hence comprises 130 head observations and 390 concentration observations, the latter mimicking three groundwater sampling events occurring at t = 500, 650 and 800 days. When the model is executed using calibrated layer-specific parameter values (in which only 3 parameters feature in the parameter estimation process, each corresponding to an entire model layer), the peak concentration at the well is below the hypothetical method-reporting limit.

[72] To obtain the base parameter sensitivity matrix for the pilot point parameters illustrated in Figure 2, and undertake the model calibration and SSMC analysis, the flow and transport models were executed for sufficient time to obtain simulated equivalents to all observations (i.e., a simulation time of 800 days). One hundred and twenty (120) superparameters were constructed through TSVD of $\mathbf{Q}^{1/2}\mathbf{X}$. Calibration of the superparameters was completed on a parallel PC network. Executing the model with the calibrated parameters results in a value for Φ of 43,700. Figure 4 illustrates weighted simulated versus observed concentrations for the layer-specific parameters and the calibrated superparameters. Concentrations measured and simulated below the hypothetical method reporting limit of 0.01 μg/L cluster in the vicinity of -21, the log of 0.01 times the observation weight. Executing the model with the best fit parameters results in a peak contaminant concentration at the production well of 205 μg/L that occurs about 1,190 days following the release. In this instance, the predicted concentration in the well is quite close to the true value. This might provide some confidence that the prediction is accurate: however, the SSMC technique is now used to evaluate the uncertainty in the concentration simulated at the well, demonstrating that the good agreement between the prediction and reality may be fortuitous.

3.1.4. SSMC Analysis: Heads and Concentrations

[73] Since the heads-only SSMC analysis indicated that two of the stochastic parameter generation techniques (unconditioned stochastic generation of pilot point parameter values and unconditioned stochastic field generation on the scale of the model cells) led to unacceptable results, only three techniques are contrasted to evaluate the effect of using head and concentration data to constrain parameters:

[74] 1. Method HC1 is node-scale stochastic parameterization, as described in section 2.5, following null-space projection of parameter differences using (18) but prior to recalibration. Hence, for each parameter set, this provides the objective function at the commencement of the following SSMC recalibrations.

[75] 2. Method HC2 is the SSMC method, incorporating node-scale stochastic parameterization as described in section 2.5, with one recalibration iteration completed using superparameters and the existing superparameter sensitivities.

[76] 3. Method HC3 is the SSMC method, incorporating node-scale stochastic parameterization as described in section 2.5, with two recalibration iterations completed using superparameters. Existing superparameter sensitivities are used for the first iteration, and recalculated during the second iteration.

[77] Thus, all three methods HC1, HC2, and HC3 employ stochastic parameter generation at the node scale. Methods HC2 and HC3 employ the Broyden rank-one update to upgrade matrix $\mathbf{X}$ during testing of alternate values of the Marquardt parameter. The results are evaluated by contrasting the objective functions calculated using each of the three approaches together with those calculated using the calibrated parameters. The concentrations simulated at the pumping well by the SSMC analysis following two recalibration iterations are compared with the concentrations simulated by the calibrated model, and those simulated using the true parameters.

[78] Figure 5 illustrates the values for Φ for the ensemble of models obtained using the 3 approaches. Again, the increasingly skewed pattern of the histograms illustrates the improvements gained through the use of the SSMC technique versus the simple null-space projection using

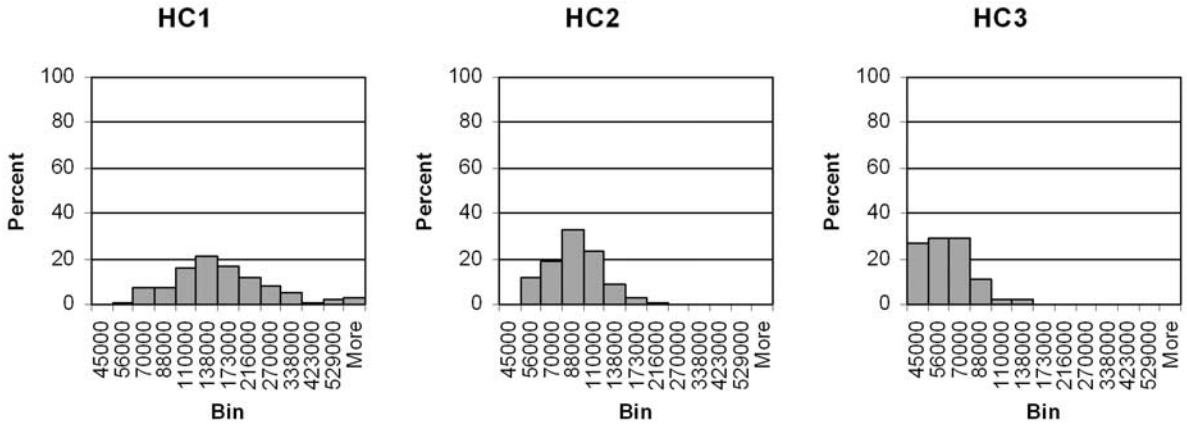

Figure 5. Comparison of objective functions obtained by parameter ensembles using the SSMC procedure.

equation (18) when undertaken with one recalibration iteration (contrasting HC2 with HC1), or, with two recalibration iterations (contrasting HC3 with HC1). While HC2 does show significant improvement over HC1, the histogram illustrating the ensemble values for Φ achieved by the two-iteration application of the SSMC technique (HC3) is located almost entirely to the left of the histogram illustrating the ensemble values for Φ achieved by the null-space-projected stochastic parameters (HC1).

[79] Over one third of the parameter sets created using the two-iteration SSMC approach led to an objective function less than 1.1 times the calibrated value. (Neither the single-iteration SSMC nor the parameter null-space projection produced any parameter sets that resulted in an objective function of less than 1.1 times the calibrated value.) Furthermore, over 80% of the parameter sets created using the two-iteration SSMC approach led to an objective function less than 1.5 times the calibrated value; this contrasts with 28% for the single-iteration SSMC, and about 10% for the parameter null-space projection. As for the heads-only case, several parameter fields generated following the second SSMC recalibration iteration led to objective functions below the minimum obtained by the calibration.

[80] Figure 6 illustrates the concentrations calculated at the well using the true parameters, and using parameters computed through superparameter calibration. Also shown are the time-varying maximum, minimum, and average concentrations calculated using stochastic parameter sets achieved at the conclusion of the two-iteration SSMC process. (Recall that the peak concentration calculated at the well using the layer-specific calibrated parameters was less than the hypothetical reporting limit.) Figure 6 illustrates that the bounds of predicted time-varying concentrations at the well generally encompass the concentrations calculated using the true parameters, although the peak concentration is not quite encompassed: this latter outcome may suggest that a larger number of parameter ensembles is required to fully encompass the peak, despite the fact that the parameter ensemble predicts very similar concentrations at earlier and later times. It is noteworthy that the time-varying concentrations calculated by the ensemble parameter sets are far more realistic than the concentrations calculated using the layer-specific parameters.

3.1.5. Discussion of Synthetic Model Results

[81] In the case that only head observations are available for calibration, Figure 3 illustrates the benefits gained through projection of differences between the stochastic

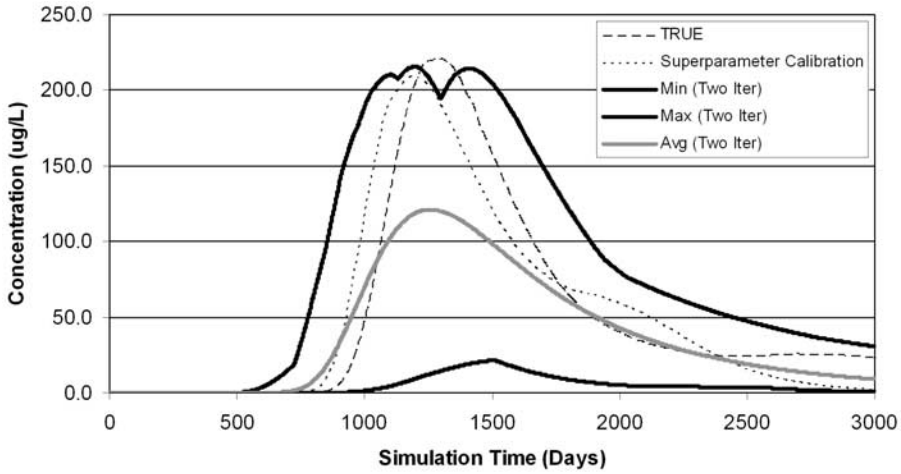

Figure 6. Concentrations at the well in the true model, superparameter calibration, and the time-varying maximum, minimum, and average concentrations from the two-iteration SSMC analysis.

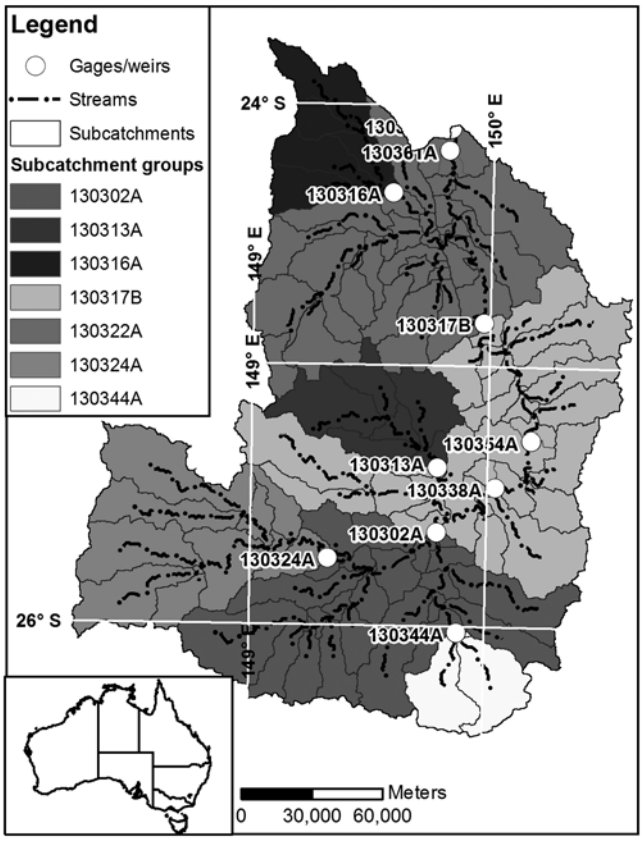

Figure 7. Location and extents of the watershed model, showing the distribution of subcatchments and location of gaging stations.

parameters and calibrated parameters onto the calibration null-space. It also illustrates the improvements gained through use of the SSMC technique after one or two recalibration iterations. In the case that both heads and concentrations are available for calibration, Figure 5 illustrates the improvements gained through use of the SSMC technique after one or two recalibration iterations. Although in the latter case the reductions in the composite objective function appear less dramatic than when conditioning to head observations only, both the first and second recalibration iterations led to substantial reductions in the composite objective function. This is a very good result, as experience with the calibration of flow-and-transport models demonstrates that the use of both groundwater head and contaminant concentration data for conditioning places far greater demands on the calibration process than the use of heads alone.

[82] Application of the SSMC procedure to the synthetic flow-only and flow-and-transport models indicated that it might be possible to identify, before commencing the first iteration of SSMC recalibration, whether some of the stochastic parameter sets might not successfully recalibrate; and, more importantly, whether at the conclusion of the first iteration it might be possible to identify which parameter sets should be rejected or retained for the second recalibration iteration, since the second iteration requires recalculation of superparameter sensitivities. While the authors have not developed any rigorous theory for these decisions at this time, some pragmatic options are now discussed, using the real-world modeling example.

3.2. Real-World Watershed Management Model

[83] Recently, the Land Resource Assessment branch of the Queensland Department of Natural Resources and Water (NR&W), Queensland, Australia, developed a hydrological and sediment transport model encompassing 40,381 km^2 of the Dawson Valley, a major branch of the Fitzroy River Catchment, in central Queensland, Australia (Figure 7). The purpose of this model is to simulate streamflow and sediment transport within the rivers that drain the Fitzroy River Catchment, as part of a wider study of the effects of land management practices on sediment transport to the Great Barrier Reef. The climate within the study area is subtropical, with an average annual rainfall of about 700 mm. Much of this rainfall occurs during summer storms driven by cyclonic influences. The primary reason for development of the model was for evaluation of sediment yield. However, since work on the sediment transport modeling is (at the time of writing) in progress, the discussion presented here focuses on the simulation of flow.

[84] The model was built using the E2 catchment modeling system [*Argent et al.*, 2006], which provides a framework for constructing node link models of arbitrary complexity over catchments of arbitrary size. A node is located at a confluence, and/or at a location where subcatchment water and/or sediment enters the system. Links route and process water and constituents as they move from node to node. In this study, the SYMHYD rainfall-runoff model, one of the rainfall-runoff models available in the E2 system [*Cooperative Research Centre for Catchment Hydrology*, 2004], was used to compute the flow of water from subcatchments to nodes. Hydrologic routing within links was simulated using the nonlinear Laurenson algorithm [*Laurenson and Mein*, 1997].

[85] The model domain is divided into 119 subcatchments (Figure 7), which range in area from 0.75 to 1274 km^2. Land use within the model domain is divided into 7 categories (Table 1), the most dominant of which is grazing (GRZ).

[86] The model is typical of many used as a basis for the study and regulation of land management practices, since it is required to represent hydrologic processes and land management decisions at a level of detail corresponding with anticipated changes. However, this requires that hundreds or thousands of parameter values be represented in the model. Although lumping of parameters is often used to achieve stable calibration and parameter identifiability, parameter aggregation of this type should ideally be aban-

Table 1. Land Use Type and Abbreviations in the Real-World Watershed Model

Land Use Type	Abbreviation
Cropping	CRP
Grazing	GRZ
Grazed state forest	GSF
Horticulture	HOR
Nature conservation	NAT
Residential	RES
Water	WTR

Table 2. Observations Used in the Calibration of the Real-World Watershed Model

Gauging Station	Location	Number of Daily Flow Observations
130302A	Dawson River at Taroom	6864
130313A	Palm Tree Creek at La Palma	2307
130316A	Mimosa Creek at Redcliffe	1885
130317B	Dawson River at Woodleigh	6647
130322AA	Dawson River at Beckers	4425
130324A	Dawson River at Utopia Downs	7305
130344A	Juandah Creek at Windamere	1347

doned when exploring predictive error since in reality hydraulic properties may show considerable variability between subcatchments that may, in turn, affect the variability of model predictions.

3.2.1. Calibration Data

[87] Historic records of daily flow are available through all or part of a 20 year calibration period between and including the years 1986 to 2005 at seven gauging stations (Figure 7 and Table 2). The model calibration employed a multicomponent objective function composed of (1) daily flows weighted according to the inverse square root of observed flows, (2) monthly volumes computed on the basis of these flows, and (3) flow exceedence time abscissae at a flow log interval of 2 per decade (Table 2).

3.2.2. Parameterization

[88] The number of separate SYMHYD instances within the model equals the number of ''functional units'' (FUs), where a single functional unit comprises a specific land use within a specific subcatchment. The Fitzroy Catchment model is composed of 119 $\times$ 7 = 833 FUs, together with 119 links. Since the SIMHYD model employs 9 parameters, in addition to the two parameters employed by the Laurenson routing algorithm (Table 3), the model notionally employs 7,735 parameters: 9 parameters for each of 833 FUs, plus 2 parameters for each of 119 links. The number of parameters requiring estimation in this study was reduced by directly assigning values to the impervious fraction for all land uses excepting residential, and to the water (WTR) land use type, which occupies a very small fraction of the land use within the model domain. Nonetheless, regularization is required to stabilize the inverse problem since this notionally leaves 6,188 parameters that require values be assigned.

3.2.3. Calibration

[89] Parameter estimation was accomplished using the hybrid Tikhonov-TSVD technique. During calibration SYMHYD parameters of the same type pertaining to the same land use were grouped (i.e., assumed to be equal) according to the most immediate downstream gauging station. Similarly, all links were temporarily assumed to possess identical Laurenson parameters. The number of parameters requiring estimation was thereby reduced to 345. These groupings were abandoned during subsequent SSMC analyses.

[90] Tikhonov regularization was implemented by including prior estimates of parameter values, developed in collaboration with NR&W personnel, in the inversion process. The weights assigned to these Tikhonov regularization equations were updated throughout the constrained optimization inverse process that sought to maximize the goodness of fit between model outputs and field measurements while minimizing the departure of parameters from their preferred values [*Tikhonov and Arsenin*, 1977; *Doherty and Skahill*, 2006]. The combined use of mathematical (Tikhonov) and manual (parameter grouping) regularization was implemented in an effort to obtain a maximum likelihood parameter set. Such a parameter set is, of necessity, a simplified set from which any complexity that is not inferable from the calibration data set is eliminated. However, the SSMC evaluation that follows restores parameter complexity.

[91] The base parameter sensitivity matrix, $\mathbf{X}$, was calculated using a parallel PC-based computer network, and $\mathbf{Q}^{1/2}\mathbf{X}$ was then formed. TSVD of $\mathbf{Q}^{1/2}\mathbf{X}$, in accordance with equation (9), was used to construct 35 superparameters. A solution space dimensionality of 35 was adopted after testing alternatives: the use of fewer superparameters prevented the attainment of an acceptable fit between model outputs and historical flows, while use of more superparameters resulted in high condition numbers, and a numerically unstable inverse problem.

[92] Calibration was undertaken on the same parallel PC-based computer network. The value of composite objective function was reduced from 22,800 at the commencement of the hybrid Tikhonov-TSDV superparameter inversion process, where parameters were assigned their preferred values, to an optimized value of 6,800. Figure 8 illustrates the typical correspondence between model-calculated and observed flows, using a 2 year segment of the calibration period at the 130322A gauging station.

Table 3. Parameters Calibrated in the Real-World Watershed Model

Parameter Description	Units	Range	Abbreviation	Model
Base flow coefficient	none	0.0–1.0	BASE	SYMHYD
Impervious threshold	none	0.0–5.0	IMPT	SYMHYD
Infiltration coefficient	d^{-1}	0.0–400.0	INFC	SYMHYD
Infiltration shape	none	0.0–10.0	INFS	SYMHYD
Interflow coefficient	d^{-1}	0.0–1.0	INTC	SYMHYD
Pervious fraction	none	0.0–1.0	PERF	SYMHYD
Rainfall interception store capacity	mm	0.0–5.0	RISC	SYMHYD
Recharge coefficient	d^{-1}	0.0–1.0	RECO	SYMHYD
Soil moisture store capacity	mm	1.0–1000.0	SMSC	SYMHYD
K	s	0 to several days	K	LINK ROUTING[a]
M	none	0.6–1.0	M	LINK ROUTING[a]

[a]Employs the Laurenson nonlinear lag procedure which uses the storage-outflow relationship $S(Q) = KQ^m$ [*Laurenson and Mein*, 1997].

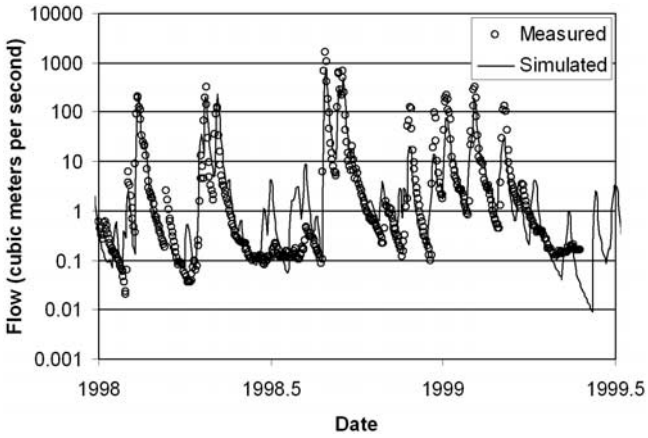

Figure 8. Correspondence between simulated and measured flow at 130322A in the watershed model.

3.2.4. SSMC Analysis

[93] In undertaking the SSMC analysis, parameter grouping that was employed during the calibration process to reduce the number of parameters from 6,188 to 345 was abandoned. In consultation with NR&W personnel, a C(**p**) matrix was constructed to characterize the potential variability of the 6,188 base parameters employed by the model, throughout the model extent. This C(**p**) matrix was developed to respect understanding of natural hydraulic property variability, together with ordering relationships between parameters of the same type pertaining to different land uses.

[94] For this study, a calibration threshold objective function value of 7,000 was selected on the basis of visual evaluation of many model results. Hence, the SSMC analysis focused on obtaining parameter sets that respect C(**p**) while resulting in an objective function no greater than 7,000. Since the model required 6 min to execute, efficiency was an important consideration, and pragmatic decisions were made during recalibration. The following strategy was adopted:

[95] 1. Parameter realizations were generated using the C(**p**) covariance matrix.

[96] 2. Differences between each realization and the calibrated parameter set were projected onto the calibration null-space using equation (18), and added to calibrated parameter values.

[97] 3. The value of the objective function was then calculated: if the objective function was greater than 10,000 that parameter set was abandoned on the basis that it was unlikely to achieve recalibration in two iterations.

[98] 4. Recalibration of the retained parameter sets was undertaken using 35 superparameters constructed using sensitivities obtained using the calibrated model.

[99] The first iteration of the recalibration process was undertaken using superparameters constructed at the completion of the initial model calibration (Figure 1). For parameter sets that required a second calibration iteration, 35 model runs were required to update sensitivities and undertake a modified Gauss-Newton optimization of the superparameters. If the threshold objective function was not achieved within two iterations, the parameter set was abandoned and recalibration of the next parameter realization commenced.

[100] Three hundred (300) stochastic parameter sets were generated and subjected to steps 1–4 above. Of these, 133 parameter sets achieved an objective function below 7,000. Hence, application of the SSMC approach, following model calibration, resulted in the production of 134 parameter sets: one of these 134 comprised the calibrated parameters, while the remaining 133 consisted of random parameter realizations that are hydrologically reasonable, respect necessary ordering relationships embodied in C(**p**), and result in a model that is as well calibrated as the "calibrated model."

[101] On average, the acquisition of each new calibration-constrained parameter set, requiring the assignment of values to 6,188 parameters, required about 15 model runs. Figure 9 and Table 4 illustrate that obtaining the new calibration-constrained parameter sets benefited greatly from the null-space projection step of the SSMC technique. Figure 9a illustrates the distribution of objective functions computed with 100 stochastic parameter realizations before null-space projection using equation (18) and with no recalibration. Figure 9b illustrates the distribution of objective functions for the same parameter sets following null-space projection using equation (18) but again prior to recalibration.

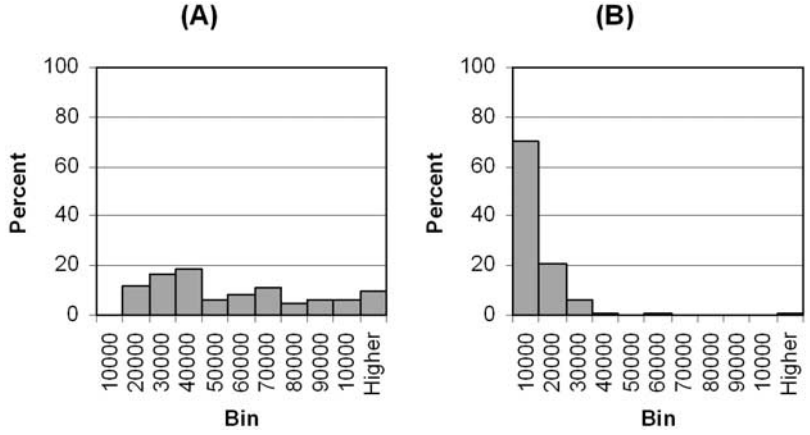

Figure 9. Distribution of objective functions computed with 100 stochastic parameter realizations (a) before null-space projection and recalibration and (b) following null-space projection.

Table 4. Objective Function Values Obtained Before and After Null-Space Projection

Summary Statistic	Objective Function Before Null-Space Projection	Objective Function After Null-Space Projection
Average Φ	57,017	11,822
Median Φ	40,161	8,398
Minimum Φ	11,209	7,152
Maximum Φ	269,420	108,830

[102] It is evident from comparison of Figures 9a and 9b that the projection of the differences between the calibrated and stochastic parameters onto the calibration null-space using equation (18) led to significant improvement in the objective function, and ensured that recalibration commenced from a much improved starting condition than from simple stochastic parameter generation. Given that following recalibration 133 parameter sets resulted in objective functions at or below the target value of 7,000 after only two parameter estimation iterations based on 35 super parameters, the first iteration of which required no model runs for calculation of sensitivities, it is evident that the recalibration process leads to further, significant, reduction in the objective function at only a mild computational cost.

[103] Figure 10 illustrates the fit between the measured flows for a part of the calibration period at one of the gaging stations used in the calibration process, and the time-varying minimum and maximum flows calculated by the calibration-constrained stochastic parameter sets at the conclusion of the SSMC procedure. Figure 11 illustrates a histogram of the variability in values for a single SYMHYD parameter in one particular FU for the 133 parameter sets.

3.2.5. Discussion of Real-World Model Results

[104] The 133 sets of parameters obtained using the SSMC approach described above could each be used with the model to evaluate a variety of model outputs. For this study, the model was used to evaluate streamflow at gage 130322A, located at the outflow (or pour point) of the model domain, during a 6-week period from December 1973 through January 1974. During this period, more rainfall occurred within the overall catchment than occurred over any interval of the same length within the 20 year model calibration period. Streamflow at gage 130322A was calculated using the calibrated parameters, using the 133 calibration-constrained SSMC parameter sets, and then compared to measured flows over this period.

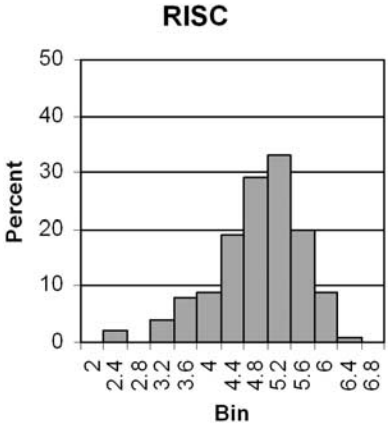

Figure 11. Histogram of RISC parameter values for the GRZ land use type at subcatchment 33.

[105] Figure 12 illustrates flows at gage 130322A as simulated on the basis of 10 of the 133 parameter realizations, including those that gave rise to maximum and minimum flow peaks, together with the observed flows over this period. It is apparent from Figure 12 that there is a high degree of uncertainty associated with the prediction of peak flow at gage 130322A. Nonetheless, the uncertainty interval encompasses the observed peak of the dominant flow event. The uncertainty interval does not encompass later, smaller, events. Discussions with NR&W personnel suggest that this may be an outcome of the fact that the management of water storage in dams, including one located upstream of gage 130322A, is not well represented within the model during this period. (This is in addition to errors accrued through use of a lumped parameter model in the first place.)

[106] Ongoing use of the parameter sets computed through SSMC analysis for examination of predictive un-

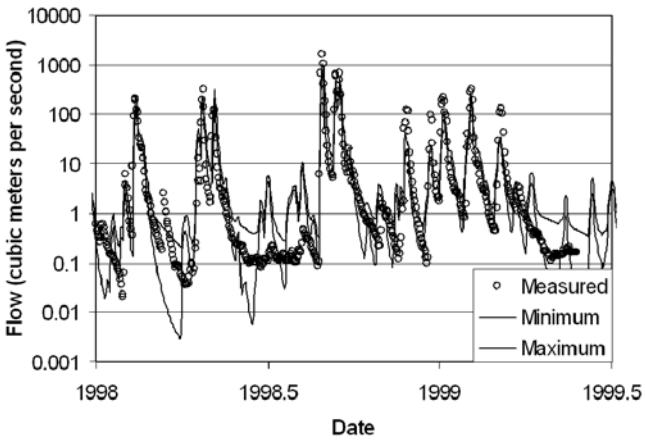

Figure 10. Correspondence between measured flows and the time-varying minimum and maximum flows calculated at the conclusion of the SSMC procedure at gage 130322A.

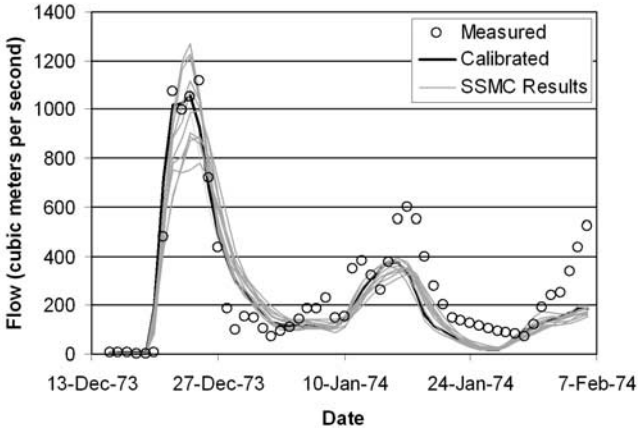

Figure 12. Observed flows at gage 130322A together with calibrated flows and flows calculated at the conclusion of the SSMC procedure, including those that gave rise to the maximum and minimum flow peaks.

certainty has demonstrated that the uncertainty associated with *changes* in flow and sediment yield resulting from *changes* in land management practices show less uncertainty that do the absolute flows demonstrated here. This is encouraging since it suggests that in spite of the considerable uncertainty associated with absolute predictions made by models of this type, relative predictions resulting from changes in management practices can be successfully explored.

4. Concluding Remarks

[107] Under ideal circumstances the single parameter set obtained through model calibration yields model predictions that approach maximum likelihood. This does not guarantee that these predictions are correct: it means that to the extent that they are likely to be incorrect, the simulated prediction represents a central tendency, and variability in the prediction is relatively evenly distributed about this value. Techniques developed to enable calculation of the potential error in model predictions are often computationally intensive when implemented with highly parameterized models and, perhaps as a result, are rarely implemented in common modeling practice.

[108] This paper presents an efficient technique for investigating the uncertainty associated with the outputs of highly parameterized models which blends error variance theory with a Monte Carlo implementation. The SSMC technique complements the use of highly parameterized models by enabling efficient computation of many different parameter realizations that (1) provide an acceptable fit between model outputs and measured data and (2) are physically reasonable, as measured by stochastic descriptors of real-world parameter variability.

[109] Model calibration always involves some form of parameter parsimony to achieve uniqueness: indeed, the superparameter technique of *Tonkin and Doherty* [2005] implements parsimony to achieve a stable solution. Whatever form the parsimony takes, if it is retained after calibration for predictive uncertainty analysis it may result in the calculation of narrower likelihood intervals than are reasonable. The SSMC procedure enables simplifications required during calibration, regardless of the parameterization technique used, to be abandoned when evaluating uncertainty, allowing fine-scale parameter detail to be incorporated.

[110] Using the SSMC technique, acceptable model-to-measurement fits are ensured by a procedure that includes the projection of differences between stochastic and calibrated parameter sets onto the calibration null-space, followed by enforcement of calibration constraints through recalibration. Efficient recalibration is achieved by using existing solution space sensitivities to undertake the first calibration iteration, and perturbing and estimating a limited number of superparameters during subsequent iterations. This approach possesses two distinct advantages: first, no model runs are required to obtain sensitivities during the first iteration; second, the reduced dimensionality requires that only a limited number of model runs be undertaken to calculate superparameter sensitivities during subsequent iterations, if these are required.

[111] Although the approach is nonlinear, linearity-based concepts are employed in the definition of orthogonal calibration solution and null subspaces that underpins the methodology. Use of subspaces defined using sensitivities calculated with the simplified calibrated parameter set to explore the potential error in predictions made by highly nonlinear models may incur errors in the assignment of confidence limits to predictive ranges. However, these errors are unlikely to be much greater than those incurred through assuming that any particular $\mathbf{C}(\mathbf{p})$ or $\mathbf{C}(\varepsilon)$ matrix provides a suitable characterization of innate hydraulic property variability, or of the stochastic character of measurement noise, respectively, both of which are required in any method of uncertainty analysis.

[112] Valuable outcomes of an SSMC analysis following regularized inversion of a highly parameterized model include (1) a single set of calibrated parameters that contains the variability that can be inferred from the calibration data and can be used to make predictions that may approach maximum likelihood and (2) many different parameter sets that provide an acceptable fit to the calibration data, but incorporate parameter variability on a finer scale than considered during calibration, more closely approaching the scale of true property variability.

[113] The example applications illustrate that the SSMC approach is general in the sense that it is not limited to any particular model or any particular parameterization scheme. The second application demonstrates that the SSMC technique is applicable to real world model applications.

[114] The SSMC procedure is a logical extension to the hybrid regularized inversion technique described by *Tonkin and Doherty* [2005], and a companion to the predictive error analysis technique described by *Tonkin et al.* [2007]. Indeed, as for these methods, the SSMC procedure is founded upon the subdivision of parameter space into orthogonal subspaces, restricting computations that impose calibration constraints to a limited subspace of parameter space, using a number of superparameters that is considerably smaller than the number of parameters represented in the model. With the SSMC technique, to the extent that predictions of interest are sensitive to fine-scale parameterization detail that is uninformed by the calibration data, that detail is incorporated in the predictive analysis through stochastic field generation. Once an ensemble of parameter sets is obtained it can be employed to make any prediction required of the model, and to calculate approximate confidence intervals associated with predictions using frequency analysis. Hence, the potential for error in model outputs can be incorporated into decision making.

[115] **Acknowledgments.** The authors wish to thank the South Florida Water Management District (SFWMD), United States, and the Queensland Department of Natural Resources and Water (NR&W), Australia, for support while developing these techniques. The approach described in this study is documented and encapsulated in programs provided as part of the PEST suite [*Doherty*, 2007], available at http://www.sspa.com/pest.

References

Argent, R. M., G. M. Podger, R. B. Grayson, K. Fowler, and N. Murray (2006), E2 catchment modelling software, user guide, Coop. Res. Cent. for Catchment Hydrol., Canberra, A.C.T., Australia.

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

Bard, J. (1974), *Nonlinear Parameter Estimation*, 341 pp., Academic, New York.

Beven, K. J., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*, 279–298, doi:10.1002/hyp.3360060305.

Broyden, C. G. (1965), Class of methods for solving nonlinear simultaneous equations, *Math. Comput.*, *19*(92), 577–593, doi:10.2307/2003941.

Certes, C., and G. de Marsily (1991), Application of the pilot point method to the identification of aquifer transmissivities, *Adv. Water Resour.*, *14*(5), 284–300, doi:10.1016/0309-1708(91)90040-U.

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

Cooley, R. L., and R. L. Naff (1990), *Regression Modeling of Ground-Water Flow, U.S. Geol. Surv. Tech. Water Resour. Invest., Book 3, Chap. B4*, 232 pp.

Cooperative Research Centre for Catchment Hydrology (2004), *Rainfall Runoff Library*, Dep. of Civ. Eng., Monash Univ., Clayton, Vic., Australia.

Dennis, J. E., and E. S. Marwil (1982), Direct secant updates of matrix factorizations, *Math. Comput.*, *38*, 459–474.

Deutsch, C., and A. Journel (1992), *GSLIB: Geostatistical Software Library and Users Guide*, 340 pp., Oxford Univ. Press, New York.

Doherty, J. (2007), *Users Manual for PEST Version 11*, 339 pp., Watermark Numer. Comput., Brisbane, Queensl., Australia.

Doherty, J., and B. Skahill (2006), An advanced regularization methodology for use in watershed model calibration, *J. Hydrol.*, *327*(3–4), 564–577, doi:10.1016/j.jhydrol.2005.11.058.

Gallagher, M., and J. Doherty (2007), Parameter estimation and uncertainty analysis for a watershed model, *Environ. Modell. Softw.*, *22*(7), 1000–1020, doi:10.1016/j.envsoft.2006.06.007.

Golub, G. H., and F. C. Van Loan (1989), *Matrix Computations*, 642 pp., Johns Hopkins Univ. Press, Baltimore, Md.

Gómez-Hernandez, J. J., A. Sahuquillo, and J. E. Capilla (1997), Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data—1. Theory, *J. Hydrol.*, *203*, 162–174, doi:10.1016/S0022-1694(97)00098-X.

Gómez-Hernández, J. J., H. J. Hendricks Franssen, and A. Sahuquillo (2003), Stochastic conditional inverse modeling of subsurface mass transport: A brief review of the self-calibrating method, *Stochastic Environ. Res. Risk Assess.*, *17*, 319–328, doi:10.1007/s00477-003-0153-5.

Guadagnini, A., and S. P. Neuman (1999), Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly non-uniform domains: 1. Theory and computational approach, *Water Resour. Res.*, *35*(10), 2999–3018, doi:10.1029/1999WR900160.

Haber, E. (1997), Numerical strategies for the solution of inverse problems, Ph.D. thesis, Univ. of B. C., Vancouver, B. C., Canada.

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 pp.

Hendricks Franssen, H.-J., J. Gómez-Hernandez, and A. Sahuquillo (2003), Coupled inverse modeling of groundwater flow and mass transport and the worth of concentration data, *J. Hydrol.*, *281*(4), 281–295, doi:10.1016/S0022-1694(03)00191-4.

Hernandez, A. F., S. P. Neuman, A. Guadagnini, and J. Carrera (2006), Inverse stochastic moment analysis of steady state flow in randomly heterogeneous media, *Water Resour. Res.*, *42*, W05425, doi:10.1029/2005WR004449.

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

Kitanidis, P. K. (1996), On the geostatistical approach to the inverse problem, *Adv. Water Resour.*, *19*(6), 333–342, doi:10.1016/0309-1708(96)00005-X.

Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, *J. Hydrol.*, *211*, 69–85, doi:10.1016/S0022-1694(98)00198-X.

Laurenson, E. M., and R. G. Mein (1997), *RORM–Version 4, Runoff Routing Program—User Manual*, 2nd ed., Dep. of Civ. Eng., Monash Univ., Clayton, Vic., Australia.

LaVenue, A. M., B. S. RamaRoa, G. De Marsily, and M. G. Marietta (1995), Pilot point methododolgy for automated calibration of an ensemble of conditionally simulated transmissivity fields: 2. Application, *Water Resour. Res.*, *31*(3), 495–516.

Lavenue, M., and G. De Marsily (2001), Three-dimensional interference test interpretation in a fractured-unfractured aquifer using the pilot point inverse method, *Water Resour. Res.*, *37*(11), 2659–2676, doi:10.1029/2000WR900289.

Madsen, K., and H. B. Nielsen (2004), *Methods for Non-Linear Least Squares*, Inf. and Math. Modell., Tech. Univ. of Denmark, Lyngby, Denmark.

Menke, W. (1989), *Geophysical Data Analysis: Discrete Inverse Theory*, 2nd ed., 289 pp., Academic, New York.

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

Oliver, D. S., L. B. Cunha, and A. C. Reynolds (1997), Markov Chain Monte Carlo methods for conditioning a permeability field to pressure data, *Math. Geol.*, *29*, 61–91, doi:10.1007/BF02769620.

RamaRoa, B. S., A. M. LaVenue, G. de Marsily, and M. G. Marietta (1995), Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments, *Water Resour. Res.*, *36*(9), 2795–2797.

Rubin, Y., and G. Dagan (1987), Stochastic identification of transmissivity and effective recharge in steady state groundwater flow: 1. Theory, *Water Resour. Res.*, *23*(7), 1192–1200.

Sambridge, M., P. Rickwood, N. Rawlinson, and S. Sommacal (2007), Automatic differentiation in geophysical inverse problems, *Geophys. J. Int.*, *170*(1), 1–8, doi:10.1111/j.1365-246X.2007.03400.x.

Seber, G. A. F., and C. J. Wild (1989), *Nonlinear Regression*, 768 pp., John Wiley, New York.

Skahill, B. E., and J. S. Baggett (2006), More efficient derivative-based watershed model calibration, *Eos Trans. AGU*, *87*(52), Fall Meet. Suppl., Abstract H43B-0499.

Tikhonov, A., and V. Arsenin (1977), *Solutions of Ill-Posed Problems*, 258 pp., V.H. Winston, Washington, D. C.

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

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

Vecchia, A. V., and R. L. Cooley (1987), Simultaneous confidence and prediction intervals for non-linear regression models with application to a groundwater flow model, *Water Resour. Res.*, *23*(7), 1237–1250, doi:10.1029/WR023i007p01237.

Woodbury, A. D., and T. J. Ulrych (2000), A full-Bayesian approach to the groundwater inverse problem for steady state flow, *Water Resour. Res.*, *36*(8), 2081–2094, doi:10.1029/2000WR900086.

Yeh, T.-C. J., M. Jin, and S. Hanna (1996), An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields, *Water Resour. Res.*, *32*(1), 85–92, doi:10.1029/95WR02869.

Zheng, C., and P. Wang (1999), MT3DMS: A modular three-dimensional multi-species transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems: Documentation and users guide, Contract Rep. SERDP-99-1, Eng. Res. and Dev. Cent., U.S. Army, Vicksburg, Miss.

J. Doherty, Watermark Numerical Computing, 336 Cliveden Avenue, Corinda, Qld 4075, Australia. (jdoherty@gil.com.au)

M. Tonkin, S. S. Papadopulos and Associates, 7944 Wisconsin Avenue, Bethesda, MD 20814, USA. (matt@sspa.com)