

# **APPENDIX J**

## **LINEAR PREDICTIVE UNCERTAINTY ANALYSIS**

DRAFT

Page Intentionally Blank

**Linear Predictive Uncertainty Analysis**  
**applied to the**  
**NFSEG Groundwater Model**

**Watermark Numerical Computing**

**August, 2016**

## Table of Contents

1. Introduction .....	1
2. Theory .....	2
2.1 General.....	2
2.2 Uncertainty Analysis and Sensitivity Analysis .....	2
2.3 Theory .....	4
2.3.1 Uncertainty Interval of a Prediction.....	4
2.3.2 Adding Value to Linear Uncertainty Analysis.....	5
2.3.3 Parameters Corresponding to Uncertainty Limits .....	5
2.3.4 Uncertainty Intervals and Predictive Intervals.....	5
2.3.5 Subspace Analysis .....	7
3. Linear Analysis and the NFSEG Model .....	8
3.1 Observations and Parameters.....	8
3.2 Prior Parameter Covariance Matrix .....	11
3.3 Measurement Noise Covariance Matrix .....	12
4. Some Parameter Outcomes.....	14
4.1 General.....	14
4.2 Dimensions of Solution Space.....	14
4.3 Parameter Identifiabilities .....	15
4.4 Relative Parameter Uncertainty Variance Reduction .....	18
5. Some Predictive Outcomes .....	23
5.1 General.....	23
5.2 Predictive Uncertainties.....	23
5.3 Parameter Contributions to Predictive Uncertainty.....	25
5.4 Observation Worth .....	26
6. Some Predictive Difference Outcomes .....	29
6.1 General.....	29
6.2 Predictive Differences through Direct Linear Analysis.....	29
6.3 Predictive Difference Uncertainties: Further Examination.....	34
6.4 Predictive Noise .....	38
7. Conclusions .....	40
8. References .....	41

## **1. Introduction**

This document reports the outcomes of linear uncertainty analysis undertaken on the NFSEG model constructed by St Johns River and Suwannee River Water Management Districts. Analyses were focussed on a small number of parameters and predictions; with little difficulty analyses of the type documented herein could be extended to include a greater number of parameters and predictions.

## 2. Theory

### 2.1 General

The theoretical basis for linear uncertainty analysis is discussed at length in Doherty (2015). Except for one, all of the methodologies employed in the present study are discussed in that text, and implemented by utility programs that support the PEST suite of software (Doherty, 2016). However, a new methodology has been applied for the first time in the present study to overcome difficulties encountered in evaluating the uncertainties associated with model-predicted system changes. This methodology is described in the present document following a brief overview of the equations that underpin linear parameter and predictive uncertainty analysis.

### 2.2 Uncertainty Analysis and Sensitivity Analysis

It is useful to point out the relationship between so-called “sensitivity analysis” and analyses that are documented in the present text; the latter are described by the general term “linear uncertainty analysis” herein.

Linear predictive uncertainty analysis makes use of two sets of sensitivities. These are as follows:

- The sensitivities of model outputs used in the calibration process to all adjustable parameters. These are encapsulated in the matrix  $\mathbf{Z}$  which features in the following equations. This matrix is often referred to as the “Jacobian matrix”.
- The sensitivities of one or a number of predictive model outcomes to all adjustable parameters. These sensitivities are encapsulated in the vector  $\mathbf{y}$  in the equations that follow.

Sensitivities contained in the matrix  $\mathbf{Z}$  and the vector  $\mathbf{y}$  are “local”. This is because they are calculated to approximate derivatives with respect to parameters at current parameter values, the latter being those achieved through the model calibration process. In the study documented herein, these are calculated by PEST using finite parameter differences based on a three-point stencil.

The uncertainty of any prediction made by a calibrated model is determined by:

- The parameters to which it is sensitive, these generally representing system hydraulic properties, boundary condition specifications, and system stresses;
- The innate uncertainties of these parameters, this being an expression of expert knowledge as it pertains to system heterogeneity and spatial hydraulic property variability;
- The extent to which this so-called “prior uncertainty” is reduced through the history-matching process.

It is apparent that sensitivities feature strongly in linear uncertainty analysis. However other quantities besides uncertainties also feature strongly in this type of analysis. This, it is hoped, renders the outcomes of such an analysis of greater use in contexts of model-based decision-making than simply “sensitivity analysis”.

The strengths of linear uncertainty analysis include the following:

- It is relatively easy to undertake using public domain software such as PEST and PyEMU (White et al, 2016).

- The numerical burden of linear uncertainty analysis is small compared with that of nonlinear uncertainty analysis, especially in highly parameterized contexts.
- Linear analysis can be readily extended to include “parameters” which would not normally be estimated through model calibration, including those that characterize the nature and disposition of historical system stresses and model boundary conditions.
- Recent advances in linear analysis allow quantification of calibration-induced bias arising from model defects and simplifications; see White et al (2014).
- As well as supporting calculation of uncertainties associated with predictions of management interest, various “value-added” quantities can also be calculated. These include:
  - dimensions of the calibration solution and null spaces;
  - parameter identifiability;
  - relative parameter uncertainty variance reduction;
  - the contributions made to the uncertainties of predictions of interest by different parameters and/or groups of parameters;
  - the worth of existing, or as-yet-uncollected, data in reducing parameter and predictive uncertainty.

The weaknesses of linear uncertainty analysis include the following:

- Environmental processes are not linear. Hence linear analysis is approximate. Nevertheless, as shown by Dausman et al (2010), linear uncertainty analysis can nevertheless yield informative results, even when used in conjunction with highly nonlinear models.
- Where sensitivities are calculated through finite parameter differencing (as is done in the study documented herein), numerical errors can arise which detract from the integrity of calculated predictive uncertainties and associated quantities. The integrity of these quantities deteriorates even further where model predictions of interest are themselves calculated through differencing.
- The integrity of linear uncertainty analysis is further eroded where model solver convergence difficulties threaten the precision of finite-difference sensitivities required for filling of the  $\mathbf{y}$  and  $\mathbf{Z}$  matrices.

Unfortunately, the NFSEG model is not immune from solver convergence issues. These appear to have their source in the MNW2 package used for simulation of multi-layer well extraction. While convergence of the NFSEG model is still tight (it is generally able to achieve a head closer criterion just above 1E-3), convergence should ideally be somewhat tighter than this (1E-4 or better) to support finite-difference derivatives calculation.

Regardless of how uncertainty analysis is undertaken, it lacks credibility if important sources of predictive uncertainty are not represented in the analysis. Hence, where used to assess the uncertainties of predictions made by a groundwater model, it must be undertaken in a highly parameterized context. Parameters employed in the analysis must represent the potential for hydraulic property heterogeneity beyond that which is inferable through the model calibration process; after all, it is the parameters that cannot be estimated that are potentially the main contributors to predictive uncertainty, and not those that can. The NFSEG model includes 8800 adjustable parameters, all of which were included in analyses described herein.

## 2.3 Theory

### 2.3.1 Uncertainty Interval of a Prediction

The equations on which linear predictive uncertainty analysis are based are presented briefly in this subsection. For more information see Doherty (2015).

Let the vector  $\mathbf{k}$  represent parameters employed by a model. Let the vector  $\mathbf{h}$  represent members of the calibration dataset; let the vector  $\boldsymbol{\varepsilon}$  represent measurement noise associated with this data. Let the matrix  $\mathbf{Z}$  represent sensitivities of model outputs used in the calibration process to model parameters. Then, if a model is linear:

$$\mathbf{h} = \mathbf{Z}\mathbf{k} + \boldsymbol{\varepsilon} \quad (2.1)$$

Here, for convenience,  $\mathbf{k}$  is assumed to be  $\mathbf{0}$  where parameter values are endowed with their pre-calibration “expected values” from an expert knowledge point of view. This protocol makes the above and following equations simpler, while having no bearing on their outcomes.

Let  $s$  (a scalar) denote a prediction of management interest. Let the vector  $\mathbf{y}$  denote sensitivities of this prediction to parameters employed by the model. Then:

$$s = \mathbf{y}^t \mathbf{k} \quad (2.2)$$

where the “ $t$ ” superscript signifies matrix and vector transposition. Let  $\mathbf{C}(\mathbf{k})$ , a positive-definite matrix in which the number of rows and columns is equal to the number of elements of  $\mathbf{k}$ , denote the prior covariance matrix of  $\mathbf{k}$ . This covariance matrix thus pertains to the prior probability distribution of  $\mathbf{k}$ ; as such, it characterizes expert knowledge pertaining to parameter variability, much of which arises from hydraulic property heterogeneity within the subsurface. Let  $\mathbf{C}(\boldsymbol{\varepsilon})$  denote the covariance matrix of measurement noise.

Use of Bayes theorem to calculate the posterior variance of uncertainty of the prediction  $s$  computed by a linear model results in the following equation, provided that both the prior parameter probability distribution and the probability distribution describing measurement noise are multi Gaussian.

$$\sigma_s^2 = \mathbf{y}^t \mathbf{C}(\mathbf{k}) \mathbf{y} - \mathbf{y}^t \mathbf{C}(\mathbf{k}) \mathbf{Z}^t [\mathbf{Z} \mathbf{C}(\mathbf{k}) \mathbf{Z}^t + \mathbf{C}(\boldsymbol{\varepsilon})]^{-1} \mathbf{Z} \mathbf{C}(\mathbf{k}) \mathbf{y} \quad (2.3)$$

This equation can be shown to be equivalent to the following equation.

$$\sigma_s^2 = \mathbf{y}^t [\mathbf{Z}^t \mathbf{C}^{-1}(\boldsymbol{\varepsilon}) \mathbf{Z} + \mathbf{C}^{-1}(\mathbf{k})]^{-1} \mathbf{y} \quad (2.4)$$

The square root of variance is standard deviation. Equations 2.3 and 2.4 can thus be used to calculate the posterior (i.e. post-calibration) uncertainty interval of a calibrated model. For a Gaussian distribution, this spans the interval  $\underline{s} - \sigma_s$  to  $\underline{s} + \sigma_s$  at the 67% confidence level, and  $\underline{s} - 2\sigma_s$  to  $\underline{s} + 2\sigma_s$  at the 95% confidence level.

If the vector  $\mathbf{y}$  is comprised of zeros, except for its  $i$ 'th element which is assigned a value of 1.0, then equations 2.3 and 2.4 can be used to calculate the posterior uncertainty of parameter  $i$ .

The first term on the right of equation 2.3 (i.e.  $\mathbf{y}^t \mathbf{C}(\mathbf{k}) \mathbf{y}$ ) expresses the pre-calibration uncertainty variance of the prediction  $s$ ; the second term quantifies the reduction in this variance accrued through history-matching.

### 2.3.2 Adding Value to Linear Uncertainty Analysis

Through manipulation of equations 2.3 and 2.4 other useful quantities can be calculated as an adjunct to linear uncertainty analysis. If perfect knowledge of one or a number of parameters is attained, these equations can be used to compute the reduction in predictive uncertainty variance thereby accrued. This is referred to herein as the “contribution made to the uncertainty of prediction  $s$  by the pertinent parameter or parameter group”. Equations 2.3 and 2.4 can also be used to quantify the information content, or “worth” of individual observations, or groups of observations, with respect to a prediction of interest. This can be assessed in a number of ways. If, for example, all observations but those which comprise a particular observation group are removed from the calibration dataset (and equations 2.3 and 2.4 are re-formulated accordingly), the decrease in uncertainty variance of prediction  $s$  from its pre-calibration level constitutes a measure of the information content of that group with respect to that prediction. On the other hand, if that observation group is removed from an otherwise complete calibration dataset, the rise in uncertainty variance of prediction  $s$  from its post-calibration level quantifies the uniqueness of the information content of that observation group with respect to that prediction.

It is interesting to note that neither equation 2.3 nor equation 2.4 features the value of a parameter or the value of an observation; these equations only cite sensitivities of model outputs to model parameters. They can thus be deployed to assess the ability of data that has not yet been acquired to lower the uncertainties of predictions of interest. They can thereby form a basis for optimisation of data acquisition.

### 2.3.3 Parameters Corresponding to Uncertainty Limits

Let the vector  $\mathbf{k}$  denote the set of parameters which are deemed to “calibrate” a model, these thereby constituting the minimum error variance solution to the inverse problem of model calibration. Let  $\underline{s}$  denote the value of a model prediction calculated using this parameter set. Suppose that we wish to calculate a parameter set  $\mathbf{k} + \delta \mathbf{k}$  for which the value of the prediction is  $\underline{s} + \sigma_s$ , where  $\sigma_s$  is calculated using equation 2.3 or 2.4; that is, we wish to find a parameter set for which a particular model prediction is one standard deviation greater than its post-calibration expected value. It can be shown that the  $\delta \mathbf{k}$  of smallest norm which achieves this outcome is:

$$\delta \mathbf{k} = \sigma_s \mathbf{y} / \mathbf{y}^t \mathbf{y} \quad (2.5)$$

Obviously, the same equation with a leading factor of 2.0 can be used to find the parameter set of minimum perturbation from  $\mathbf{k}$  which gives rise to a predictive value of  $\underline{s} + 2\sigma_s$ . A leading negative sign finds parameter sets which give rise to predictions of  $\underline{s} - \sigma_s$  and  $\underline{s} - 2\sigma_s$ .

### 2.3.4 Uncertainty Intervals and Predictive Intervals

A distinction is often made between the so-called “uncertainty interval” associated with a model prediction and the so-called “predictive interval” associated with that same prediction. The former is an outcome solely of parameter uncertainty. It is calculated using equation 2.3 or 2.4. The latter is expanded to include the “noise” associated with the prediction. It thus attempts to define the interval that an “observation of the prediction” could span, accounting for the fact that observations

of system state are accompanied by measurement noise. If a prediction is similar in nature to an observation comprising the calibration dataset, then the standard deviation associated with this noise can, in theory, be calculated from the level of model-to-measurement fit achieved through the model calibration process for observations of similar type. Where it is not, the level of “predictive noise” associated with a particular prediction must be estimated by other means.

In some circumstances the distinction between the uncertainty and the predictive interval of a prediction can be ignored. In other cases it cannot. In practice, most model-to-measurement misfit results from so-called “structural noise” arising from model imperfections. These imperfections should be reduced as much as possible through the model construction and calibration processes. The use of many parameters, and inference of values for these parameters through highly parameterized inversion (as was done for the NFSEG model), helps in this regard. However structural noise cannot be completely eliminated, especially where a model domain is regional and its cell sizes are large.

Model imperfections which are exposed through the calibration process persist when a model is used to make predictions. Hence these should be included in the model predictive uncertainty analysis process; thus predictive intervals should be used in preference to uncertainty intervals. Unfortunately this is a difficult undertaking, as “predictive noise” is not really “noise” at all when it has structural origins. It cannot therefore be treated as a homoscedastic, random quantity with no spatial correlation that is endowed with random values drawn from a known probability distribution. Nor does structural noise “cancel” when many observations are used to infer parameters through the history-matching process. Nor, indeed, can the calibration process be used to quantify the statistical structure of model output errors, as these errors are not random quantities at all.

In practice, there is little choice but to assume that expressions of model structural deficiencies do indeed have noise-like qualities during the model calibration process, this enabling the use of equations such as 2.3 and 2.4 to quantify post-calibration parameter and predictive uncertainty intervals. Furthermore, when making a model prediction, the fact that model structural imperfections have been exposed through the model calibration process necessitates their expression when quantifying predictive uncertainty. Thus uncertainty intervals must be expanded to predictive intervals through the addition of a predictive noise component. However the prediction-specific nature of model structural defects, and the fact that “predictive noise” is not a random quantity, makes this a difficult undertaking.

For predictions that pertain to measurement types and locations used in the calibration process, it may be possible to quantify the magnitude, and possibly the sign, of predictive error through inspection of pertinent calibration residuals. That is, it may be possible to assume that structural noise in the future will be similar to structural noise in the past. An alternative means of accommodating model defects is to focus on predictive differences rather than on predictive absolutes, for example on model-predictive alterations to the head or spring flow at a particular location rather than on the actual value of this head or flow. If model structural defects lead to consistent errors in the making of a certain type of prediction at a certain location, then these errors will cancel when pertinent model outputs are differenced. The “predictive noise” term thus becomes zero. In practice it is unrealistic to expect exact cancellation. However it is not unrealistic to expect

that the structural errors associated with predictive differences will be much lower than those associated with individual predictions.

### 2.3.5 Subspace Analysis

Let  $\mathbf{Q}$  denote the weighting matrix used in the calibration process. Ideally this should be proportional to the inverse of  $\mathbf{C}(\boldsymbol{\epsilon})$ , the covariance matrix of measurement noise. In practice, some account must be taken of contributions to model-to-measurement misfit made by model structural defects in establishing an appropriate  $\mathbf{C}(\boldsymbol{\epsilon})$  before (conceptually) taking its inverse to devise an appropriate weighting strategy (Doherty and Christensen, 2011; White et al, 2014).

If the matrix  $\mathbf{Q}^{1/2}\mathbf{Z}$  is subjected to singular value decomposition, we obtain:

$$\mathbf{Q}^{1/2}\mathbf{Z} = \mathbf{U}\mathbf{S}\mathbf{V}^t = \mathbf{U}\mathbf{S}_1\mathbf{V}_1^t + \mathbf{U}\mathbf{S}_2\mathbf{V}_2^t \quad (2.6)$$

$\mathbf{U}$  is an orthonormal matrix whose columns are unit vectors that span model output space while  $\mathbf{V}$  is another orthonormal matrix whose unit vectors span parameter space;  $\mathbf{S}$  is a diagonal matrix. Partitioning of  $\mathbf{S}$  into  $\mathbf{S}_1$  and  $\mathbf{S}_2$ , and  $\mathbf{V}$  into  $\mathbf{V}_1$  and  $\mathbf{V}_2$ , should be such as to effectively subdivide parameter space into mutually orthogonal calibration-specific solution and null spaces; the elements of  $\mathbf{S}_2$  are thereby zero or small. The null space defines parameter combinations that are inestimable due to insensitivity, excessive post-calibration correlation, and/or undue amplification of measurement noise in their estimation. The number of dimensions of the solution space can therefore be regarded as defining the number of pieces of useable information contained within the calibration dataset.

### 3. Linear Analysis and the NFSEG Model

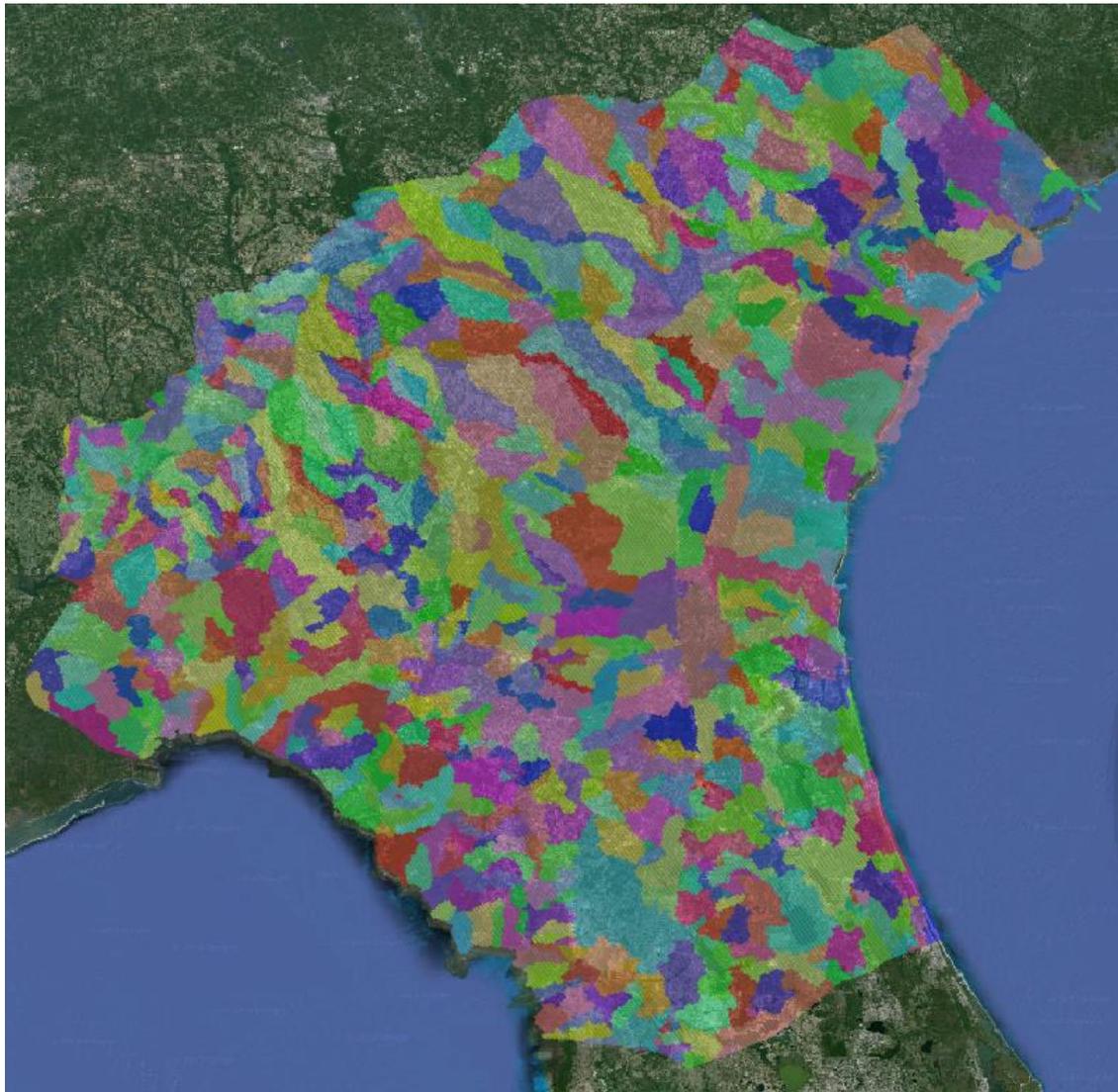
#### 3.1 Observations and Parameters

Details of parameters employed by the NFSEG model, and of observations comprising the NFSEG model calibration dataset, are described elsewhere. In undertaking linear analysis for the NFSEG model, parameters and observations used in computation of the Jacobian matrix (i.e. the  $\mathbf{Z}$  matrix introduced in the previous section) remained unchanged from those employed in the calibration process, with the following exceptions.

- “Wetting penalty observations”, used for prevention of groundwater inundation of the land surface in certain parts of the model domain, were removed from the calibration dataset. (It is worthy of note that few of these incurred a penalty during the calibration process anyway.)
- Recharge multiplier parameters were decreed as adjustable. These were fixed at a value of 1.0 during the previous calibration process. A multiplier was associated with each of the recharge zones depicted in figure 3.1, these corresponding to sub-basins used for HSPF modelling.
- A suite of EVT rate multiplier parameters was introduced using the same zonation as that used for recharge multipliers; see figure 3.1 again.

All recharge and EVT rate multiplier parameters were endowed with a value of 1.0. All other NFSEG model parameters retained the same values as were estimated during the calibration process.

It is important to note that, notwithstanding the fact that linear analysis explores the extent of post-calibration parameter (and hence predictive) variability, parameters are not actually varied through this process, except for the provision of incremental changes required for finite-difference derivatives calculation. The degree to which parameters may vary within the confines of expert knowledge is expressed through the prior parameter covariance matrix. This is the  $\mathbf{C}(\mathbf{k})$  matrix featured in equations presented in the previous section.



**Figure 3.1** Zones used for definition of recharge and EVT rate multiplier parameters.

Table 3.1 lists parameter groups employed in the linear analysis process. All parameters within all of these groups were log-transformed for the purpose of calculating parameter sensitivities and for the purpose of expressing parameter variability. Table 3.2 lists observation groups employed in the linear analysis process.

Parameter group name	Parameterization device	Number of parameters	Description
k1x	pilot points	518	horizontal hydraulic conductivity – layer 1
k3x	pilot points	1767	horizontal hydraulic conductivity – layer 3
k5xk3x	pilot points	201	horizontal hydraulic conductivity multiplier outside MCU – layer 5
k5x	pilot points	364	horizontal hydraulic conductivity – layer 5
k7x	pilot points	55	horizontal hydraulic conductivity – layer 7
k2z	pilot points	556	vertical hydraulic conductivity – layer 2
k2zk3z	pilot points	333	vertical hydraulic conductivity multiplier outside ICU – layer 2
k4zk3z	pilot points	139	vertical hydraulic conductivity multiplier outside MCU – layer 4
k4z	pilot points	230	vertical hydraulic conductivity – layer 4
k6z	pilot points	68	vertical hydraulic conductivity – layer 6

vanis1	entire layer	1	vertical anisotropy – layer 1
vanis2	zoned according to ICU/non-ICU	2	vertical anisotropy – layer 2
vanis3	entire layer	1	vertical anisotropy – layer 3
vanis4	zoned according to MCU/non-MCU	2	vertical anisotropy – layer 4
vanis5	zoned according to MCU/non-MCU	2	vertical anisotropy – layer 5
vanis6	entire layer	1	vertical anisotropy – layer 6
vanis7	entire layer	1	vertical anisotropy – layer 7
lcm	zoned according to lakes	257	multiplier applied to lakebed conductance
rcm	zoned according to river reaches	1871	multiplier applied to river reach conductance
sc	zoned according to springs	377	GHB conductance at springs
rechmul	zones (see fig 3.1)	904	multiplier applied to recharge rates
evtrmul	zones (see fig 3.1)	904	multiplier applied to maximum EVT rates
lkzmul	zoned according to lakes	246	vertical conductivity multiplier under lakes

**Table 3.1** Parameter groups used in linear analysis. A total of 8800 parameters collectively comprise these groups.

Observation group name	Number of observations with non-zero weight	Description
h2001_lay1	228	Heads in layer 1: 2001
h2001_lay2	96	Heads in layer 2: 2001
h2001_lay3	979	Heads in layer 3: 2001
h2001_lay4	13	Heads in layer 4: 2001
h2001_lay5	39	Heads in layer 5: 2001
h2001_lay7	2	Heads in layer 7: 2001
h2009_lay1	239	Heads in layer 1: 2009
h2009_lay2	111	Heads in layer 2: 2009
h2009_lay3	990	Heads in layer 3: 2009
h2009_lay4	10	Heads in layer 4: 2009
h2009_lay5	41	Heads in layer 5: 2009
h2009_lay7	2	Heads in layer 7: 2009
hd2001_lay3	289	Lateral head gradients in layer 3: 2001
hd2009_lay3	263	Lateral head gradients in layer 3: 2009
td_lay1	0	Temporal head differences: layer 1
td_lay2	0	Temporal head differences: layer 2
td_lay3	639	Temporal head differences: layer 3
td_lay4	0	Temporal head differences: layer 4
td_lay5	34	Temporal head differences: layer 5
td_lay7	0	Temporal head differences: layer 7
vd_1to3	231	Vertical head differences: layer 1 to 3
vd_3to5	36	Vertical head differences: layer 3 to 5
qr01	92	Inflow to river segments between one or more gages: 2001
qr09	85	Inflow to river segments between one or more gages: 2009
qspring01	374	Inflow to springs: 2001
qspring09	376	Inflow to springs: 2009
qs_spring01	7	Inflow to spring groups: 2001
qs_spring09	7	Inflow to spring groups: 2009
qs01	10	Cumulative inflow to river upstream of a gage: 2001
qs09	6	Cumulative inflow to river upstream of a gage: 2009
qlake01	257	Flow to/from lakes: 2001
qlake09	257	Flow to/from lakes: 2009

**Table 3.2** Observation groups used in linear analysis. A total of 5713 non-zero-weighted observations collectively comprise these groups.

Equations 2.3 and 2.4 feature 2 covariance matrices, namely  $C(\mathbf{k})$  and  $C(\boldsymbol{\epsilon})$ . Filling of these matrices is now described.

### 3.2 Prior Parameter Covariance Matrix

For the analyses described herein,  $C(\mathbf{k})$  was built as a block-diagonal matrix. In fact, submatrices pertaining to many of these blocks are diagonal, this denoting statistical independence of parameters represented by these blocks.

Parameter groups comprising pilot point parameters were assigned a full covariance matrix based on spatially variable variograms. All variograms are exponential, and hence specified by the equation:

$$\Upsilon(h)=C(0)[1 - \exp (-h/a)] \quad (3.1)$$

In equation 3.1  $h$  is distance and  $C(0)$  is the overall variance of the hydraulic property in question, this being equal to the sill of the variogram. The range of an exponential variogram is often characterized as  $3a$ .

For all pilot point parameters, the variogram “ $a$ ” value was decreed to be pilot-point-specific. This reflects the fact that pilot points are not capable of representing the natural heterogeneity of a complex carbonate aquifer. Instead, their use implies a degree of upscaling, with the extent of upscaling decreasing with increasing spatial density of pilot points. Where spatial density of pilot point emplacement is high, they are capable of representing short range hydraulic property heterogeneity. Alternatively, where it is low, they can only represent long range hydraulic property heterogeneity. To reflect this, the variogram range associated with each pilot point was calculated in the following way.

1. The average separation between the pilot point to which an “ $a$ ” value must be assigned and its 20 closest neighbours was calculated.
2. The “ $a$ ” value ascribed to the variogram associated with that pilot point was designated as twice this average separation.

Admittedly, this manner of establishing a local variogram range is somewhat arbitrary. However by repeating some of the analyses documented herein using variations of this methodology in which different “ $a$ ” values were assigned to pilot points, it was established that outcomes of these analyses are not unduly sensitive to variogram range unless it is unrealistically large or unrealistically small.

For all pilot point parameter groups except one, the variogram sill (applied to the logarithm of hydraulic conductivities associated with pilot points) was denoted as uniformly 0.25, this implying a standard deviation of parameter variability of 0.5, or half an order of magnitude. The exception was parameter group k4zk3z for which the sill was ascribed a value of 1.0, this reflecting gaps in knowledge of the disposition and details of the MCU. Variograms assigned to all pilot points are isotropic, except for a number of pilot points in the k3x and k3xk5x parameter groups to which a horizontal anisotropy of 2.0 with a bearing of 354 degrees was ascribed.

Covariance matrix construction was undertaken using the PPCOV\_SVA utility supplied with the PEST Groundwater Data Utility suite; see Doherty (2014).

For all other parameter types except recharge and EVT multipliers, within-group statistical independence was assumed in filling the pertinent blocks of the  $C(\mathbf{k})$  matrix. A uniform standard deviation was applied to each member of the group. Details are provided in table 3.3.

Parameter group name	Standard deviation ascribed to each parameter within group
vanis1	0.25
vanis2	0.25
vanis3	0.25
vanis4	0.25
vanis5	0.25
vanis6	0.25
vanis7	0.25
lcm	1.0
Rcm	0.6
Sc	1.0
lkzmul	0.6

**Table 3.3 Standard deviations assigned independently to each parameter within each respective parameter group. Note that these are actually applied to the log (to base 10) of each parameter for the purpose of linear analysis.**

As has already been discussed, a recharge rate multiplier parameter and an EVT rate multiplier parameter were assigned to each polygon depicted in figure 3.1. These parameters were assumed to be statistically independent of each other between polygons. However a high degree of negative correlation (-0.99) is assumed to exist between the recharge rate multiplier and the maximum EVT rate multiplier in each polygon. This reflects the fact that they are both calculated by the same HSPF model which is calibrated to reproduce baseflow, and that they can vary in opposite directions to produce the same average baseflow. Use of this correlation coefficient is equivalent to assuming that the recharge rate multiplier is equal to a random number plus a second random number whose standard deviation is about 10% of that of the first, while the EVT rate multiplier is equal to the same first random number minus a third random number which also has a standard deviation of 10% of that of this first random number. For the present study, the standard deviation of the first random number was chosen such that two standard deviations span a range in log space that is equivalent to multiplying and dividing the “calibrated” value of 1.0 for these recharge and EVT multiplier parameters by a factor of 1.25. A joint covariance matrix for recharge and EVT rate multipliers was constructed accordingly.

### 3.3 Measurement Noise Covariance Matrix

The  $C(\epsilon)$  matrix featured in equations 2.3 and 2.4 is assumed to be diagonal. Diagonal elements were given values such that if a diagonal weight matrix  $\mathbf{Q}$  is calculated as the inverse of the diagonal  $C(\epsilon)$  matrix, then each observation-group-specific component of the objective function achieved through the calibration process is equal to the number of non-zero-weighted observations within the respective observation group. Meanwhile, relatively of weighting within each group as employed in the actual model calibration process was preserved for linear analysis, as this reflects the credibility of each observation as judged by those who built and calibrated the model. This manner of calculating  $C(\epsilon)$  respects the fact that if weights are indeed awarded values which are the inverse square root of the standard deviation of measurement noise, then each weighted residual achieved through the calibration process is on average equal to 1.0. The expected value of the total objective

---

function, and of observation-group-specific components thereof, is thus equal to the total number of observations, or the number of observations comprising the group.

This method of calculating  $C(\epsilon)$  recognizes that one of the important roles played by the calibration process is quantification of the ability (or otherwise) of the model to fit different components of the calibration dataset. Reduction of pertinent weights to reflect a diminished ability on the part of the model to fit certain types of measurements provides implicit recognition of the structural defects of the model that are responsible for this diminished fit. Parameter and predictive uncertainties will rise accordingly. Unfortunately however, it does not recognize that, where misfit is caused by model structural defects, the resulting “structural noise” does not possess a diagonal covariance matrix; see Doherty and Welter (2010). In fact, as was discussed in the previous section, it is not really a statistical quantity at all. Sadly, this is almost impossible to take into account; furthermore, accommodation of misfit arising from model inadequacies is made all the more difficult because the covariance matrix of structural noise is, in fact, singular (Doherty and Welter, 2010), and hence does not possess an inverse from which an appropriate weighting strategy can emerge.

## 4. Some Parameter Outcomes

### 4.1 General

This subsection presents a few outcomes of linear analysis as they pertain to parameters.

The outcomes of linear analysis presented in this section are not meant to be complete; they should be considered as samples only. A broader range of results can be readily provided based on the same Jacobian matrix (i.e. the  $\mathbf{Z}$  matrix of equations 2.3 and 2.4) that was used to calculate the results presented herein. Alternatively, it may be better to postpone provision of a more complete set of linear analysis results as these pertain to estimated parameters until further model refinements have been undertaken, and/or until discussions with stakeholder groups indicate satisfaction with the manner in which the  $\mathbf{C}(\mathbf{k})$  and  $\mathbf{C}(\epsilon)$  covariance matrices were constructed.

### 4.2 Dimensions of Solution Space

The PEST SUPCALC utility was used to estimate the dimensions of the calibration solution space. SUPCALC calculates a lower and an upper limit for this quantity. The lower limit coincides with the point at which “overfitting” may occur. This is the point at which reductions in parameter uncertainties accrued through improved matching of model outputs to observations are more than offset by increases in the potential for parameter error born of amplification of measurement noise as values are inferred for parameters through solution of the inverse problem of model calibration. The upper limit of solution space dimensionality provided by SUPCALC is the point at which not only measurement noise, but numerical noise associated with solution of the inverse problem, can contribute significantly to errors in estimated parameters.

SUPCALC-calculated lower and upper limits for the dimensionality of the calibration solution space are 1123 and 1429 respectively. As has already been discussed, this implies that the calibration dataset contains between 1123 and 1429 useable items of information. Each such item allows unique estimation (albeit with uncertainty arising from measurement/structural noise) of a single linear combination of parameters. These linear combinations are specified by the orthogonal vectors comprising the columns of the matrix  $\mathbf{V}_1$  of equation 2.6.

The calibration null space is the orthogonal complement of the calibration solution space. This is spanned by orthogonal combinations of parameters (specified as the columns of the  $\mathbf{V}_2$  matrix of equation 2.6) that are completely inestimable on the basis of the calibration dataset, or are estimable only with unacceptable propensity for error forthcoming from the presence of measurement/structural noise in the calibration dataset. For the NFSEG model, the dimensions of the null space are between 7371 and 7677; the solution space and null space dimensions must add to the total dimensionality of parameter space, this being 8800 in the present case.

It is important to note that the uncertainties of predictions that are sensitive only to combinations of parameters that lie entirely within the calibration null space are not reduced through the history-matching process, regardless of how good a fit is achieved between model outputs and measurements comprising the calibration dataset.

---

### 4.3 Parameter Identifiabilities

Doherty and Hunt (2009) define the identifiability of a parameter as the square of the cosine of the angle between a parameter and its projection onto the calibration solution space. This ranges between zero and one. If a parameter has an identifiability of zero, then no model output that is employed in the calibration process is sensitive to that parameter. If a parameter has an identifiability of one, then it is uniquely estimable on the basis of the calibration dataset. Its estimation will be accompanied by uncertainty; however this uncertainty arises only from contamination of the calibration dataset by measurement/structural noise, and not from a deficit of information in the calibration dataset.

If a parameter has an identifiability that is between zero and one, this indicates that information pertinent to that parameter resident in the calibration dataset is shared between this parameter and at least one other parameter. Because the parameter has a non-zero projection onto the calibration null space, it cannot be estimated uniquely.

Figure 4.1a maps the identifiability of  $k3x$  parameters; also depicted in this figure are the pilot points with which these parameters are associated. Figure 4.1b shows  $k3x$  parameter identifiabilities together with layer 3 observation wells. Unsurprisingly, identifiabilities are highest where observation data density is greatest.

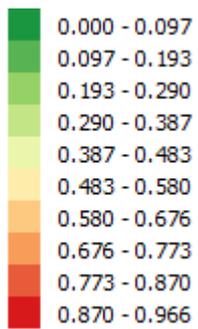
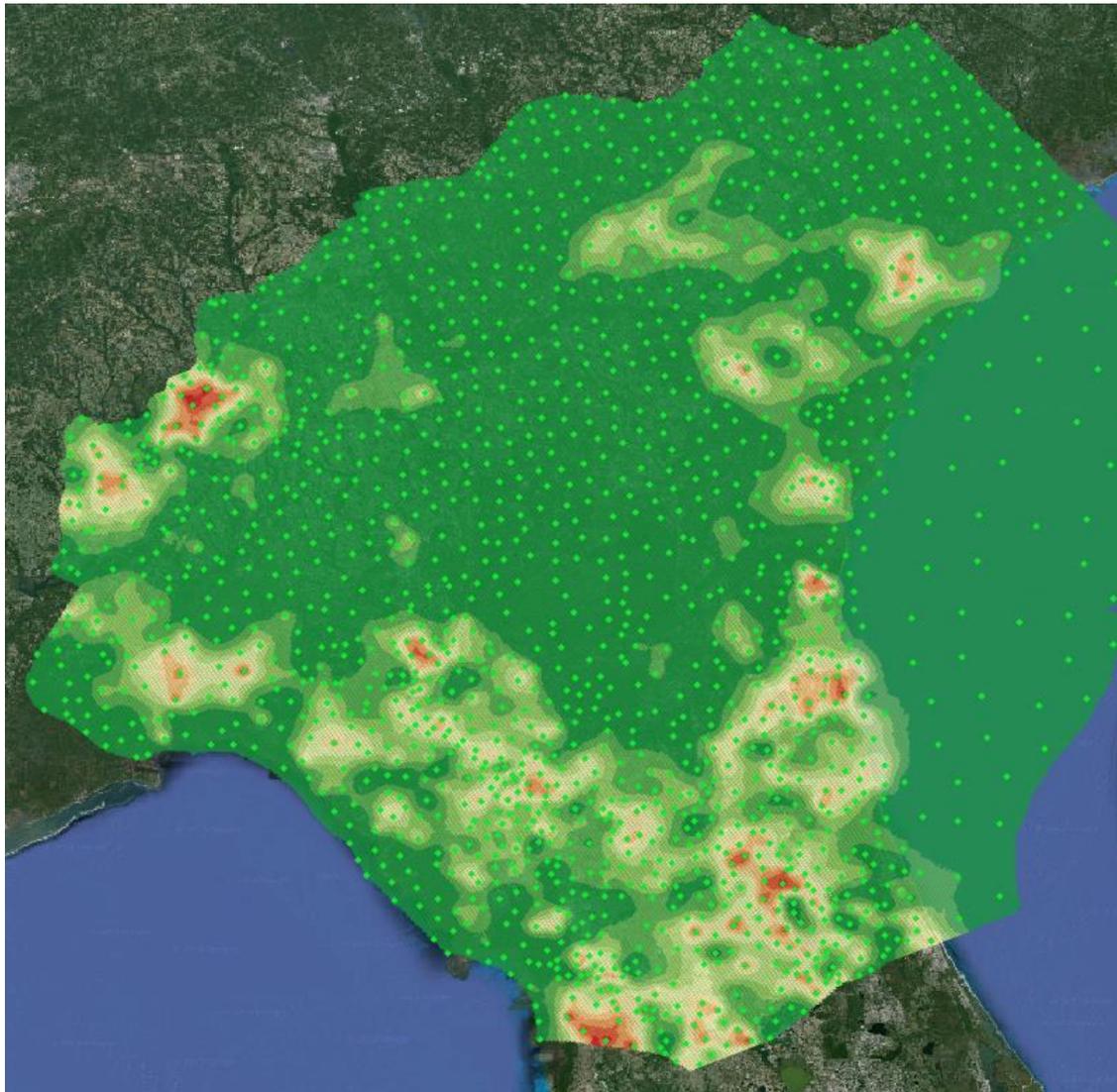
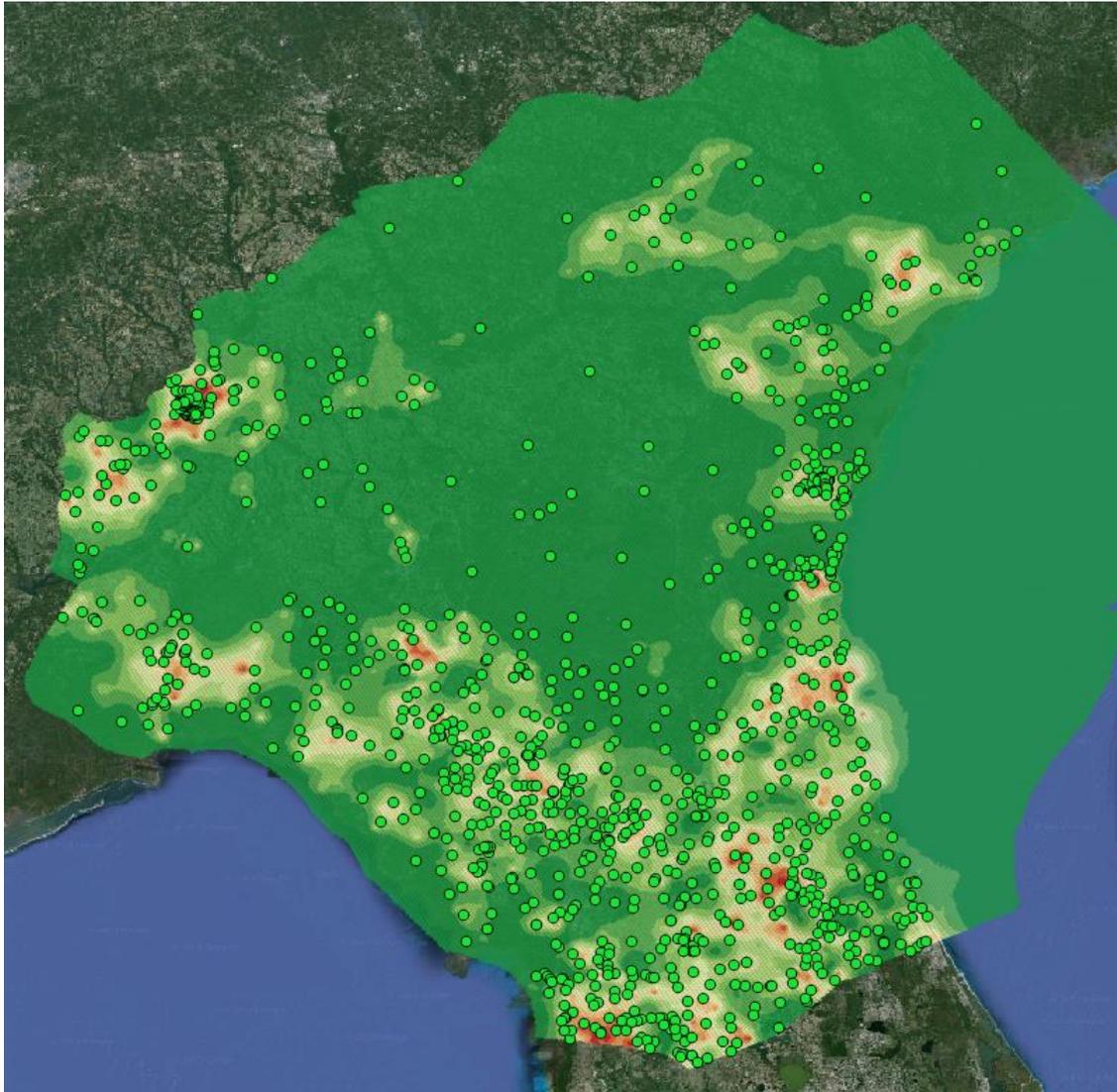
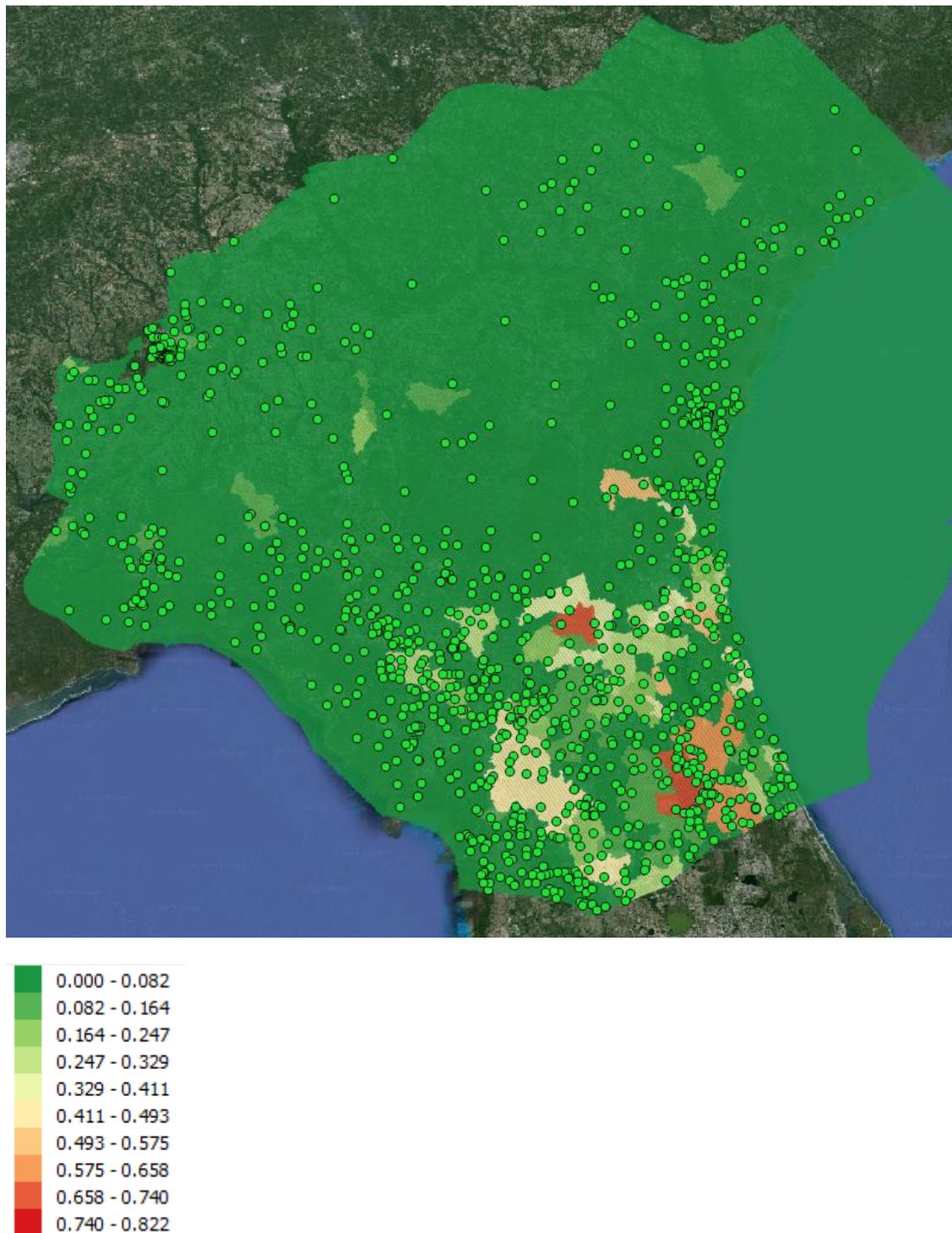


Figure 4.1a Identifiabilities of  $k3x$  parameters together with  $k3x$  pilot points.



**Figure 4.1b** Identifiabilities of  $k_{3x}$  parameters together with observation wells in layer 3; see figure 4.1a for colour scale.

Identifiabilities of recharge multiplier parameters are mapped in figure 4.2.



**Figure 4.2** Identifiabilities of recharge multiplier parameters together with observation wells in layer 3.

In areas of small parameter identifiability, parameters are relatively uninformed by the calibration dataset, and hence retain their prior uncertainties. Knowledge from outside of the calibration process must form the basis for parameter value assignment in these areas.

#### 4.4 Relative Parameter Uncertainty Variance Reduction

Like identifiability, the relative uncertainty variance reduction of a parameter is a number between zero and one. For the  $i$ 'th parameter it is calculated as:

---

$$r_i = \frac{(\sigma_{ip}^2 - \sigma_{ic}^2)}{\sigma_{ip}^2} \quad (4.1)$$

where

$\sigma_{ip}^2$  is the prior uncertainty variance of parameter  $i$ ; and

$\sigma_{ic}^2$  is the posterior (i.e. post-calibration) uncertainty variance of parameter  $i$ .

Prior and posterior parameter uncertainty variances are calculated using equation 2.3 or 2.4 with  $\mathbf{y}$  in these equations tailored to a specific parameter in the manner discussed above. This statistic takes more explicit account of the presence of measurement/structural noise in the calibration dataset than does identifiability; it also takes greater account of prior parameter spatial correlation expressed in the  $\mathbf{C}(\mathbf{k})$  matrix.

Maps of relative parameter uncertainty variance reduction for  $k3x$  and recharge multiplier parameters are presented in figures 4.3 and 4.4.

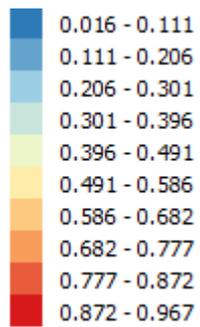
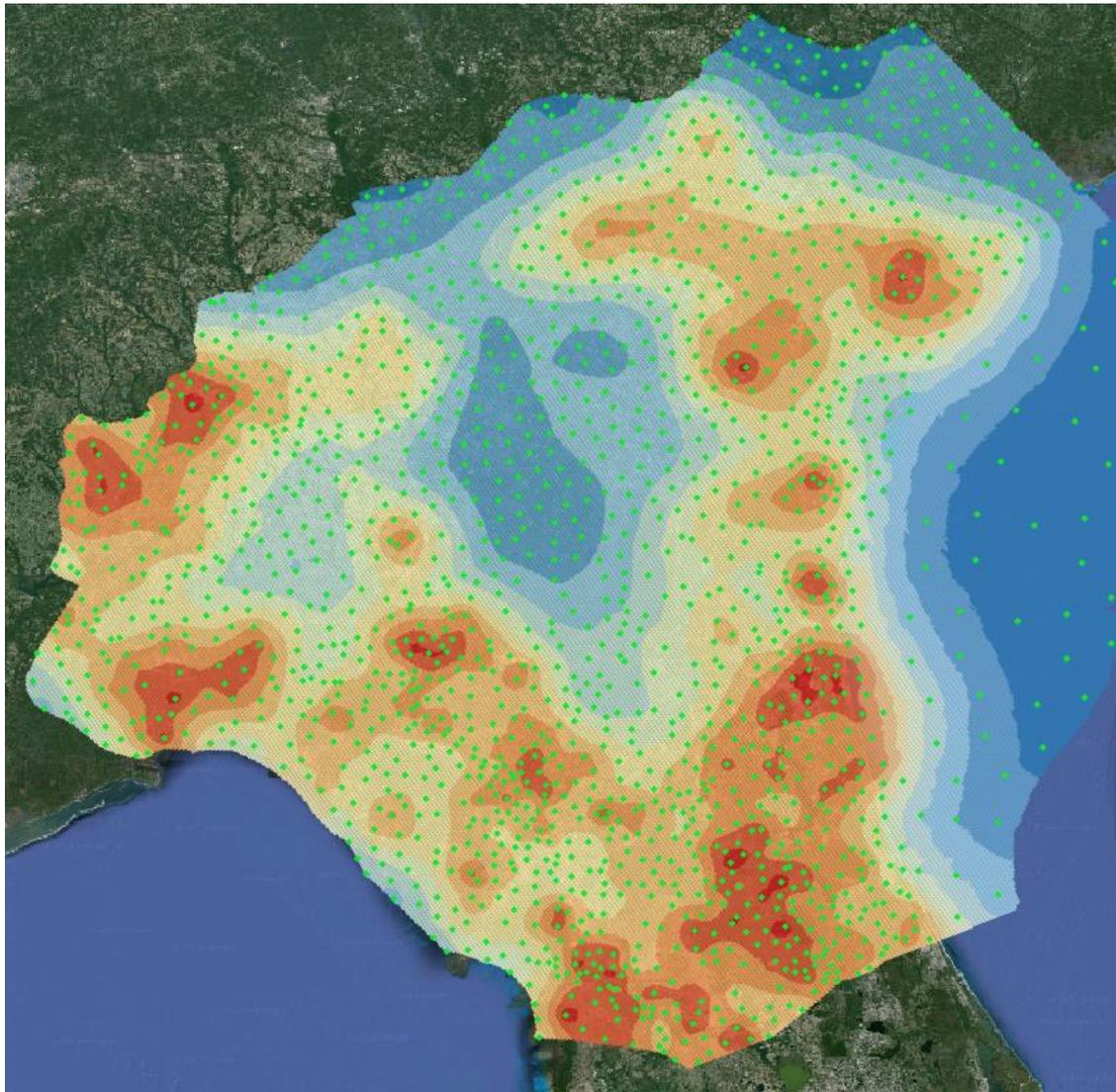


Figure 4.3a Relative uncertainty variance reduction of  $k3x$  parameters together with pilot points associated with these parameters.

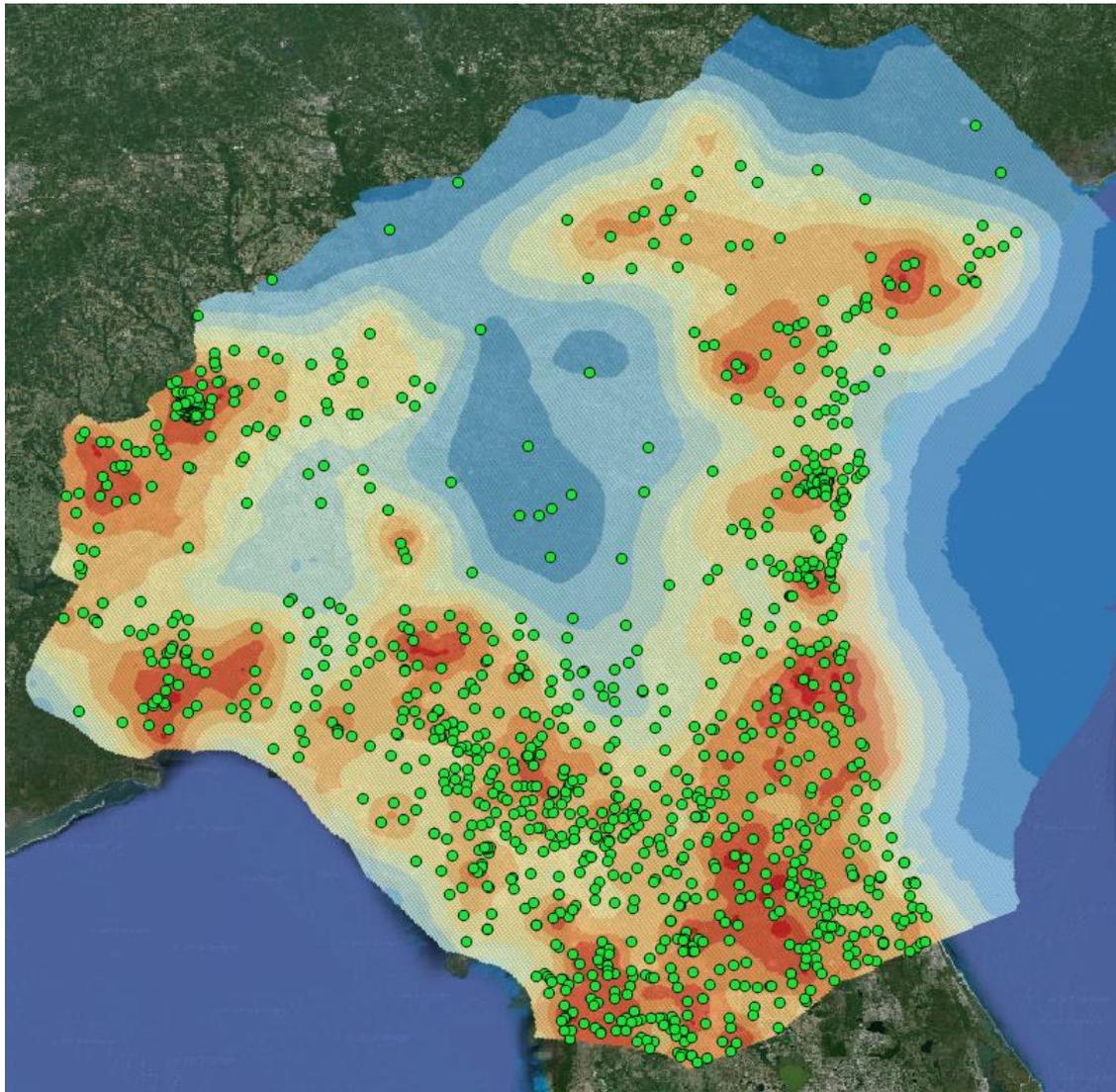


Figure 4.3b Relative parameter uncertainty variance reduction of  $k_{3x}$  parameters together with observation wells in layer 3; see figure 4.3a for colour scale.

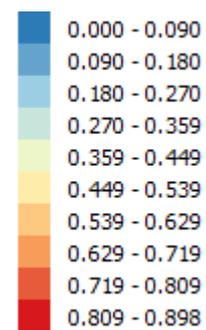
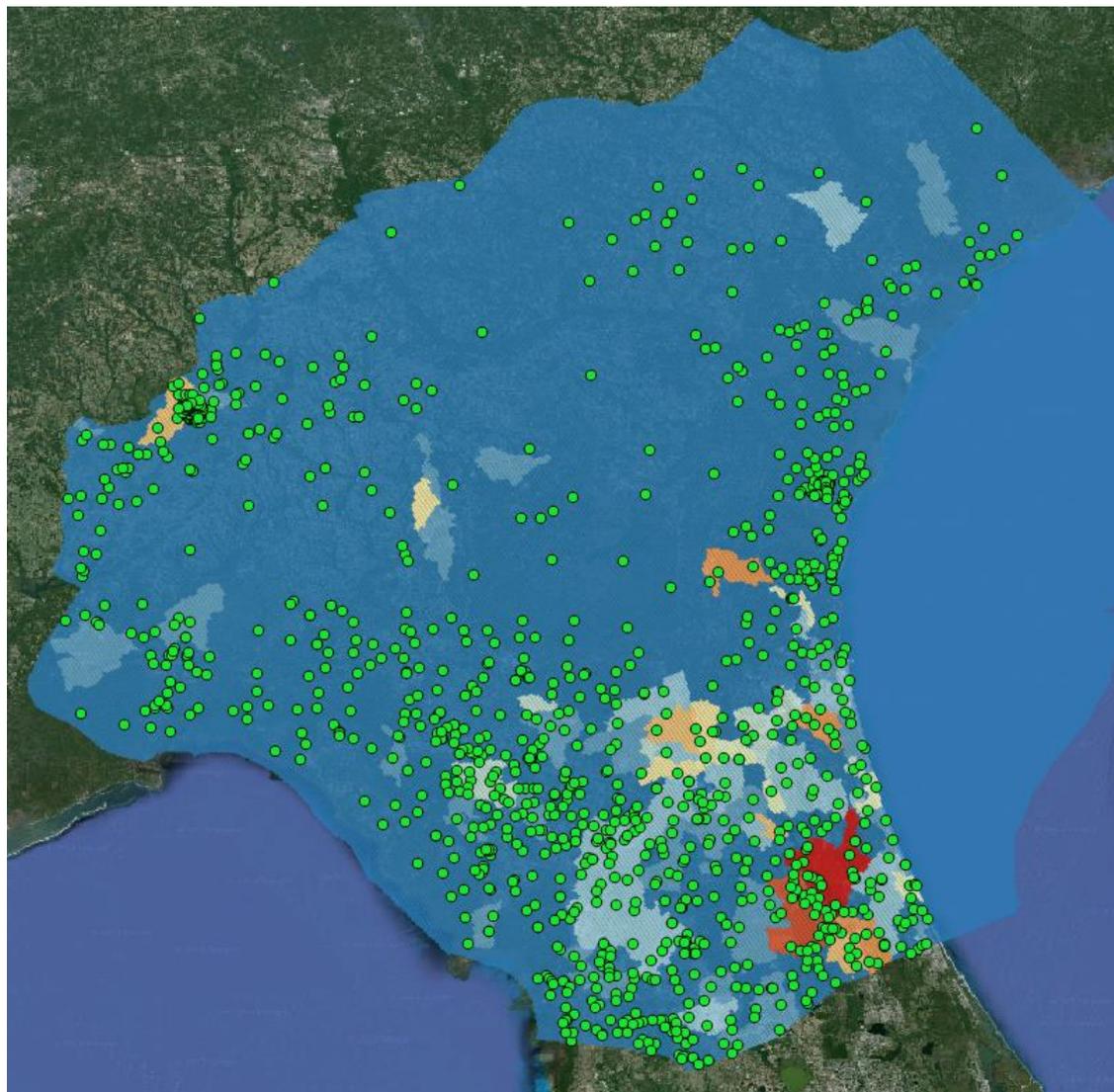


Figure 4.4 Relative parameter uncertainty variance reduction of recharge multiplier parameters together with observation wells in layer 3.

## 5. Some Predictive Outcomes

### 5.1 General

This section examines the uncertainties associated with a number of predictions made by the NFSEG model. For making these predictions, pumping within the North Florida water supply planning region was set to rates projected to occur 2035, while in the rest of the model domain pumping was assumed to be the same as in 2009. Recharge was assumed to be the same in both 2009 and 2035.

All of the predictions considered in this report have measured counterparts in one or both of 2001 and 2009 which are featured in the model calibration dataset. It is a general rule of model usage that uncertainties associated with predictions that resemble in place and nature observations comprising the calibration dataset possess less post-calibration uncertainty than those that have no counterparts in the calibration dataset. Furthermore, their uncertainties are likely to be determined more by noise in the calibration dataset than by lack of information in the calibration dataset; that is, they are determined more by  $C(\epsilon)$  than by  $C(k)$ .

This section also reports the outcomes of ancillary linear analyses conducted for two predictions. These analyses yield parameter contributions to predictive uncertainty and the information content of different subsets of the calibration dataset as it pertains to these predictions.

As for the previous section, the results presented herein are representative only. Analyses that give rise to these results can be readily extended to other predictions. Contributions to predictive uncertainty by different groupings of parameters, or by individual parameters, can be readily calculated. Similarly, the information content of different groupings of observations, or of individual observations (including, as has already been discussed, observations which have not as yet been made) can also be explored.

### 5.2 Predictive Uncertainties

Table 5.2 lists pre- and post-calibration uncertainty variances and standard deviations (variance is the square of standard deviation) for a number of predictions of management interest calculated using equation 2.3. Table 5.1 identifies these predictions. Note that the names used for these predictions correspond to observed counterparts in 2009, the latter forming part of the calibration dataset.

Prediction name	Description
w00202_09	UFA observation well near Lake Brooklyn
w00258_09	UFA observation well near Lake Geneva
w00878_09	UFA observation well near Putnam County MFL lakes
qr09_2319500	Baseflow pickup in reach upstream of the Suwannee River near Ellaville, Florida
qr09_2320700	Baseflow to the Santa Fe River near Graham, Florida
qr09_2323500	Baseflow pickup in reach upstream of the Suwannee River near Wilcox, Florida
qr09_2324000	Baseflow to the Steinhatchee River near Cross City, Florida
qr09_2326000	Baseflow to the Econfina River near Perry, Florida
qr09_2326550	Baseflow pickup in the reach upstream of the Aucilla River near Nutall Rise,
qspring09_s121610002	Blue Spring near Bronson
qspring09_s101429027	Little Fanning Springs near Fanning Spring

qspring09_s101429001	Fanning Springs near Wilcox
qspring09_n011117008	Madison Blue Spring near Blue Springs
qr09_lsf_sprgrp	Lower Santa Fe Springs Group
qr09_iche_sprgrp	Ichetucknee Springs Group
qr09_wacissa_sprgrp	Wacissa Springs Groups
qr09_silver_sprgrp	Silver Springs Group
qs09_2315500	Baseflow to the Suwannee River near White Springs, Florida
qs09_2317620	Baseflow to the Alapaha River near Jennings, Florida
qs09_2319000	Baseflow to the Withlacoochee River near Pinetta
qs09_2320500	Baseflow to the Suwannee River near Branford, Florida
qs09_2321500	Baseflow to the Santa Fe River near Worthington Springs
qs09_2322500	Baseflow to the Santa Fe River near Fort White

**Table 5.1 Names of predictions featured in this section.**

Prediction	Modelled value	Pre-calibration variance	Post-calibration variance	Pre-calibration std dev	Post-calibration std dev
w00202_09	77.0313	196.9804	0.9583	14.0350	0.9790
w00258_09	75.7702	215.1865	1.3180	14.6692	1.1481
w00878_09	28.3804	34.1637	0.9218	5.8450	0.9601
qr09_2319500	-853.1482	135237.1000	12735.7200	367.7459	112.8527
qr09_2320700	-4.6485	10.7651	0.6726	3.2810	0.8201
qr09_2323500	-343.1368	11145.7900	634.8607	105.5736	25.1964
qr09_2324000	-78.7990	8360.8590	2006.6430	91.4377	44.7956
qr09_2326000	-65.0888	864.3049	321.6057	29.3991	17.9334
qr09_2326550	-695.6051	259435.4000	386.9768	509.3480	19.6717
qspring09_s121610002	-2.8528	23.8719	0.2567	4.8859	0.5067
qspring09_s101429027	0.0218	0.0025	0.0025	0.0503	0.0503
qspring09_s101429001	-64.0665	3050.6520	72.4080	55.2327	8.5093
qspring09_n011117008	-103.3659	14328.3200	15.4598	119.7010	3.9319
qr09_lsf_sprgrp	-730.1142	82023.9900	1453.0130	286.3983	38.1184
qr09_iche_sprgrp	-242.7187	13918.4300	306.3752	117.9764	17.5036
qr09_wacissa_sprgrp	-521.8089	183672.7000	244.2470	428.5706	15.6284
qr09_silver_sprgrp	-474.2603	71803.8700	714.6787	267.9624	26.7335
qs09_2315500	-2.1171	993.6093	492.8851	31.5216	22.2010
qs09_2317620	-1106.8150	52990.1000	4909.4280	230.1958	70.0673
qs09_2319000	-1078.8640	96024.9500	6404.1510	309.8789	80.0259
qs09_2320500	-4001.6930	474455.0000	23073.2300	688.8070	151.8987
qs09_2321500	-35.8504	695.9704	72.8482	26.3813	8.5351
qs09_2322500	-653.2566	29968.5400	1700.9140	173.1143	41.2421

**Table 5.2 Pre- and post-calibration uncertainty variances and standard deviations for selected predictions.**

An outcome of the linearity assumption that underpins use of equations 2.3 and 2.4 is that sometimes a calculated uncertainty standard deviation may imply a predictive value that falls outside of its realistic range. This applies particular to predictions of spring flow which, generally speaking, cannot change sign. Where a model prediction is such that its behaviour changes abruptly

at a certain threshold (for example where the piezometric surface falls below the ground surface at the site of a particular spring), then the assumption of model linearity with respect to this prediction is obviously violated. Predictive uncertainty intervals implied by linear standard deviations must therefore be reduced to forestall implications of impossible system behaviour. In circumstances such as these, one side of a predictive uncertainty interval inferred through linear analysis remains intact while the other side must be contracted.

The role of structural defects in contributing to the uncertainties of model predictions was discussed in section 2.3.4. As was stated in that section, these should be recognized through addition of a “predictive noise” term to predictive uncertainty intervals calculated using equations 2.3 and 2.4. For the predictions listed in table 5.2 this requires a (probably subjective) assessment of the magnitude of model-to-measurement misfit experienced during the calibration process for the 2001 and 2009 counterparts to these predictions. This is beyond the scope of the present study, and hence was not carried out. Furthermore, the use of predictive differences rather than predictive absolutes as a basis for model-based decision-making (see the next section) should reduce the size of this term considerably. It should be noted, however, that despite the absence of this term in equations used for calculation of the predictive uncertainties listed in table 5.2, account is taken of calibration residuals as they pertain to different observation types in construction of the  $C(\epsilon)$  matrix that is used in these equations; see section 3.3.

### 5.3 Parameter Contributions to Predictive Uncertainty

Figures 5.1 and 5.2 depict contributions made by different parameter groups to the uncertainty variances of two 2035 predictions. In each of these figures, the back row depicts pre-calibration contributions to predictive uncertainty while the front row depicts post-calibration contributions to predictive uncertainty. The effect of the calibration process in reducing predictive uncertainty is obvious from these figures.

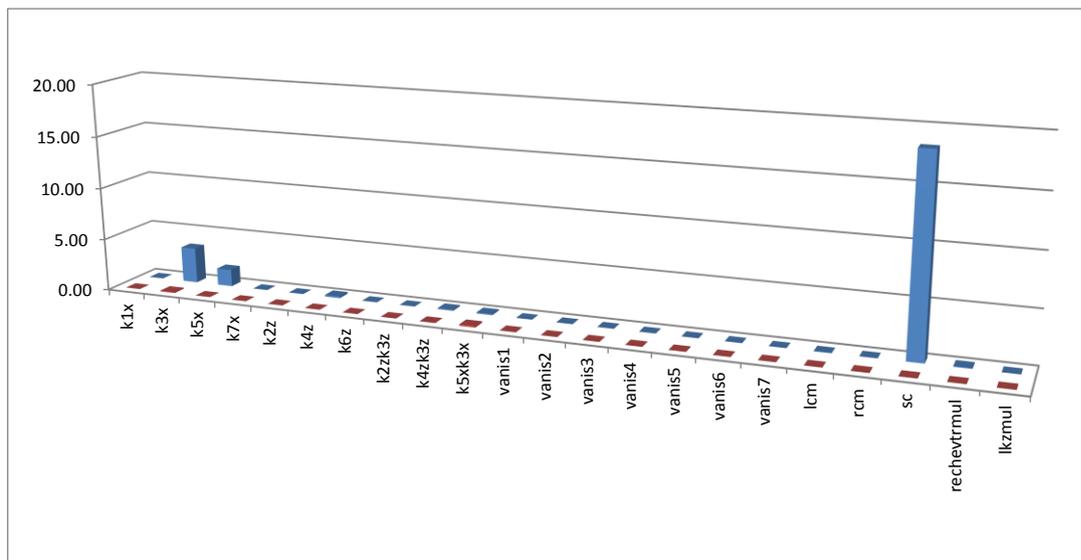


Figure 5.1 Contributions made by different parameter groups to the uncertainty variance of prediction *qspring09\_s12161002*. Pre- and post-calibration predictive variances are 23.87 (ft/day)<sup>2</sup> and 0.257 (ft/day)<sup>2</sup> respectively.

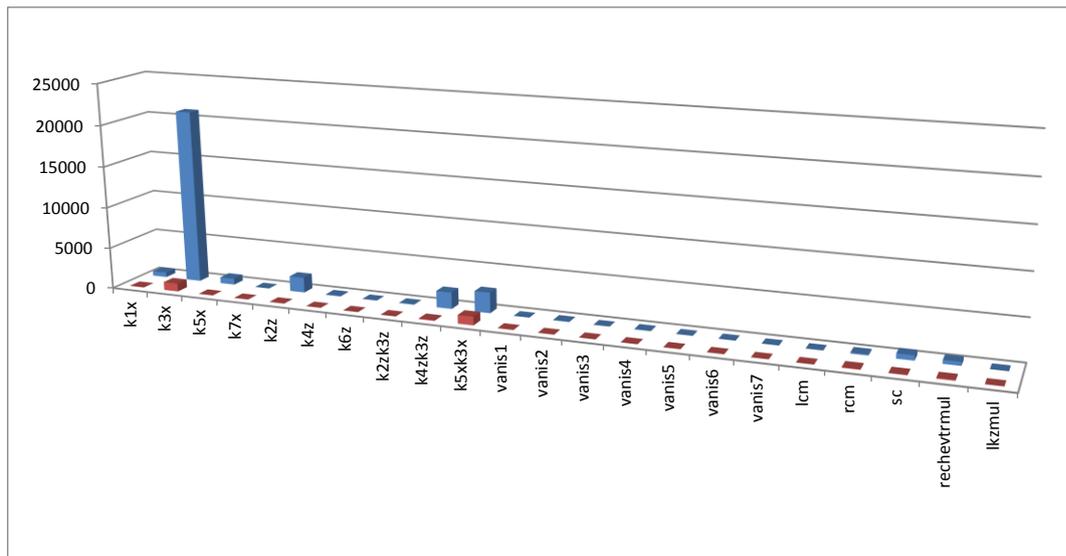


Figure 5.2 Contributions made by different parameter groups to the uncertainty variance of prediction *qs09\_2322500*. Pre- and post-calibration predictive variances are  $29968 \text{ (ft/day)}^2$  and  $1701 \text{ (ft/day)}^2$  respectively.

#### 5.4 Observation Worth

Equations 2.3 and 2.4 were employed to evaluate the worth with respect to two predictions of each of the observation groups featured in table 3.2. As has already been discussed, in the present study data worth is assessed in two ways. In the first of these ways, members of the observation group whose worth is being assessed comprise the entirety of the calibration dataset. Reduction in the uncertainty variance of a prediction from its pre-calibration level is then calculated. This constitutes one measure of the worth of that observation group with respect to that prediction. A second option is to remove the observation group from the full calibration dataset, leaving the rest of the calibration dataset intact. The resulting increase in uncertainty variance of a prediction above its post-calibration level constitutes a second measure of the worth of that observation group. This measure quantifies uniqueness of the information content of the observation group with respect to that prediction.

Both of these analyses were carried out for each of the predictions that have already been discussed in this section, namely predictions *qspring09\_s12161002* and *qs09\_2322500* (both pertaining to 2035 conditions). See figures 5.3 and 5.4.

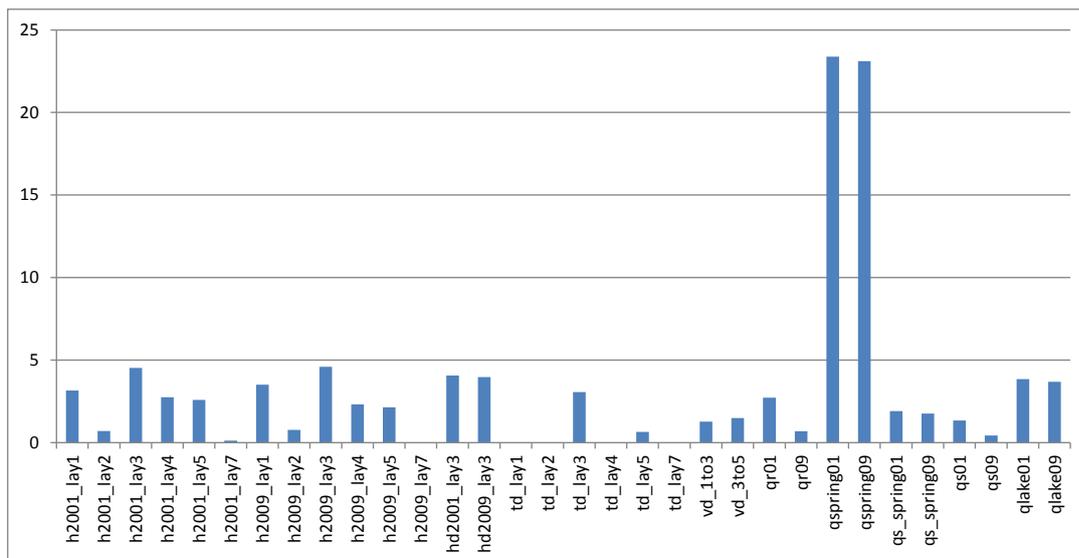


Figure 5.3a Decrease in the uncertainty variance of prediction *qspring09\_s12161002* from its pre-calibration value of 23.87 (ft/day)<sup>2</sup> accrued if each observation group comprises the entirety of the calibration dataset.

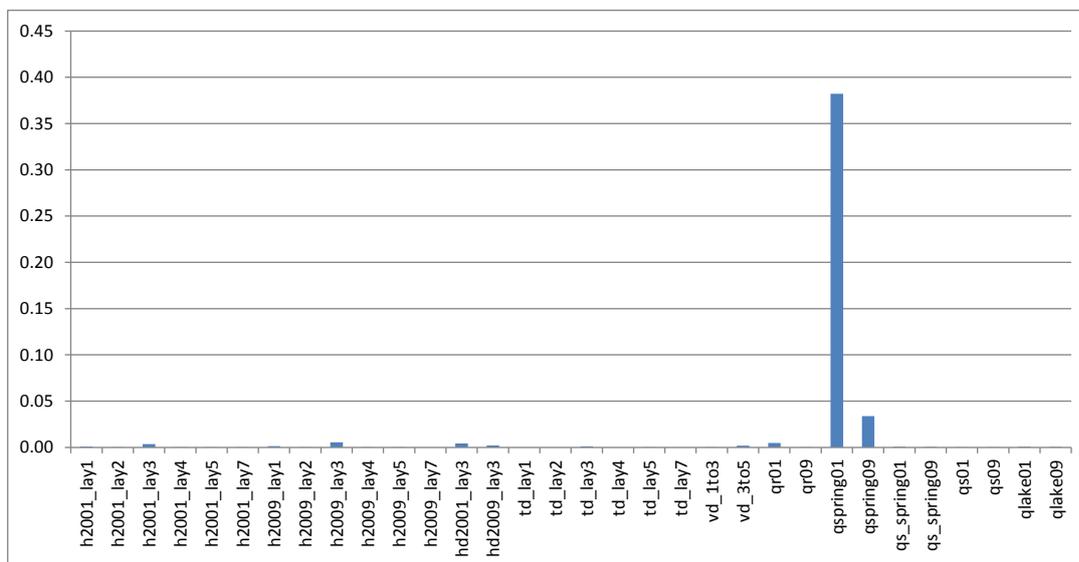
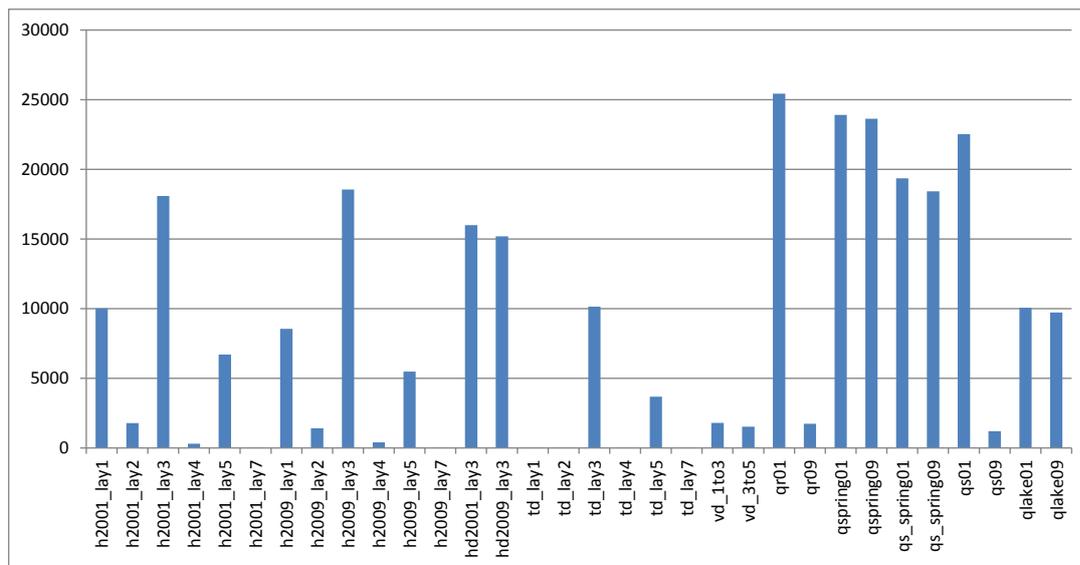
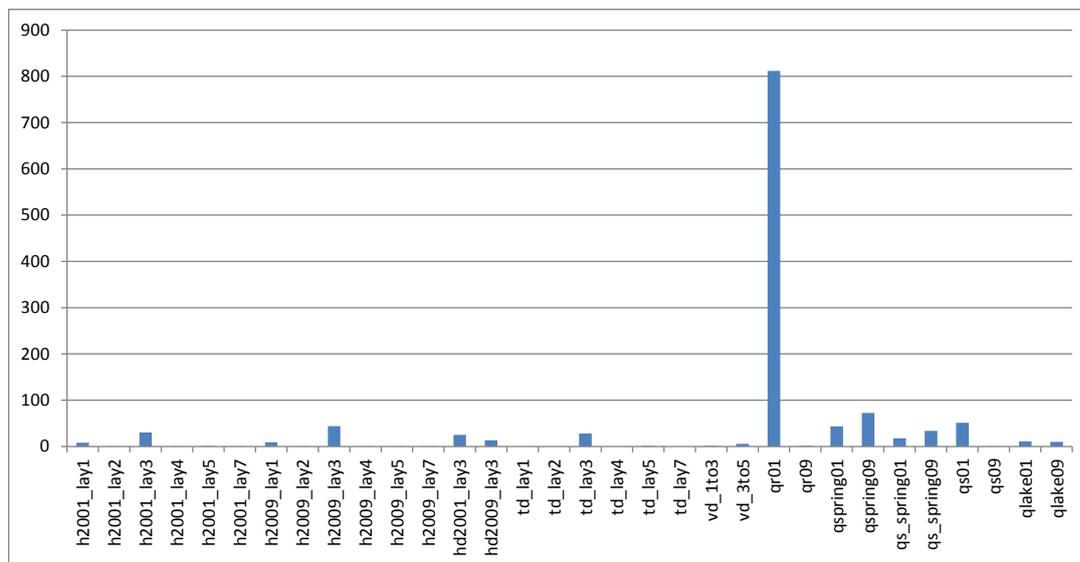


Figure 5.3b Increase in the uncertainty variance of prediction *qspring09\_s12161002* from its post-calibration value of 0.257 (ft/day)<sup>2</sup> incurred if each observation group is removed from the calibration dataset.



**Figure 5.4a** Decrease in the uncertainty variance of prediction *qs09\_2322500* from its pre-calibration value of 29968 (ft/day)<sup>2</sup> accrued if each observation group comprises the entirety of the calibration dataset.



**Figure 5.4b** Increase in the uncertainty variance of prediction *qs09\_2322500* from its post-calibration value of 1701 (ft/day)<sup>2</sup> incurred if each observation group is removed from the calibration dataset.

It is apparent from the above figures, and from those provided in the previous subsection, that observations which are similar in nature to a prediction contain unique information pertaining to that prediction. Where the prediction is flow from a spring, this is an outcome of that observation’s ability to inform the conductance of the boundary condition that governs changes in the amount of water that flows from the spring in response to changes in groundwater head. However other parameters also influence where and how much water emerges from the groundwater system from springs and rivers. In particular, regional hydraulic conductivity parameters influence the direction of groundwater flow, and hence the locations of its emergence. During the history-matching process, hydraulic conductivities are illuminated by a number of different types of observations, including measured heads and head differences.

## 6. Some Predictive Difference Outcomes

### 6.1 General

It is generally accepted that models are better at predicting changes than absolutes. Defects in their construction arising from their simplified numerical representation of complex environmental processes can inculcate consistent biases in various model predictions. Defect-induced biases of this type will often “cancel out” as the value of a prediction pertaining to one simulation time is subtracted from its value at another simulation time in order to predict the change in system behaviour precipitated by alterations to human management of that system. For the same reason, the use of predictive differences can also ameliorate the effects of erroneous parameter values, or of parameterization devices that do not accurately represent the nature of hydraulic property heterogeneity within a system. Values ascribed to such parameters during the calibration process often reflect the surrogate roles that they must play in order to compensate for unavoidable model simplification if a good fit between model outputs and historical system behaviour is to be attained.

For the above reasons, model-based environmental management can often benefit from a reliance on predictive differences rather than on predictive absolutes for formulation of management policy, and for implementation of management strategies arising from that policy.

Unfortunately, the use of linear methods to quantify an improvement of a model’s ability to predict changes rather than absolutes is hampered by the fact that finite-difference calculation of the predictive sensitivity vector (i.e. the  $\mathbf{y}$  vector of equation 2.2) is degraded where sensitivities are calculated for model output differences. This occurs because differences must then be taken of model-calculated differences. Loss of leading significant figures is incurred at each differencing stage. For the NFSEG model, this problem is compounded by solver convergence issues associated with use of the MNW2 package.

This problem is exacerbated by the fact that the values of predictive differences are generally much smaller than the values of predictions themselves. The numerical smallness of management-salient quantities that then become the focus of uncertainty quantification requires that calculations undertaken at all stages of the uncertainty quantification process be as precise as possible. Errors in the  $\mathbf{Z}$  matrix featured in equations 2.3 and 2.4 arising from problematic solver convergence under calibration conditions can thus contribute to errors in quantification of predictive uncertainty where predictions of interest are model output differences.

The present section examines the uncertainties associated with a number of differences in model outputs between 2009 and 2035 arising from alterations to pumping over that period. As will be discussed, these uncertainties are explored in two ways – firstly through standard linear analysis based on equations 2.3 and 2.4, and secondly through direct evaluation of predictive differences based on models runs undertaken using parameter sets calculated using equation 2.5. The latter method was employed for the first time in this study.

### 6.2 Predictive Differences through Direct Linear Analysis

Table 6.1 lists the same predictions as does table 5.2; the predictions are described in table 5.1. Pre- and post-calibration uncertainty variances associated with differences in model outputs between 2009 and 2035 are listed in columns 3 to 6 of table 6.1. The values of the predictive differences

themselves are listed in column 2. Predictive difference uncertainty calculations were carried out using equation 2.3, with the  $\mathbf{y}$  vector in each case encapsulating the sensitivities of the pertinent predictive difference to parameters employed by the NFSEG model.

Predictive difference	Modelled value	Pre-calibration variance	Post-calibration variance	Pre-calibration std dev	Post-calibration std dev
w00202_09	-1.3383	0.2893	0.0539	0.5379	0.2322
w00258_09	-1.3206	0.3830	0.0880	0.6189	0.2967
w00878_09	1.2176	1.0274	0.0604	1.0136	0.2459
qr09_2319500	49.2268	1458.5180	278.2862	38.1906	16.6819
qr09_2320700	0.0110	0.0049	0.0036	0.0697	0.0601
qr09_2323500	6.8752	1258.7760	396.4176	35.4792	19.9102
qr09_2324000	-0.0154	192.9498	144.3981	13.8906	12.0166
qr09_2326000	0.1290	0.1500	0.0357	0.3873	0.1889
qr09_2326550	0.1105	499.9878	172.9847	22.3604	13.1524
qspring09_s121610002	0.0545	0.4491	0.1421	0.6702	0.3769
qspring09_s101429027	0.0000	0.0000	0.0000	0.0001	0.0001
qspring09_s101429001	1.0914	38.7342	23.8690	6.2237	4.8856
qspring09_n011117008	0.6568	21.8880	4.5077	4.6785	2.1231
qr09_lsf_sprgrp	10.5886	5450.0760	1333.9000	73.8246	36.5226
qr09_iche_sprgrp	12.8240	486.5039	235.8756	22.0568	15.3582
qr09_wacissa_sprgrp	0.0606	407.8221	141.1818	20.1946	11.8820
qr09_silver_sprgrp	1.1722	6775.8200	677.4674	82.3154	26.0282
qs09_2315500	0.1182	0.0632	0.0298	0.2514	0.1726
qs09_2317620	0.0540	0.1252	0.1147	0.3539	0.3386
qs09_2319000	3.8130	47.7346	10.6169	6.9090	3.2584
qs09_2320500	65.3840	2749.0020	400.2897	52.4309	20.0072
qs09_2321500	0.1455	0.5582	0.4184	0.7472	0.6468
qs09_2322500	23.6933	3379.7580	1190.6590	58.1357	34.5059

**Table 6.1 Pre- and post-calibration variances and standard deviations of uncertainty for selected predictive differences.**

A comparison of table 6.1 with table 5.2 reveals that while the uncertainties associated with many predictive differences are smaller than those associated with their absolute counterparts, the uncertainties of others have not been reduced by much. See for example predictions *qspring09\_s12161002* and *qs09\_2322500* which were discussed in detail in the previous section. Figures 6.1 to 6.4 show parameter contributions to the uncertainties of these predictive differences, and the worth of different observation types with respect to these predictive differences.

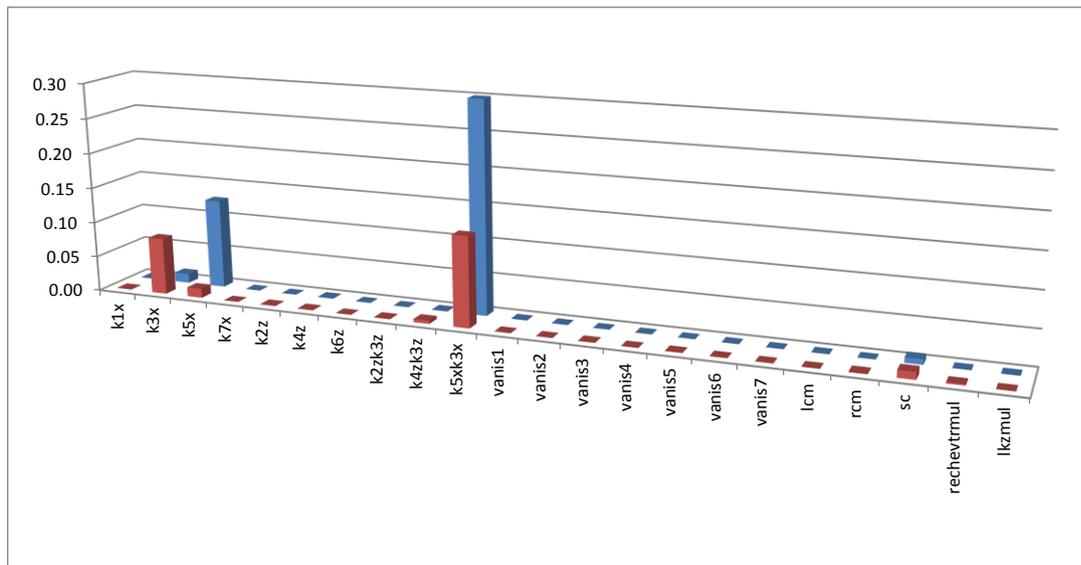


Figure 6.1 Contributions made by different parameter groups to the uncertainty variance of predictive differences for *qspring09\_s12161002*. Pre- and post-calibration predictive difference variances are 0.449 (ft/day)<sup>2</sup> and 0.142 (ft/day)<sup>2</sup> respectively.

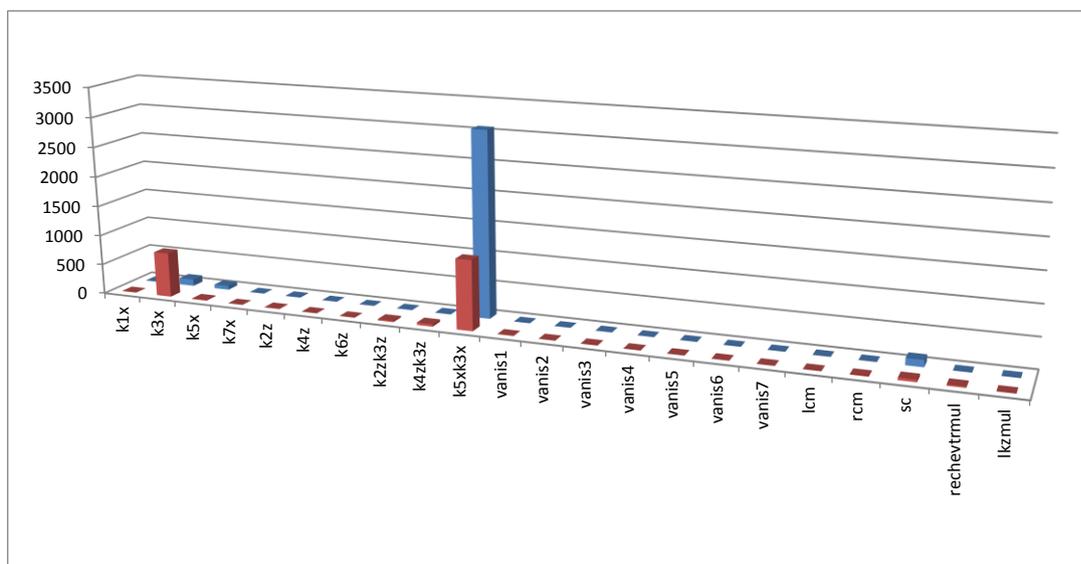


Figure 6.2 Contributions made by different parameter groups to the uncertainty variance of predictive differences for *qs09\_2322500*. Pre- and post-calibration predictive difference variances are 3380 (ft/day)<sup>2</sup> and 1191 (ft/day)<sup>2</sup> respectively.

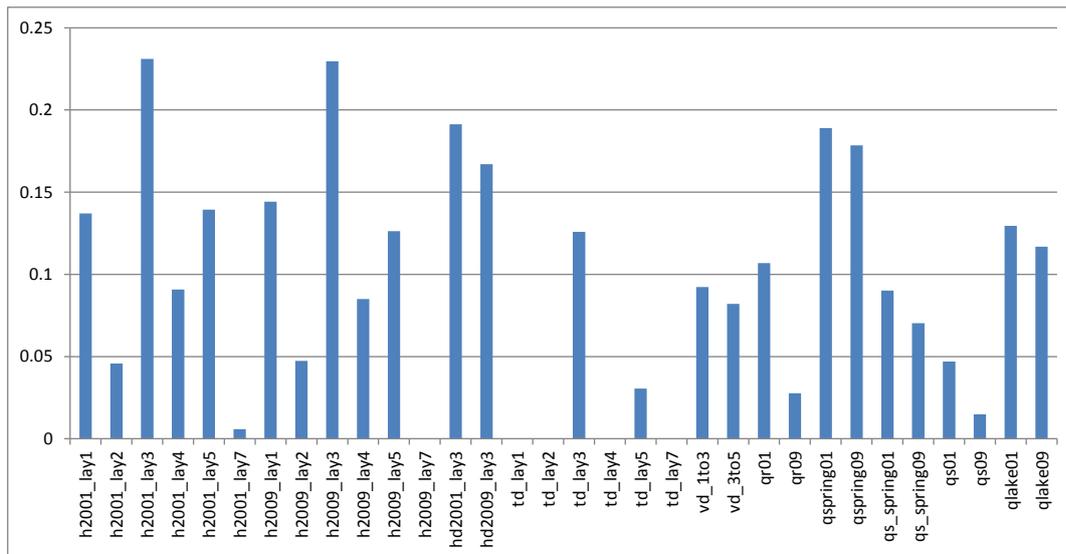


Figure 6.3a Decrease in the uncertainty variance of predictive difference *qspring09\_s12161002* from its pre-calibration value of 0.449 (ft/day)<sup>2</sup> accrued if each observation group comprises the entirety of the calibration dataset.

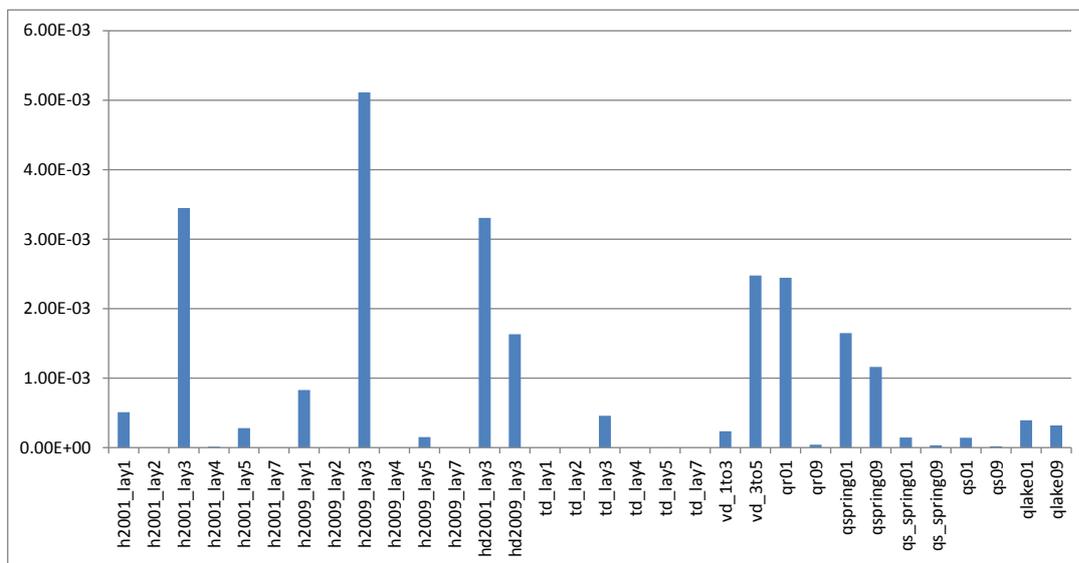


Figure 6.3b Increase in the uncertainty variance of predictive difference *qspring09\_s12161002* from its post-calibration value of 0.142 (ft/day)<sup>2</sup> incurred if each observation group is removed from the calibration dataset.

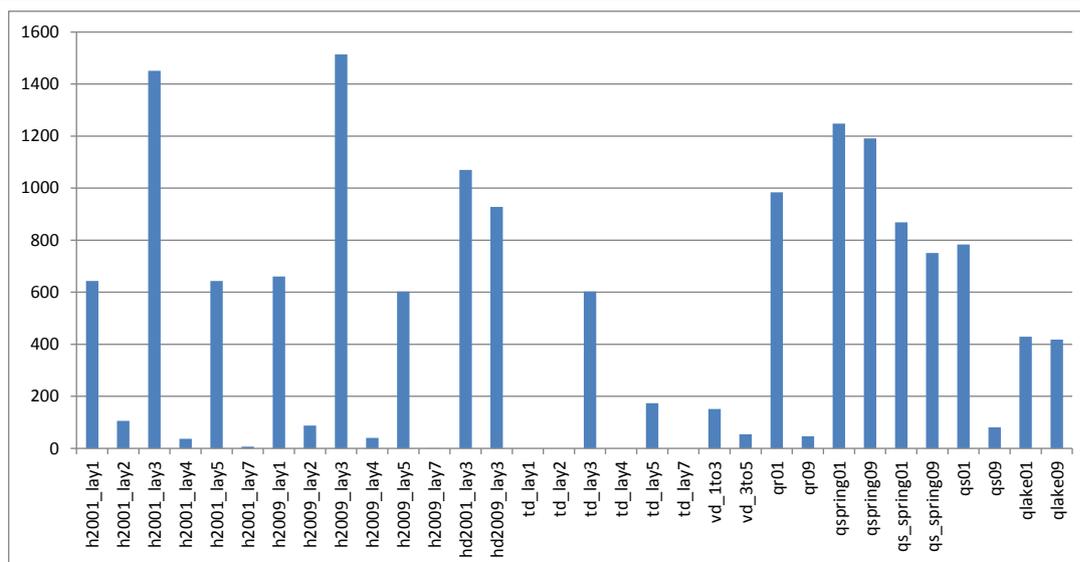


Figure 6.4a Decrease in the uncertainty variance of predictive difference *qs09\_2322500* from its pre-calibration value of 3380 (ft/day)<sup>2</sup> accrued if each observation group comprises the entirety of the calibration dataset.

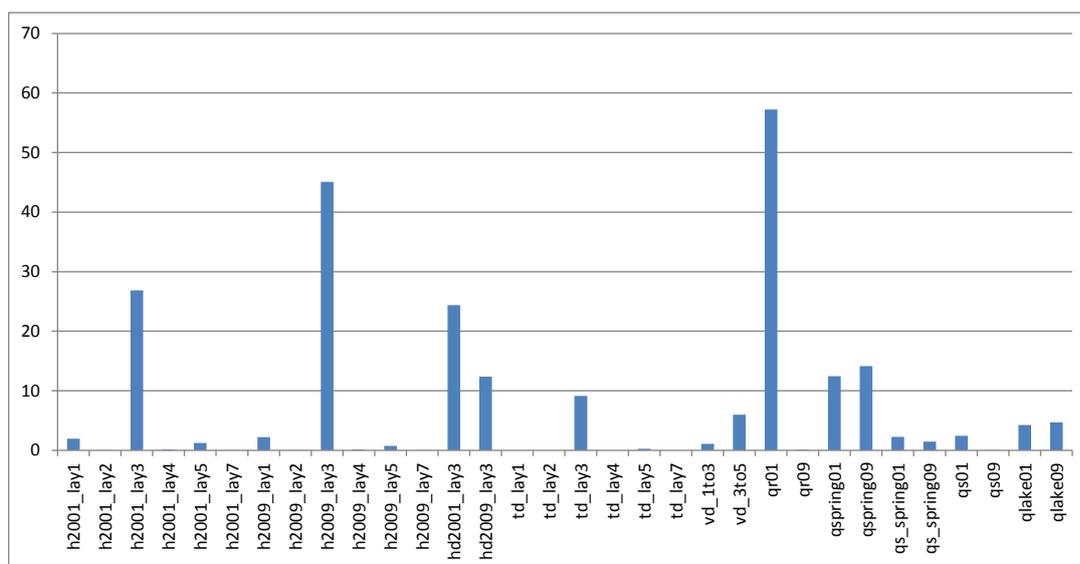


Figure 6.4b Increase in the uncertainty variance of predictive difference *qs09\_2322500* from its post-calibration value of 1191 (ft/day)<sup>2</sup> incurred if each observation group is removed from the calibration dataset.

Figures 6.1 and 6.2 illustrate a phenomenon that sometimes occurs when exploring parameter contributions to predictive uncertainty, namely that the post-calibration contribution of some parameter groups to the uncertainty of a prediction is greater than the pre-calibration contribution of those same parameter groups to the prediction’s uncertainty. This is an outcome of the definition of “contribution to predictive uncertainty”. Recall from section 2 of this report that this is defined as the decrease in predictive uncertainty accrued through acquisition of perfect knowledge of the parameter group in question. A prediction may not be sensitive to any member of a particular parameter group. However it may be sensitive to other parameters with which members of the same parameter group are highly correlated in the post-calibration context due to sharing of information contained in the calibration dataset between members of the different groups; that

information may pertain to a combination of parameters rather than to parameters individually. Suppose that a prediction is sensitive to parameter A but not to parameter B. Suppose also that parameters A and B show a high degree of post-calibration correlation. Acquisition of knowledge of parameter B therefore reduces the post-calibration uncertainty of parameter A, and with it the post-calibration uncertainty of any prediction that is sensitive to parameter A, even if that prediction is not sensitive to parameter B.

The graphs presented so far in this section (which pertain to predictive differences) are somewhat different in character from their counterparts in the previous section (which pertain to predictive absolutes), particularly those pertaining to observation worth. Some of this may, of course, be attributable to corruption of sensitivities by double finite differencing, a matter which is explored below. However even if this is the case, it does not necessarily follow that assessments of parameter contributions to predictive uncertainty, and assessments of data worth in relation to specific predictions, are entirely corrupted as a consequence. Problematical finite-difference derivatives calculation may indeed promulgate errors in a number of bars featured in the charts presented above; however other aspects of these plots probably retain integrity. If this is the case, then the importance of hydraulic conductivity parameters in figures 6.1 and 6.2, and of data types that support estimation of hydraulic conductivity parameters in figures 6.3 and 6.4, suggest that uncertainties associated with groundwater outflow predictive differences arises more from local nuances in directions of groundwater flow than from uncertainties in gross controlling mechanisms such as the conductances of groundwater outflow boundaries.

### 6.3 Predictive Difference Uncertainties: Further Examination

As stated above, the uncertainties associated with some predictive differences listed in table 6.1 are suspiciously high. To explore the integrity of these uncertainties, the following methodology was adopted.

- For a number of predictions featured in table 6.1,  $\delta\mathbf{k}$  was calculated using equation 2.5, together with the pertinent predictive difference standard deviations listed in that table.
- Two parameter sets were created, these deviating from the calibrated parameter set by  $\pm\delta\mathbf{k}$ .
- The model was run under both calibration and predictive conditions using these parameter sets, together with the parameter set  $\mathbf{k}$  pertaining to the calibration model. Because of convergence difficulties, the model was run multiple times with updated initial heads to ensure model solver convergence under each scenario.
- Differences in model predictions made using the calibrated parameter set  $\mathbf{k}$ , and the parameter set  $\mathbf{k}-\delta\mathbf{k}$  on the one hand and  $\mathbf{k}+\delta\mathbf{k}$  on the other hand, were used to assess the predictive difference uncertainty interval.
- Meanwhile, it was verified that the calibration objective functions calculated using both the  $\mathbf{k}-\delta\mathbf{k}$  and  $\mathbf{k}+\delta\mathbf{k}$  parameter sets differ by only a small amount from that calculated using the  $\mathbf{k}$  parameter set.

Objective function components obtained using the  $\mathbf{k}$  parameter set, as well as those calculated using the  $\mathbf{k}-\delta\mathbf{k}$  and  $\mathbf{k}+\delta\mathbf{k}$  parameter sets with  $\delta\mathbf{k}$  tuned to specified predictions, are listed in tables 6.2a and 6.2b. In most cases, the addition or subtraction of  $\delta\mathbf{k}$  to/from the calibrated parameter set makes little difference to the objective function and to all of its components. In some cases,

however, there are differences, especially to objective function components that pertain to flow. This is almost certainly an outcome of solver convergence difficulties.

	calibration	w00202_09	w00202_09	w00878_09	w00878_09	qr09_iche_sprgrp	qr09_iche_sprgrp
Objective function components	<u>k</u>	<u>k-δk</u>	<u>k+δk</u>	<u>k-δk</u>	<u>k+δk</u>	<u>k-δk</u>	<u>k+δk</u>
total	5712.8	6267.2	6053.4	5753.9	5734.2	6168.3	5921.5
h2001_lay1	227.99	232.1	226.01	227.32	229.85	228.44	227.59
h2001_lay2	95.987	97.104	95.505	96.212	95.823	96.109	95.879
h2001_lay3	976.61	1009.4	957.29	990.48	967.49	979.93	974.24
h2001_lay4	12.987	13.129	12.858	13.065	12.923	12.956	13.019
h2001_lay5	39.001	39.745	38.9	38.848	39.193	39.115	38.937
h2001_lay7	1.9947	2.0314	1.9667	2.0239	1.9709	1.9943	1.9956
h2009_lay1	238.94	240.46	237.84	234.42	244.04	239.51	238.44
h2009_lay2	110.98	111.49	110.62	111.11	110.88	111.13	110.84
h2009_lay3	992.3	997.03	1002.3	987.31	1000.9	993.3	991.71
h2009_lay4	10.006	10.037	10.009	9.9537	10.069	9.9948	10.018
h2009_lay5	41.026	41.337	41.093	40.889	41.183	41.108	40.969
h2009_lay7	2.0185	1.9175	2.129	1.9423	2.0928	2.0157	2.0202
hd2001_lay3	288.96	292.75	286.77	287.84	290.5	290.93	287.45
hd2009_lay3	262.99	265.62	261.36	263.46	262.69	263.5	262.68
td_lay1	0	0	0	0	0	0	0
td_lay2	0	0	0	0	0	0	0
td_lay3	639.29	640.11	638.89	639.89	639.68	638.31	640.33
td_lay4	0	0	0	0	0	0	0
td_lay5	34.027	34.111	33.953	34.056	34.011	34.023	34.035
td_lay7	0	0	0	0	0	0	0
wp_wet_2001	231.17	228.53	240.92	226.72	237.71	230.97	231.53
wp_wet_2009	36.005	35.946	36.113	35.961	36.113	36.006	36.004
vd_1to3	91.583	579.7	428.38	93.855	89.818	529.21	288
vd_3to5	85.002	85.093	84.902	84.999	85.002	87.262	83.342
qr01	374.03	375.61	372.66	376.38	375.5	376.38	380.81
qr09	375.93	379.02	373.12	377.3	381.62	374.52	385.21
qspring01	7.0077	6.1279	8.0224	6.7343	7.3146	12.002	6.6449
qspring09	6.9885	7.4952	6.6231	7.5335	6.4628	8.6893	10.515
qs_spring01	10	10.002	10.009	9.9898	10.011	9.97	10.042
qs_spring09	5.9983	6.0493	5.9466	6.0177	5.9792	5.998	5.9979
qs01	257.06	261.75	268.33	305.99	239.85	256.25	258.14
qs09	256.92	263.5	260.89	243.65	275.45	258.72	255.44

**Table 6.2a Objective function components calculated when varying parameters to explore predictive difference uncertainty margins – part A.**

Objective function components	qs09_2320500	qs09_2320500	qs09_2321500	qs09_2321500	qs09_2322500	qs09_2322500
	$k-\delta k$	$k+\delta k$	$k-\delta k$	$k+\delta k$	$k-\delta k$	$k+\delta k$
Total	5720.00	5717.20	7300.80	7324.20	5763.80	5757.60
h2001_lay1	228.07	227.9	228.18	227.8	226.41	229.73
h2001_lay2	96.049	95.925	96.302	95.605	95.663	96.324
h2001_lay3	987.03	967.56	977.04	976.29	979.76	977.33
h2001_lay4	13.057	12.92	12.988	12.987	12.693	13.26
h2001_lay5	39	39.026	39.018	38.983	37.175	40.91
h2001_lay7	2.02	1.9722	1.995	1.9946	2.0104	1.9812
h2009_lay1	239.19	238.68	239.42	238.47	238.7	239.48
h2009_lay2	111.01	110.95	111.37	110.51	110.74	111.24
h2009_lay3	986.9	999.12	992.45	992.13	992.45	996.12
h2009_lay4	9.9869	10.031	10.006	10.006	9.8461	10.161
h2009_lay5	40.941	41.125	41.035	41.016	40.245	41.871
h2009_lay7	1.9656	2.0742	2.0175	2.0194	1.9887	2.0448
hd2001_lay3	288.55	289.41	289.02	288.91	287.26	291.43
hd2009_lay3	263.31	262.71	263.02	262.96	263.26	263.06
td_lay1	0	0	0	0	0	0
td_lay2	0	0	0	0	0	0
td_lay3	640	638.78	639.25	639.36	639.1	639.53
td_lay4	0	0	0	0	0	0
td_lay5	34.088	33.977	34.028	34.028	34.023	34.037
td_lay7	0	0	0	0	0	0
wp_wet_2001	230.72	231.71	231.26	231.07	233.12	229.76
wp_wet_2009	35.964	36.053	36.006	36.004	35.919	36.368
vd_1to3	91.61	92.134	1675.9	1706.9	99.035	116.31
vd_3to5	83.625	87.014	86.657	82.978	84.802	85.407
qr01	377.68	373.9	373.97	374.09	386.34	371.92
qr09	375.34	379.73	375.69	376.17	402.85	359.99
qspring01	7.2425	6.8252	7.07E+00	6.9504	2.484	24.801
qspring09	6.7754	7.2456	6.9743	7.0039	17.666	11.477
qs_spring01	10.029	9.973	10.021	9.9782	10.339	10.136
qs_spring09	5.992	6.0009	6.0027	5.9934	6.0176	5.974
qs01	255.82	258.48	257.11	257.04	255.26	260.17
qs09	258.01	255.95	256.93	256.93	258.68	256.79

**Table 6.2b Objective function components calculated when varying parameters to explore predictive difference uncertainty margins – part B.**

Table 6.3 shows the outcomes of predictive model runs undertaken to assess predictive difference uncertainty intervals.

Prediction	Value of prediction calculated using $\underline{k}$	Value of predictive change from 2009 to 2035 calculated using $\underline{k}$	Value of predictive change from 2009 to 2035 calculated using $\underline{k}-\delta\mathbf{k}$	Value of predictive change from 2009 to 2035 calculated using $\underline{k}+\delta\mathbf{k}$
w00202_09 <sup>1</sup>	78.37	1.3383	1.2311	1.4223
w00878_09 <sup>1</sup>	27.16	-1.2176	-1.4508	-0.9865
qr09_iche_sprgrp <sup>2</sup>	-255.54	-12.8240	-13.1784	-12.4925
qs09_2320500 <sup>2</sup>	-4067.08	-65.3840	-65.4740	-63.9390
qs09_2321500 <sup>2</sup>	-36.00	-0.1455	-0.1453	-0.1440
qs09_2322500 <sup>2</sup>	-676.95	-23.6933	-24.3685	-22.9294

<sup>1</sup>Values are in feet. Positive predictive changes mean drawdown.

<sup>2</sup>Values are in cubic feet per second. Negative predictive changes mean reduction in flows.

**Table 6.3 Predictive difference uncertainty intervals obtained by direct running of the NFSEG model.**

Predictive difference uncertainty standard deviations implied by the last two columns of table 6.3 are significantly smaller than those presented in table 6.1, especially for groundwater outflow predictions. This, together with the unexpectedly large uncertainty values featured in table 6.1, suggests that the uncertainty intervals listed in table 6.3 are better approximations to the true uncertainty intervals of these predictive differences than those forthcoming from table 6.1. It could be argued, however, that calculation of  $\delta\mathbf{k}$  using equation 2.5 may itself be corrupted by the same numerical problems as those that beset calculation of  $\sigma^2$ , using equation 2.3 or 2.4, and that this may invalidate the predictive difference intervals provided in table 6.3. It is nevertheless suggested, however, that the estimates of predictive difference uncertainty listed in table 6.3 are improvements over those listed in table 6.1 because their calculation involves running of the model, and does not rely completely on the integrity of sensitivity vectors and matrices.

If nothing else, the above analyses demonstrate that evaluation of the uncertainties of predictive differences is a numerically difficult procedure. Two possible alternatives to the linear methodology pursued herein are as follows.

- Calibration-constrained Monte Carlo methods such as the Null Space Monte Carlo (NSMC) methodology supported by PEST could be employed. A problem with the NSMC method however is that, while being numerically efficient when compared to other methods used for post-calibration uncertainty analysis in highly parameterized contexts, its numerical burden is nevertheless very high. Furthermore, for the NFSEG model, time-consuming manual intervention may be required to ensure that each parameter field emerging from the NSMC process is valid according to a suite of pertinent metrics (see Sepulveda and Doherty, 2014).
- A far cheaper, but still somewhat approximate, methodology based on equation 2.5 could be employed whereby a line search is undertaken along the direction of the predictive sensitivity vector. In undertaking this line search, the prediction of interest would be maximized or minimized subject to the constraint that the calibration objective function remains below a user-specified threshold.

As a partial assessment of the credibility of predictive difference uncertainty figures presented in table 6.3, absolute (rather than differential) predictions listed in table 6.1 were calculated using the calibrated parameter set  $\underline{k}$ , as well as the  $\underline{k}+\delta\mathbf{k}$  and  $\underline{k}-\delta\mathbf{k}$  parameter sets pertinent to each respective predictive difference. While the  $\delta\mathbf{k}$  parameter fields are pertinent to predictive differences rather

than to predictive absolutes, use of these parameter fields to make absolute predictions can be expected to expose at least some of the uncertainties associated with them. Table 6.4 shows the outcomes of these calculations. The second column lists the average absolute alteration to the pertinent prediction as made by the calibrated model that is incurred when altering the calibrated parameter field by  $\delta\mathbf{k}$ , and then  $-\delta\mathbf{k}$ , to make the same prediction. The post-calibration standard deviation calculated for the uncertainty of each of these predictions (taken from table 5.2) fills the final column of table 6.4. Agreement between the two columns is not unreasonable.

Prediction	Average predictive difference calculated using $\mathbf{k}-\delta\mathbf{k}$ and $\mathbf{k}+\delta\mathbf{k}$	Standard deviation of prediction from table 5.2
w00202_09	2.3281	0.9790
w00878_09	1.20885	0.9601
qr09_iche_sprgrp	21.086	17.5036
qs09_2320500	63.98	151.8987
qs09_2321500	0.689	8.5351
qs09_2322500	34.712	41.2421

**Table 6.4 Comparison between approximate predictive uncertainties calculated using parameter fields geared toward exploration of predictive difference uncertainties, and predictive difference standard deviations computed using linear analysis.**

## 6.4 Predictive Noise

This section has focussed on examining the uncertainties associated with the ability of a calibrated model to predict changes in the state of a system resulting from changes in the way that the system is managed. Results presented above suggest that where these predictions are similar in nature and location to those at which measurements were made for use in the calibration process, then the uncertainties associated with these predictive differences are very small. A similar conclusion was reached by Sepulveda and Doherty (2014). However these authors based their analyses on the considerably more expensive (because it is nonlinear) null space Monte Carlo methodology.

The uncertainty intervals that are quantified in the present section using linear and semi-linear methods are reflective of parameter uncertainty only. No predictive noise term has been included in these analyses. In accordance with concepts presented in section 2.3.4 of this document, these can thus be classified as “uncertainty intervals” rather than “predictive intervals”. For predictions considered in the present section, failure to consider predictive noise is a less serious omission than for predictions considered in the previous section, as simplification-induced errors in model predictions are likely to largely cancel where differences are taken between model outputs calculated for different times at the same location. Furthermore, as was discussed in the previous section, assessment of the effect of model defects on the uncertainties associated with specific model predictions is likely to be fraught with difficulty. While indicators of the magnitude of predictive noise may be gained from an inspection of residuals incurred for model outputs of similar type and location employed in the calibration process, it is a difficult matter to translate these indicators into useable estimates of predictive noise in characterization of predictive difference uncertainty.

It is therefore, in a way, inconvenient that the above analyses suggest that model predictive uncertainty intervals are very small for the model predictive differences that were considered in these analyses. It follows that if residuals for calibration counterparts to model outputs used in the

---

differencing process are considerably higher than these uncertainties, and if the structural noise whose existence these residuals suggest does not quite cancel through differencing, then this uncanceled model error will constitute the principle source of uncertainty for these particular predictive differences. Sadly, the magnitude of this possible error cannot be quantified.

## 7. Conclusions

Linear predictive uncertainty analysis has been applied to calculation of the uncertainties associated with some predictions made by the NFSEG model. These predictions include both the values of heads and groundwater outflows, as well as alterations to these heads and outflows following alterations to pumping. Linear analysis appears to perform well when assessing the former types of predictions. Its performance appears to be poor, however, when assessing uncertainties associated with the latter types of predictions. Problems in using linear analysis in conjunction with predictive differences probably have their roots in numerical contamination of finite-difference derivatives which form the centrepiece of linear uncertainty analysis; where calculation of values for predictions requires that one model output be subtracted from another, differences of model output differences must be taken in formulation of these derivatives.

To overcome numerical difficulties associated with computation of the uncertainties associated with predictive differences, a new method was applied that relies partially on finite-difference sensitivities and partially on carrying out model runs to actually compute the predictive differences. While the method is unlikely to be completely immune from the numerical problems that beset predictive difference uncertainty evaluation using purely linear methods, it appears to provide more realistic estimates of the uncertainties associated with predictive differences. It is probable that the method could be further developed to improve quantification of predictive difference uncertainties while still retaining a high degree of numerical frugality.

Analyses reported herein suggest that uncertainty intervals pertaining to differences of model outputs that are similar to model outputs used in the calibration process are very small. The uncertainties of these differential predictions are therefore likely to be dominated by unquantifiable effects of model structural defects. Hopefully, however, these will be small in magnitude, as their expressions are supposedly largely eliminated through the model output differencing process.

Indicators of the magnitude of the effects of structural defects on predictions that are similar in nature to model outputs used in the calibration process can be gleaned from the magnitudes of calibration residuals associated with these model output types. To some extent, the smaller are these residuals, the smaller will be the magnitude of potential predictive noise associated with related predictions (and predictive differences). However caution must be exercised in attempting to reduce residuals by too much as a structural noise eradication strategy, especially for a coarse-gridded regional model such as the NFSEG model. White et al (2014) and Doherty and Christensen (2011) show that while it may be possible to reduce residuals to low levels in a highly parameterized calibration context, parameters may be forced to adopt roles which compensate for model defects in achieving this level of model-to-measurement fit. The predictive outcomes of the surrogate roles that parameters may thereby be forced to play may be worse than the original problem of high residuals; an unknown level of unquantifiable bias may be introduced to many model predictions.

Finally, this document has also exemplified other quantities that can be calculated using linear analysis as an adjunct to highly parameterized inversion. These include parameter identifiability, relative parameter uncertainty variance reduction, contributions made by different parameter groups to the pre- and post-calibration uncertainties of predictions of interest, and the information that members of the calibration dataset carry that is pertinent to these same predictions.