Journal of Hydrology (2007) 334, 513–533

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jhydrol

Predictive error analysis for a water resource management model

Mark Gallagher [a,b,*], **John Doherty** [b,c,*]

[a] *Natural Resource Sciences, Department of Natural Resources and Water, Indooroopilly 4068, Australia*
[b] *Department of Civil Engineering, University of Queensland, St. Lucia, Qld 4072, Australia*
[c] *Watermark Numerical Computing, Brisbane, Australia*

Received 31 March 2006; received in revised form 12 October 2006; accepted 31 October 2006

KEYWORDS
Predictive uncertainty;
Ground water;
Numerical models;
Basin management

Summary In calibrating a model, a set of parameters is assigned to the model which will be employed for the making of all future predictions. If these parameters are estimated through solution of an inverse problem, formulated to be properly posed through either pre-calibration or mathematical regularisation, then solution of this inverse problem will, of necessity, lead to a simplified parameter set that omits the details of reality, while still fitting historical data acceptably well. Furthermore, estimates of parameters so obtained will be contaminated by measurement noise. Both of these phenomena will lead to errors in predictions made by the model, with the potential for error increasing with the hydraulic property detail on which the prediction depends. Integrity of model usage demands that model predictions be accompanied by some estimate of the possible errors associated with them.

The present paper applies theory developed in a previous work to the analysis of predictive error associated with a real world, water resource management model. The analysis offers many challenges, including the fact that the model is a complex one that was partly calibrated by hand. Nevertheless, it is typical of models which are commonly employed as the basis for the making of important decisions, and for which such an analysis must be made. The potential errors associated with point-based and averaged water level and creek inflow predictions are examined, together with the dependence of these errors on the amount of averaging involved. Error variances associated with predictions made by the existing model are compared with ''optimized error variances'' that could have been obtained had calibration been undertaken in such a way as to minimize predictive error variance. The contributions by different parameter types to the overall error variance of selected predictions are also examined.

* Corresponding authors. Address: Department of Civil Engineering, University of Queensland, St Lucia, Qld 4072, Australia. Tel.: +61 7 3371 3159 (M. Gallagher); tel.: +61 7 3379 1664 (J. Doherty).
E-mail addresses: s307783@student.uq.edu.au (M. Gallagher), johndoherty@ozemail.com.au (J. Doherty).

0022-1694/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2006.10.037

Introduction

It is now commonplace for groundwater models to be employed as a basis for management of water resources. These models are often complex and of large areal extent, covering areas of poorly known and non-uniform geology. They often simulate exchanges of water between ground and surface water bodies. They often include the use of one or more soil water balance submodels to simulate water use by crops, demand for irrigation water, and recharge to the groundwater system. They are often data-intensive models, for information on water usage, groundwater levels, streamflow and climate are commonly available over many years to assist in their calibration.

Predictions required by such models can be many and varied, and can encompass many scales. They can range in scope from the volume of water within an entire aquifer, to the effect of one user's pumping on the level in another's well; or from losses incurred by an entire river system to the streamflow depletion incurred by one user's pumping.

Those who build models such as these, and those who employ them for management, are aware of the fact that their predictions may be in error. There is a general belief that the size of this potential error is likely to be smaller in a relative sense for predictions that are of larger scale, or involve a higher degree of averaging, than for predictions that pertain to system detail. Furthermore, it is normally assumed that a model's propensity for predictive error is somewhat mitigated by calibration, though the extent of this mitigation as it applies to different model predictions is almost never tested. In fact, rarely is any attempt made to quantify the potential error associated with key model predictions in spite of the obvious importance of such knowledge. Instead, most modelling reports include a post-calibration ''sensitivity analysis'' that is carried out for reasons which are unclear, and whose pertinence to the evaluation or model predictive error is never properly explained.

Changes to the way in which model outcomes are reported is overdue. Increasingly, government agencies and stakeholder groups are demanding that model predictions of system behaviour be accompanied by solid estimates of the potential errors associated with those predictions. However, at the time of writing, these demands are not being met. The principal reason for this is the lack of a suitable, cost-effective methodology for implementation of rapid and robust model predictive error analysis as an adjunct to routine model-based groundwater management. The present paper attempts to rectify this inadequacy by demonstrating such a methodology. This methodology can be employed for calculation of the potential error of any prediction, at any scale, made by a calibrated model.

Existing techniques for quantification of the uncertainty and/or error variance (these two terms will be distinguished shortly) associated with predictions made by a calibrated groundwater model fall into a number of different categories, of which only a few will be mentioned herein. Where parameterisation of such models is parsimonious in order to allow formulation of a stable over-determined inverse problem, linear methods such as those discussed by Hill (1989) and available in Poeter and Hill (1998) are sometimes employed. Model non-linearity can be accommodated using the constrained maximisation/minimisation approach developed by Vecchia and Cooley (1987) and implemented by Christensen and Cooley (1999). Accommodation of the inherent complexity of real-world systems, and of the spatial averaging involved in definition of a parsimonious, often zonal, parameter set, is discussed extensively by Cooley (2004).

A disadvantage of these approaches in data-rich environments is that calibration based on zones of piecewise constancy, or on other pre-calibration parameter parsimonising devices, can be cumbersome to implement, and often fails to capture the full information content of a calibration dataset; see Moore and Doherty (2006) for a full discussion of this issue. In practical terms, the design of a zonation scheme that, on the one hand, is flexible enough to be responsive to information contained within a calibration dataset on the existence of heterogeneity within the model domain, while on the other hand does not require that more parameters be estimated than can be done with numerical stability (thus invalidating post-calibration predictive error analyses of the types discussed above), is a difficult procedure. As will be discussed below, the regularised inversion approach, based on an abundance of parameters rather than just a few, in which parsimony is enforced through the calibration process itself, provides greater flexibility in calibrating models of this type.

The last two decades have seen extensive use of probabilistic parameter characterizations in groundwater modelling; see, for example, reviews by Carrera and Glorioso (1991), Yeh (1992), Gutjahr and Bras (1993), Kitanidis (1995), Zimmerman et al. (1998), Gomez-Hernandez et al. (2003) and Carrera et al. (2005). Probabilistic methods differ from those discussed above in that they do not require the introduction to the model of artificialities such as zones of piecewise constancy for the sake of achieving a stable calibration process; in fact in many cases they do not require that a ''calibrated model'' actually exist, for no single set of parameters is deemed worthy of perpetual model usage. However in many cases they require either extensive modifications to the model for their implementation (for example for formulation and solution of stochastic differential equations), or large run times (for example to accomplish multiple deformations of stochastic ''seed fields'' to satisfy calibration constraints).

The present paper discusses and demonstrates a methodology through which the potential error associated with any prediction made by a calibrated model can be calculated. It is based on theory developed by Moore and Doherty (2005). Though designed for use as an adjunct to calibration achieved through regularized inversion, it can also be applied to models whose parameterization is based on zones of piecewise constancy or some other parameter lumping device. Its implementation does not require that the model be re-programmed; in fact, as demonstrated in our example, the model can be comprised of a number of executable programs run consecutively through a batch or script file. Furthermore, it takes account of contributions to model predictive uncertainty originating in both measurement noise, and in system complexity beyond that which can be represented with integrity in a calibrated model. Because the latter is often the dominant contributor to predictive error (especially for predictions involving ''fine system de-

tail'' or local scale), and because its contribution often diminishes with the amount of averaging involved in a prediction, the methodology can be used to inform modellers and managers of the predictive scale at which the model can be used with integrity, and of the costs in terms of heightened probability for error that are incurred if predictions are attempted at a smaller scale than this.

As is demonstrated herein, the methodology is easily extended to the investigation of related aspects of model parameterisation and performance. For example, the contributions made to the potential error of a particular prediction by different parameter types employed by a model can be evaluated, and/or the efficacy of different data acquisition strategies in reducing potential predictive error can be compared. The chief disadvantage of the methodology is that model linearity is assumed. Thus estimates of model predictive error variance are likely to be approximate rather than exact. However this does not detract from its usefulness in providing such approximations. Nor is it likely to degrade the quality of comparisons made between the uncertainties of predictions of different types, or of the same types with/without the inclusion of notional extra data in the calibration dataset for the purpose of assessing its worth prior to committing financial resources to its acquisition. Nor is it likely to invalidate estimates of the contributions made to predictive uncertainty by different parameter types. All of these quantities are extremely difficult to calculate without the assumption of model linearity.

Methodology

Calculation of predictive error variance

Let the vector $\mathbf{p}$ denote the ''true parameters'' (equivalent, for the present purpose, to the true hydraulic properties) of a simulated system. It is assumed that these properties are *represented* in a model with a degree of spatial and/or temporal detail that is commensurate with their true variability, or at least enough of their variability as is necessary to make important predictions with integrity. (As will be discussed below, they cannot necessarily be *estimated* to this level of detail, with the result that these predictions will indeed lack integrity.)

$\mathbf{p}$ is never known exactly. However something is known of it, albeit with uncertainty, from site characterisation studies which include, perhaps, direct measurements of hydraulic properties at certain points. Let $C(\mathbf{p})$ denote the covariance matrix of the parameters $\mathbf{p}$ determined through such studies (or estimated on the basis of expert knowledge or simply ''informed intuition''). $C(\mathbf{p})$ provides a description of the ''innate variability'' of the parameters $\mathbf{p}$, and the extent to which parameters of different types, or of the same type at different locations, are statistically interdependent. In some cases detailed geostatistical studies may allow accurate characterization of $C(\mathbf{p})$, though rarely is such information available to support groundwater model parameterisation. (Of course, to the extent that $C(\mathbf{p})$ is in error, then so too are estimates of possible predictive error; this is explored later in this paper. However the nature of predictive error is such that analysis of its magnitude by any means is impossible without making some assumption of the structure and magnitude of $C(\mathbf{p})$).

Let s characterize a model prediction of interest. Let the sensitivities of this prediction to the parameters $\mathbf{p}$ be encapsulated in the vector $\mathbf{y}$. Then, assuming model linearity and ignoring prediction and parameter offsets for the sake of simplicity, s is calculated from $\mathbf{p}$ using the equation

$$s = \mathbf{y}^t\mathbf{p} \tag{1}$$

Using basic matrix manipulation the variance of s, σ_s^2, can be calculated as

$$\sigma_s^2 = \mathbf{y}^t C(\mathbf{p})\mathbf{y} \tag{2}$$

Eq. (2) provides a full description of model predictive uncertainty if parameter estimates are not constrained through a calibration process (El Harrouni et al., 1997). However, most models employed in groundwater management are, in fact, calibrated. Suppose that an observation dataset $\mathbf{h}$ is used for calibration purposes, and that measurement noise ε is associated with this dataset, the covariance matrix of which is $C(\varepsilon)$. Then (once again assuming model linearity and neglecting parameter offsets)

$$\mathbf{h} = \mathbf{X}\mathbf{p} + \varepsilon \tag{3}$$

where the $\mathbf{X}$ matrix represents the action of the model under calibration conditions.

Suppose that the calibration process results in estimation of a parameter set $\underline{\mathbf{p}}$. It is assumed that some kind of regularization strategy is employed in the estimation of $\underline{\mathbf{p}}$ as model parameterization density is assumed to be high. As Moore and Doherty (2006) point out, the use of regularised inversion to overcome parameter detail inestimability leads to a calibrated parameter field $\underline{\mathbf{p}}$ that is a simplified (normally smoothed) form of the real parameter field $\mathbf{p}$. Nevertheless if the regularised inverse problem is properly formulated, this process will result in maximum information content being extracted from the calibration dataset $\mathbf{h}$. Also, as will be demonstrated, because the parameterisation detail on which predictions of interest depend is actually represented in the model, the uncertainty associated with those predictions can be properly analysed, even if the predictions themselves cannot be made with integrity due to lack of complete knowledge of $\mathbf{p}$.

A number of regularisation methodologies are commonly used in groundwater and other contexts for data interpretation and model calibration, including Tikhonov regularisation, truncated singular value decomposition, and the highly efficient ''SVD-assist'' scheme; see Doherty (2003, 2005), and Tonkin and Doherty (2005) for details.

Suppose that $\underline{\mathbf{p}}$ is calculated from $\mathbf{h}$ using the equation

$$\underline{\mathbf{p}} = \mathbf{G}\mathbf{h} \tag{4}$$

where $\mathbf{G}$ depends on the regularisation methodology employed. From (3) it follows that

$$\underline{\mathbf{p}} = \mathbf{G}\mathbf{X}\mathbf{p} + \mathbf{G}\varepsilon = \mathbf{R}\mathbf{p} + \mathbf{G}\varepsilon \tag{5}$$

where $\mathbf{R}$ is the so-called ''resolution matrix'' calculated as a by-product of the regularized inversion process. (Examples of $\mathbf{R}$ and $\mathbf{G}$ are provided later in the paper.)

The error in $\underline{\mathbf{p}}$ as an estimate of $\mathbf{p}$ can be calculated as

$$\mathbf{p} - \underline{\mathbf{p}} = (\mathbf{I} - \mathbf{R})\mathbf{p} - \mathbf{G}\varepsilon \tag{6}$$

where **I** is the identity matrix. Because **p** is unknown, so too is parameter error. However its covariance matrix, and hence its statistical structure, can be calculated from the covariance matrices of **p** and ε using the equation

$$C(\mathbf{p}-\underline{\mathbf{p}}) = (\mathbf{I}-\mathbf{R})C(\mathbf{p})(\mathbf{I}-\mathbf{R})^{t} + \mathbf{G}C(\varepsilon)\mathbf{G}^{t} \quad (7)$$

If the calibrated model is employed to make a prediction then, following Eq. (1), that prediction is calculated as

$$\underline{s} = \mathbf{y}^{t}\underline{\mathbf{p}} \quad (8)$$

Predictive error (which can never be known) is given by

$$s - \underline{s} = \mathbf{y}^{t}(\mathbf{p}-\underline{\mathbf{p}}) \quad (9)$$

From Eqs. (6) and (7) predictive error variance $\sigma^2_{s-\underline{s}}$ is calculated as

$$\sigma^2_{s-\underline{s}} = \mathbf{y}^{t}(\mathbf{I}-\mathbf{R})C(\mathbf{p})(\mathbf{I}-\mathbf{R})^{t}\mathbf{y} + \mathbf{y}^{t}\mathbf{G}C(\varepsilon)\mathbf{G}^{t}\mathbf{y} \quad (10)$$

The first term of Eq. (10) represents the contribution to predictive error variance incurred through the inability of the calibration process to ''capture'' more than a certain amount of parameterisation detail. The second term describes the contribution to error variance made by measurement noise. As discussed by Moore and Doherty (2005) the calibration process can be viewed as subdividing parameter space into two separate subspaces. Parameter combinations lying within the ''calibration null space'' are not informed by the calibration process. The uncertainty associated with these parameter combinations is unchanged from what it was prior to calibration, and is thus characterised completely by $C(\mathbf{p})$. For parameter combinations lying within the ''calibration solution subspace'' potential parameter error is reduced through the calibration process, for these parameter combinations are informed by the calibration dataset. However this information is ''contaminated'' to some extent by measurement noise. A prediction will, in general, depend on parameter combinations lying within both subspaces; hence its potential error is a combination of these two factors.

Note that the term $\sigma^2_{s-\underline{s}}$ is properly referred to as ''error variance'' rather than ''predictive uncertainty''. The latter term is a function solely of $C(\mathbf{p})$, $C(\varepsilon)$ and the action of the model; it must be calculated using some type of Bayesian analysis. The term ''error variance'', however, presupposes the existence of a calibrated model, parameterised with a single parameter set that will form the basis of future predictions of system behaviour.

Calculation of R and G matrices

Formulas for calculation of **R** and **G** pertaining to two different regularization strategies are now listed. See Doherty (2005) for more details and for formulas pertaining to other regularization mechanisms.

Tikhonov regularisation

$$\mathbf{R} = (\mathbf{X}^{t}\mathbf{Q}\mathbf{X} + \beta^2\mathbf{T}^{t}\mathbf{S}\mathbf{T})^{-1}\mathbf{X}^{t}\mathbf{Q}\mathbf{X} \quad (11a)$$

$$\mathbf{G} = (\mathbf{X}^{t}\mathbf{Q}\mathbf{X} + \beta^2\mathbf{T}^{t}\mathbf{S}\mathbf{T})^{-1}\mathbf{X}^{t}\mathbf{Q} \quad (11b)$$

Truncated singular value decomposition

$$\mathbf{R} = \mathbf{V}_1\mathbf{V}_1^{t} \quad (12a)$$

$$\mathbf{G} = \mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^{t}\mathbf{X}^{t}\mathbf{Q} \quad (12b)$$

Terms used in these equations are as follows:

- **X** is the sensitivity of model outputs corresponding to **h** to parameters **p**
- **Q** is the observation weight matrix
- $\mathbf{V}_1$ is a matrix whose columns are pre-truncation eigenvectors achieved through singular value decomposition of $\mathbf{X}^{t}\mathbf{Q}\mathbf{X}$ according to the equation

$$\mathbf{X}^{t}\mathbf{Q}\mathbf{X} = [\mathbf{V}_1\mathbf{V}_2]\begin{bmatrix}\mathbf{E}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{E}_2\end{bmatrix}[\mathbf{V}_1\mathbf{V}_2]^{t} \quad (13)$$

- $\mathbf{V}_2$ is a matrix whose columns are post-truncation eigenvectors
- $\mathbf{E}_1$ is the diagonal matrix of eigenvalues corresponding to $\mathbf{V}_1$
- $\mathbf{E}_2$ is the diagonal matrix of eigenvalues corresponding to $\mathbf{V}_2$
- **T** is a vector whose rows are regularisation constraints on parameter values
- **S** is the regularisation weight matrix
- β^2 is the regularisation weight factor (determined through the calibration process)

Where regularization is implemented using truncated SVD, and where a measurement weight matrix **Q** is provided which is proportional to the inverse of $C(\varepsilon)$ such that

$$C(\varepsilon) = \sigma_h^2\mathbf{Q}^{-1} \quad (14)$$

Eq. (10) becomes

$$\sigma^2_{s-\underline{s}} = \mathbf{y}^{t}\mathbf{V}_2\mathbf{V}_2^{t}C(\mathbf{p})\mathbf{V}_2\mathbf{V}_2^{t}\mathbf{y} + \sigma_h^2\mathbf{y}^{t}\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^{t}\mathbf{y} \quad (15)$$

Optimisation of predictive error variance

For most regularisation methodologies, one or a number of ''regularisation variables'' exist that determine the strength with which regularisation is applied in the inversion process. For Tikhonov regularisation it is the ''regularisation weight factor'' β^2 of Eqs. (11a) and (11b). For truncated singular value decomposition it is the number of singular values before truncation. If regularisation is too strongly enforced a poor fit between model outputs and field measurements is obtained; furthermore estimated parameter fields will be too smooth, possibly lacking detail that they may otherwise have captured through the calibration process. However if regularisation is too light, then the fit between model outputs and field data may be ''too good''. Such ''over-fitting'' can result in wildly erroneous parameter values that are more reflective of measurement noise than information contained in the calibration dataset. Moore and Doherty (2005) demonstrate that, for any model prediction, predictive error variance as calculated using Eq. (10) or (15) is minimized at an optimal level of regularisation between these two extremes.

Where off-diagonal components of $C(\mathbf{p})$ are zero, where parameters are scaled by their uncertainties (so that $C(\mathbf{p})$ becomes the identity matrix), and where the parameters represented in $C(\mathbf{p})$ are normally distributed and the model is linear, parameter sets achieved through truncated singu-

lar value decomposition are of optimal likelihood due to the fact that they approach minimum norm (because parameter projections onto the calibration null space are zero). It follows that if the error variance of a particular prediction is computed using Eq. (15) on the basis of scaled parameters at enough singular values to allow determination of the minimum error variance for that prediction, this minimized error variance constitutes a good approximation to the minimum that can be achieved by a calibrated model under any circumstances on the basis of the available calibration dataset. An assessment of the performance of a calibration process implemented by other means can then be made by comparing the error variance of various predictions computed on the basis of that calibrated model with minimized error variance as computed through a notional singular value decomposition exercise carried out as described above on scaled parameters.

Adding parameters to an existing model

Where a model is calibrated using a standard technique for regularised inversion, **R** and **G** matrices are readily available as an outcome of the inversion process through equations such as (11a) and (11b), and (12a) and (12b). However the situation will often arise (such as in the example presented below), where a model has not been calibrated using one of these standard methods. In these cases approximate **R** and **G** matrixes must be computed.

It will also often be necessary to add parameters to an existing model when undertaking predictive error variance analysis for that model. The purpose of adding these parameters is to allow better representation of the calibration null space in this analysis, and thus to avoid underestimation of the contribution made to predictive error variance by the first term of Eq. (10). New **R** and **G** matrixes pertaining to the expanded parameter set will result from this process.

Let $\mathbf{R}'$ and $\mathbf{G}'$ represent **R** and **G** matrices calculated on the basis of the parameter set $\underline{\mathbf{q}}$ employed by an existing model. Let $\underline{\mathbf{p}}$ represent an expanded parameter set, the values for the elements of $\underline{\mathbf{p}}$ being such that the model provides the same outputs under both calibration and predictive conditions when $\underline{\mathbf{p}}$ is employed in place of $\underline{\mathbf{q}}$. Let $\underline{\mathbf{p}}$ be calculable from $\underline{\mathbf{q}}$ using the relationship

$$\underline{\mathbf{p}} = \mathbf{L}\underline{\mathbf{q}} \tag{16a}$$

and $\underline{\mathbf{q}}$ be calculable from $\underline{\mathbf{p}}$ using the relationship

$$\underline{\mathbf{q}} = \mathbf{N}\underline{\mathbf{p}} \tag{16b}$$

Moore (2006) shows that selection of **L** does not provide a unique determination of **N**. However **N** and **L** must satisfy the relationship

$$\mathbf{N} = \mathbf{NLN} \tag{17}$$

Moore also shows that **R** and **G** for the expanded parameter set can be calculated from $\mathbf{R}'$ and $\mathbf{G}'$ for the smaller parameter set using the relationship

$$\mathbf{R} = \mathbf{L}\mathbf{R}'\mathbf{N} \tag{18a}$$

and

$$\mathbf{G} = \mathbf{L}\mathbf{G}' \tag{18b}$$

For zone based parameters complementary pairs of **N** and **L** matrices can be computed exactly. For pilot point parameters (Certes and de Marsily, 1991; Doherty, 2003) it is an easy matter to define pairs of **L** and **N** matrices such that Eq. (17) is obeyed. However definition of $\underline{\mathbf{p}}$ such that model outputs are unchanged from those of $\underline{\mathbf{q}}$ is approximate, with the approximation improving with the number of added pilot points. (Errors in the present analysis incurred by failure of an expanded parameter set to exactly match model outputs computed on the basis of an existing, more limited, parameter set are an outcome of model non-linearity – for if the model were perfectly linear, sensitivities computed on the basis of any parameter set would be the same; it is only sensitivities, and not model parameter values or outputs, that are employed in Eqs. (10) and (15). To the extent that non-linearity-induced errors in computed predictive error variance exist, these are expected to be incurred more by differences in model outputs between predictive and calibration conditions, than by the much smaller differences in model outputs between an existing and expanded set of pilot points.)

Some parameters of types that are declared as fixed during calibration of an existing model may be declared as adjustable for the purpose of predictive error variance analysis. The calibration process provides no information on such parameters; hence they belong to the calibration null space. Thus any contribution that their variability makes to the error variance of a particular prediction is described by the first term of Eqs. (10) and (15), with the submatrix of **R** pertaining to these parameters being **0**.

The addition of new parameters to a model for the purpose of predictive error variance analysis can be a very useful device for analysing the effects of assumptions made in assigning boundary or other conditions to models on potential errors associated with model predictions. In the example discussed below, lateral inflow to the model domain is one such quantity. Predictive errors associated with misassignment of this quantity cannot be tested purely through post-calibration sensitivity analysis (as is often attempted in practice), for such an analysis ignores the fact that values assigned to many model parameters through the calibration process may be dependent on the values assigned to these boundary conditions; if the latter change, then the former must change as well to maintain the model in a calibrated state. Use of Eq. (10) automatically takes this interdependence into account.

An example

The model

The model used in this example is typical of many models that are used for groundwater management. It was not built or calibrated by the authors of this paper; rather it was supplied to us with the request that the error margins of important predictions be calculated. It is a composite model, comprised of a groundwater model plus a number of ancillary lumped parameter soil water balance models which calculate demand for irrigation water (much of which is then extracted from the groundwater system) and recharge to that system. It was calibrated partly by hand and partly

using parameter estimation software. The spatial variability of some parameter types was characterised using pilot points, while for other parameter types zones of piecewise constancy were employed. Variability of soil water balance model parameters was effected through choice of model instance based on soil type and land use, rather than through explicit incorporation of spatial variability in definition of these parameters.

The Pioneer River Valley (see Fig. 1) is an area of irrigated agriculture in which sugar cane is the dominant crop. It is situated about 250 miles north of the Tropic of Capricorn in eastern Australia. Alluvial deposits associated with a number of creeks and streams comprise an aquifer from which about 40GL per year is drawn to support irrigation; irrigation is also drawn from surface water sources, for which a number of storages have been constructed. Both alluvial and non-alluvial areas are included within the model domain. Dominant non-alluvial rocks include granite and tuffaceous sediments of the Permian Camilla beds. Rocks within these non-alluvial areas are generally of low hydraulic conductivity and support use of water only for stock and domestic purposes. Recharge to the alluvial aquifer is from rainfall (of which there is an average of 1500 mm/year), irrigation, and river replenishment during times of high flow. However over most of the year the rivers and creeks which drain the area receive water from the groundwater system rather than supply it to that system.

The Pioneer Valley model was developed by Water Assessment Group, Queensland Department of Natural Resources and Water (NRW) to provide a basis for short and long term management of the area. Construction details of this model are provided in Kuhanesan et al. (2005). Background information on the characteristics of the model domain, its hydrogeology, details of surface water–groundwater interaction, recharge processes, and water management can be found in Murphy et al. (2005).

Groundwater flow within the Pioneer Valley is simulated using a single layer MODFLOW (Harbaugh et al., 2000) model comprised of 10,695 active cells, each of size 250 m × 250 m. Exchange of water with surface water bodies is simulated using the MODFLOW *river* package. Recharge processes and vegetative water use is simulated using a lumped parameter soil water balance model named SPLASH; see Arunakumaren (1997) for details. Different instances of SPLASH are employed in each of 10 rainfall zones into which the model domain is subdivided, for 4 land use types and 10 soil types in each such zone; one SPLASH instance is also employed for urban land. Simulated land use types are irrigated cane, non-irrigated cane, scrub/grassland/pasture, and rainforest. Recharge calculated by all SPLASH instances is transferred to relevant model cells (where it serves as input to the MODFLOW *recharge* package) by another submodel component named STRESGEN. The component of irrigation demand met by groundwater is extracted from the MODFLOW model using the latter's *well* package (with management of this transfer again being handled by STRESGEN). In all, 401 incidences of the SPLASH model comprise the soil water component of the Pioneer Valley model.

Lateral inflow is assumed to occur through parts of the side of the model domain; zero flow is assumed to occur over other lateral segments. This matter is further discussed below. Fixed heads are assigned to a number of small boundary segments in the west, northwest and southwest of the model domain (these representing inflow from upland alluvial areas), and along the coastal and estuary segments of the model boundary at its eastern edge.

Observations

The groundwater and soil water components of the Pioneer Valley model were calibrated separately, the former using PEST (Doherty, 2005) and the latter completely by hand.

The calibration dataset for the groundwater submodel was comprised of 3825 head measurements made in 147 bores over a five year period beginning in June 1998, as well as 60 monthly estimates of groundwater inflow to Sandy Creek. Measurement bore locations are depicted in Fig. 1; the reach over which inflows to Sandy Creek were measured (i.e. upstream of the gauging station) is depicted in this same figure.

Calibration of SPLASH submodels was effected in the following way:

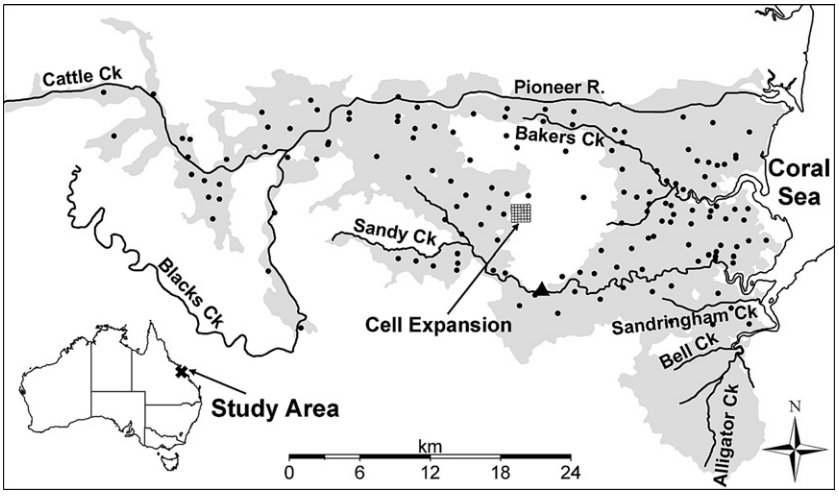

Figure 1 Locality map of the Pioneer Valley featuring Sandy Creek gauging station (triangle), measurement bores (circles) and alluvium (shaded grey). Model cell expansion for water table averaging is also shown.

- Estimates (obtained from soil chloride profile analysis) of long term ''leaching fraction'' (ratio of recharge to applied water) under irrigated cane for all soil types occurring within the Pioneer Valley were compared with leaching fractions calculated by SPLASH for these soil types over a 33 year simulation period spanning the years 1970–2003.
- Total yearly groundwater extraction volumes over the period 1999–2001 in part of the Pioneer Valley area where this pumping is metered, and where irrigation water is sourced solely from groundwater, were compared with total irrigation demands calculated by all SPLASH instances operating within this area over the same period.
- Estimates of total annual ground and surface water use in the Pioneer Valley based on cane production tonnages over the years 1998–2003 were compared with cumulative SPLASH-calculated irrigation demand over this same period.

Eqs. (10) and (15) do not employ observations directly in the calculation of predictive error variance. However they do employ the uncertainties associated with these observations as encapsulated in the $C(\varepsilon)$ matrix. In the present case $C(\varepsilon)$ was assumed to be diagonal, with the uncertainties pertaining to different observations thus assumed to be independent of each other. The magnitude of these uncertainties for each observation type discussed above was estimated on the basis of the nature of the observations themselves, as well as on the degree of model-to-measurement misfit achieved through calibration of the model. See Doherty and Gallagher (2005) for further details.

Note that in standard model calibration practice, the user selects a form for $C(\varepsilon)$ (usually diagonal) and then estimates a ''reference variance'' through the calibration process by which this covariance matrix is multiplied. As Moore (2006) shows, when regularised inversion is undertaken, determination of the magnitude of $C(\varepsilon)$ in this fashion is not as easy, as goodness of fit in model calibration can be played against acceptability of heterogeneity introduced to a calibrated parameter field to attain this fit; thus a reference variance for $C(\varepsilon)$ cannot be estimated separately from a reference variance for $C(\mathbf{p})$. Uncertainty in the estimation of $C(\varepsilon)$ is further exacerbated by the fact that the contribution to this term from model structural error is normally large and, as Cooley (2004) demonstrates, is likely to show a high degree of spatial correlation. Only for some aspects of model design – for example the use of zones of piecewise constancy to represent innately heterogeneous media – can the contribution to total structural error be computed; however, despite the fact that structural error originating in the necessity for a model to simplify reality is often the dominant contributor to $C(\varepsilon)$, it is generally ignored. It follows that in all calibration contexts selection of an appropriate $C(\varepsilon)$ will be somewhat approximate. It also follows that the greater the extent to which structural contributions to $C(\varepsilon)$ can be reduced through the use of a multiplicity of parameters in conjunction with regularised inversion as a calibration device (thereby endowing the calibration process with the capacity to transfer any information on spatial heterogeneity that may be present within the data to the calibrated parameter field), the more is the magnitude of $C(\varepsilon)$ likely to be reduced.

Parameters

A total of 198 pilot points were employed for representation of hydraulic conductivity in the Pioneer Valley model. Their locations are shown in Fig. 2a. These were assigned to two different zones depending on whether or not they represent alluvium; neither spatial interpolation from pilot points to the finite difference grid, nor regularisation constraints (see below) were enforced across zone boundaries. The calibrated hydraulic conductivity distribution is shown in Fig. 2b. Six hundred and eight new pilot points were added to the model to allow better representation of hydraulic property detail for the purpose of predictive error variance analysis. New pilot points are shown in Fig. 2a.

A total of 471 pumping tests have been carried out in both alluvial and non-alluvial materials within the model area. (Most of these tests were between 6 and 24 h in duration and analysed using a modified Sternberg method as documented in Eden and Hazel, 1973). A variogram of log hydraulic conductivity within alluvial material was constructed on the basis of values inferred from these tests (insufficient tests had been carried out in other areas for construction of a variogram). Fig. 3 shows the variogram. It is apparent from this figure that only limited spatial correlation is evident. Visual inspection of pumping test results from non-alluvial areas suggest the same for these as well. Hence the submatrix of the $C(\mathbf{p})$ matrix used to characterize spatial variability of log hydraulic conductivity was assumed to be diagonal. Based on the analyses of pumping tests, variances (square of standard deviation) of 0.25 and 0.42 were assigned to pilot point log hydraulic conductivity parameters within alluvial and non-alluvial areas respectively. (It should be noted that the use of pilot points to characterize hydraulic conductivity, combined with the fact that grid-based hydraulic conductivities employed by the model are spatially interpolated from these, results in some spatial correlation between the latter, notwithstanding the assumption of no spatial correlation between pilot point parameters themselves.)

Fig. 4a shows the 147 pilot points employed for representation of log specific yield in the Pioneer Valley model; as for hydraulic conductivity, these were subdivided into alluvial and non-alluvial parameters. The calibrated distribution of specific yield is shown in Fig. 4b. A total of 356 new specific yield pilot points were added to the model for the purpose of predictive error variance computation. New pilot points are shown together with old pilot points in Fig. 4a. The $C(\mathbf{p})$ submatrix pertaining to log specific yield was assumed to be diagonal, a variance of 0.3 being assumed for alluvial material, and a variance of 0.6 being employed for non-alluvial material.

In calibration of the Pioneer Valley model, riverbed hydraulic conductivity was represented by 44 zones of piecewise constancy. A pre-processor calculated riverbed conductance on the basis of these conductivity parameters (which were log-transformed for the purpose of parameter estimation) and river length and width within each model cell. Eight new zones were added to increase spatial parameterisation

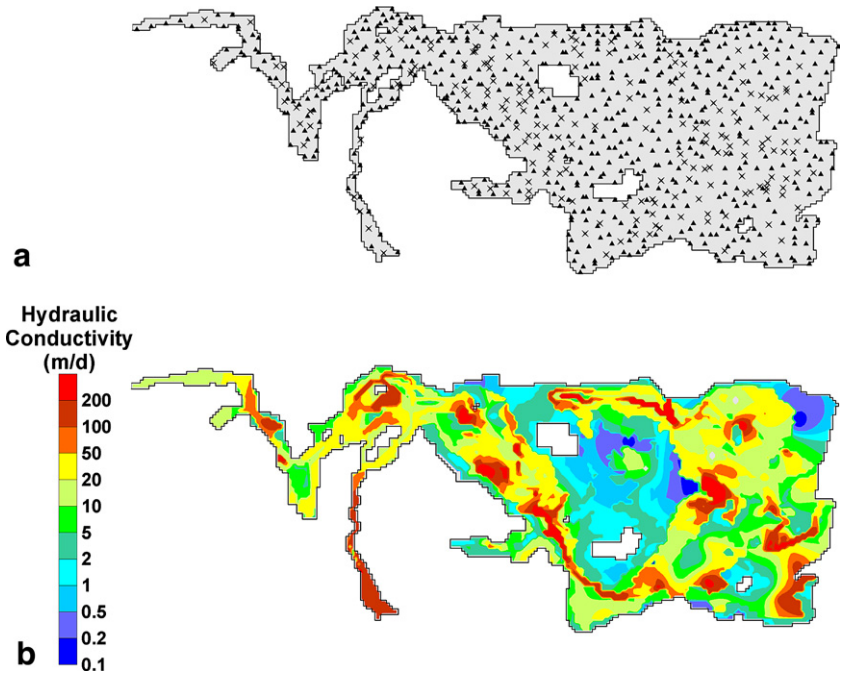

Figure 2 (a) Locations of original (crosses) and new (triangles) pilot points used for representation of hydraulic conductivity; (b) Distribution of calibrated hydraulic conductivity.

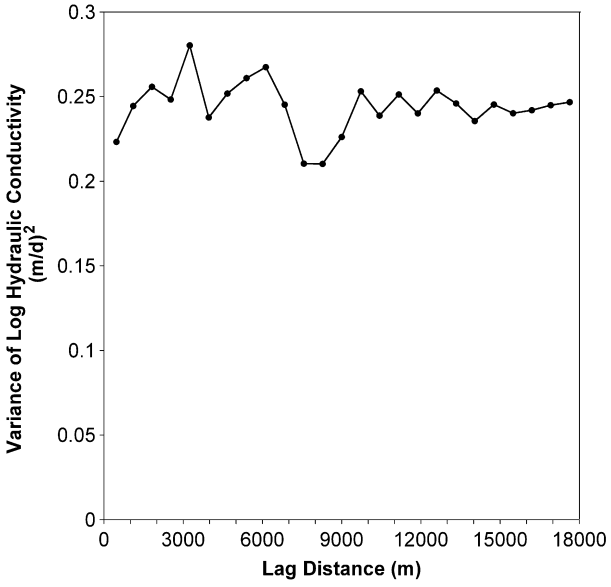

Figure 3 Variogram of log hydraulic conductivity in alluvium.

detail for the purpose of predictive error variance analysis. All 52 zones are depicted in Fig. 5. For the purpose of building the $C(\mathbf{p})$ submatrix pertaining to these parameters, they were assumed to be independent and characterized by a log variance of 3.0 (reflecting the fact that very little is known about spatial variability of this parameter, except that it is likely to be high based on geological and riverbed morphology considerations).

Drain cells with high conductance values were employed to represent surface seepage of groundwater from some non-alluvial areas. Much of these areas are waterlogged much of the time due to the low hydraulic conductivities of the rocks which underlie them. Drain conductances were not treated as parameters in the present study as they are very high in order to allow unobstructed seepage of groundwater where the water table intersects the land surface.

Two new parameter types were added to the Pioneer Valley model that were not considered parameter types for calibration purposes. The benefits of this strategy have already been discussed. Lateral inflows to parts of the model domain boundary were calculated on the basis of regional recharge estimates, local hydraulic gradients and estimates of the hydraulic conductivities of rocks lying just outside of the model domain. For the purpose of predictive error variance analysis, uncertainties were assigned to these inflow estimates, and to the possibility that non-zero inflow occurs over other parts of the model domain where no such inflow was assumed to exist in the original model. A total of 42 lateral inflow parameters were defined based on the zonation shown in Fig. 6. Diagonal elements of the $C(\mathbf{p})$ submatrix pertaining to these zones were assigned on the basis of expected uncertainties in lateral inflow estimates (about 10% of estimated inflow for non-zero inflows, and about 1% of peak inflow for zero inflows.)

Another assumption that was ''built in'' to the original model, but was tested during the predictive error analysis process documented herein, was the head assigned to fixed head cells along some of the model boundaries. Along the coastal and estuarine boundaries on the eastern edge of the model head uncertainty arises from lack of exact knowledge of average estuary water levels, and from a ''tidal overheight affect'' incurred by daily and seasonal sea and estuarine water level variations and the breaking of waves

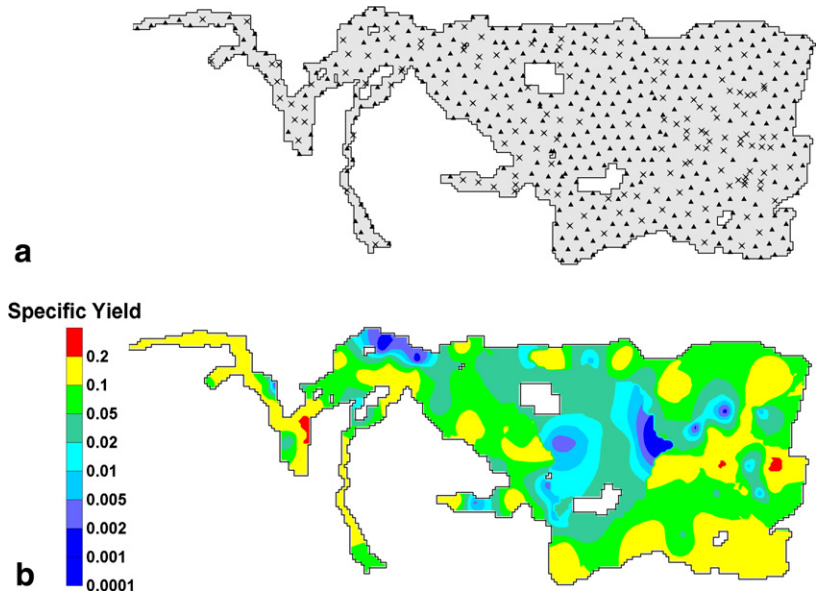

Figure 4 (a) Location of original (crosses) and new (triangles) pilot points for representation of specific yield; (b) Distribution of calibrated specific yield.

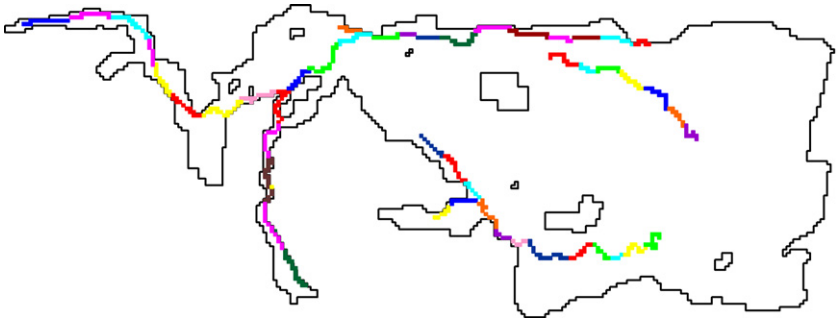

Figure 5 Riverbed hydraulic conductivity parameter zones.

on sandy shores of moderate slope. A total of 23 fixed head parameters were introduced to the model domain, their locations being shown in Fig. 7. The model was modified so that as the values of these parameters were varied for the purpose of sensitivity calculation, so too were initial heads in adjoining parts of the model domain so that equilibrium at these boundaries was maintained. See Doherty and Gallagher (2005) for more details. A variance of 0.09 m^2 was assumed for coastal and estuarine fixed head parameters while a variance of 1 m^2 was assumed for upland fixed head parameters representing inflow from creek alluvial sediments not represented in the model.

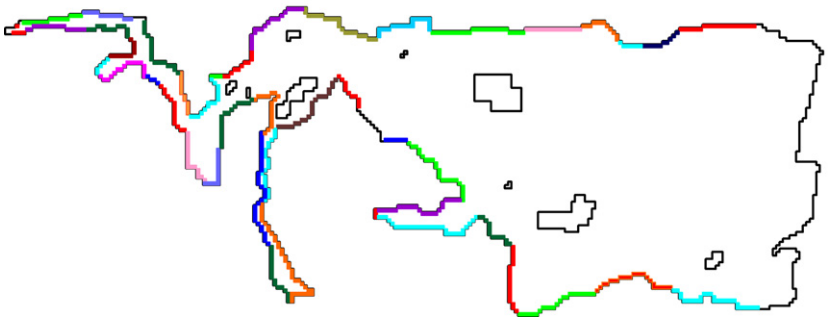

Figure 6 Lateral inflow parameter zones.

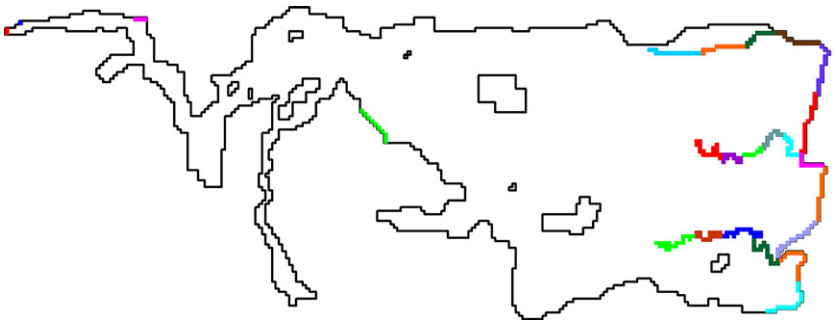

Figure 7 Fixed head parameter zones.

A summary of soil water model parameters involved in the predictive error analysis process is now presented. It is not essential that the role of each of these be fully understood in order for the reader to appreciate the nature of the error variance computation methodology which it is the central purpose of the current paper to convey; hence the following discussion is brief. The interested reader is referred to Doherty and Gallagher (2005) for further details.

A total of 158 soil water model parameters were declared as adjustable for the purpose of predictive error variance analysis. These were as follows:

- Eleven ''capillary moisture store volume'' parameters. One such parameter was employed in all SPLASH incidences operating on any one of the 10 soil types (plus urban land use) occurring within the model domain. This parameter affects soil drainage as well as the demand for irrigation water.
- Eleven ''gravity moisture store volume'' parameters — one for each soil type. This parameter affects soil drainage.
- Eleven ''saturated vertical hydraulic conductivity'' parameters — one for each soil type. This parameter affects rate of drainage of water from saturated soil to the water table.
- Eleven ''power factor'' parameters — one for each soil type. This parameter affects drainage of water below the root zone as a function of soil water content.
- Forty-one ''alpha'' and forty-one ''beta'' parameters, one for each vegetation type/non-urban soil type combination, and one for urban land use. These parameters affect the reduction in evapotranspiration that occurs as the water content of the soil decreases.
- Seven ''crop coefficient'' parameters. These determine the ratio of crop evapotranspiration to class ''A'' pan evaporation. One such parameter was employed for each vegetation type but sugar cane. Two were required for sugar cane to describe its seasonal variation. One was required for urban areas.
- Twenty ''farm management factors''. These determine the ratio between water demanded and water actually used in sugar cane areas. One such parameter was employed for each of irrigated and non-irrigated sugar cane growing on each soil type.
- Five other parameters required to accommodate certain features of Pioneer Valley water management that are described in Doherty and Gallagher (2005).

Based on advice received from those experienced in its use, values were assigned to elements of $C(\mathbf{p})$ corresponding to these different SPLASH parameter types; these values are thought to be reflective of the ''innate variability'', or pre-calibration uncertainty, of these parameter types under conditions prevailing in the study area.

Calculation of R and G matrices

As has already been discussed, calculation of **R** and **G** matrices is easily undertaken using equations such as Eqs. (11) and (12) where a model is calibrated using a standard method of regularised inversion. In the more common case where it is not, but where predictive error variance estimation is nevertheless required, approximate **R** and **G** matrices must be computed.

Because the groundwater and soil water model components of the Pioneer Valley model were calibrated independently of each other, their **R** and **G** matrices are also independent. Let $\mathbf{R}_g$ and $\mathbf{G}_g$, and $\mathbf{R}_s$ and $\mathbf{G}_s$ pertain to groundwater and soil water submodels respectively. **R** and **G** matrices for the entire model are then calculated as

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_g & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_s \end{bmatrix} \tag{19a}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_g & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_s \end{bmatrix} \tag{19b}$$

Tikhonov regularisation was employed in calibration of the groundwater submodel component of the Pioneer Valley model by those who built the Pioneer Valley model. Regularisation constraints encapsulated in the **T** matrix of Eq. (11) were of the ''preferred difference'' type. Thus each parameter was linked to one or more of its neighbours by a relationship of the type

$$\log(p_i) - \log(p_j) = 0.0 \tag{20}$$

Relative weighting of these relationships was determined using the method described by Doherty (2003) in which greater weights are applied to differences between parameters which are closer together than those which are further apart. A total regularisation weight factor (β^2 of Eq. (11)) was calculated in accordance with the modellers' desire to achieve a level of model-to-measurement fit considered compatible with measurement noise; see Doherty (2003) and De Groot-Hedlin and Constable (1990) for details. Once $\mathbf{R}'$ and $\mathbf{G}'$ matrices pertaining to the original parameter set

were computed using Eq. (11), $\mathbf{R}_g$ and $\mathbf{G}_g$ were computed for the expanded parameter set employed for error variance analysis using Eqs. (18a) and (18b).

Computation of $\mathbf{R}_s$ and $\mathbf{G}_s$ was less straightforward as the soil water submodel was calibrated by hand. Hence a notional calibration of this submodel (i.e. one iteration of a least squares parameter estimation process in which sensitivities were calculated on the basis of manually derived parameters) was carried out using truncated SVD with truncation at six singular values. Application of Eq. (12) then yielded $\mathbf{R}_s$ and $\mathbf{G}_s$. Selection of this number of singular values was somewhat arbitrary, its selection being based on the fact that the fit between SPLASH outputs and calibration data was not particularly good, leaving one with the impression that less information was captured from the data than could have been captured in a more formal calibration exercise undertaken under software control. However at six singular values, the root mean squared difference between SPLASH outputs and the measurement dataset was roughly the same as that achieved through manual calibration. Thus it is hoped that error variances computed through this mechanism are commensurate with those associated with the actual model; in practice they are likely to be somewhat lower due to the error variance minimization characteristics of singular value decomposition as a regularisation device.

Total **R** and **G** matrices for the entire Pioneer Valley model were then calculated using Eqs. (19a) and (19b).

For a number of the predictions examined during the current study, additional, ''optimal'' **R** and **G** matrices were computed through minimization of σ^2_{s-s} with respect to truncation limit using Eq. (15). It is to be noted, however, that **R** and **G** matrices computed in this manner are prediction-dependent, for the truncation limit may be different for different predictions. This was found to be the case in the current study; however for many of the predictions examined the minimum occurred at about 290 singular values.

The fact that the calibration dataset has enough information to warrant the inclusion of 290 degrees of freedom in parameter space for minimisation of the error variance of many model predictions demonstrates the benefits of regularised inversion in calibration of models of this type in areas of simular or greater data availability. For 290 degrees of freedom to be achieved, a much larger number of parameters should be represented in the model. The ability to accommodate such a large number of parameters is only forthcoming through the use of regularised inversion as a calibration methodology.

Predictive simulation

All predictions discussed herein were calculated using a model run of 60 months duration with monthly stress periods. Measured rainfall and evaporation spanning the period 1st July 1900 to 1st July 1905 was employed as input to the composite soil water model which, in turn, calculated irrigation demand and recharge for the groundwater model. The amount of land devoted to sugar cane production was assumed to be that prevailing during the 2003/2004 water year.

Predictions discussed below pertain to the end of stress period 54, this occurring 6 months before the end of the simulation, and coinciding with the end of a long dry period. Hence the analysis presented herein provides information on the capacity of the model to make useable predictions in times when water availability is low, but demand is high.

Only two types of prediction will be reported herein, viz. heads and stream inflow from groundwater. For application of the analysis to other types of model prediction see Doherty and Gallagher (2005).

Water levels

Fig. 8a shows water levels computed over the domain of the Pioneer Valley model at the end of stress period 54. In broad terms the water table rises to the west. However in finer detail (which is difficult to see at the scale of the figure) cell-by-cell water table variations are induced by spatially variable pumping.

The standard deviation of predicted head error in every cell of the model domain is contoured in Fig. 8b. This error is locally as high as 10 m, with largest errors occurring where calibration data is scarce, both in the central, low-conductivity, non-alluvial parts of the model domain and in the area at the central north of the model domain where unknown amounts of stream gains or losses could have a large effect on water levels. Over large parts of the alluvium, head error standard deviation is 2—3 m, though this lessens to the east toward the fixed head boundaries (where pumping is also less). Also of interest are a number of small areas of almost zero head error variance situated in the central part of the model in non-alluvial areas. These are areas of low hydraulic conductivity and high MODFLOW drain conductance where the water table is at the surface (and hence where water is flowing from MODFLOW *drains*). In these areas the water table elevation is determined by the land surface elevation; the uncertainty associated with predicted groundwater levels is thus minimal.

Predicted head error variance, and predicted head error variance contributions from the first and second terms of Eq. (10) are shown in Fig. 9a—c. (Note that variance, unlike standard deviation, is an additive quantity.) It is apparent from these figures that the contribution to predictive error variance made by the second term of Eq. (10) is much smaller than that of the first term. Thus the chief contributor to predictive error for this predictive type is not measurement noise per se, as much as the inability of the calibration process to ''capture'' sufficient hydraulic property details to make precise cell-by-cell head predictions. This applies especially in areas of historical data scarcity, and in areas of high groundwater pumping where the response of the water table to local pumping is very dependent on local hydraulic property details.

Minimized standard deviations of model-predicted head errors over the entire model domain, calculated on the basis of optimized **R** and **G** matrices, are shown in Fig. 10a. The difference between actual predictive standard deviations and optimized predictive standard deviations is shown in Fig. 10b. While differences of between 0.5 m and 1.0 m are common, there are a few areas where the difference is greater than this.

It is commonly stated that predictions made by a model such as the Pioneer Valley model cannot be expected to have precision on a cell-by-cell basis, but that average

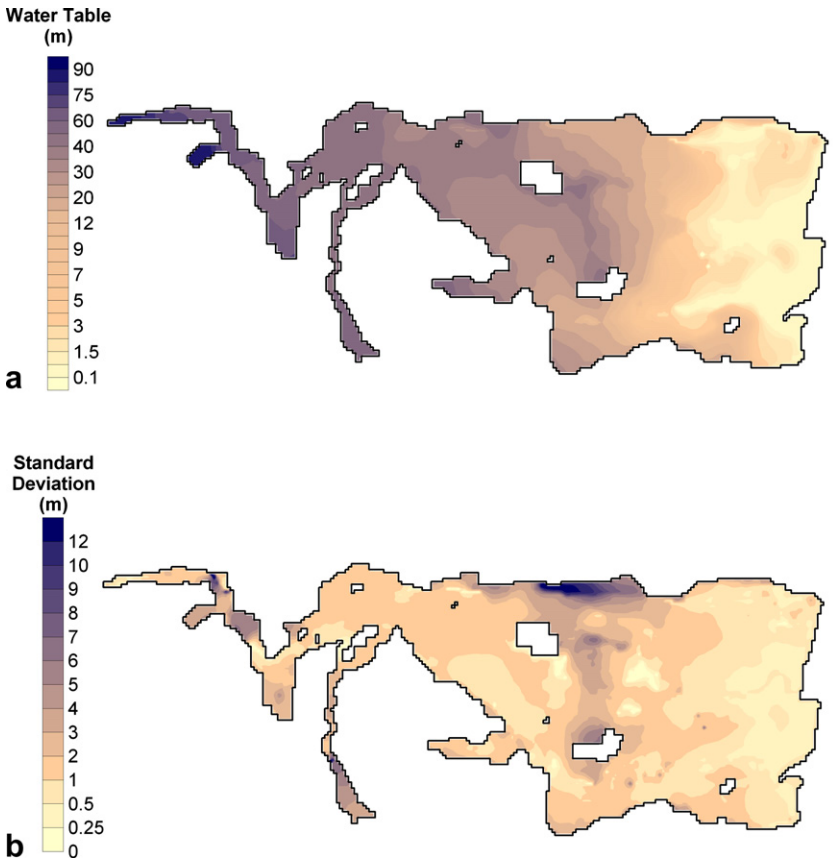

Figure 8 (a) Model predicted water table elevation, and (b) contoured standard deviation of predictive error in water table elevation.

heads over areas considerably larger than that of an individual cell are likely to be predicted with greater precision. To test this hypothesis, average heads, and the potential errors associated with these average heads, were calculated over a number of expanding square areas, starting with 1 cell and increasing to 36. It was found that in some cases predictive error variance decreases whereas for others it increases, depending on conditions encountered by the expanding averaging area. In general, if extra pumping wells are encountered, and if these wells are strongly pumped, then error variance tends to rise rather than fall with increasing area of averaging. However other factors can also contribute. The location of one of the cell expansions that was investigated is depicted in Fig. 1. Fig. 11 shows a plot of error variance versus number of cells for both actual and optimized **R** and **G** matrices. Expansion takes place from the top left to the bottom right of the square depicted in Fig. 1 as cell numbers are increased from 1 to 36. The expansion takes place in a non-alluvial area from a cell that is immediately adjacent to alluvium (where hydraulic property variance terms in the $C(\mathbf{p})$ matrix are lower than those for non-alluvial areas) into an area of zero density of calibration observations.

Fig. 12a shows the contributions to optimized error variance from the first and second terms of Eq. (15) plotted against singular value number for each of the six different head predictions comprising averages over 1, 4, 9, 16, 25 and 36 cells. As expected, the first (null space) term decreases with increasing singular value whereas the second (measurement noise) term increases. The total error variance (Fig. 12b) shows a minimum, in this case at about 290 singular values. As explained by Moore and Doherty (2005) the variance at zero singular value corresponds to the square of pre-calibration predictive uncertainty. These plots show that the reduction in predictive error variance attained through calibration is greater for smaller averaging areas than for larger average areas. This is a direct result of proximity to alluvium and head measurement locations for cells employed in low-area head averaging in this particular example.

Fig. 13 shows the contribution to the optimized error variance of the predicted average head over nine cells by different parameter types employed by the Pioneer Valley model. These contributions were calculated by fixing pertinent parameter types (thereby assuming perfect knowledge of them) and repeating the predictive error variance minimization process based on Eq. (15). The abbreviations employed for the different parameter types cited in this plot are listed in Table 1. Five parameter types dominate pre-calibration predictive uncertainty (the back row variances depicted in this plot), two of which are SPLASH parameters and three of which are MODFLOW parameters. The calibration process reduces to almost zero the contributions of all of these but hydraulic conductivity and specific yield. (It must be remembered that the examined prediction was made at the end of a dry period. If it had been made at

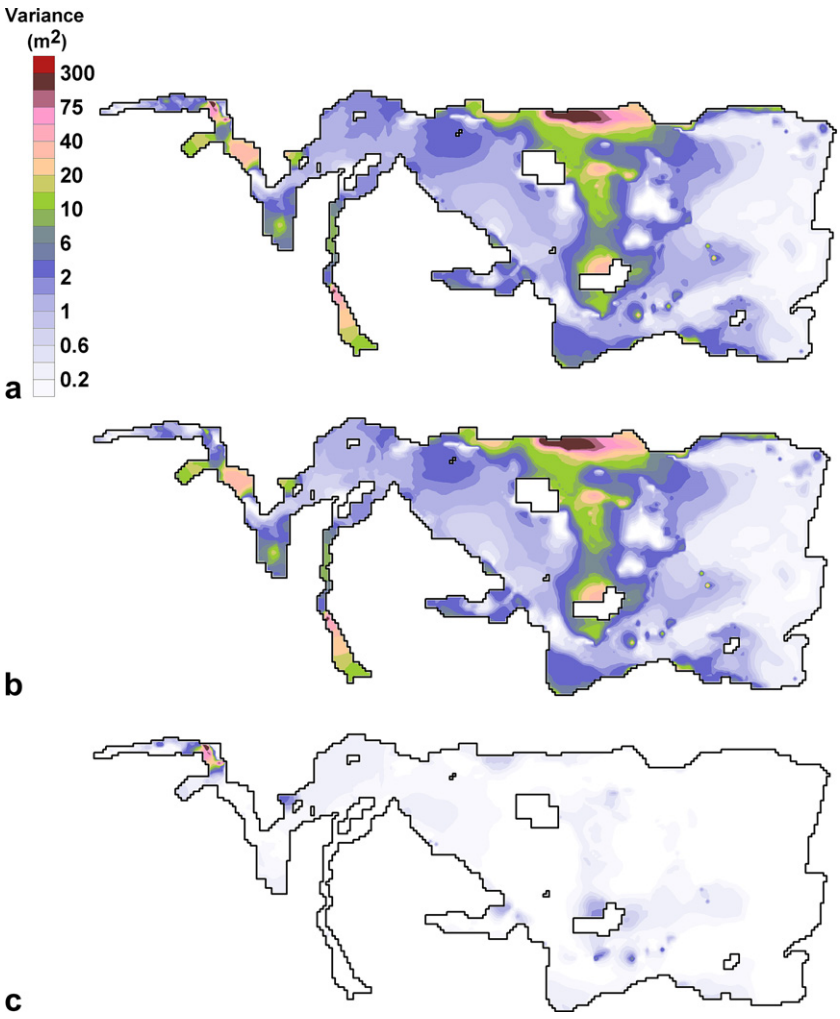

Figure 9 (a) Error variance of water table elevation predictions; (b) contribution from first term of Eq. (10) to predicted water table error variance; (c) contribution from second term of Eq. (10) to predicted water table error variance.

the end of a wet period, the contribution by SPLASH parameters may have been much higher.)

River inflow

Fig. 14 depicts predicted inflow on a cell-by-cell basis to Sandy Creek plotted against cell number for the segment of MODFLOW *river* cells upstream of the gauging station shown in Fig. 1; cell numbering begins at the upstream end of Sandy Creek. Also shown in Fig. 14 is riverbed conductance and cell inflow error standard deviation. It is apparent that calibrated riverbed conductance varies considerably along the length of the reach. Predictive error standard deviation is dependent on cell conductance, with smaller errors being calculated in areas of smaller cell conductance. This may, in fact, be an outcome of the linearity assumption on which Eq. (10) is based. It is of interest to note that the standard deviation of cell inflow predictive error is many times higher than the actual inflow in such low conductance regions, but is of the same order of magnitude as inflow in high conductance regions.

In order to explore the affects of predictive averaging, average cell inflow to Sandy Creek was calculated for an increasing number of cells, starting from 1 and increasing in groups of 5 downstream to a total of 26 cells along the creek segment indicated in Fig. 14. The error variance of average Sandy Creek inflow is depicted in Fig. 15 for different numbers of cells over which averaging takes place. A decrease in predictive error variance with increasing averaging is readily apparent from this figure.

Sensitivity analysis

Errors in the assessment of predictive error variance can arise from a number of sources. Some of these are inherent in any methodology through which predictive error variance is computed, most notably selection of suitable $C(\mathbf{p})$ and $C(\boldsymbol{\varepsilon})$ matrices to employ in the analysis. (No analysis of error variance which recognises the inherent complexity of a natural system can escape the use of some kind of stochastic descriptor of this variability; nor can it forego characterisation of the possible noise associated with measurements from which system parameters are inferred.) A benefit of the methodology described herein is that computations of predictive error variance can be readily repeated with different incidences of $C(\mathbf{p})$ and $C(\boldsymbol{\varepsilon})$ so that the robustness of error variance calculations in the face of limited knowledge of $C(\mathbf{p})$ and $C(\boldsymbol{\varepsilon})$ can be tested. This is undertaken

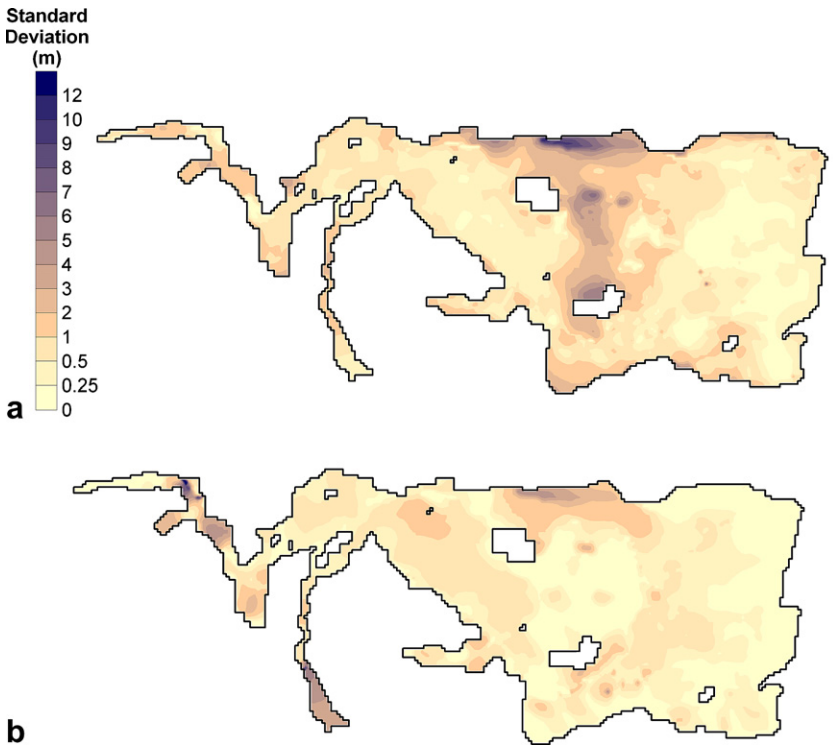

Figure 10 (a) Contoured standard deviation of predictive error in water table elevation for optimal regularization; (b) contoured difference between actual predictive standard deviation and optimized predictive standard deviation of water table elevation.

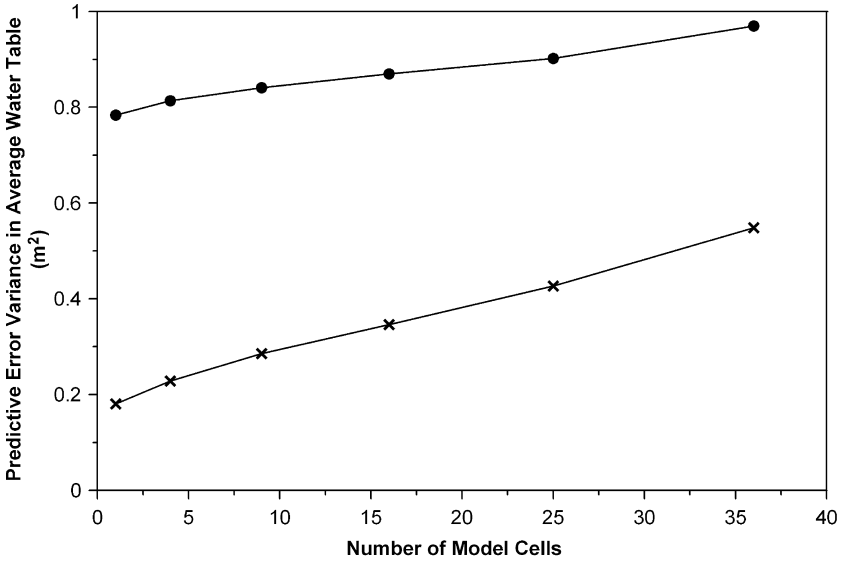

Figure 11 Actual (circles) and optimized (crosses) predictive error variance in average water table elevation for five areal expansions.

herein for average water table predictions made at an expanding number of cells as previously outlined. Whilst not shown for brevity, a similar analysis performed on Sandy Creek inflow predictions produces similar outcomes.

Fig. 16a shows the optimized variance of predicted average water level versus number of model cells for the following cases:

- the same $C(\mathbf{p})$ and $C(\varepsilon)$ matrices as were employed in the previous analysis;
- a $C(\varepsilon)$ in which all terms are doubled;
- a $C(\varepsilon)$ in which all terms are halved;
- a $C(\varepsilon)$ in which both temporal and spatial correlation exists for measurements taken in the 12 closest wells to the cell expansion area, but for which diagonal terms

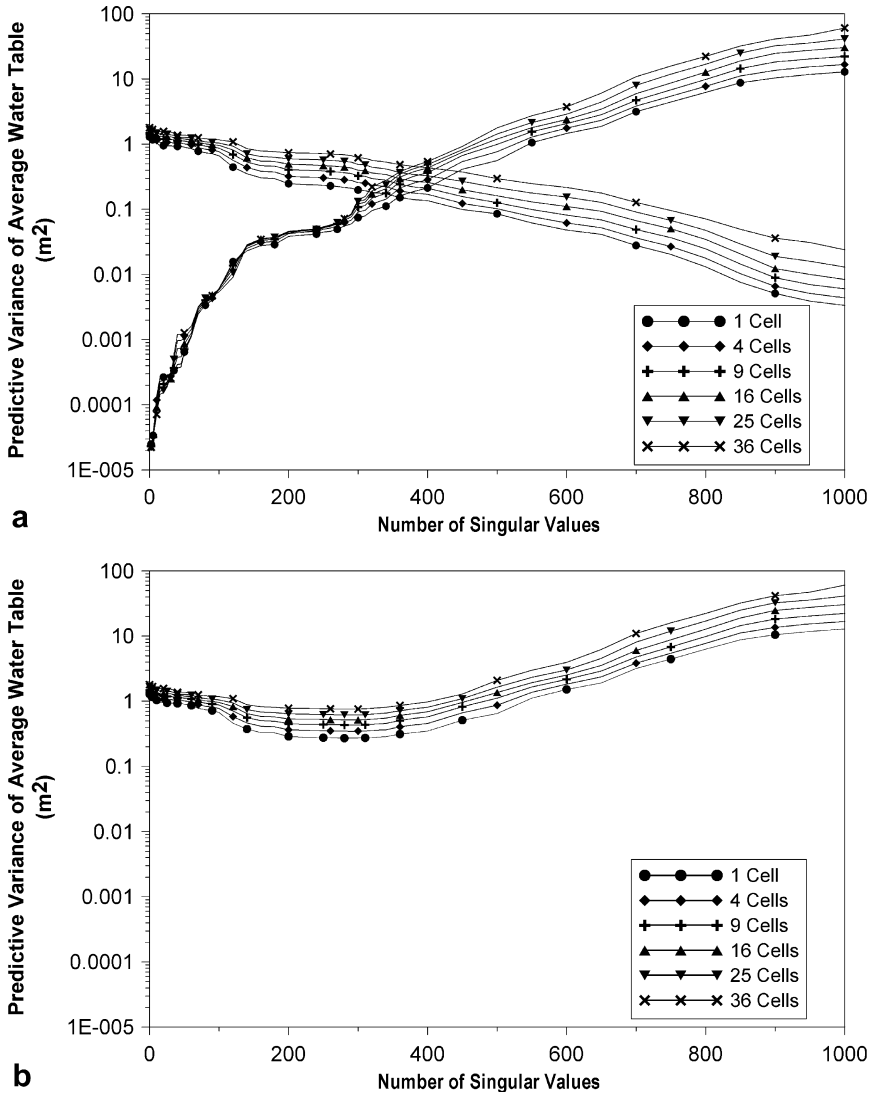

Figure 12 (a) Contributions to the optimized error variance of averaged water table elevation from the first and second terms of Eq. (15). Downward sloping lines are the null space term, upward sloping lines are the measurement noise term; (b) total error variance of averaged water table elevations.

are the same as in the previous analysis. (For simplicity, correlation between measurements decays with distance from the diagonal of the $C(\varepsilon)$ matrix with a decay rate of 0.95 per element; this simplistic correlation scheme is employed solely to demonstrate the effect of a high level of correlation between measurement errors on computed predicted error standard deviations, and is not meant to be a valid representation of true measurement noise correlation.)

Fig. 16b shows the optimized variance of predicted average water level versus number of model cells for the following cases:

- the same $C(\mathbf{p})$ and $C(\varepsilon)$ matrices as were employed in the previous analysis;
- a $C(\mathbf{p})$ in which all terms are doubled;
- a $C(\mathbf{p})$ in which all terms are halved;
- a $C(\mathbf{p})$ in which hydraulic conductivity and specific yield have the same diagonal elements as in the previous analysis, but for which spatial correlation shows an exponential decay with distance, the decay constant being 1000 m.

It is apparent from Fig. 16, that the error variance of predicted head is somewhat sensitive to assumptions pertaining to $C(\mathbf{p})$ and $C(\varepsilon)$, but not excessively so. The trend of increasing error variance with increased head averaging area is maintained. Further analysis (not shown herein) demonstrates that computations of parameter contributions to predictive error variance, and the extent to which these are reduced through the calibration process, are reasonably insensitive to the shape and magnitude of $C(\mathbf{p})$ and $C(\varepsilon)$ when varied by the same amount.

Another source of error in the computations presented herein, and one of potentially greater concern as it is an

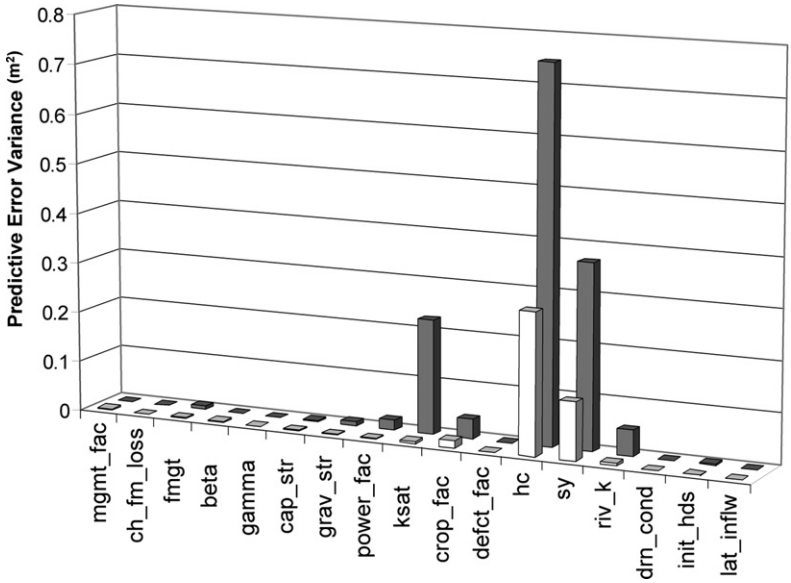

Figure 13 Contributions made by different parameter types to the optimized error variance of the predicted average water table elevation for a 9 cell expansion.

Table 1 Explanation of parameter types appearing in Fig. 13

Parameter type	Description
mgmt_fac	Coastal and inland restriction factors (—)
ch_fm_loss	Channel and farm loss factor (—)
fmgt	Farm management factor (—)
beta	Consumptive use parameter (—)
gamma	Consumptive use parameter (—)
cap_str	Capillary moisture store volume (mm)
grav_str	Gravity moisture store volume (mm)
power_fac	Percolation power factor (—)
ksat	Vertical hydraulic conductivity (mm/d)
crop_fac	Monthly crop factors (—)
defct_fac	Monthly crop deficit factors (—)
hc	Horizontal hydraulic conductivity (m/d)
sy	Specific yield (—)
riv_k	Riverbed hydraulic conductivity (m/d)
drn_cond	Drain conductance (m^2/d)
init_hds	Initial heads (m)
lat_inflw	Lateral inflows (m^3/d)

The first 11 of these are SPLASH parameter types. The others are groundwater model parameter types.

outcome of the methodology itself rather than the intrinsic nature of predictive error, is the effect of model non-linearity. When undertaking linear error variance analysis as an adjunct to solution of an over-determined inverse problem, a Beale type analysis is often employed to characterise the possible effects of model non-linearity on these analyses; see for example Christensen and Cooley (2004) and references cited therein.

A Beale's measure suitable for use in the regularised inversion context could be developed, for example by sampling parameter space on the basis of $C(\mathbf{p} - \underline{\mathbf{p}})$ as calculated using Eq. (7) rather than the parameter covariance matrix determined through traditional over-determined parameter estimation. Through exploration of departures between outcomes of linear and non-linear analysis in a variety of modelling contexts, a Beale's measure threshold could then be proposed above which the results of linear predictive error variance analysis could be considered to be questionable. This was not done in the present investigation because

- it would be a lengthy exercise which is beyond the scope of the present study;
- there are many aspects of an analysis of the present type which may be of use even if exact representation of predictive error variance is compromised; these aspects include contributions to predictive error variance by different parameter types, contributions to total predictive error variance by null space and measurement noise components, etc.

Nevertheless, in a simple numerical experiment, undertaken in order to explore the potential magnitude of non-linearity-induced errors on computed predictive error variance, the following steps were taken.

- Each model parameter was increased by one standard deviation from its calibrated value, with standard deviation being computed on the basis of $C(\mathbf{p} - \underline{\mathbf{p}})$.
- Sensitivities of all calibration and predictive model outputs were computed with all parameters simultaneously perturbed in this fashion.
- The predictive error variance of the average head over the expansion area of Fig. 1 was computed.
- The process was repeated with all parameters simultaneously decremented by one standard deviation.

Fig. 17 shows the error variance of predicted average heads at various levels of expansion for both of these cases, as well as for the original case depicted in Fig. 11. In this rather harsh test of the effects of non-linearity (for the model

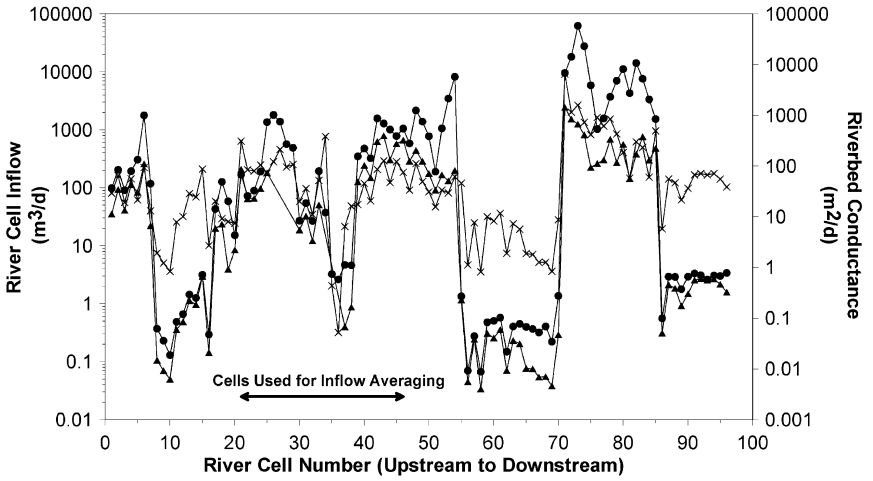

Figure 14 Predicted river cell inflow (triangles), standard deviation of river cell inflow error (crosses) and riverbed conductance (circles) for Sandy Creek.

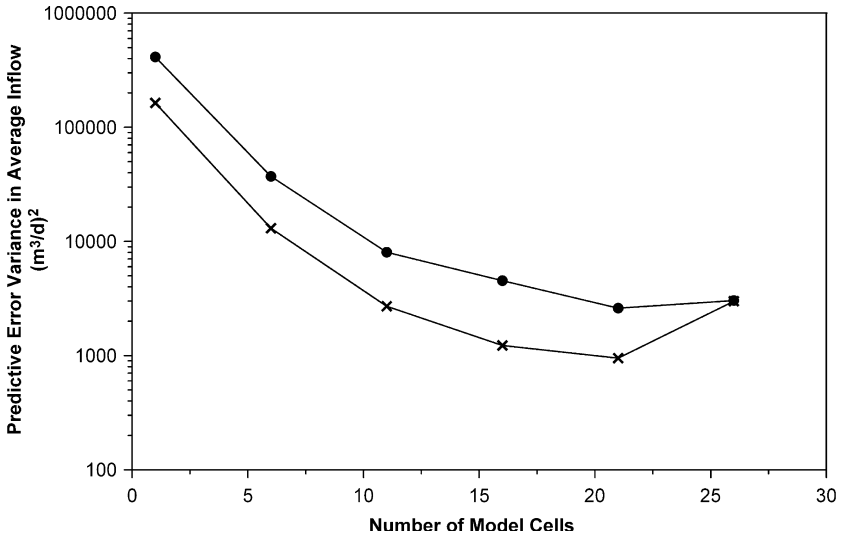

Figure 15 Actual (circles) and optimized (crosses) predictive error variance of average Sandy Creek inflow over an increasing number of cells.

can no longer be considered to be even approximately calibrated with all parameters simultaneously varied by the amounts indicated), robustness of predicted head error variance is reasonably good. Furthermore, the shape of the curves is maintained. Further computations (not shown herein) demonstrate that other qualitative aspects of the analysis (such as relativity of parameter contributions to pre- and post-calibration predictive error variance, the reduction in these contributions incurred by the calibration process, the number of singular values at which optimum predictive error variance is achieved, etc) are also maintained. Notwithstanding these results, the effects of model linearity on error variance computation (both in its quantitative and qualitative aspects) is a subject that is worthy of further investigation.

Discussion

Integrity in model usage requires that some estimate of potential predictive error be supplied with model predictions themselves. The methodology discussed herein allows such estimates to be made with a high level of numerical efficiency. This is beneficial both in itself, and as a means of inquiring into other aspects of the model construction and calibration process. The authors of this paper feel that the ability to undertake such calculations relatively inexpensively as a routine adjunct to model construction and calibration could have profound implications for the way in which models are employed in environmental management. Some of these implications are now briefly discussed.

The current trend in environmental modelling is for models to become increasingly complex. Thus many models attempt to simulate many different environmental processes, and the interactions between these processes, at a high level of spatial and temporal detail. During the model calibration process a simplified parameterization scheme is often then ''draped'' over the complexities of the processes simulated by the model, this being done in order to reduce the number of model parameters to that which can

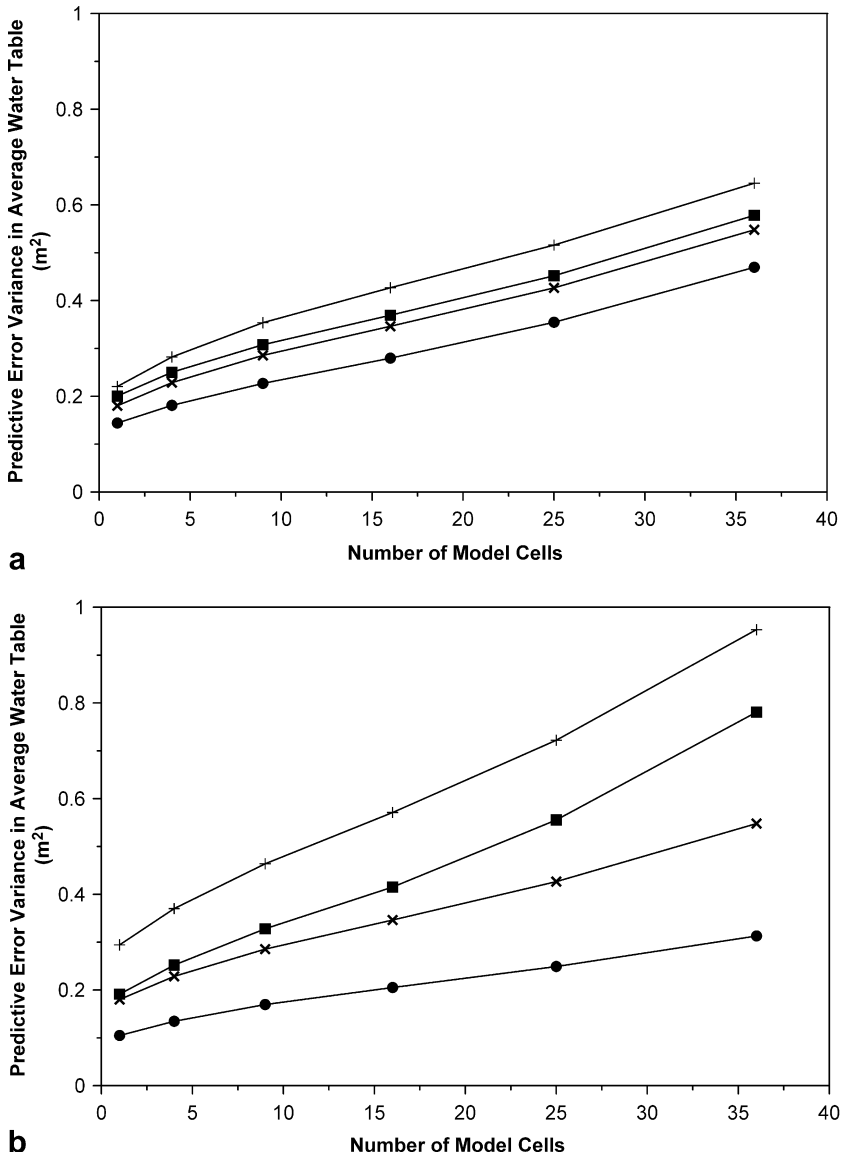

Figure 16 Optimized predictive error variance in average water table elevation for five areal expansions using the original $C(\varepsilon)$ and $C(\mathbf{p})$ matrices (crosses) compared with (a) doubled $C(\varepsilon)$ terms (pluses), halved $C(\varepsilon)$ terms (circles) and a $C(\varepsilon)$ for which temporal and spatial correlation exists (squares); (b) doubled $C(\mathbf{p})$ terms (pluses), halved $C(\mathbf{p})$ terms (circles) and a $C(\mathbf{p})$ for which spatial correlation exists in hydraulic conductivity and specific yield (squares).

be estimated on the basis of an often sparse calibration dataset. The present study, and previous studies cited herein, demonstrate that the ''cost of uniqueness'' incurred by doing this is potentially a high level of model predictive error. In fact it is not impossible that a seemingly complex model, on which is superimposed a simplistic parameterization scheme that ignores the complexity and variability of the real world, may in fact be no better a predictor of system behaviour than a simpler model, constructed at a fraction of the expense.

In spite of this, too often the decision is made to build a complex model rather than a simple model, this decision commonly being based more on wishful thinking than on a proper understanding of the returns that will be gained from construction of the complex model. In fact a complex model does not guarantee integrity of predictions at all. What it can guarantee, however, is integrity of predictive error analysis (for the null space term of Eq. (10) can then be properly defined). If such an analysis is not carried out, then the case for building a complex model may indeed be questionable in many modelling contexts. However if it is carried out in at least one modelling exercise which is considered typical of others undertaken by a particular institution or agency, then that analysis may allow a suitable ''engineering safety margin'' to be determined for future modelling exercises based on simpler, less expensive models. Alternatively, because such an analysis does not require the explicit use of measurements of system state, but only of the uncertainties associated with such measurements, and because use of Eq. (15) does not require that the model be calibrated, it can be conducted before an expensive calibration exercise is commenced, in order to judge the worth of that exercise.

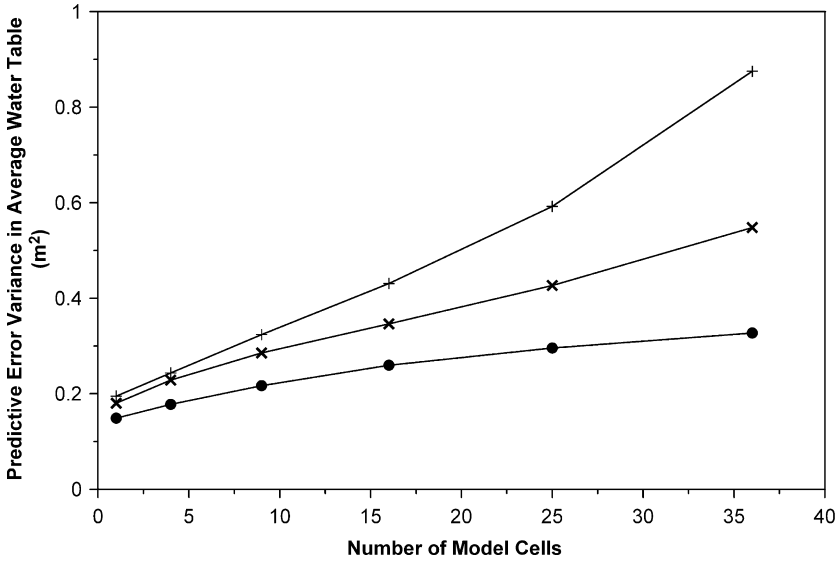

Figure 17 Optimized predictive error variance in average water table elevation for five areal expansions using base parameter values (crosses), a one standard deviation increase in base parameter values (circles) and a one-standard deviation decrease in base parameter values (pluses).

Where a complex and expensive model is built as part of an environmental management process involving different stakeholder groups, it is essential that the model gain acceptance by these different groups if it is to fulfil the purposes for which it was constructed. The process of gaining acceptance often commences with the model construction process itself, as discussions are held between, for example, modellers and regulators on the subject of what should be included in the model, and what can be omitted from the model, or represented only simply, without compromising its predictive integrity. Using the methodologies presented herein, such discussions can now take place on a much better scientific basis than in the past. For example, if it can be established through a relatively inexpensive pre-calibration analysis, that the contribution to the error variance of a key prediction made by a particular boundary condition is relatively small, then this may put an end to arguments about how that boundary need be represented in the final model. It may allow agreement to be reached that a simplistic representation of that boundary is justified, based on assumptions whose possible invalidity has few repercussions for predictive integrity. This could result in reduced model cost, both financial and emotional.

Use of the methodology discussed herein for optimization of data acquisition has already been mentioned. It is an easy matter to include in Eq. (15) terms pertaining to measurements of system state, or of system properties, that have not actually been made. As has been stated, the actual values of these system states or properties are not actually required in this analysis, only the uncertainties associated with their measurement. In this manner, the efficacy of different measurement strategies can be compared in terms of their ability to reduce the potential errors of key model predictions. Spending on data acquisition can thereby be placed on a more firm scientific footing than it otherwise would be.

Perhaps the largest contribution that the methodology presented herein can make to model-based environmental management is through its ability to promote informed discussion between different stakeholder groups. No longer can one group assert to another that ''our model says this and therefore it must be true''. Groups that are subject to environmental management by governing authorities have the right to know the potential error associated with assertions based on model outputs. With the availability of a rapid means for quantification of possible model error, there is no longer any excuse for failing to provide such an assessment. If, after predictive error variance analysis has taken place, predictive integrity can be shown to be high, then a model's claim to special status in environmental management is sound; if it is not, then it will have become clear to all involved in the decision-making process that the making of a definitive prediction of at least one aspect of environmental behaviour may not be possible. If this is the case, then decisions will have to be made by other means; if informed intuition rather than model outputs must form the basis for important management decisions, this too will gain greater acceptance if model-based analysis indicates that there is no other alternative.

Conclusions

A methodology for calculation of the magnitude of potential error of any prediction made by a calibrated model has been presented. The method is perfectly general and requires no alterations to the model; it only requires that (a) model outputs are not affected by any numerical misbehaviour on the part of the model that compromises calculation of derivatives of these outputs with respect to adjustable parameters (normally using finite parameter differences), and (b) that outcomes of the analyses of most interest are not contaminated unduly by model non-linearity. Because qualitative aspects of the analysis will often be just as important as quantitative predictive error computations, the latter is not expected to be a problem in most groundwater modelling

contexts (and does not indeed appear to be a problem in the present context), though further analysis of this matter is required. The methodology can be applied to a model that has already been calibrated and is actually being employed to make predictions. Or it can be applied to a model that has not yet been calibrated, or has been only ''approximately calibrated'' using a simplistic parameterization scheme. This is possible because calculation of predictive error variance using this methodology does not require knowledge of actual field measurements, only of the uncertainties associated with those measurements.

The methodology takes account of the two major contributors to model predictive error, namely the effects of measurement error on parameter estimation error, and the fact that there are strict limits on hydraulic property detail that can be inferred through the calibration process. It is the authors' experience, and it is indeed the case in the example presented in this paper, that the latter contribution can be substantial. In general, the more that a prediction depends on ''system fine detail'', the more that errors in its prediction are likely to be dominated by this term.

The methodology is useful in its own right as a basis for testing the veracity of predictions that may form the basis for environmental management. As such it allows the use of modelling in environmental management to take place with more integrity than has been done in the past. It can also be used to judge the worth of undertaking an expensive modelling exercise, in terms of the predictive accuracy that is achievable on the basis of currently available data, before that exercise is actually undertaken. Alternatively, it can be used to optimize the acquisition of data in support of a modelling exercise, ensuring that newly acquired data is maximally effective in raising the integrity of key model predictions.

Software

Software through which predictive error variance analysis of the type described herein can be implemented is supplied with the PEST suite. It can be downloaded free of charge from the following site: http://www.sspa.com/pest.

Acknowledgements

The authors would like to express their gratitude to NRW Water Assessment staff for numerous discussions and critical revision undertaken during the course of the documented study. Funding and logistical support provided by NRW is also gratefully acknowledged.

References

Arunakumaren, N.J., 1997. A lumped parameter model for simulating the temporal behaviour of moisture in the plant root zone and in the unsaturated zone below the root zone. LWRRDC R&D Project QP127. Department of Natural Resources and Mines, Queensland Government, Brisbane.

Carrera, J., Glorioso, L., 1991. On geostatistical formulations of the groundwater flow inverse problem. Adv. Water Resour. 14 (5), 273–283.

Carrera, J., Alcolea, A., Medina, A., Hidalgo, A., Slooten, L., 2005. Inverse problem in hydrogeology. Hydrogeol. J. 13, 206–222.

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

Christensen, S., Cooley, R.L., 1999. Evaluation of prediction intervals for expressing uncertainties in groundwater flow predictions. Water Resour. Res. 35 (9), 2627–2639.

Christensen, S., Cooley, R.L., 2004, User guide to the UNC process and three utility programs for computation of nonlinear confidence and prediction intervals using MODFLOW-2000: US Geological Survey Techniques and Methods Report 6-A10, US Geological Survey, Reston, VA, 195 pp.

Cooley, R.L., 2004. A theory for modeling ground-water flow in heterogeneous media. Professional Paper 1679, US Geological Survey, Reston, VA, 229pp.

De Groot-Hedlin, C., Constable, S., 1990. Occam inversion to generate smooth, 2-dimensional models from magnetelluric data. Geophysics 55 (12), 1613–1624.

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

Doherty, J., 2005. PEST: Model independent parameter estimation. Watermark Numerical Computing, Corinda, Australia. <http://www.sspa.com/pest>.

Doherty, J., Gallagher, M., 2005. Analysis of predictive error of the Pioneer Valley groundwater model. Watermark Numerical Computing, Brisbane, Australia. A copy of this report in PDF format can be supplied by either of the authors on request.

Eden, R.N., Hazel, C.P., 1973. Computer and graphical analysis of variable discharge pumping tests of wells. The Institution of Engineers Australia. Civil Engineering Transactions, pp. 5–10.

El Harrouni, K., Ouazar, D., Wrobel, L.C., Cheng, A.H.-D., 1997. Uncertainty analysis in groundwater flow with DRBEM. Eng. Anal. Bound. Elem. 19, 217–221.

Gomez-Hernandez, J.J., Franssen, H.J.W.M.H., Sahuquillo, A., 2003. Stochastic conditional inverse modeling of subsurface mass transport: a brief review of the self-calibrating method. Stoch. Environ. Res. Risk A 17, 319–328.

Gutjahr, A., Bras, R.L., 1993. Spatial variability in subsurface flow and transport – a review. Reliab. Eng. Syst. Safe. 42 (2–3), 293–316.

Harbaugh, A.W., Banta, E.R., Hill, M.C., McDonald, M.G., 2000. The US Geological Survey Modular Ground-Water Model – User Guide to Modularization Concepts and the Ground-Water Flow process. US Geological Survey Open File Report 00-92. Reston, VA.

Hill, M.C., 1989. Analysis of accuracy of approximate, simultaneous, nonlinear confidence intervals on hydraulic heads in analytical and numerical cases. Water Resour. Res. 25 (2), 177–190.

Kitanidis, P.K., 1995. Recent advances in geostatistical inference on hydrogeological variables. Rev. Geophys. 33, 1103–1109.

Kuhanesan, S., Durick, A., Werner, A., Weeks, S., Murphy, S., 2005. Groundwater amendment to the Pioneer Valley Water Resources Plan. Report 3: Groundwater flow modelling of the Pioneer Valley. Water Assessment, Resource Condition and Trend. Natural Resource Sciences, Queensland Department of Natural Resources, Mines and Water.

Moore, C., 2006. The use of regularised inversion in groundwater model calibration and prediction uncertainty analysis. Ph.D. Thesis, University of Queensland, Australia.

Moore, C., Doherty, J., 2005. Role of the calibration process in reducing model predictive error. Water Resour. Res. 41, W05020. doi:10.1029/2004WR00350.

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

Murphy, S., Kuhanesan, S., Foster, L., Durick, A., 2005. Groundwater amendment to the Pioneer Valley Water Resources Plan. Report 1: Conceptualisation of groundwater resources for the Pioneer Valley flow model. Water Assessment, Resource Condi-

tion and Trend, Natural Resource Sciences, Queensland Department of Natural Resources, Mines and Water.

Poeter, E.P., Hill, M.C., 1998. Documentation of UCODE, a computer code for universal inverse modelling. US Geological Survey Water. Resources Investigation Report 98-4080, p. 122.

Tonkin, M., Doherty, J., 2005. A hybrid regularised inversion methodology for highly parameterised models. Water Resour. Res. 41, W10412. doi:10.1029/2005WR00399.

Vecchia, A.V., Cooley, R.L., 1987. Simultaneous confidence and prediction intervals for nonlinear regression models with application to a groundwater flow model. Water Resour. Res. 23 (7), 1237—1250.

Yeh, T.C.J., 1992. Stochastic modeling of groundwater-flow and solute transport in aquifers. Hydrol. Process. 6 (4), 369—395.

Zimmerman, D.A., de Marsily, G., Gotway, C.A., Marietta, M.G., Axness, C.L., Beauheim, R.L., Bras, R.L., Carrera, J., Dagan, G., Davies, P.B., Gallegos, D.P., Galli, A., Gomez-Hernandez, J., Gorelick, S.M., Grindrod, P., Gutjahr, A.L., Kitanidis, P.K., LaVenue, A.M., McLaughlin, D., Neuman, S.P., RamaRao, B.S., Ravenne, C., Rubin, Y., 1998. A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. Water Resour. Res. 34 (6), 1373—1413.