Efficient Calibration and Uncertainty Analysis Using Surrogate Models Conjunctively with a Complex Groundwater Model.

Wesley A. Burrows

School of the environment

Flinders University

A thesis submitted for the degree of

Doctor of Philosophy

JULY 5, 2016

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Wesley Burrows

Coauthorship

This PhD thesis has been structured to incorporate work that has been either published or accepted in international scientific journals as well as work that has not been provided for peer review. Two research journal articles have been produced from the work undertaken during the course of research. At the time of thesis submission these two articles had been published. These published articles are listed below.

I am the lead author in both of these articles and as such was responsible for leading the research contained within them. Also I declare that I was responsible for the majority of the final writeup, submission and publication of these articles. I sincerely appreciate and warmly acknowledge the valuable contribution made by my co-author and co-supervisor, Dr. John Doherty in preparation of these articles.

I also acknowledge the contribution made to these publications by the anonymous peer reviewers whose critique of these works provided much useful feedback for improvements that were subsequently incorporated.

Paper 1:

Burrows, W.A. and Doherty, J., 2014. "Efficient Calibration/Uncertainty Analysis Using Paired Complex/Surrogate Models". *Ground Water*, Vol. 53, no. 4, pp. 531-541. doi: 10.1111/gwat.12257.

Paper 2:

Burrows, W.A. and Doherty, J., 2016. "Gradient-based model calibration with proxy-model assistance". *Journal of Hydrology*, Vol. 533, pp. 114-127. doi: 10.1016/j.jhydrol.2015.11.033.

Acknowledgements

I wish to acknowledge the contributions of Dr. John Doherty in acting as co-supervisor and providing guidance. John has been a wonderful mentor, and the opportunity to learn from him has been tremendously rewarding. I hope that in some small way at least I have contributed to the ongoing development of the PEST suite of software.

I also wish to thank Dr. Craig Simmons for the support and guidance he has provided through his role as principal supervisor for my candidacy.

Without the unwavering support of my immediate and extended family, pursuit of this research would not have been possible. In particular, my wife Jane has been astonishingly dependable and the naively apposite words of encouragement from my sons Finlay and Oliver have brought clarity of vision when the path seemed impaired.

It is likely that I would never have undertaken this research if not for the opportunity granted me by the National Centre for Groundwater Research and Training (NCGRT). I am sincerely grateful for the funding provided by NCGRT and Australian Postgraduate Awards (APA) that enabled pursuance of this endeavour.

Summary

This PhD research addresses some of the impediments to calibration and uncertainty analysis of complex groundwater models, whose long runtimes and sometimes questionable numerical stability often renders those analyses intractable. Research documented herein explores the use of surrogate and proxy models in overcoming these impediments. More specifically this research focuses on the application of surrogate and/or proxy models conjunctively with an original complex model, in facilitating and expediting gradient-based calibration and uncertainty analysis.

Gradient methods have the advantage over so-called global methods in that they are generally much faster and can readily be adapted to include formal mathematical regularisation which can accommodate large numbers of adjustable parameters. This supports calibration and uncertainty analysis for a broad range of physically-based, distributed models wherein complex environmental processes are simulated within heterogeneous media. Gradient methods are, however, highly dependent on the so-called Jacobian matrix which is comprised of derivatives of all model generated equivalents to calibration dataset observations with respect to all adjustable parameters. These derivatives are usually calculated from model outputs using finite-differencing methods. When parameters are many and model runtimes are long, population of the Jacobian matrix can be extremely computationally demanding. Also, the integrity of finite-difference derived derivatives can be severely degraded by numerical inconsistencies that often attend complex model outputs. Research documented herein demonstrates the novel use of a faster running and more numerically stable surrogate model for population of the Jacobian matrix that can overcome these difficulties, therefore promulgating calibration and uncertainty assessment of problematic complex models when it was otherwise not possible.

This study exemplifies a number of differing simplification strategies that can be implemented in this novel approach including: (1) use of a single model based on a coarser grid; (2) use of multiple surrogate models based on parameter-specific grid

coarsening; (3) use of a model that employs an alternative simulation algorithm and; (4) use of a large suite of observation-specific analytical proxies.

Results from these demonstrations give great cause for optimism that the surrogate-enabled gradient methods have a bright future in modern groundwater modelling. As models become more complex, and as decision makers and stakeholders increasingly demand that predictions of future environmental outcomes made by models are accompanied by estimates of the uncertainties associated with those predictions, the need for parameterisation complexity will grow. So too will be the requirement that calibration and uncertainty analysis be based on gradient methods. It is anticipated that in the next generation of modelling in support of the decision-making process, the role of surrogate and proxy models in that process will expand.

Contents

List of Figures

List of tables

Chapter 1

Introduction

1.1 Overview

The use of complex physically-based numerical models for the prediction of future system behaviour has become common place in recent years in both environmental and engineering contexts. Decisions of political, economic or environmental importance are often based on predictions made by such models and the integrity of these decisions is underpinned by the assumed accuracy of those model generated predictions. It is often the belief, that the calibration process is what endows a model with the ability to predict future system behaviour and that through inclusion of a high degree of detail (complexity) in a numerical model, that the assumed accuracy of a prediction will be enhanced. The inconvenient truth is that the most a calibrated, complex model can hope to achieve is: (1) the quantification of the model's potential for predictive error; and (2) the reduction of that potential for predictive error to its theoretical minimum. It must therefore be a matter of moral and ethical diligence that the potential in a model for predictive "uncertainty" be assessed, duly accounted for and appropriately considered when decisions are to be based on these predictions. Perhaps the greatest impairment to the correct reporting of numerical model uncertainty, outside a lack of knowledge of the implementation of such analyses, is the large run times and sometimes questionable numerical stability of the "complex" models conceived from a desire to make better predictions. There is therefore an increasing need for methodologies that can promulgate conservative, while at the same time tractable, estimates of predictive uncertainty in circumstances where it may otherwise not be achievable due to excessive computational burdens and/or numerical model misbehaviour. This research attempts to address the impediments to calibration and predictive uncertainty assessment of problematic complex models, when such analyses are performed within gradient-based frameworks.

In groundwater or reservoir simulation models, complexity is accrued as a consequence of attempts to include in the model conceptual detail commensurate with expert knowledge of the system under investigation. It may arise as algorithmic complexity in attempting to compute solutions to more involved governing equations that better represent underlying physical processes. Complexity also arises from attempts to include in the model: detail pertaining to complex geological structures and their continuity (or lack thereof); knowledge of the degree and spatial extents of geological heterogeneity; knowledge of important water exchanges (which are often spatially and time variable) and; knowledge of spatially and time-varying stress factors. Numerical implementation and representation of these physical and geological processes then often requires fine scale spatial and temporal model discretisation, inevitably resulting in long model runtimes. When analysis such as calibration and predictive uncertainty assessment is to be performed, these processes require repeated model runs. The huge computational burdens that then ensue can be particularly onerous in the rigorous treatment of these analyses.

In addition, models that have been constructed to include a large amount of complexity often display a propensity for numerical instability. These numerical instabilities may be an outcome of round-off errors that propagate from within the algorithm/s employed by the model. They may also arise from the non-linear nature of many physically-based distributed models. Normally, use of a particular numerical algorithm will demand that specific discretisation criteria are met to ensure numerical stability is achieved, however strict adherence to these criteria cannot always be attained or guaranteed in construction of a distributed physically-based model. Adherence to these criteria becomes particularly problematic when hydraulic properties assigned to the model (at least in some areas of the model domain) are allowed or required to display large variances in values assigned to adjacent grid cells or elements. While sophisticated solvers can employ devices such as adaptive time stepping that are able to mitigate the onset of numerical oscillations and facilitate solution convergence, criteria defining attainment of solution convergence is often a matter left to the discretion of the modeller. When solution convergence is not strictly achieved, numerical errors will attend model outputs. The deleterious effects of model numerical instability become particularly onerous when the model is

subjected to analysis schemes that require systematic (or random) adjustment of model parameters. In some cases, even where parameters are altered by only small increments from values employed in a previously stable model run, model solution convergence cannot be attained. In other cases, inconsistencies in model outputs can lead to unreliable model output to parameter relationships when these are estimated from pertinent model outcomes arising from incremental changes in model parameters. Both model run failure and unreliable model output to parameter relationships can be inimical to any calibration or uncertainty analysis scheme.

The question of how much complexity should be included in construction of a groundwater model such that the model is able to fulfil the objectives of a particular modelling exercise is elegantly addressed by *Doherty and Simmons* (2013) who discuss optimal groundwater model design as it relates to the decision making process. In that text it is argued that optimality of model complexity should be measured on two metrics: (1) the necessity to include sufficient model parameterisation, process and structural detail to ensure that predictive uncertainty is not underestimated; and (2) that justification for increased model complexity is only forthcoming when rejection of the hypothesis that an undesirable outcome can occur, cannot be concluded through use of a simpler model. Unfortunately, the matter of optimal model simplification is rarely considered in development of many project terms of reference; rather a high degree of complexity is often a mandatory requirement. Nor is model simplification usually foremost in modellers' minds when they embark on the process of assimilating into a model the many disparate datasets, expert conceptualisations and/or recommendations provided to them while also maintaining respect for key project deliverables required of them. There is therefore a tendency for modellers to over-extend from the outset in terms of the level of complexity that they build into models. It is only at the point where automated model calibration is attempted that the undesirable manifestations of model complexity as described above become fully apparent.

Where a model cannot be calibrated its value as a prediction making tool cannot be supported, nor can the benefits of appropriate model calibration be realised. That is to say, the model's ability to make predictions that have minimised potential for error

cannot be supported. At the same time the models ability to reduce a prediction's uncertainty to its theoretical lower limit (determined by model to measurement misfit) may be foregone. When calibration of a model to an acceptable level of observation misfit is not possible because of long runtimes and/or numerical instability, then a rigorous analysis of predictive uncertainties will certainly be more challenging if not abandoned altogether. Subsequently, the use of the model to inform the decision making process will be severely compromised.

1.2 Calibration and Uncertainty Analysis

Model calibration is the process, normally undertaken prior to the model being used for predictive purposes, in which a unique parameter set is found that enables the model to adequately replicate historical measurements of system state to within limits defined by measurement noise. Calibration of physically-based distributed models such as those used to simulate sub-surface, hydraulic processes constitutes an inversion problem that is inherently ill-posed and therefore has a non-unique solution; this issue has been recognised as early as in the works of *Carrera and Neuman* (1986a, 1986b). Inverse problem theory and implementations are detailed comprehensively in the works of *Oliver et al.* (2008), *Aster et al.* (2012), *Doherty* (2015b) or any good textbook on the topic. Recent developments in the history matching process (model calibration) have been compiled by *Oliver and Chen* (2011) with specific areas including (1) parameterisation; (2) form of the objective function; (3) minimisation algorithms; and (4) uncertainty quantification, forming the focus of this work. *Oliver and Chen* (2011) also identify the application of the ensemble Kalman Filter (enKF) as a significant area of recent advancement.

While the often stated reason for inclusion of a high degree of model complexity is to provide more accurate predictions on which important management decisions can be based, an *accurate* prediction is unfortunately something that a model cannot substantively claim. *Moore and Doherty* (2005, 2006) show that even if a model is free of defects and is perfectly calibrated, uncertainties associated with many

predictions of interest made by the model may still be accompanied by a high degree of error. This is primarily because while the calibration process may achieve reductions in uncertainties of those parameters to which model output counterparts to calibration data are sensitive, the calibration process may achieve nothing in reducing the uncertainty in parameters to which some important predictions may be sensitive. Those authors explain that parameter uncertainty, and therefore predictive uncertainty, is a consequence of the so-called calibration "null-space". This calibration "null-space" is born of an information deficit in most calibration datasets. When predictions made by the model are sensitive to parameters and/or combinations of parameters that reside in the calibration "null-space", the potential for predictive error may remain high. This is particularly likely in cases where the predictions required of the model differ in spatial and/or temporal proximity to the elements of the calibration dataset, and where measurements that comprise the calibration dataset are acquired under a stress regime that differs to that under which predictions are to be made. Unfortunately, this is precisely the backdrop under which many groundwater models are framed when they are used to support the decision making process.

In spite of the fact that non-uniqueness prevails in most, if not all hydrological and hydrogeological settings, a unique solution is often still sought so that a model may be deployed to make prediction/s of interest. The unique set of model parameters thus found is then deemed to "calibrate" the model; hence model prediction accuracy is then implied. However, when viewed from the perspective of Bayesian inference, what is actually achievable through model calibration becomes more apparent. When viewed in this way, historical measurements of system state to which the model is to be matched are considered to contain a certain amount of error. This error is often treated as additive noise that accompanies the true value of the measurement. Historical measurements must then be characterised by a probability distribution. Model parameters are usually only known to within a range of values on the basis of expert knowledge; they too must also be characterised by probability distributions. Inversion is then considered a conditioning process in which the prior probability distribution of model parameters is constrained by the necessity for the model to replicate historical measurements to within an acceptable tolerance. This tolerance is

normally determined by the amount of noise in the latter, although model structural error will usually require that a tolerable level of misfit must be larger than that which could be achieved if measurement noise where the only contributor to this misfit. The outcome of this conditioning process is therefore another probability distribution. This process is described by Bayes equation which can be written in its simplest form as:

$$P(\boldsymbol{k}|\boldsymbol{h}) \propto P(\boldsymbol{h}|\boldsymbol{k})P(\boldsymbol{k}) \tag{1.1}$$

where $\boldsymbol{k}$ is a vector of model parameters; $\boldsymbol{h}$ is the vector of historical measurements comprising the calibration dataset; $P(\boldsymbol{k})$ is the prior probability density function of parameters, this expressing expert knowledge as it relates to uncertain parameters; $P(\boldsymbol{h}|\boldsymbol{k})$ is the likelihood function, this increasing with the reduced level of model to measurement misfit attained through the calibration process; and $P(\boldsymbol{k}|\boldsymbol{h})$ is the posterior parameter probability density function and describes the uncertainty that remains in parameter values following conditioning. Bayes equation explicitly states that the result of this conditioning process is in fact an infinite number of different sets of parameters, all of which are able to adequately fit historical measurements to within the aforementioned tolerance. Model predictive uncertainty then exists as any one of these parameter sets when provided to the model will likely result in a different value for a particular prediction from that obtained from any other parameter set that is compatible with the posterior parameter distribution.

The second term on the right hand side of Equation (1.1), express what is most attractive about a complex model. Namely, $P(\boldsymbol{k})$ represents its ability to incorporate all facets of expert knowledge. Aside from the ability to express what we do know about a particular study site, this term allows for expression of what is unknown or uncertain about these processes. Stochastic representation of these known and unknown components of system detail allows for their bearing on a particular prediction of interest to be explored and quantified. Expression of expert knowledge or lack thereof, through the vehicle of model parameterisation detail is thus a major contributor to model complexity. Detailed expression of the innate variability of geological structures and other relevant processes at a particular study site requires

that many model parameters be employed. Use of a high level of parameterisation detail (hence complexity) affords additional benefits when the conditioning of these parameters is undertaken so that the model is better able to reproduce historical measurements of system state including:

1. It allows for more flexibility of response to information contained in the historical measurements that comprise the calibration dataset. In this way the modeller does not need to decide in advance whether a parameter/process is significant (hence inferable) or not. Meanwhile parameters are free to respond to information forthcoming from calibration data which allows for better assimilation of information contained within historical measurements of system state (*Doherty*, 2003);

2. While some (often many) parameters may not be inferable on the basis of calibration data, they may have a significant bearing on the outcome of a prediction required of the model. Omission of these non-inferable parameters may result in artificial reduction in perceived uncertainties associated with model predictions if these predictions are in fact sensitive to the omitted parameters. This topic is addressed by *Doherty and Welter* (2010);

3. Use of many parameters can reduce structural noise accompanying model outputs that are equivalents to historical measurements, thereby promulgating attainment of a better level of fit with historical measurements. When model structural error exists, its covariance is usually unknowable and this complicates definition and calculation of the likelihood function. Permissibility of parameter sets that are actually part of the posterior parameter distribution may then be compromised. Further discussions on the presence and repercussions of structural noise in the history matching process are provided in the works of *Doherty and Welter* (2010), *Beven* (2005) and *Beven et al.* (2008), to name a few; and

4. Use of many parameters, avoids some parameters taking on "compensatory" roles as the conditioning process seeks to attain goodness of fit with historical measurements. These compensatory roles introduce bias to values assumed

by parameters which can in turn lead to biased predictions. This topic is discussed and examined at length by *Doherty and Christensen* (2011), *White et al*. (2014) and *Watson et al.* (2013).

Theoretically, if rigorous sampling of the posterior parameter probability distribution, $P(\boldsymbol{k}|\boldsymbol{h})$ of Equation (1.1), can be accomplished then the probability distribution of any model prediction can be evaluated using the relationships between these samples and associated predictions, embodied in the model. However, rigorous exploration of the posterior parameter probability distribution becomes problematic when the elements of the vector $\boldsymbol{k}$ number in the hundreds or even thousands (as is often the case in complex models) and model runtimes are long. Bayesian methodologies such as Markov Chain Monte Carlo (MCMC), employed by *Oliver et al.* (1997), *Keating et al.* (2010) and *Kennedy and O'Hagan* (2001), that seek to generate samples directly from the posterior probability distribution, often become extremely difficult to implement under these circumstances. Hundreds of thousands of model runs may be required to adequately sample the posterior parameter probability distribution when parameter numbers exceed approximately 50 or so. Despite development of MCMC methodologies such as the Shuffled Complex Evolution Metropolis algorithm (SCEM-UA) proposed by *Vrugt et al.* (2003) and the Differential Evolution Adaptive Metropolis algorithm (DREAM) presented by *Vrugt et al.* (2008), that realise model run efficiencies through adaptation of the proposal distribution and include parallelisation of the process, these methods remain extremely computationally demanding in high parameter dimensionalities and where model runtimes are long.

High parameter dimensionality is not as problematic for methods that seek a unique solution to the inverse problem of calibration. A unique solution to an ill-posed inversion problem can be guaranteed through use of regularisation. Formal mathematical regularisation strategies such as Singular Value Decomposition (SVD) and Tikhonov schemes (*Tikhonov and Arsenin*, 1977) can be used to achieve a solution to an ill-posed inverse problem that is unbiased and therefore has a minimum potential for error. *Doherty* (2015), *Oliver et al.* (2008) and *Aster et al.* (2012) provide details on the use of regularisation for solution of the inverse problem.

Use of as many parameters as is required to allow stochastic representation of prior knowledge, combined with a formal mathematical regularisation strategy appropriate for the problem at hand affords other benefits such as:

1. The maximum assimilation of information contained within the calibration dataset;

2. If properly formulated, ensure a minimum error variance status for both estimable parameters (or combinations) and inestimable parameters (or combinations); and

3. Allow quantification of the role that the regularisation strategy plays in the constraining of some parameters but not others. This allows quantification of the potential for parameter and predictive error and hence parameter and predictive uncertainty.

As has already been discussed, a single solution to the inverse problem will almost certainly be in error. However use of a regularisation strategy that is appropriate for the problem at hand can promulgate a solution that lies somewhere near the centre of the posterior parameter probability distribution. The estimated parameter set thus obtained will therefore be minimally biased or can be considered to have a minimum potential for parameter error. When the model is populated with this parameter set and run under predictive conditions, a prediction of minimised error potential can also then be expected.

Although the value of parameter error can never be directly calculated (as this would require detailed knowledge of the "real" parameter set) the statistics of this potential for error can be calculated following attainment of a minimised error potential parameter set; the statistics of potential predictive error can then also be calculated. *Tonkin et al.,* (2007) broadly categorise methods which seek to evaluate the potential for model predictive error into two groups (1) predictive uncertainty analysis and (2) predictive error variance analysis. The latter of these focuses on quantification of the range of possible values that a particular prediction can take under the constraints of calibration within a certain tolerance of model generated observation to measurement

misfit. Linear and non-linear methods based on variance propagation have been implemented by *Vecchia and Cooley* (1987), *Christensen and Cooley* (1999), *Moore and Doherty* (2005) and *Tonkin et al.* (2007) in groundwater problems. The former of the two predictive error groupings is a more intrinsic concept that propagates the prior stochastic parameter definitions through a model to develop posterior parameter and prediction probability distributions. To achieve such analysis requires appropriate exploration of the parameter space and includes Bayesian, Markov-chain Monte Carlo techniques as already mentioned. It also includes calibration-constrained Monte Carlo methods such as the efficient null-space Monte Carlo (NSMC) method described by *Tonkin and Doherty* (2009) and employed by *Herckenrath et al.*, (2011). *Keating et al.*, (2010) conducted a comparison between the NSMC method and the more Bayesian DREAM method and reported consistent and similar results for parameter estimation and uncertainty analysis, arising from both these methods; in spite of the inherent difficulties that accompanied each procedure in that example problem. Other methods that seek to derive prediction probability distributions from stochastic parameter definitions include generalized likelihood uncertainty analysis (*Bevan and Binley*, 1992), a non-Bayesian parameter field deformation technique described by *Gomez-Hernandez et al.* (2003), and ensemble Kalman Filtering methods such as used by *Sarkov et al.* (2012) and *Chen and Oliver* (2013).

As is described by *Tonkin et al.* (2007), a linear approximation to the post-calibration covariance matrix of parameters can be calculated using a linearization of the model centred on a "calibrated" parameter set that is ideally of minimised error potential; further details and discussions on this topic can be found in *Dausman et al.* (2010) and *Doherty* (2015b). When combined with the calibrated parameter set, these statistics of potential parameter error may define an approximation to the posterior parameter distribution which through a process of sampling and evaluation can be used to empirically derive histograms of posterior parameter and prediction probabilities. This process, while not strictly Bayesian in nature can be an efficient and effective alternative to direct posterior parameter probability sampling. The approach to approximate definition of posterior parameter probabilities, just described, depends upon the presumption that calibrated parameters attain minimised

error status. Formal mathematical regularisation strategies, if used appropriately, can ensure that this optimal calibration status is achieved.

Gradient methods provide a run efficient means for solution of the inverse problem. They are normally much faster at arriving at a unique solution to the inverse problem than other so-called "global" methods for estimation/optimisation such as particle swarm optimization (*Kennedy and Mendes*, 2002); shuffled complex evolution (*Duan et al*., 1992); genetic and evolutionary programming (*Vrugt and Robinson*, 2007); and covariance matrix adaption algorithms (*Hansen et al.*, 2003) among many others. While these global methods are robust is the face of particularly difficult estimation problems, their model run requirements increase rapidly with the number of parameters to be estimated. Gradient methods also afford other benefits including:

1. They can readily be adapted to include mathematical regularisation devices such as Singular Value Decomposition (SVD) and Tikhonov schemes (*Tikhonov and Arsenin*, 1977), thus highly parameterised problems can be easily accommodated while also endowing the inversion problem with the other benefits forthcoming from formal mathematical regularisation strategies already mentioned;

2. As a direct outcome of calibration other post-calibration statistics are readily calculable using a local linearization of the model about the calibrated parameter set such as: parameter identifiability (*Doherty and Hunt*, 2009); parameter and predictive uncertainty (*Gallagher and Doherty*, 2007a and 2007b; *James et al.*, 2009), and the worth of existing and yet-to-be acquired data in terms of its ability to reduce the uncertainties of parameters and predictions of interest (*Dausman et al.*, 2010).

For these reason the use of gradient based methods for numerical model calibration is gaining much traction within the groundwater, surface water and reservoir modelling fraternity. PEST (*Doherty*, 2015a) implements a version of gradient based estimation which is in essence a modified version of the Gauss-Marquardt approach to non-linear parameter estimation. It also has functionality that allows inclusion of SVD and Tikhonov type regularisation strategies, or a combination of both. The suite of

software provided with PEST also has functionality that allows for parameter and predictive uncertainty analysis to be performed. Additional functionality that allows for definition of so-called "super-parameters" to be used in both calibration and assessment of predictive uncertainty assessment, provides for increased efficiency gains; *Tonkin et al.*, (2007) provide a more detailed discussion of this topic.

While efficient, gradient methods are not without their own problems. Such methods rely on parameter sensitivities that collectively form the so-called Jacobian matrix. Calculation of these sensitivities is normally achieved via finite differencing methods in which the difference in the outcome of a particular model output is calculated from an incremental variation in a particular parameter. The difference in pertinent model outputs is then divided by the parameter increment and provides an approximation to the derivative of that model output with respect to the parameter thus varied. In the calibration context, the Jacobian matrix contains one such entry for each model output equivalent to an observation in the calibration dataset with respect to each adjustable parameter in the model. Population of the Jacobian matrix thus requires that the model be run at least once for every adjustable model parameter. That is if a two-point finite differencing stencil is used. Three point or five point (or more) stencils can be used to derive better approximations to these derivatives; when they are implemented model run requirements for population of the Jacobian matrix increases accordingly. Where model runtimes are long population of the Jacobian matrix can be extremely computationally expensive. The iterative process of non-linear parameter estimation requires that re-population of the Jacobian matrix be conducted several times throughout the estimation process. Additionally, where model outputs have a propensity for inconsistency born of model numerical problems, this can lead to the calculation of unreliable parameter sensitivities. Unreliability in the Jacobian matrix can lead to extremely slow inversion solution attainment or complete failure of the process to reduce model to measurement misfit. When numerical instability is pervasive enough, model run failures can occur and prevent Jacobian matrix population altogether.

When population of a reliable Jacobian matrix becomes problematic for the reasons already mentioned, estimation of a parameter set that can claim to be close to the

mean of the posterior parameter distribution (hence approaching maximum likelihood status) may be severely impaired. Assessment of predictive uncertainty via a calibration-constrained Monte Carlo analysis that can efficiently explore post-calibration parameter variability of both estimable and inestimable components of parameter space is thus also impaired, this later process potentially requiring re-adjustment of many stochastic parameter sets to ensure calibration constraints are respected in approximation of posterior probabilities.

In spite of the persuasive reasoning for inclusion of a high degree of model complexity, a complex model may sadly become its own worst enemy when it comes to calibration and the assessment of uncertainties associated with the predictions for which it was initially constructed. This is not to say that model complexity has no place in the decision making process. Indeed the ability of a model to encapsulate all relevant parameter and process detail of a system is of necessity, not because it is estimable on the basis of a noisy, sparse calibration dataset, but precisely because it is not estimable and thus lies within the calibration null-space. If a prediction of interest is dependent on system detail that resides in the calibration null-space, then the model can still be used to quantify the contribution of this in-estimable detail to predictive uncertainty and ensure that the latter is not underestimated. Hence, new methodologies which promulgate this end in the face of computational burdens and numerical misbehaviour of such models will play a significant role in making more informed decision which deliver better economic, social and environmental outcomes.

1.3 Model simplification and Surrogate modelling

To alleviate the problems of excessive computational burdens in analysis schemes involving complex models, recent research in the field of parameter estimation and/or uncertainty quantification, has been keenly interested in the use of faster running surrogate models. *Razavi et al.* (2012a) and more recently *Asher et al.* (2015) provide reviews on the use of surrogate models within the fields of hydrology and

hydrogeology. Of primary concern in the studies compiled by those authors is the need to derive a much faster running variant of the original complex model to which much of the forward model run requirements of the analyses can be assigned. The frameworks in which this strategy is implemented vary greatly, as do the types of surrogate models deployed.

Razavi et al. (2012a) categorises types of surrogate models into 2 main groups. Firstly, there are the so called "response surface models" or "model emulators" which attempt to substitute the costly computational output landscape of the original model through some form of function approximation. They are thus data-driven versions of the original model and have been typically developed using functions such as polynomials as applied by *Fen et al.* (2009) for optimisation of soil vapor extraction design; radial basis functions used by *Regis and Shoemaker* (2004) in optimisation of a groundwater bioremediation problem; and Kriging was enlisted by *Hemker et al.* (2008) to expedite optimisation of well field design; Artificial Neural Networks are also sometimes used and fall under this category. Many other examples of the use of data-driven model emulators are provided by *Razavi et al.* (2012a) however recent and relevant applications of these techniques include *Borgonovo et al.* (2012) who employed model emulation to speed-up sensitivity analysis for predictions in a problem relating to subsurface migration of radionuclides, and *Sun et al.* (2012) who demonstrated a full emulation approach for calibration and uncertainty analysis of reaction rates of biodegradation of a trichloroethylene (TCE) chain reaction in a groundwater system.

More recently, much more complicated versions of data-driven model emulators of simulator outputs have been developed, which employ generalised Polynomial Chaos expansion theory. *Laloy et al.* (2013) provides an example of this strategy within a highly parameterised groundwater flow model. In that study, Hermite polynomials are used as the orthogonal basis describing model output response with respect to continuous hydraulic conductivity fields; assumed Gaussian to allow association with the Hermite polynomials. Analysis efficiency and tractability was obtained through use of a parameterisation scheme based on Karhunen Loeve transformation of the prior cell-by-cell covariances; the assumption of multi-Gaussian prior distribution

allows for a predictable relationship between captured field variability and the number of eigenvectors thus retained. As a precursor to that paper, *Marzouk and Xiu* (2009) had introduced a stochastic collocation scheme based on generalised Polynomial Chaos theory as an efficient alternative to purely Bayesian inference.

The second of the surrogate groupings adopted by *Razavi et al*. (2012a) is the "lower-fidelity" family of surrogates in which model simplifications are employed while attempting to preserve the physical processes encapsulated in the original model. Within this group model reductions usually take the form of a coarsened discretisation grid, parameter lumping or other simplifying assumptions applied to the various boundary conditions or the underlying physics. Examples include the works of *Kennedy and O'Hagan* (2000) who compare multiple levels of discretisation reductions within a Bayesian framework. *Efendiev et al.* (2005, 2009) and *Mondal et al.* (2010) employ lower-fidelity models in two-stage markov-chain Monte Carlo strategies applied to petroleum reservoir simulation. *Sun et al.* (2010) apply a two-stage multi-fidelity approach to optimisation of honeycomb designs in material sciences. They used the particle swarm optimisation method and combine response surface modelling techniques with error correction modelling of outputs from the lower-fidelity model. *Forrester et al.* (2007) utilise co-kriging regression to combine a few expensive "high-fidelity" model samples with many more computationally cheap "low-fidelity" model samplings, in optimisation of aircraft wing design. They use high and low level codes conjunctively in this process and show that the larger search area made possible by the computationally inexpensive low level code provides for a much improved prediction of optimised parameters over the use of a single code.

Asher et al. (2015) add a third category of surrogate models to those declared by *Razavi et al.* (2012), namely projection based surrogates. Surrogate modelling methodologies that find support in this category are those that project modelled governing equations onto a reduced model subspace defined by a limited number of eigenvectors. These eigenvectors are determined either from covariances that represent model parameter to model output relationships (generally obtained from a few "snap-shot" output values of the original model run with stochastic realisations

of model input parameters) or via solution of a generalised eigenvector problem that describes the original model. These specialised model reduction schemes include proper orthogonal decomposition (POD) used by authors such as *Vermuelen* (2005, 2006), *Saide et al.* (2010) and have proven powerful means in achieving model simplification which maintains the process and structural detail that can be supported by prior knowledge and measurement data. Projection based surrogate models also include those that achieve reductions in the dimensionality of input parameter space through application of Kahunen Loeve transformation of prior parameter covariances.

Construction of a surrogate model necessarily requires some form of original model simplification. Model simplifications have historically been employed routinely as a means of achieving a stable solution to the inverse problem as documented by *Carrera and Neuman* (1986a, 1986b) who view these simplifications as a form of regularisation. *Sivakumar* (2004, 2008) advocates the use of "dominant processes concept" (DPC) in simplification of hydrological models to avoid problems in over-parameterised models. It has long been recognised though that models which employ parameterisation schemes that are too simple may prevent the flow of valuable information from the calibration dataset as well as induce biases in predictions made on the basis of these over-simplified parameter sets. These are the sentiments relayed by *Doherty* (2003) and supported by *Hunt et al.* (2007). Recently, effects of model simplifications have been formally examined by *Watson el al.* (2012), *Doherty and Welter* (2010) and *Doherty and Christensen* (2011) who approach the topic from a sub-space point of view and treat simplification induced biases as separable error terms in the definition of the inverse problem. Those studies characterise simplification induced, model defects as the omission of parameters from the calibration process and hence their inclusion in both of the calibration solution and null-spaces. They also show that such parameter omissions can lead to estimable parameters taking on compensatory roles in order to "soak up" model to measurement misfit which can cause greater post-calibration predictive error than existed prior to calibration.

Of particular interest to this research is the apparent change in the manner in which surrogate models are deployed within analysis schemes. Traditional surrogate model

deployment has been one where an abstraction of the original complex model is used as a replacement for the complex model during the computationally expensive analysis. Examples of this approach are *Shultz et al.* (2004, 2006) and *Bliznyuk et al.* (2007) who conducted Markov-Chain Monte Carlo based uncertainty analysis on environmental models where sampling of the posterior is performed entirely by the surrogate. Also in the environmental modelling context *Borgonovo et al.* (2012) performed sensitivity analysis in which it is the surrogate model that performs the many forward model runs required in determination of sensitivity metrics. Increasingly, surrogate model deployment is progressively moving towards a more collaborative interaction between the original complex model and a simplified version of it. Examples of this are found in the studies of *Efendiev et al.* (2005, 2006), *Cui et al.* (2011) and *Mondal et al.* (2010) who employ the so-called "two-stage" MCMC method whereby a simplified surrogate model is used as a pre-screening mechanism for proposal parameter sets. The goal of that strategy is to increase posterior parameter distribution acceptance rates (as evaluated by the original model) by eliminating highly unlikely parameter sets prior to evaluation using the expensive simulator. The evolution of such conjunctive model usage strategies appears to be in response to recognition of the fact that maintenance of both terms on the right hand side of Bayes equation (as expressed by Equation 1.1) is necessary for appropriate exploration of posterior parameter (and therefore prediction) uncertainty. This is something that a simplified model cannot support to the same extent as a complex model.

It is important to point out that of all water related journal articles chosen for detailed consideration by both *Razavi et al.* (2012a) and *Asher et al.* (2015), not one is specifically focused on gradient-based methods for estimation and/or uncertainty analysis. The most relevant contribution to the use of surrogate models within gradient-based schemes is that of *Doherty and Christensen* (2011). Their approach is somewhat similar to the traditional deployment of surrogate models in which a simple model counterpart to an original complex model is used to make predictions. They develop an "offline" error correction model from repeated calibrations of the simple model to complex model outputs so that corrections for simplification induced bias and an approximation of predictive uncertainty can be applied to

prediction made by the simple model. Although authors such as *Vermeulen et al.* (2005, 2006) and *Saide et al.* (2012) have demonstrated reduced models with Quadratic Programming (QP) techniques (which is in essence a specialised gradient method) for calibration of parsimonious synthetic groundwater models, those types of methods require reformulation of the original model's set of governing equations to derive quasi-linearized systems of equations, this requiring extensive model re-programming and is highly context specific. As is noted by *Saide et al.* (2012) QP methods become impractical or even infeasible in real-world modelling contexts without application of some form of model reduction technique. In addition, QP methods have questionable reliability in highly non-linear problems and are not readily adaptable to include mathematical regularisation. Thus they are rarely used, if ever, in complex real-world hydrogeological settings.

To this author's knowledge there is no other study outside of what has been published as an outcome of this research that examines the use of surrogate models within gradient-based analysis schemes that employ sophisticated regularisation techniques suitable for highly parameterised modelling problems. This research attempts to address this gap in the scientific literature at the same time as it attempts to address the difficulties encountered by model independent gradient-based estimation schemes when seeking calibration and/or predictive uncertainty quantification of problematic complex models.

1.4 Research aims and contributions

It is evident that the majority of the recent literature, relevant to surrogate modelling, highlights research and development of Bayesian based approaches to parameter inference and uncertainty quantification, particularly in highly parameterised context. This is likely due to the completeness in sampling the model parameter space of Bayesian workflows, and the amenable implementation (though non-trivial) of the surrogate model into these frameworks as a pre-screening mechanism of proposed models. In this way the simplification induced errors can be treated as additive to the

existing measurement and structural noise and are accounted for stochastically while not directly infringing on the integrity of the a posteriori probability distribution. To this author's knowledge surrogate models have not yet been implemented in such an unobtrusive, intrinsic way within gradient based optimisation methodologies, due to the fundamental differences in the approaches and the absence of the necessary tools with which to implement it. Notwithstanding this, gradient based methods have much to gain from surrogate modelling techniques as they traditionally rely heavily on iterative processes.

This research demonstrates that one way in which gradient based methods for calibration and predictive uncertainty assessments of complex models can benefit from a surrogate-enabled approach is through the assignment of the many runs required for population of the Jacobian matrix to a much faster running and more numerically "well-behaved", simplified surrogate version of that complex model. As has already been mentioned, population of the Jacobian matrix can be extremely computationally expensive, especially when parameters are many and model runtimes are long. Model numerical misbehaviour can also inhibit population of the Jacobian and/or lead to the calculation of unreliable derivatives that degrade the progression of gradient methods.

Introduction of a surrogate model in this way is unobtrusive to the estimation process. During calibration, although the surrogate model is used for the purposes of Jacobian matrix population, the original complex model is maintained for those runs required for testing of parameter improvements, albeit calculated on the basis of surrogate model parameter sensitivities. In this way the potential for surrogate induced bias to accompany estimated parameters is mitigated, as acceptance of an improved parameter set is assessed on the basis of complex model outcomes rather than the surrogate. With attainment of a calibrated parameter set that is without bias, assessments of parameter and predictive uncertainty can then be accomplished. This process also stands to benefit from the computational gains that the conjunctive surrogate/complex model approach offers, especially when a calibration-constrained Monte Carlo analysis is implemented for that purpose.

The objective of this research has been to explore the potential and demonstrate the veracity of this proposed conjunctive usage approach, in promulgating calibration and uncertainty analysis of complex models when those processes would otherwise be intractable because of large computational burdens and numerical misbehaviours. Such an approach has not previously been explored or implemented as the software needed for its implementation had not been developed prior to this research. PEST has now been equipped with the necessary modifications to facilitate such an approach. This "observation re-referencing" functionality readily integrates with the many other features that the PEST suite of software offers for efficient calibration, predictive error variance and predictive uncertainty analysis.

The methodology developed herein further adds to the armoury that modellers have at their disposal when tasked with construction and deployment of large, complex and numerically problematic models of which there is an increasing appetite within the groundwater industry.

Chapter 2

Efficient Calibration and Uncertainty Analysis Using a Complex Model Paired with a Surrogate Model

This chapter presents work arising from this research that has been published in *Ground Water* journal (see *Burrows and Doherty*, 2014). The work presented here is in a form that is almost identical to that in which the published journal article appears. There may therefore be some repetition of material in early sub-sections of this chapter that was previously discussed in the introduction. However its inclusion in the present chapter assists in the overall flow of the chapter.

2.1 Abstract

The use of detailed groundwater models to simulate complex environmental processes can be hampered by (1) long runtimes and (2) a penchant for solution convergence problems. Collectively, these can undermine the ability of a modeller to reduce and quantify predictive uncertainty, and therefore limit the use of such detailed models in the decision-making context. We explain and demonstrate a novel approach to calibration and the exploration of posterior predictive uncertainty, of a complex model, that can overcome these problems in many modelling contexts. The methodology relies on conjunctive use of a simplified surrogate version of the complex model in combination with the complex model itself. The methodology employs gradient-based inversion techniques and is thus readily adapted for use in highly-parameterized contexts. In its most basic form, one or more surrogate models are used for calculation of the partial derivatives that collectively comprise the Jacobian matrix. Meanwhile, testing of parameter upgrades and the making of predictions is done by the original complex model. The methodology is demonstrated

using a density-dependent seawater intrusion model in which the model domain is characterized by a heterogeneous distribution of hydraulic conductivity.

2.2 Introduction

Complex models often employ many parameters. These may be required for characterization of the processes that the model simulates; they may also be used for representation of hydraulic property heterogeneity. Because parameter field details are generally not completely inferable through either calibration or direct measurement, inclusion of a high number of parameters in calibration-constrained predictive uncertainty analysis can avoid under-estimation of the uncertainty associated with predictions of management interest, particularly if these predictions are sensitive to parameterization detail. The use of many parameters also allows implementation of inversion methods and regularisation devices that maximize transfer of information from calibration datasets to the calibrated parameter field. Hence while predictions made by a calibrated model will almost certainly be accompanied by error, models calibrated using inversion methods that can accommodate many parameters have a reduced propensity for predictive bias, and can provide an appropriate platform for analysis of predictive uncertainty. See *Moore and Doherty* (2005), *Doherty and Welter* (2010), *Hunt et al.* (2007) and papers cited therein for a full discussion of these issues. Furthermore, as is explained in these works, the use of a large number of parameters allows for much finer scale model detail to be varied as the uncertainties associated with predictions of interest are investigated. In analysing predictive uncertainty, parameter combinations that are estimable on the basis of the calibration dataset are then varied over a reduced range of values, the limits of their post-calibration variability being set by the level of measurement noise associated with the calibration dataset on which basis they are estimated. Those that are inestimable (and thus comprise the calibration null space) remain variable over a range of values whose limits are set by expert knowledge as it is applied in the current geological context.

As well as being characterized by large numbers of parameters, complex models are also often characterized by long run times and a propensity for solution convergence

difficulties. Both of these are often an outcome of the highly nonlinear nature of the environmental processes (such as groundwater/surface water interaction, unsaturated and multiphase flow, concentration and temperature-dependent density, etc.) that such detailed models attempt to simulate. Long model run times and convergence difficulties makes calibration of these models a difficult undertaking, at the same time as it renders calibration-constrained predictive uncertainty analysis almost impossible.

Difficulties in working with complex models have spawned the development of uncertainty analysis methods that rely on conjunctive use of a complex model with a surrogate model that runs much faster than the complex model, and is much less prone to numerical problems than the original complex model. For example, authors such as *Efendiev et al.* (2005 and 2009), *Cui et al.* (2011) and *Mondal et al.* (2010) used Markov-chain Monte Carlo analysis to sample posterior parameter probability distributions, employing a simplified surrogate model as a screening mechanism for acceptance of proposal parameter sets prior to evaluation of complex model outcomes based on these sets. Use of a surrogate model in this way is shown to dramatically increase acceptance rates of proposal parameter sets thereby greatly reducing computational burdens. However while Markov chain Monte Carlo methods have the advantage of full compliance with Bayes equation, their use becomes problematical where parameter numbers are high, as exploration of posterior parameter and predictive uncertainties may require a prohibitively large number of model runs under these circumstances (see for example, *Keating et al.* 2010).

Gradient-based subspace methods provide an alternative means of exploring posterior parameter and predictive uncertainty through techniques such as the "null space Monte Carlo" (NSMC) methodology described by *Tonkin and Doherty* (2009) and implemented by *Herckenrath et al.* (2011) in the seawater intrusion modelling context. This methodology is available through the PEST suite (*Doherty*, 2015a). Gradient-based methods make use of sensitivities of model outputs with respect to model parameters, these being encapsulated in the so-called Jacobian matrix. Calculation of these sensitivities can be a computationally expensive process

particularly when the number of parameters to be estimated is large and where sensitivities are calculated through finite parameter differencing. What is of greater concern is that the numerical integrity of finite-difference sensitivities may be severely compromised where a model suffers from convergence difficulties.

In this paper we describe a method that employs a simplified version of a complex model as a surrogate model for calculation of sensitivities within gradient-based calibration and uncertainty analysis frameworks. It is shown that considerable computational savings can be gained through use of this methodology in many modelling contexts when performing these types of analysis. The methodology is demonstrated using a model to which a large number of parameters is assigned to ensure adequate representation of hydraulic property heterogeneity. While use of a large number of parameters does not constitute an essential context for use of the methodology described herein, the highly parameterised nature of the example serves to highlight the advantages to be gained through use of a surrogate model in many contemporary groundwater modelling applications. At the same time it is hoped that this also serves to differentiate this methodology from other examples of conjunctive complex/surrogate model usage cited in the literature. To the authors' knowledge, all previous demonstrations of complex/surrogate model usage employ parsimonious parameterization, and are implemented in conjunction with so-called "global methods" for parameter inference and enforcement of calibration constraints on parameter values during exploration of parameter and predictive uncertainty. See *Razavi et al.* (2012a) for a recent review of surrogate modelling in the water resources field. The present work is, to the authors' knowledge, the first instance of joint complex/surrogate model usage in a gradient-based inversion/uncertainty analysis context.

2.3 Concepts

2.3.1 Uncertainty Analysis using Gradient Methods

Where model predictions are sensitive to hydraulic property values that show a high degree of spatial variability, where direct field measurements of those properties are limited, and where historical measurements of system state are insufficient to allow

unique estimation of those properties, post-calibration predictive uncertainty can be large. See *Moore and Doherty* (2005 and 2006) for a discussion and examples.

Bayesian analysis provides a conceptual basis for quantification of posterior parameter and predictive uncertainty subject to the constraints of prior knowledge of the model parameters (expressed through a prior parameter probability distribution) on the one hand, and historical measurements of system state (expressed through a likelihood function) on the other hand. In practice, direct implementation of Bayes equation in the groundwater modelling context is inhibited by the large number of parameters that are required to represent spatial parametric variability in complex geological environments, and also by the long run times that are often associated with groundwater models that are capable of expressing such parameter detail.

As stated in Section 2.2, subspace methods such as NSMC can be used to achieve similar outcomes to a purely Bayesian analysis, but with greatly reduced computational burden. Through use of the NSMC methodology many random parameter sets can be efficiently generated, all of which are geologically reasonable, and all of which allow model outputs to satisfy calibration constraints. Use of this methodology requires that the model first be calibrated, and that the parameter field achieved through the inversion process incurs as little bias as possible through that process. Ideally, this can be achieved through use of Tikhonov regularisation expressing a "preferred parameter condition", accompanied by a regularisation weighting strategy that induces a preferred correlation structure in heterogeneity. Alternatively (or as well) highly parameterized inversion can be implemented using truncated singular value decomposition in conjunction with Karhunen Loeve transformation of parameters based on this same spatial correlation structure. For further discussion on the use of these regularisation strategies see *Tikhonov and Arsenin* (1977), *Moore and Doherty* (2006) and *Watson et al.* (2013). Both of these regularization strategies are offered by the PEST suite; see *Doherty* (2015a) for details.

The NSMC methodology generates a suite of stochastic parameter fields whose solution space projections (inferable components) are close in parameter space to the parameter field achieved through model calibration, but whose null space projections

(non-inferable components) are variable in accordance with a prior knowledge of system properties. Each such stochastic parameter field "almost" immediately achieves calibration conditions, this being an outcome of the proximity of their solution space projections to those of the calibrated model parameter set. However model non-linearities and a less than perfect delineation of the calibration solution and null spaces requires that a small adjustment of the solution space components, of each stochastic field, be undertaken to ensure that calibration constraints are honoured. Respect for these constraints lowers post-calibration parameter uncertainty and thus reduces the uncertainties of model predictions.

Central to the operation of the NSMC methodology is the so-called Jacobian matrix. This is used for (1) implementation of the inversion exercise prior to undertaking NSMC analysis, (2) decomposition of parameter space into orthogonal solution space and null space components, and (3) adjustment of stochastic parameter fields generated through the NSMC process so that calibration constraints are better respected (i.e. a better fit is obtained between field measurements comprising the calibration dataset and pertinent model outputs).

Sensitivities comprising the Jacobian matrix are normally calculated using a finite difference method in which each parameter in succession is varied incrementally from its current value and model outputs are calculated accordingly. Differences in model outputs divided by the parameter increments are then taken as approximations to corresponding derivatives. Hence at least two forward model runs are required for each adjustable parameter in computation of these finite-difference derivatives (a process that must be repeated many times under non-linear estimation conditions). Filling of the Jacobian matrix is normally by far the largest contributor to the numerical burden associated with gradient-based inversion and uncertainty analysis methodologies.

2.3.2 Reducing Computational Burden through use of a Surrogate Model

Though efficient, the NSMC process may also become computationally demanding where model run times are long and where parameters are many. As already stated,

most of the numerical burden associated with implementation of the NSMC methodology is incurred in calculation of the Jacobian matrix, a matrix that must be re-populated many times during calibration and subsequent approximation of the posterior parameter probability distribution. The computational burden of Jacobian matrix calculation can be greatly reduced by using a fast-running, surrogate model in place of the more complex original model for the many forward model runs required for the filling of this matrix.

While a surrogate model can be used for filling of the Jacobian matrix, use of the original complex model is retained for those model runs that require the full process representation and numerical precision that the complex model is capable of providing. These model runs are those required for testing of parameter upgrades (during the original model calibration process, and then in adjustment of NSMC stochastic parameter sets to respect calibration constraints), and for computation of predictions based on the suite of NSMC-generated, calibration-constrained, random parameter sets.

Use of a simplified surrogate model for calculation of derivatives can provide advantages other than speed of computation. As has already been mentioned, highly detailed, complex models tend to be more susceptible to solution convergence difficulties than their more simple counterparts and such difficulties may prevent or invalidate calculation of finite-difference derivatives. Use of a simplified model that does not have the same problematical numerical behaviour may thus enable calculation of derivatives where this would otherwise be impossible through use of the complex model alone.

Conjunctive usage of a simplified surrogate model with a complex model has been implemented in the latest version of PEST; see the "observation re-referencing" functionality described by *Doherty* (2015a). While simple in concept, its algorithmic implementation is complex, particularly where more than one simple model can be used for derivatives calculation, and where simple and complex model runs can be undertaken in parallel. Both of these capabilities are provided through PEST. Further run management complexity follows from the fact that an additional simple model run must be undertaken prior to filling the Jacobian matrix so that reference values of

calibration-pertinent model outputs can be obtained prior to their use in finite-difference derivatives computation. Where the same model is used for both sensitivity calculation and for testing of parameter upgrades, this "re-referencing model run" is not required, as model outputs calculated during previous parameter upgrade tests can provide these reference values. This additional model run becomes computationally insignificant when the surrogate model has a much faster runtime.

Figure 2.1 presents a conceptual overview of PEST's implementation of complex/surrogate model usage. As indicated in this figure, these capabilities are compatible with PEST's "SVD-assist" functionality, in which so called "super parameters" comprised of parameter solution space projections can be directly estimated, this requiring the filling of a much smaller Jacobian matrix than that based on native model parameters; see *Tonkin and Doherty* (2005) for details. This functionality is integral to the numerical efficiency of the NSMC method. It can also be employed in a stand-alone calibration process, as is demonstrated in the example presented later in this paper.

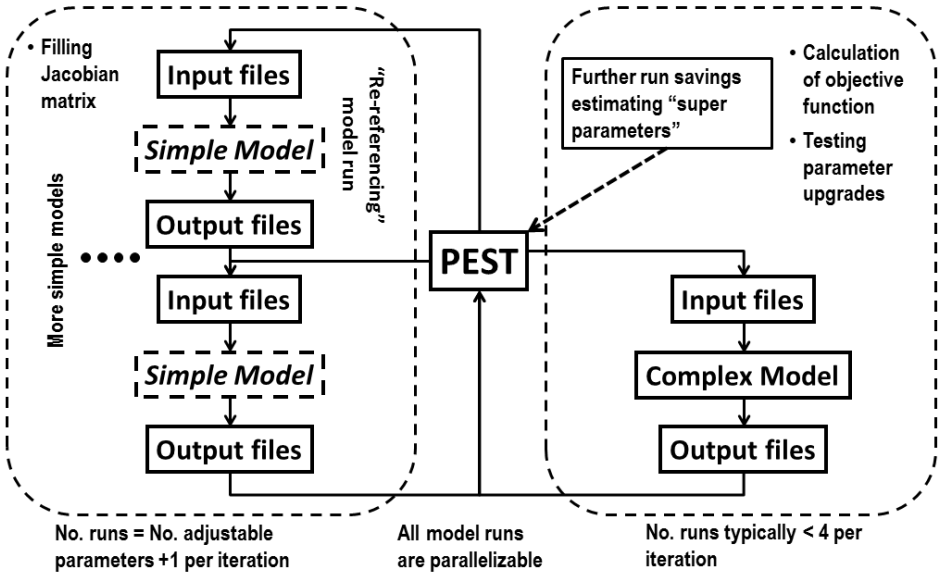

Figure 2.1: Conceptual overview of complex/surrogate model functionality implemented by PEST.

2.3.3 Considerations for Construction of a Surrogate Model

Naturally, certain conditions must be met by a surrogate model if it is to be used in place of a complex model for computation of derivatives. First it is necessary that the simplifications which underpin construction of the surrogate model do not degrade the surrogate model's performance to the point where its outputs are not reasonably consistent with those of the original complex model. Second, the surrogate model must be capable of computing equivalents to all pertinent outputs computed by the complex model (or at least those employed in the calibration process). Third, the parameters employed by the original and surrogate models must play similar roles. Provided these constraints are met there can be considerable latitude in design of a surrogate model.

In practice, some differences in the dependence of model outputs on model parameters between a complex and surrogate model must be expected. After all, simplified surrogate model outputs cannot be expected to have the same quality as complex model outputs. However the deterioration of model output quality that accompanies use of a surrogate model may not necessarily degrade the quality of parameter sensitivities calculated by it. If simplification-induced model output error is reasonably consistent as a parameter is incrementally varied, that error will cancel as model outputs are differenced to calculate the parameter sensitivities. The success of the complex/surrogate methodology as demonstrated in the example presented later in this paper suggests that this may be the case, for that particular problem at least.

Furthermore, when parameters play similar roles within the surrogate and complex models, then parameters (and parameter combinations) to which model outputs are relatively insensitive, will generally be consistent between model versions. Hence, even if surrogate model sensitivities are a somewhat compromised version of the sensitivities that would be calculated using a complex model, they may still have enough integrity to allow definition of calibration solution and null spaces. Their subsequent use in definition of so-called "super parameters" based on singular value decomposition of the sensitivity matrix, in an inversion process that is restricted to solution space parameter components only, will thus have integrity.

Strategies for construction of a surrogate model that can be used in place of a more complex model for derivatives calculation are manifold. For example the surrogate model may employ a less accurate but more computationally efficient algorithm for simulation of pertinent physical processes. Alternatively, faster run times may be achieved through increasing the length of simulation time steps and/or increasing the dimensions of model grid cells (as is done in the example presented below). Another alternative is to employ data-driven strategies such as those used to design metamodels or model emulators, whereby the complex model's response landscape is approximated using analytical functions; see *Razavi et al.* (2012a, 2012b) for further details.

Where the surrogate model remains a numerical simulator of underlying physical processes, strategies other than those mentioned above that may lead to faster run times, with some deterioration in numerical accuracy (but not enough deterioration to invalidate the calculation of finite-difference derivatives) may include (1) decoupling of fully coupled groundwater-surface water exchange, (2) calculation of approximate solute concentrations based on particle tracking, (3) assuming confined, rather than unconfined, conditions to avoid numerical problems associated with cell desaturation, and (4) use of a simplified "lumped parameter" recharge model in place of a more complex unsaturated zone simulator.

Naturally, there can be no guarantee that any particular simplification strategy is universally applicable; whether any one of them "works" or not in any particular context must be established in that context. A condition on which success will depend however is that the surrogate model be free of any numerical problems or algorithmic artefacts that hamper the integrity of finite-difference derivatives calculations.

The specifics of PEST's implementation of the complex/surrogate model methodology are described in *Doherty* (2015a). Outwardly, PEST setup is little different from that which is required for standard model calibration. PEST requires that model input files and model output files have the same format for both the complex and surrogate models; hence the same template and instruction files can be used with both of these models. This may require that a model pre-processor be run

as part of the surrogate or complex model batch/script file to undertake appropriate data translation tasks. Where pilot points are employed as a parameterization device, and where model simplification involves grid coarsening, the batch or script files that encapsulate the complex and surrogate models will need to provide different pilot-point-to-grid interpolators.

PEST also requires that it be supplied with the commands to run the complex and surrogate models. As previously stated, PEST allows different surrogate models to be employed for calculation of sensitivities with respect to different parameters. Where this is the case, surrogate model execution commands must be linked to pertinent adjustable parameters in the PEST control file (via the DERCOM variable).

2.4 Henry Problem Test Case

2.4.1 The Complex Model

We developed a highly parameterised, non-linear, synthetic model to test and demonstrate the use of a simplified surrogate model for calculation of derivatives during calibration, and in subsequent generation of calibration-constrained random parameter fields using the NSMC methodology. The test case presented herein is fashioned on the well-known Henry problem *(Henry,* 1964). The model seeks to predict the change in position of a seawater interface in a heterogeneous hydraulic conductivity field under altered flow conditions. Figure 2.2 shows a schematic of the test case.

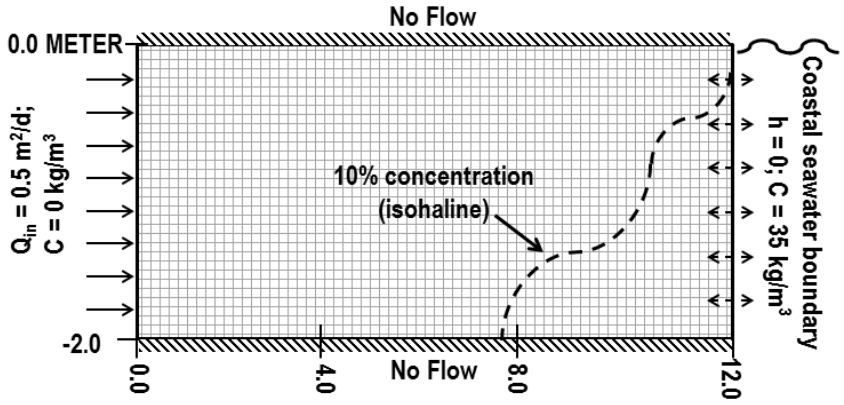

Figure 2.2: Schematic of the synthetic test case.

SEAWAT *(Langevin et al*., 2008) is used to simulate density-dependent flow and transport. Mechanical dispersion effects are neglected (dispersivities α_L and α_T are set to zero). Porosity is specified uniformly as 0.35 and a uniform diffusion coefficient of 0.1 m^2/d is employed. The model domain is vertical and two-dimensional. It is discretised into 100 layers and 600 columns resulting in 60,000 cells with dimensions 2 cm x 2 cm and a nominal width of 100 cm. A constant inflow of freshwater from the western boundary is implemented through the use of injection wells to uniformly distribute a total inflow of 0.5 m^3/d across this boundary. A coastal boundary is defined on the eastern side of the model domain using general head boundary cells with a reference hydraulic head of 0.0 m and a constant salinity of 35 kg/m^3. Areal recharge and leakage is specified as zero. Advective transport is solved using the "time variation diminishing" (TVD) scheme. This model is referred to henceforth as the "fine model" or "complex model".

Two stress periods are simulated. Firstly, a steady-state period is employed to establish an initial system state; model calibration is undertaken under these conditions. Following this, a transient stress period of 1.5 days is introduced in which flow of freshwater into the western boundary ceases, this simulating a dramatic reduction in net inland recharge inducing westward movement of the seawater interface. The average runtime of this fine model is approximately 7.5 minutes for simulation of the steady state calibration period on an i7-720QM machine clocking at 1.6 GHz.

As described below, the model domain is characterized by hydraulic property heterogeneity on a cell-by-cell scale. However for calibration purposes the heterogeneous hydraulic conductivity field is represented using 600 pilot points; a calibration-adjustable parameter is associated with each of these. Interpolation from pilot points to individual model cells employs simple kriging based on a log exponential variogram with range of 7.5 m horizontally and 3.75 m vertically, and a log-mean of 2.301 (200 m/d prior to log10 transformation); variance in the log domain is 0.5. Figure 2.3shows the distribution of pilot points throughout the model domain and also depicts the locations of observation sites used in the calibration process.

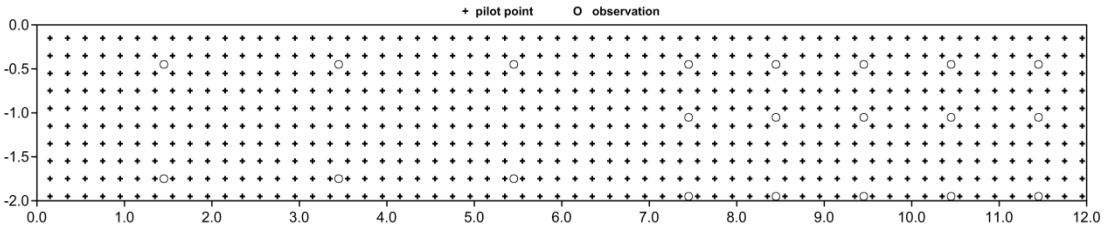

Figure 2.3: Pilot points used for parameterization of the inversion process are shown as dots; observation sites are shown as open circles.

Prior to deciding on a "reality" hydraulic conductivity field for use in the following analysis, the complex model was initially populated with 500 stochastic realisations of hydraulic conductivity generated using a sequential Gaussian simulation algorithm engine based on the SGSIM code supplied with the GSLIB geostatistical library (*Deutsch and Journal*, 1998). The variogram on which these fields are based is identical to that used for kriging from pilot points (see above). We chose as the "reality field" the hydraulic conductivity field that gives rise to the maximum change in position of the interface toe, defined by the 10% seawater concentration contour, when recharge at the western boundary is reduced to zero during the transient, prediction, stress period. The hydraulic conductivity field chosen as reality is shown in Figure 2.4. Choice of a parameter field which leads to a prediction that lies at one of the extremes of the prior predictive probability distribution assists in validation of the integrity of the NSMC uncertainty analysis process, as the reality outcome must

lie within the uncertainty limits calculated using the NSMC methodology. This thus presents a challenging test-case for the proposed conjunctive use methodology.

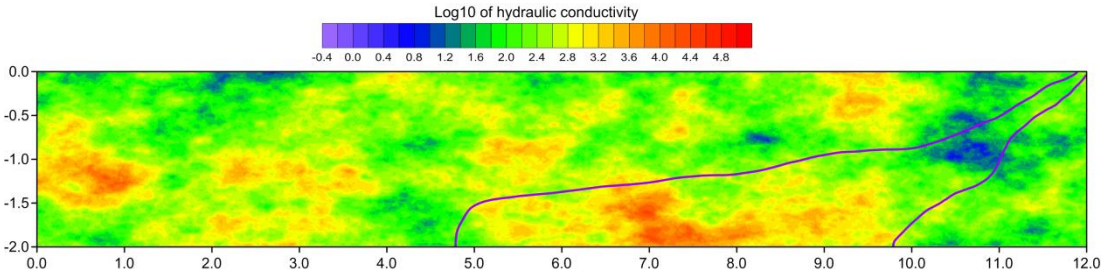

Figure 2.4: Hydraulic conductivity field chosen as "reality". The position of the 10% concentration isohaline at the end of the calibration period (stress period 1) is shown on the right. Its position at the end of the transient predictive period is shown on the left.

Head measurements at the end of the steady state stress period were calculated at all observation sites depicted in Figure 2.3. These head measurements comprised part of the model calibration dataset. Meanwhile concentration measurements for use in the calibration dataset were calculated at the 15 observation points in the right half of the model domain. Random, Gaussian "measurement" noise with a standard deviation of 0.3% of the total range of heads was added to all head observations; similarly randomised "concentration errors" with a standard deviation of 1% of "observed" concentration values were added to concentration observations comprising the calibration dataset.

2.4.2 The Surrogate Model

The surrogate model that was built to complement the complex model described in Section 2.4.1 has an identical domain, and identical boundary conditions, to that of the complex model, and is also 2D vertical. However it employs a much coarser numerical grid, this consisting of only 20 layers and 120 columns (2400 cells in all). To further increase execution speed, advective transport is calculated using the implicit finite difference (IFD) scheme in place of the computationally more

expensive TVD scheme employed by the complex model. The number and location of pilot points used by the surrogate model is, however, identical to that employed by the complex model; the two models thus employ identical parameterization schemes.

To maintain water balance consistency between surrogate and complex models, freshwater inflow into each western boundary cell of the coarse-gridded surrogate model is calculated as the summation of inflows over the five complex model cells which each coarse cell replaces. Similarly, the conductance value employed by each coarsened model general head boundary cell at the eastern end of the model domain is five times that of each of the five fine scale general head boundary cells which it encapsulates. The average runtime of the surrogate model over the steady state calibration period is 10.2 seconds. This represents a factor of 44.5 increase in execution speed over that of the complex model.

2.5 Calibration and Predictive Uncertainty Analysis

2.5.1 Parameter Estimation

As stated in Section 2.3.1 the starting point for implementing NSMC-based generation of calibration-constrained random parameter fields is a calibrated parameter field that is hopefully without bias. Predictions based on the calibrated parameter field are therefore presumably also lacking in bias. This reference parameter field is generally calculated through regularised inversion. In the present case, Tikhonov constraints were applied to the 600 pilot points estimated through the inversion process. Through this mechanism each pilot point was assigned a "preferred value" equal to the mean hydraulic conductivity value discussed above. This assignment takes place through the use of a set of prior information equations to which PEST-calculable weights can be applied. A covariance matrix based on the variogram that was used for generation of hydraulic property heterogeneity was inverted to form the weight matrix applied to these prior information equations. In accordance with its implementation of Tikhonov regularisation, PEST calculates an overall multiplication factor for this weight matrix, subject to the constraint that a

user-specified “target measurement objective function” be achieved. This target was calculated as the expected value of the objective function based on “measurement noise” accompanying the calibration dataset described above. See *Doherty et al.* (2010) and *Doherty and Hunt* (2010) for further discussions on the use of pilot points and regularised inversion in the groundwater modelling context.

To increase calibration speed, PEST’s “SVD-assist” methodology as described by *Tonkin and Doherty* (2005) was employed to estimate “super parameters”. Hence calculation of a Jacobian matrix based on 600 pilot points was required only once. For all iterations of the inversion process only super parameters were varied, these being defined through singular value decomposition of the original Jacobian matrix as those combinations of parameters which are uniquely estimable on the basis of the calibration dataset. The use of a limited number of super parameters in place of actual parameters as a basis for model calibration means that filling of the Jacobian matrix during every iteration of the inversion process requires only as many runs as there are super parameters, instead of requiring as many model runs as there are actual model parameters. A total of 36 such super parameters were estimated; this number being equal to the number of observations in the calibration dataset. Both of the surrogate and fine models used the same parameters for derivatives calculation and testing of parameter improvement respectively. Meanwhile Tikhonov constraints on base parameters were maintained. Estimation of super parameters was implemented using truncated singular value decomposition in order to maintain unconditional numerical stability.

Throughout the inversion process, the surrogate model was used for calculation of all derivatives. These included those required to fill the initial 600 pilot point Jacobian matrix on which basis super parameters were defined, as well as those undertaken for calculation of sensitivities of super parameters themselves as required for implementation of the super parameter estimation process. The fine model was run only to calculate the initial objective function, and to calculate objective function improvements based on improved (in the sense of providing a better fit between fine model outputs and the observation dataset) estimates of values for super parameters. Meanwhile the target measurement objective function sought through PEST’s

implementation of Tikhonov regularisation was applied to the objective function calculated on the basis of fine model outputs, this providing a guarantee that this target was actually achieved.

The target measurement objective function was achieved after 17 iterations. Only 74 runs of the fine model were required. At this stage the model was deemed to be calibrated. The calibrated hydraulic conductivity field is shown in Figure 2.5. Clearly, this field does not show the detail of the reality hydraulic conductivity field depicted in Figure 2.4. As *Moore and Doherty* (2006) explain, this is the "cost of uniqueness", for a calibrated parameter field cannot claim to represent reality. However it can claim to represent the projection of the real hydraulic conductivity field onto a parameter subspace of relatively small dimensions, this subspace spanning only combinations of parameters that are inferable on the basis of the calibration dataset. The seawater interface position calculated using the calibrated hydraulic conductivity field is shown in Figure 2.5. The prediction made by the calibrated model is obviously in error. However, as explained by *Moore and Doherty* (2006), this does not invalidate the model nor the calibration process, but is simply an outcome of the necessarily low dimensionality of the parameter subspace in which the calibrated parameter field lies. The challenge of post-calibration uncertainty analysis is to define predictive uncertainty intervals which encompass this error.

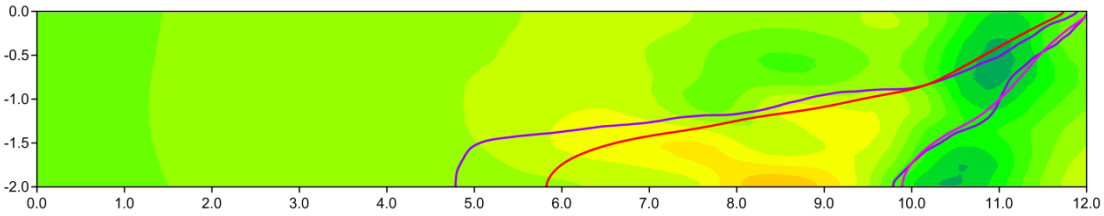

Figure 2.5: Calibrated hydraulic conductivity field arising from the conjunctive model calibration process (see Figure 2.4 for hydraulic conductivity scale). The calculated seawater interface positions at calibration time (right) and prediction time (left) arising from this field are shown as red lines. Also shown as purple lines are the interface positions calculated using the reality hydraulic conductivity field.

For validation and comparative purposes we repeated the calibration process using the fine model only. In this case, estimation of values for the 600 pilot point parameters required 13 iterations and a total of 8493 computationally expensive fine model runs. (The SVD-assist methodology was not employed.) The calibrated hydraulic conductivity field emerging from this exercise is shown in Figure 2.6. It is visually very similar to that presented in Figure 2.5 and achieves only a slightly better fit with the calibration dataset. An inspection of individual parameter values reveals that hydraulic conductivities assigned to pilot points through the two inversion exercises differ in some cases. However these differences are small compared with the prior geological variability of these parameters. (Differences are mostly less than 5% of prior parameter variability, with an occasional difference of up to 20% between individual estimated pilot point values). The use of a surrogate model for derivatives calculation may be responsible for some of these differences; conjunctive use of the SVD-assist inversion methodology may also be responsible for some of them.

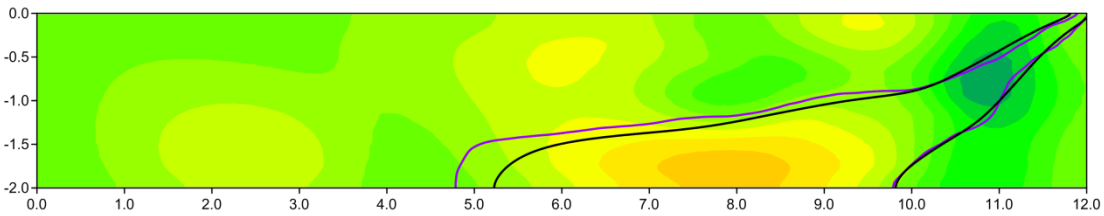

Figure 2.6: Hydraulic conductivity field achieved through calibration of the complex model without use of the simple model for derivatives calculation and without use of the SVD-assist methodology.

A summary of comparative computational costs associated with the calibration of the complex model with and without the use of a simple model for derivatives calculation is provided in Table 2.1.

Table 2.1: Summary of model runs required for parameter estimation and uncertainty analysis using 600 adjustable parameters.

Operation	Complex model runs	Simple model runs
Parameter estimation - complex model only	8493	0
Parameter estimation - surrogate enabled with SVDA	74	3084
Parameter re-adjustment during NSMC analysis	10 (per field average)	74 (per field average)

2.5.2 Null-Space Monte Carlo Analysis

The purpose of the null space Monte Carlo (NSMC) methodology is to generate many different parameter fields, centred on the calibrated parameter field, that achieve a good fit with the calibration dataset, while encapsulating geologically realistic detail that is necessarily missing from the calibrated parameter field because it cannot be estimated uniquely on the basis of measurements comprising the calibration dataset.

Doherty (2015b) and *Tonkin and Doherty* (2007) describe how NSMC-generated random parameter fields can include the same level of cell-by-cell variability as that which characterizes a "reality" hydraulic conductivity field (used by the fine model in the present instance), despite the fact that calibration constraints on these fields are actually enforced through adjustment of pilot point parameters. Briefly, random pilot point values are obtained through sampling of stochastic cell-by-cell parameter fields generated using the sequential Gaussian method described above. Sampling is done in such a way that kriging between the sampled pilot points gives rise to a cell-by-cell parameter field that provides an optimal least squares fit to the detailed stochastic parameter field. The "difference field" that is obtained by subtracting one from the other is then "carried" by the pilot point field following replacement of its solution space component with that of the calibrated model during subsequent readjustment of pilot point parameter values to ensure a good fit between model outputs and the calibration dataset.

First we generated 400 stochastic hydraulic conductivity fields using the same approach as employed to generate the "reality" hydraulic conductivity field. By undertaking model runs based on a few of these fields it was verified that model-calculated steady state heads and concentrations bear little resemblance to those comprising the calibration dataset selected as above. These fields do not therefore provide a good fit between model outputs and the calibration dataset. Next pilot point samples of these fields were taken in the manner described above, and subjected to null space projection and re-calibration.

In implementing this procedure the surrogate model was used for sensitivity calculations, while the fine model was used only for computation of the objective function corresponding to improved parameter sets, with parameter improvements being calculated on the basis of surrogate model parameter sensitivities. Whenever the fine model was run for prediction purposes, the cell-by-cell difference field of hydraulic property heterogeneity was provided to the model as an addition to the cell-by-cell pilot point interpolated field. However whenever a surrogate model run was undertaken, the assignment of hydraulic conductivities to coarsened grid model cells was based purely on interpolation from pilot points with no added difference field.

On average, the production of each new calibration-constrained hydraulic conductivity field in which hydraulic conductivity heterogeneity is represented on a cell-by-cell level required 10 fine model runs. Meanwhile sensitivities needed for:

- solution and null space separation based on hydraulic conductivities achieved through the previous calibration process; and
- improving model-to-measurement fit during successive random field adjustment processes (this done using "super parameters");

were all calculated using finite differences based on surrogate model runs. The efficiency of the NSMC parameter space exploration technique, combined with SVD-assist and conjunctive use of complex and surrogate models, is evidenced by

the small number of complex model runs required for construction of each stochastic field.

The outcome of the entire NSMC analysis was 359 stochastic hydraulic conductivity fields all of which (1) are as detailed in their representation of spatial heterogeneity as the reality parameter field, (2) vary greatly between each other (and hence collectively are likely to represent a reasonably comprehensive sample of the posterior parameter distribution), and (3) fit the calibration dataset to a level that is commensurate with measurement noise.

A sample of these stochastic fields is provided in Figure 2.7. Also depicted in this figure are corresponding seawater interface locations calculated by the fine model on the basis of these fields, together with predicted interface locations. Calibrated and predicted interface locations for all parameter fields are shown in Figure 2.8. Figure 2.9 shows a histogram of the predicted position of the seawater interface, calculated on the basis of the calibration-constrained random fields forthcoming from the NSMC process. This histogram quantifies the uncertainty in the prediction of westward movement of the seawater interface arising from post-calibration uncertainties in model parameters.

It is apparent from Figure 2.9 that “the right answer” is indeed covered by the empirical predictive probability distribution forthcoming from NSMC analysis.

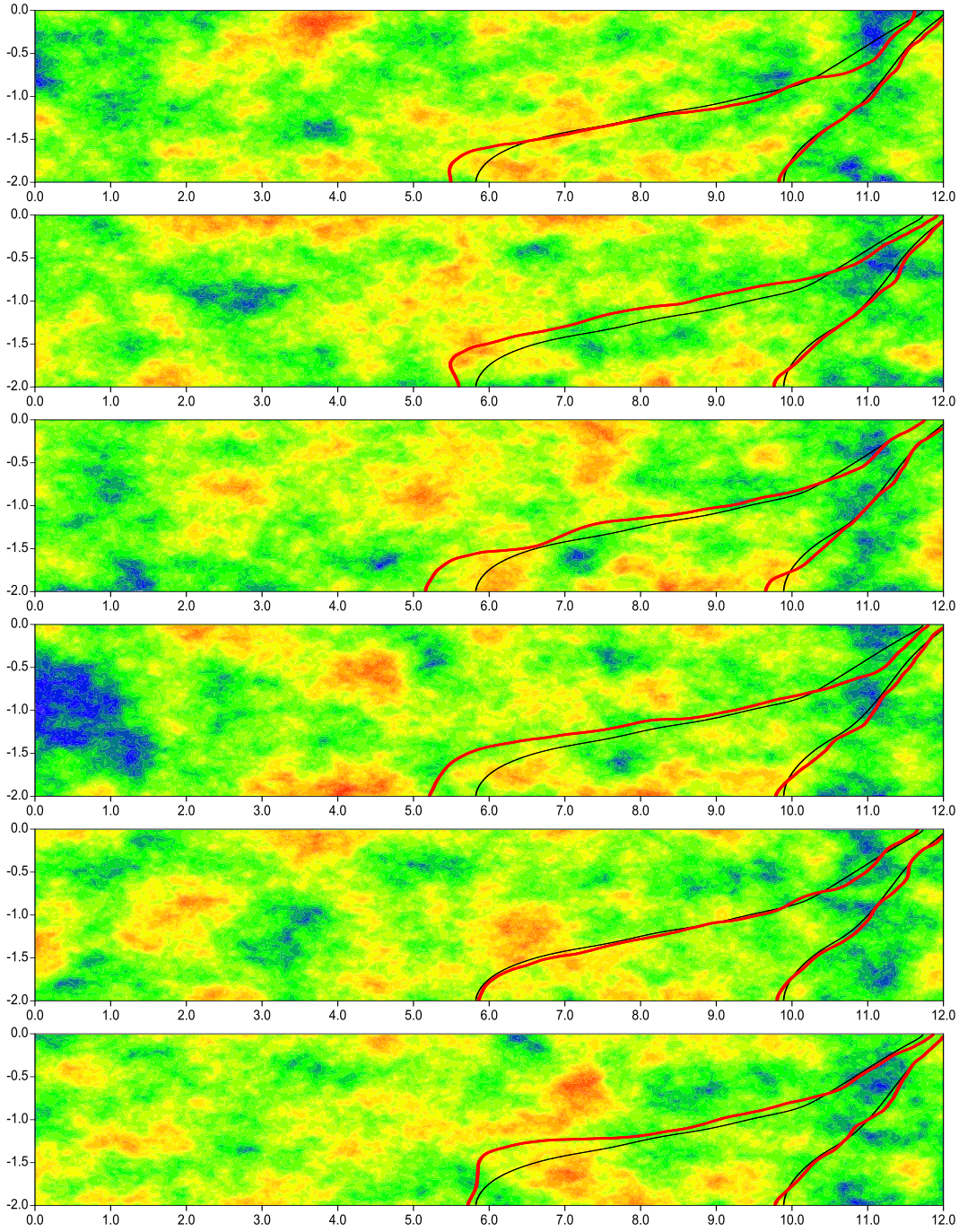

Figure 2.7: Six randomly chosen stochastic hydraulic conductivity fields obtained through the NSMC process. Also shown are the steady state and predicted seawater interface positions calculated using each stochastic field (red lines), together with the interface positions calculated using the calibrated hydraulic conductivity field (black lines).

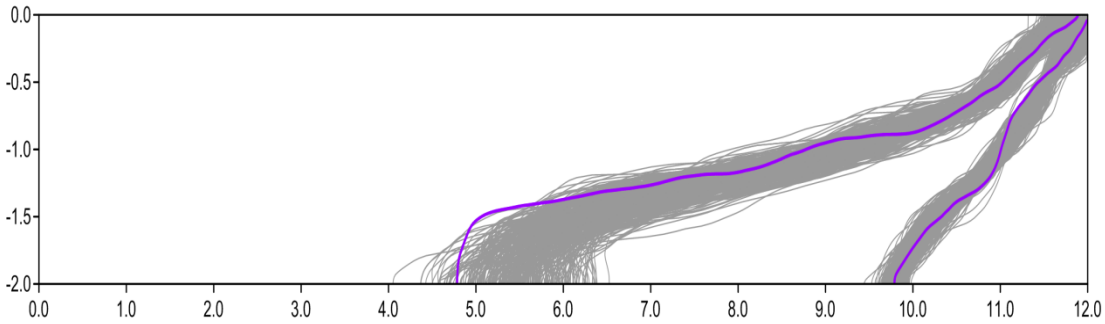

Figure 2.8: Seawater interfaces calculated using NSMC-derived hydraulic conductivity fields. Calibration-time interfaces are shown on the right while predicted interfaces are shown at left. Interface positions arising from the "reality" hydraulic conductivity field are shown in purple.

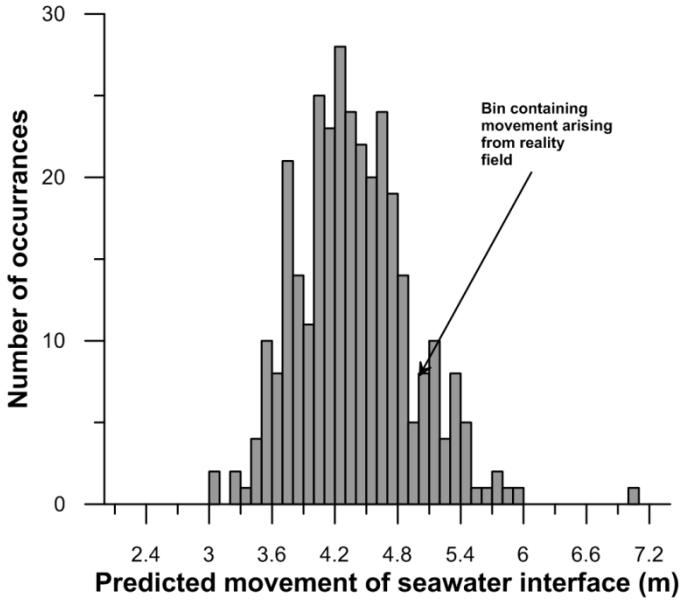

Figure 2.9: Histogram of the predicted seawater interface movement arising from 359 calibration-constrained, hydraulic conductivity fields.

2.6 Discussion

The example presented in Sections 2.4 and 2.5 demonstrates how a considerable reduction in the computational burden of post-calibration uncertainty analysis can be achieved through strategic use of a surrogate model in combination with a more complex original model. In this example the high computational speed of the surrogate model in comparison to that of the original model is achieved through use of a coarsened grid, and through use of the implicit finite difference scheme in place of the TVD scheme for solution of advective transport. The cost of these simplifications is a less precise calculation of salt concentrations. However because the surrogate model is used for calculation of differences in outputs resulting from parameter perturbations, simplification induced output errors tend to cancel to some extent in the differencing process so that the finite differences retain enough of their integrity to underpin gradient-based parameter estimation and calibration-constrained uncertainty analysis.

The example used in the present paper is "benign" in some respects. Use of a reasonably large molecular diffusion coefficient provides for a seawater interface which is relatively wide when compared with the distance between pilot point parameters. This maintains the sensitivity of individual pilot point parameters over a broader range of values than would be the case for a narrow interface. Model output nonlinearity with respect to parameter values is thus reduced. So too is degradation of quality of model outputs with use of a coarser grid. However if the saltwater interface were narrower, as would be the case with a low level of diffusion/dispersion, this problem could be overcome by formulating the objective function differently. For example it could be formulated in terms of differences between modelled and observed distances of the 10% isohaline contour from the eastern boundary of the model domain in different model layers. Alternatively a methodology similar to that proposed by *Schwede and Cirpka* (2009) could be implemented, whereby concentration observations are ascribed to an expanded area of the model domain, with the ascribed area decreasing as the inversion process progresses.

In the example presented above, model simplification is achieved in a relatively straightforward way, namely through grid coarsening. In many modelling contexts a similar simplification strategy can be readily pursued. In other modelling contexts other simplification strategies may need to be explored. Regardless of the simplification strategy selected for a particular modelling context, the work documented herein provides cause for optimism that there are many cases where use of one or more surrogate models for computation of parameter sensitivities may allow implementation of inversion and/or calibration-constrained uncertainty analysis where it would otherwise be numerically difficult, if not unachievable.

It is not impossible that with a slightly defective sensitivity matrix populated by model runs based on a surrogate model, parameters that are estimated through the calibration process may incur some calibration-induced bias. The Section 2.7 that follows, explores this issue through examining the outcomes of calibration exercises conducted with surrogate models similar to that presented in the above example, but with progressively coarsened grids. Naturally, the modeller should be aware of this possibility and reject unreasonable parameter fields that may emerge from use of an over-simplified surrogate model. The extent and type of simplification which will give rise to such fields can only be determined on a case-by-case basis. On the other hand, where parameter fields that emerge from the use of a complex/surrogate model pair in the manner described herein are reasonable, and where use of a surrogate model enables implementation of a calibration process that would otherwise be impossible because of high run times and/or questionable numerical convergence of a complex model, a small degree of parameter bias incurred in this way may be a small price to pay. This will be especially the case if parameter error thus incurred is small in relation to overall post-calibration parameter uncertainty. The efficiency gains accrued through use of a surrogate model in the manner described herein make exploration of these uncertainties possible.

2.7 Short Analysis of the Impacts of Increasing Simplification on Paired Model Calibration

The following work was submitted to *Ground Water* journal as supporting information to the published paper *Burrows and Doherty* (2014). It presents a brief analysis of the potential for parameter bias to be imparted upon estimated parameters due to the use of simplification degraded derivatives. It therefore provides useful insights into this phenomenon, hopefully thereby providing guidance to the modeller as to a suitable level of discretisation to use in construction of a surrogate model for the purpose of calculation of derivatives.

2.7.1 Introduction

The paper which this document supports describes how a simplified surrogate version of an original complex model can be used to efficiently calculate model output derivatives that underpin gradient-based parameter estimation and posterior predictive uncertainty analysis. An important issue that arises from such conjunctive use of a complex/surrogate model pair is this:

"*What level of surrogate model simplification can be tolerated before derivatives calculated by this surrogate model no longer retain sufficient integrity for their use in gradient-based complex model calibration, and in calibration-constrained uncertainty analysis of complex model predictions?"*

Two factors must be considered in assessing the outcomes of a calibration exercise. These are (a) the level of model-to-measurement fit attained through the calibration process, and (b) the extent to which estimated parameter values respect sensible ranges for these values as assessed by expert knowledge. Quantitative measures are available for making these assessments. An objective function based on weighted model-to-measurement residuals is often applied in the first case. Residuals may then be subjected to further statistical analysis for assessment of bias, temporal/spatial correlation, and other unwanted properties; see, for example, *Draper and Smith* (1998) for details. In everyday modelling practice, however, model-to-measurement misfit is often assessed visually, as the extent to which an objective function can be

reduced, and the extent to which residuals can be endowed with desired characteristics, is often limited by the inadequacies of even the most detailed numerical model as a simulator of complex environmental processes.

Assessment of estimated parameter values can be even more complex. In the highly parameterized context, the integrity of estimated parameters cannot be assessed through direct comparison between estimated and true parameter values. As *Moore and Doherty* (2006) explain, the calibration process yields estimates of projections of real parameter values onto a subset of parameter space that is the orthogonal complement of the null space of the model operator. It follows that, even in the absence of measurement noise, the calibrated parameter field will not be "correct"; it can only provide a simplified representation of the true parameter field. Furthermore this representation is subject to error as an outcome of its estimation on the basis of a calibration dataset which is contaminated by measurement noise. Ideally, the calibration process should leave null-space-projected parameter values unchanged. Sadly, however, this cannot be guaranteed because of the simplified nature of even the most complex model. As *Watson et al*. (2013) point out, to the extent that null space parameter components are "entrained" through the model calibration process, parameter and predictive bias is introduced to the calibrated model through the calibration process. This bias is in addition to that introduced through the model construction process itself. As is done for model-to-measurement misfit, statistical measures can be employed to assess the integrity of an estimated parameter field. In practice, however, such quantitative analysis is difficult because (a) it is rarely possible to endow real-world parameters with a stochastic description from which an objective function and bias-related statistics can be calculated, and (b) as all models are simplified representations of reality a small degree of calibration-induced bias is inevitable; see *White et al*. (2014) for details. Hence the assessment of calibrated parameter field integrity is often made visually.

In what follows, we present a brief analysis of the integrity of calibration outcomes, in terms of bias imparted on parameter values, when calibration is implemented using a simple model surrogate of a more complex model for the purpose of derivatives calculation. The results that follow extend the analysis presented in the

main body of the paper that comprises this chapter, wherein surrogate-enabled, highly parameterized inversion is employed in calibration of a two-dimensional sea water intrusion model. As the surrogate model used for derivatives calculation is progressively made more and more simple, we examine the level of fit attained through the calibration process, and the values estimated for parameters through that calibration process.

2.7.2 Method

In the analysis presented herein, the grid of the surrogate model is progressively coarsened. Calibration outcomes are compared with those of a "baseline" calibration. The latter is achieved through calibration of the complex model without use of a simple model surrogate for derivatives calculation; that is, the complex model itself is used for derivatives calculation. In contrast to the surrogate models which use the implicit finite difference scheme to solve for solute transport, the complex model used for baseline calibration employs the computationally intensive TVD scheme, as does the "reality model" described in the paper which was used to generate the calibration dataset. Regularization is achieved through application of Tikhonov constraints to pilot point parameter values, subject to a parameter covariance matrix that reflects true spatial parameter variability. The PEST "SVD-Assist" scheme is not used. Hence, at a very high numerical cost, the baseline calibration process avoids model and parameter simplification as much as possible.

Successive grid coarsening is achieved for each surrogate model by increasing the dimensions of model grid cells by factors of 10, 20 and 33.333. Calibration processes undertaken through use of these progressively more simplified models are named "scenario 2", "scenario 3" and "scenario 4" herein. Meanwhile "scenario 1" refers to the calibration example presented in the manuscript wherein the surrogate model grid is coarsened by a factor of 5 over that of the complex model. Simplification scenarios 1 to 4 thus embody grids of 20 x 120, 10 x 60, 5 x 30 and 3 x 18 layers and columns respectively. These are to be compared with the 100 x 600 grid employed by the complex model. The resulting surrogate model grids, together with the distribution of pilot point parameters in relation to these grids, are depicted in Figure 2.10.

To facilitate reader analysis of calibration outcomes achieved through use of these progressively simplified surrogate models, the reality hydraulic conductivity field as described previously in this chapter is reproduced in Figure 2.11.

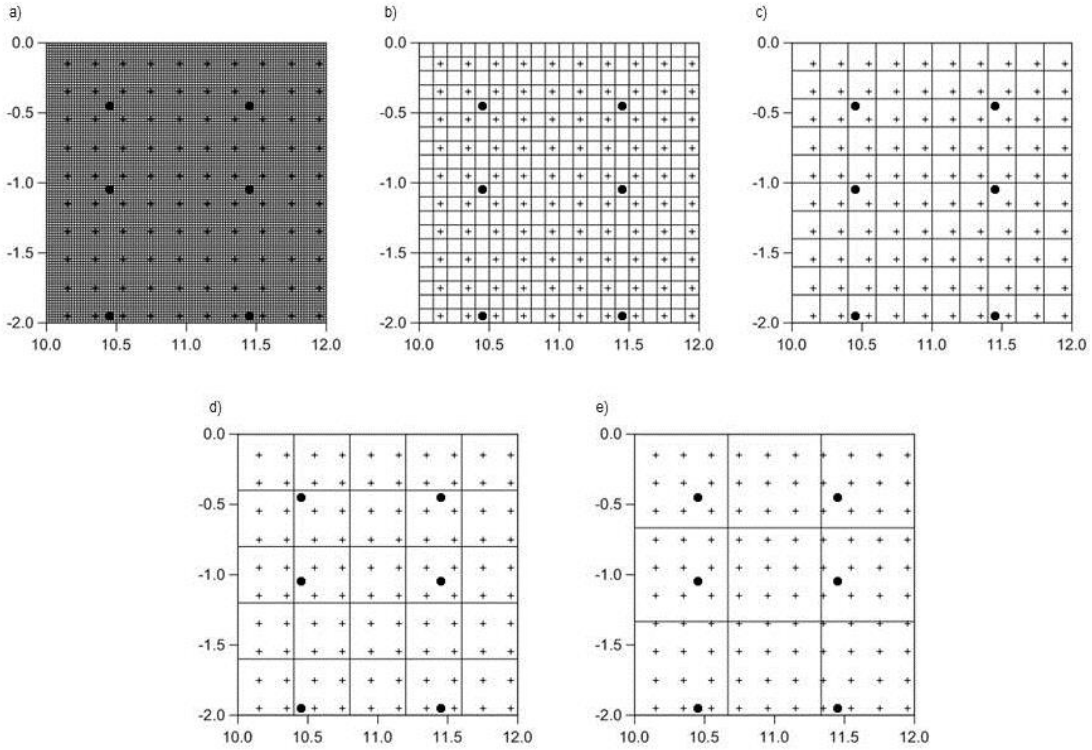

Figure 2.10: A small section of the model domain depicting model grids employed by: (a) the complex model; (b) the surrogate model described in Section 2.4–scenario 1; (c) surrogate model scenario 2; (d) surrogate model scenario 3; and (e) surrogate model scenario 4. Pilot point parameter locations are shown as crosses. Observation locations are shown as full circles.

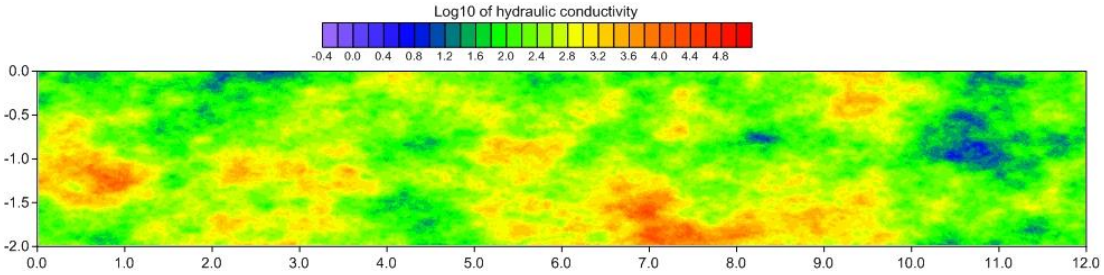

Figure 2.11: The reality hydraulic conductivity field; see the previous part of this chapter for full details.

2.7.3 Results

As is described earlier, the calibration process assigns values of log hydraulic conductivity to pilot points. In order to compare calibrated pilot point values with "real" pilot point values, the latter where calculated by sampling the reality log hydraulic conductivity field of Figure 2.11 at pilot point locations. Sampling was of the least squares type. That is, the values assigned to the set of 600 pilot points that are used to parameterize the "real" model are those that minimize the discrepancy between the field of Figure 2.11 and a parameter field interpolated from pilot points to the complex model grid using kriging based on the same variogram as that used to generate the reality parameter field in the first place. A least-squares objective function based on residuals between the real and interpolated field at every model cell was minimized in computing the "reality" pilot point parameter set. "Reality" pilot point values achieved through this process are shown in Figure 2.12 as a dashed line connecting the points. In this and other figures, the log of pilot point hydraulic conductivity is plotted against pilot point index. The latter is obtained by numbering pilot points sequentially; counting proceeds along each row, starting at the top, and then progressing downwards.

Pilot point values calculated through baseline calibration are also shown in Figure 2.12. As is expected, calibrated parameter values track reality parameter values in a subdued manner, albeit with some small errors arising from the presence of measurement noise in the calibration dataset. Both sets of parameters exhibit a

similar mean value of about 200 m/d, this being the mean value employed in stochastic generation of the original reality field. The hydraulic conductivity field arising from this baseline calibration is shown in Figure 2.13 in a spatial setting (this is a repeat of Figure 2.5), along with the calculated position of the saltwater interface arising from this field; the "real" interface position is also shown. Notwithstanding the sparse and noise-degraded observation dataset on which calibration is based, the calculated interface position is subjectively very good; at the same time the calibrated hydraulic conductivity field reflects the broad-scale characteristics of the reality field.

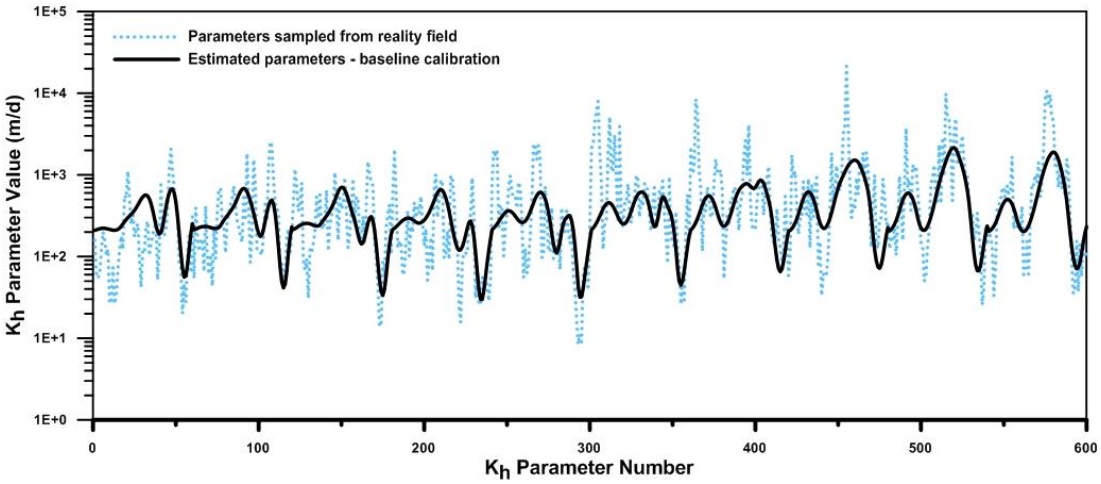

Figure 2.12: Comparison of baseline estimated pilot-point parameters, with values sampled from the reality log hydraulic conductivity field.

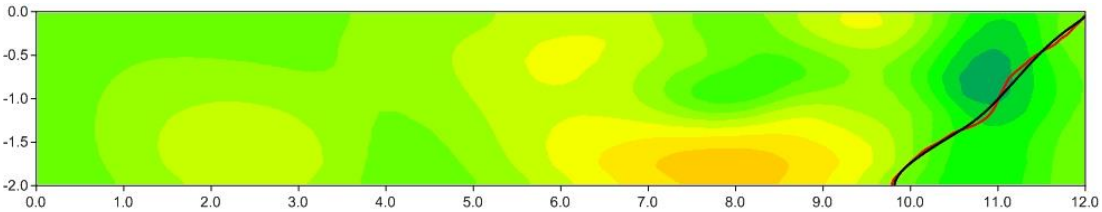

Figure 2.13: Kriging-interpolated log hydraulic conductivity field arising from baseline calibration. The seawater interface associated with this baseline estimated parameter field is shown as a black line, while the true position of the interface is depicted using a red line.

Outcomes of surrogate-enabled parameter estimation exercises are compared with baseline calibration outcomes in Figure 2.14 through to Figure 2.21. Calibration scenarios 1, 2 and 3 all achieve a high degree of fit between model outputs and calibration observations, as is illustrated by proximity of observed and "reality" sea water interface locations in these figures. However Figure 2.18 illustrates the onset of some spurious parameter values with surrogate model grid coarsening at scenario 3. Visual inspection of the log hydraulic conductivity field (Figure 2.19) also suggests that the geological plausibility of this parameter field is questionable.

Calibration scenario 4 fails to reduce model-to-measurement misfit to the same extent as the other calibration exercises, notwithstanding an improvement in misfit over that associated with the initial parameter field used in this exercise; the latter is comprised of uniform pre-calibration expected parameter values of 200 m/d. At the same time, it is apparent from Figure 2.20 and Figure 2.21 that parameter values achieved through the calibration process deviate erratically and erroneously from baseline calibration outcomes.

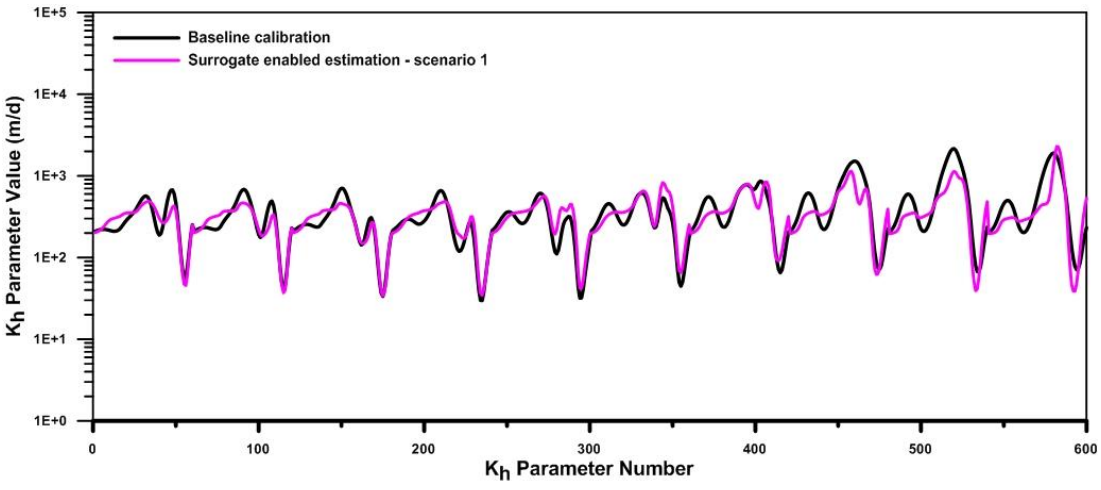

Figure 2.14: Comparison of pilot point parameter values arising from surrogate-enabled calibration scenario 1 with those obtained from the baseline calibration.

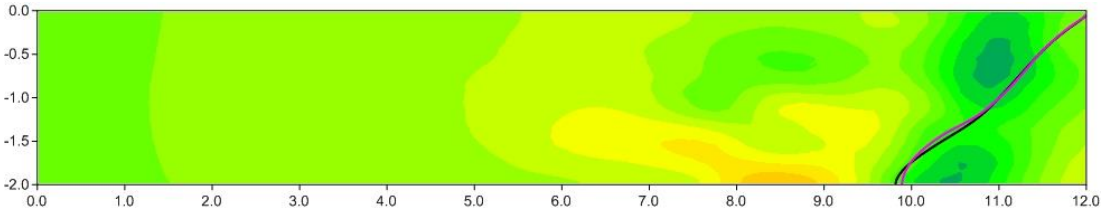

Figure 2.15: Kriging-interpolated hydraulic conductivity field arising from calibration scenario 1. The calculated position of the seawater interface using this estimated field is shown in purple while the interface position calculated using the baseline estimated field is shown in black.

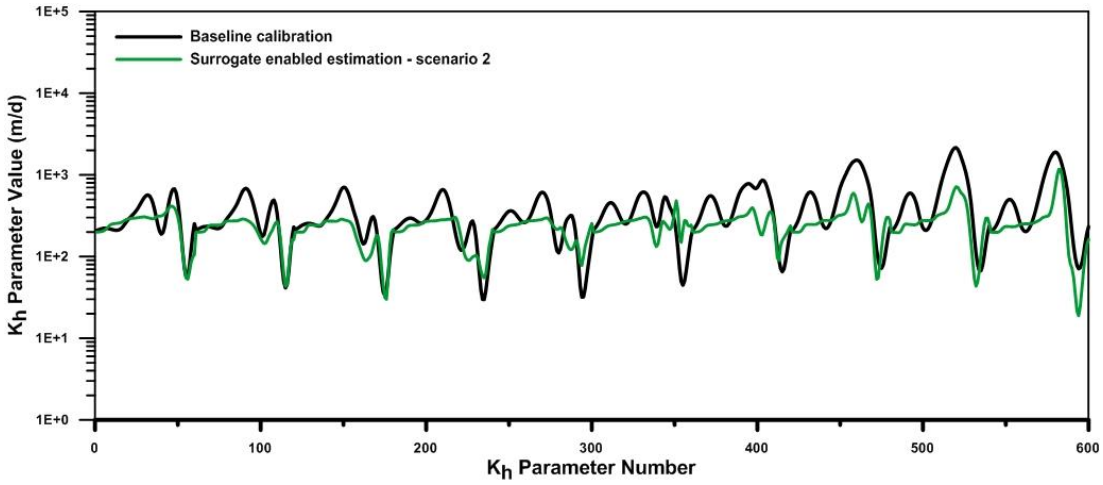

Figure 2.16: Comparison of pilot point parameter values arising from surrogate-enabled calibration scenario 2 with those obtained from the baseline calibration.

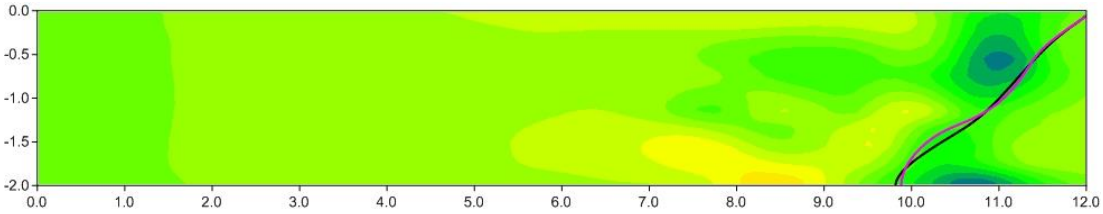

Figure 2.17: Kriging-interpolated hydraulic conductivity field arising from calibration scenario 2. The calculated position of the seawater interface using this estimated field is shown in purple while the interface position calculated using the baseline estimated field is shown in black.

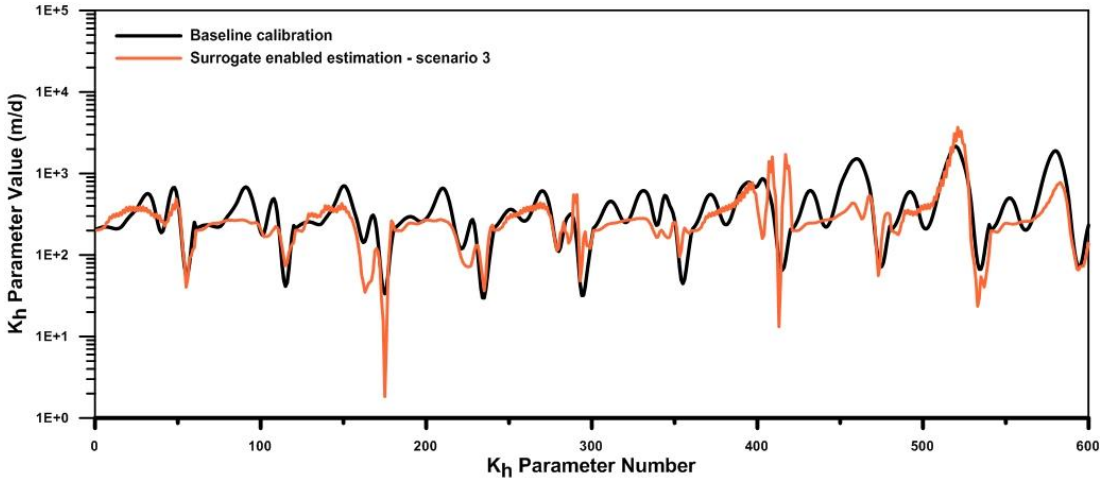

Figure 2.18: Comparison of pilot point parameter values arising from surrogate-enabled calibration scenario 3 with those obtained from the baseline calibration.

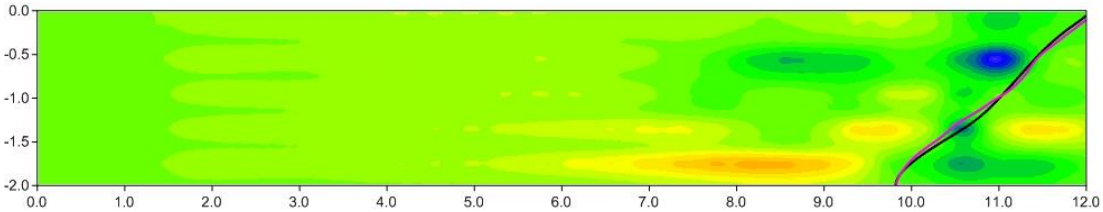

Figure 2.19: Kriging-interpolated hydraulic conductivity field arising from calibration scenario 3. The calculated position of the seawater interface using this estimated field is shown in purple while the interface position calculated using the baseline estimated field is shown in black.

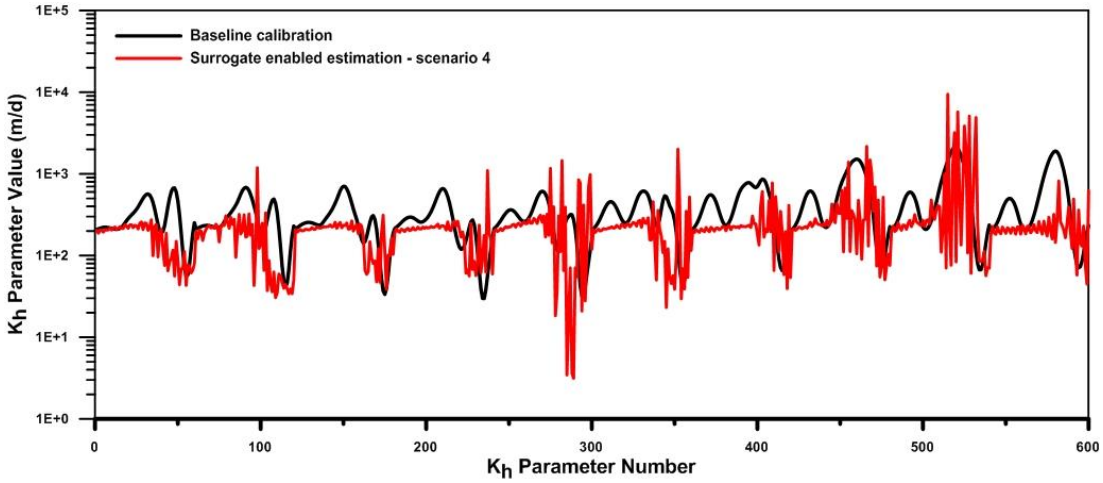

Figure 2.20: Comparison of pilot point parameter values arising from surrogate-enabled calibration scenario 4 with those obtained from the baseline calibration.

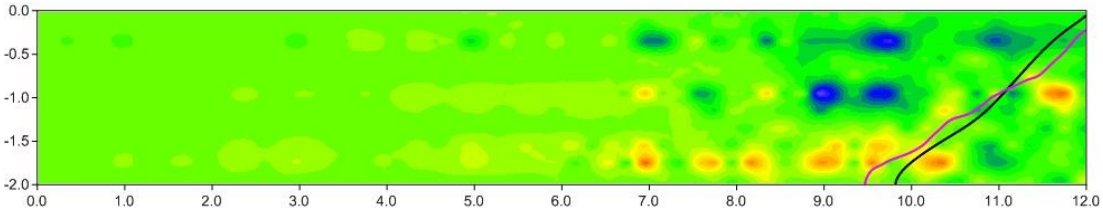

Figure 2.21: Kriging-interpolated hydraulic conductivity field arising from calibration scenario 4. The calculated position of the seawater interface using this estimated field is shown in purple while the interface position calculated using the baseline estimated field is shown in black.

2.7.4 Conclusions

This analysis demonstrates that, for the specific case presented as the example in this chapter, the proposed method of conjunctive complex/surrogate model usage is successful in achieving estimated parameter sets that are realistic and which allow the model to replicate historical system behaviour, when surrogate model grid coarsening is much greater than that used in the original example. However with greater model simplification, the propensity for the calibration process to estimate unrealistic parameter values increases. If surrogate model simplification proceeds too

far, then not only are unrealistic parameter values calculated; the level of model-to-measurement fit achievable during calibration is also severely compromised. It is worthy of note, however, that for the present case at least, there is little need for simplification to proceed this far. The run times of the simplified models associated with scenarios 2 to 4 are only marginally smaller than that associated with scenario 1 for which calibration outcomes are very good indeed.

For the model which forms the basis of the present example, the onset of spurious parameter values coincides with a level of grid coarsening which is large enough to allow multiple pilot points to reside in single grid cells. At this point a second level of parameter non-uniqueness is superimposed on that which attends the highly parameterized inversion process itself. At the same time, model outputs employed in derivatives calculation are compromised by the simplicity of the model, this creating a form of "structural noise" that is superimposed on those outputs. Undocumented calibration runs undertaken by the author demonstrates that the level of parameter spuriousness that emerges from use of such a coarse surrogate model grid can be significantly reduced if the level of model-to-measurement fit sought through the Tikhonov-regularised inversion process is reduced to a level that is supposedly commensurate with that of simplicity-induced structural noise. However an analysis of trade-off between parameter and model-to-measurement fit integrity under these circumstances is beyond the scope of the present analysis.

In summary, the analysis presented herein demonstrates that calculation of derivatives using a surrogate model with a coarser grid than that employed by the model that is actually used for simulation purposes does not compromise the integrity of the calibration process, provided the surrogate model grid is not "too coarse". The metric for "too coarse" is problem-specific. However the present analysis suggests that it can readily be made coarse enough to promulgate significant run time gains, before being so coarse as to compromise the integrity of the calibration process.

Chapter 3

Gradient-Based Model Calibration with Proxy-Model Assistance

This chapter presents an extension of the surrogate-enabled calibration strategy which employs an analytical proxy model developed from output equivalents to elements of the calibration data set obtained from the original model when supplied with variations in input parameters. The surrogate model thus developed is therefore a data-driven proxy for the original model in terms of the calibration dataset. This is in contrast to the surrogate model simplification strategy used in Chapter 2 whereby grid coarsening is used to derive a faster running version of the original model. As is explained herein, this strategy has particular application when finite-difference derived parameter sensitivities calculated from outputs of the complex model are compromised by numerical inconsistencies that emanate from the model. This chapter is presented in almost identical form to that in which the work was published in *Journal of Hydrology* (see *Burrows and Doherty*, 2016).

3.1 Abstract

Use of a proxy model in gradient-based calibration and uncertainty analysis of a complex groundwater model with large run times and problematic numerical behaviour is described. The methodology is general, and can be used with models of all types. The proxy model is based on a series of analytical functions that link all model outputs used in the calibration process to all parameters requiring estimation. In enforcing history-matching constraints during the calibration and post-calibration uncertainty analysis processes, the proxy model is run for the purposes of populating the Jacobian matrix, while the original model is run when testing parameter upgrades; the latter process is readily parallelized. Use of a proxy model in this fashion dramatically reduces the computational burden of complex model calibration and

uncertainty analysis. At the same time, the effect of model numerical misbehaviour on calculation of local gradients is mitigated, this allowing access to the benefits of gradient-based analysis where lack of integrity in finite-difference derivatives calculation would otherwise have impeded such access. Construction of a proxy model, and its subsequent use in calibration of a complex model, and in analysing the uncertainties of predictions made by that model, is implemented in the PEST suite of software.

3.2 Introduction

Environmental models that simulate the details of complex physical and chemical processes over domains wherein the properties which govern those processes are spatially and temporally heterogeneous are often characterized by long runtimes and a propensity for problematic solver convergence. Furthermore, it is not uncommon for models of these types to exhibit good numerical behaviour when provided with one set of parameters, but suffer serious degradation of numerical performance when supplied with another set of parameters. Where this occurs, calibration and uncertainty analysis become very difficult undertakings. This can erode the use of such models in environmental decision-support.

In the present chapter we focus on those aspects of a model's performance which compromise the ability of a model-independent inversion package such as PEST (*Doherty*, 2015a) to calculate derivatives of model outputs with respect to the parameters which require adjustment during calibration, and calibration-constrained uncertainty analysis. In PEST, derivatives are calculated using a finite-difference methodology based on a two, three or five point stencil. Model outputs are computed based on values of a particular parameter which are varied incrementally in accordance with the selected stencil; differences in these outputs form the basis for approximation of local partial derivatives with respect to that parameter. These derivatives are housed in a so-called Jacobian matrix. The Jacobian matrix is then employed in calculation of an improved set of parameters. Jacobian matrix and

parameter upgrade calculations are undertaken repeatedly in an iterative process whose outcome is a set of parameter values that produce an acceptable level of fit between model outcomes and field observations of system state. Where a model is being calibrated, a set of parameters which constitute a minimum error variance solution to the inverse problem is sought through this process. Where calibration-constrained uncertainty analysis is being undertaken, multiple sets of parameters are sought, all of which are considered to be reasonable expressions of system properties, and all of which fit field measurements to within limits that reflect the noise content of those measurements.

A variety of numerical methods have been developed to expedite calibration and calibration-constrained uncertainty analysis. Many of these methods do not, in fact, require calculation of a Jacobian matrix. However, use of so-called "gradient methods" which do make use of partial derivatives of model outputs with respect to adjustable parameters to perform the above tasks accrues certain benefits. A major benefit that gradient methods have over other methods is their speed; see, for example *Keating et al.* (2010). Another benefit is that gradient-based inversion algorithms are easily extended to include mathematical regularisation schemes that readily accommodate parameter nonuniqueness (*Aster et al.*, 2013; *Menke*, 1989). A further benefit is that, once a Jacobian matrix has been filled, it can be used in calculation of post-calibration statistics such as parameter identifiability (*Doherty and Hunt*, 2009), parameter and predictive uncertainty (*Gallagher and Doherty*, 2007a and 2007b; *James et al.*, 2009), and the worth of existing and yet-to-be acquired data in terms of its ability to reduce the uncertainties of parameter and predictions of interest (*Dausman et al.*, 2010).

Use of gradient methods is not without its problems, however. Their performance may be hampered where the relationship between model outputs and parameters is highly non-linear (*Duan et al.*, 1992). Even worse, it may not be possible to use these methods at all where model performance is such that elements of the Jacobian matrix lose their integrity. Where these elements are calculated using finite parameter differences, loss of integrity can occur when incremental changes in model outputs employed in finite-difference derivatives calculation reflect more than simply

incremental changes in parameter values. This is not an uncommon situation, particularly where the complex nonlinear environmental processes simulated by a model challenges its solver. While strategies such as adaptive time stepping that alter the solution procedure when convergence becomes problematic, may mitigate these problems as far as the model is concerned, they may exacerbate them as far as calculation of finite-difference derivatives is concerned, for model outputs may then become somewhat dependent on solution path.

Examples of modelling contexts in which calculation of finite-difference parameter derivatives may be compromised are not hard to find. The handling of "dry cells" in MODFLOW (*Harbaugh et al*., 2000) is a common example. *Kavetski et al.* (2006) discuss how algorithmic design of models that simulate surface water movement can lead to similar problems in these kinds of models. Other contexts in which model numerical behaviour can compromise finite-difference derived gradients include:

- simulation of the effects of mining and tunnelling operations on groundwater systems;
- interaction of ground and surface waters near streams and wetlands;
- high temperature geothermal reservoir simulation where water phase is a discontinuous function of temperature and pressure; and
- chemical reactions in mobile contaminant plumes.

In the difficult numerical circumstances that these modelling contexts present, so-called "global methods" which do not rely on calculation of derivatives of model outputs with respect to adjustable parameters, provide an alternative option for software-controlled history-matching. Examples of non-gradient based calibration and calibration-constrained uncertainty analysis algorithms include (among many others) particle swarm optimization (*Kennedy and Mendes*, 2002), shuffled complex evolution (*Duan et al.*, 1992), genetic and evolutionary programming (*Vrugt and Robinson*, 2007), and covariance matrix adaption algorithms (*Hansen et al.*, 2003). All of these replace the need to calculate derivatives with respect to adjustable parameters with intelligent random sampling of parameter values. While delivering

robustness in the face of problematical numerical behaviour, the cost of this robustness is the requirement for a greater number of model runs than that required by gradient methods. This differential between model run requirements of the two different approaches tends to grow with the number of parameters that require estimation or adjustment.

To ease the computational burden of applying global methods to the problems of model calibration and calibration-constrained uncertainty analysis, increasing use is being made of fast-running model surrogates. In recognition of the fact that the complex simulator is the most accurate replicator of reality available, in many applications the surrogate model does not completely replace the original simulator. Rather it is strategically substituted for the simulator on many occasions that a model run is required. The greater is the ratio of surrogate to simulator runs, the greater is the efficiency of the overall process. The surrogate may be a simulator that runs much faster than that which it replaces because of its simpler algorithmic design. One example is the SWI package for MODFLOW, (*Bakker et al.*, 2013) which replaces mass conservative governing equations with equations based on continuity of flow, avoiding the need for fine-scale vertical discretisation. Another example is MODFLOW-USG, described by *Panday et al.* (2013), which is able to represent flow in grids with highly irregular spatial discretisation thereby reducing the number of simultaneous equations required in solution. More sophisticated model reduction strategies may be employed as are used by *Efendiev et al.* (2005, 2009) and *Mondal et al.* (2010) whereby a coarse-gridded simulator whose parameterization is based on single-phase upscaling procedures, surrogates for a fine scale, dual-phase reservoir model.

Alternatively the surrogate may undertake data-driven reproduction of simulator outputs, or interpolate between samples of simulator outputs to non-sampled parts of parameter space using devices such as radial basis functions, kriging or artificial neural networks; see for example, *Regis and Shoemaker* (2004), *Bliznyuk et al.* (2007), and *Alam et al.* (2004) respectively. More recently, *Laloy et al.* (2013) and *Elsheikh et al.* (2014) deploy polynomial chaos expansion theory to develop interpolators of simulator outputs. Alternatively the statistical characteristics of

simulator outputs can be modelled using Gaussian process theory; see *Johnson et al.* (2011) and *Conti et al.* (2009) as examples. Data-driven surrogates such as these are commonly known as model emulators or proxy models.

One analysis scheme that can benefit enormously from simulator run reductions through strategic use of surrogate models is Markov Chain Monte Carlo (MCMC). Various adaptations of the so-called "two-stage MCMC" approach have been documented (see *Efendiev et al.,* 2005 and 2009; *Mondal et al.,* 2010 and *Cui et al.*, 2011 for examples) that seek to reduce the need for expensive simulator runs in assessment of low-probability proposal parameter fields. In these example studies calibration and/or uncertainty assessment of complex reservoir simulators is undertaken wherein a surrogate model is used in "stage one" of the process as a pre-screening mechanism. The goal is to increase the acceptance rate of proposed parameter fields in "stage two" where acceptance/rejection of the proposal is determined on the basis of the simulator. The studies just mentioned use surrogate models based on simplified algorithms, as has already been mentioned. The studies cited earlier in relation to polynomial chaos expansion theory (that is *Laloy et al.*, 2013 and *Elsheikh et al.*, 2014), also deployed their model emulators within the two-stage MCMC framework. Two-stage MCMC consistently demonstrates several fold savings in computational costs over full/direct MCMC, effected primarily through inexpensive pre-screening of proposals.

Of course use of a surrogate model, either as a direct substitute or as a companion to a more accurate simulator, will undoubtedly incur some cost on the analysis undertaken. Put simply, a simplified model cannot be expected to replicate the same level of accuracy at all spatial and temporal locations of a modelled domain as can a simulator. It is readily acknowledged in the literature of two-stage MCMC cited above, that there exists potential for rejection of parameter proposals in the pre-screening stage of the process when assessed by the surrogate, that would otherwise find support through the simulator. To accommodate surrogate model "malperformace" in two-stage MCMC, it is common to employ elevated measurement uncertainty when calculating proposal likelihood with the surrogate model. Unfortunately this effectively admits a larger number of low-probability

proposals to pass through the pre-screening step and erodes efficiency gains accrued through use of the surrogate. As will be explained further in the next section of this paper, where surrogate models are conjunctively used in gradient methods for approximation of derivatives degraded efficiency may also occur. However accommodation of surrogate model "malperformance" can largely be achieved through maintenance of the Marquardt parameter. The use of elevated measurement uncertainty as a means to accommodate so-called structural error is also a strategy that is routinely applied in gradient based uncertainty analyses.

To the authors' knowledge, the only documented gradient-based parameter adjustment process that does not completely replace the original simulator with a surrogate, is that of *Burrows and Doherty* (2014). These authors undertook calibration and calibration-constrained uncertainty analysis of a sea-water intrusion model. The former was implemented using Tikhonov regularisation while the latter was implemented using the null space Monte Carlo method (described by *Tonkin et al.*, 2009) and supported by PEST. Jacobian matrix calculation was undertaken using the surrogate - a coarse-gridded version of the original model. The work undertaken by these authors suggests that, in some modelling contexts at least, although system states calculated by a simplified model may be approximate, incremental changes in these calculated states arising from incremental changes in parameter values, may have the integrity required to support calculation of derivatives of model outputs with respect to parameters in a highly parameterized inversion context.

The present paper documents an extension of the work undertaken by these authors in which the issue of problematical model numerical behaviour is addressed. Where model outputs are contaminated by "numerical granularity" arising from factors such as those discussed above, use of gradient-based methods for calibration and uncertainty analysis becomes difficult or impossible. In such circumstances global methods may not present a viable alternative, particularly if model runtimes are high, unless strategic use is made of a surrogate model. The alternative presented here is that an emulator can be used in conjunction with the simulator within a gradient based approach. This is in contrast to the previous study undertaken by these authors, in which the surrogate is a coarser-gridded version of the same model. This approach

can be employed where outputs provided by a simplified version of the complex model lack sufficient integrity to be used in calculation of derivatives. Alternatively, the approach discussed herein can be used where a modeller has insufficient time to construct a simpler version of a complex model and test its integrity. As steps in the emulator construction and training process are automated (and supported by the PEST suite), complex model calibration and uncertainty analysis can therefore proceed with relatively little user-difficulty. A deficiency of the approach presented herein however, is that unlike the surrogate model strategy presented by *Burrows and Doherty* (2014), the number of parameters that can be adjusted is limited to a few tens, rather than hundreds or even thousands. This should not, however, be seen as a deficiency of the method, rather it is reflective of numerical and/or runtime difficulties associated with the complex model.

In the approach documented herein a series of analytical proxy models, each of which emulates the relationship between one model output used in the history-matching process and all parameters that are adjusted through that process, replace the simulator for the purpose of derivatives calculation. In the present study these proxies employ second order polynomials; however the methodology is general enough for these to be replaced by more complex analytical functions where appropriate. Meanwhile the original simulator is used for testing and adjusting parameter upgrades calculated using the proxy-derived Jacobian matrix, this ensuring the integrity of those upgrades. Proxy model training, through which the proxy model ensemble is taught to replicate the behaviour of the simulator, is implemented through an automated sequence of numerically cheap calibration processes undertaken prior to calibration of the simulator itself. The level of proxy-to-simulator fit achieved through the training process can be user-adjusted so that the proxy ensemble is able to replicate broad-scale simulator behaviour while ignoring local expressions of its numerical difficulties. Proxy-assisted parameter adjustment can therefore support calibration of a simulator that is beset with solver convergence problems. It can also support post-calibration Monte Carlo analysis in which many different parameter fields are generated, all of which reflect expert knowledge of parameter variability, and all of which respect calibration constraints. Furthermore, because the proxy ensemble replaces the simulator in calculation of the Jacobian

matrix, calibration and calibration-constrained uncertainty analysis become numerically inexpensive.

As far as the authors are aware, the methodology presented herein is novel in a number of respects. Firstly, it illustrates use of a data-driven model emulator in a gradient-based parameter estimation framework. In so doing, it demonstrates how some of the benefits of gradient-based parameter adjustment can be realised in contexts where the simulator on its own cannot support the use of these methods. Secondly it employs not one, but a series of emulators, each emulator being specific to a model output for which there is a matching field measurement. Thirdly, as will be demonstrated by example, a form of proxy-assisted, Tikhonov-regularized inversion is employed to efficiently impose calibration constraints on samples of an approximation to the posterior parameter probability distribution, without compromising the coverage of posterior parameter space established by those samples.

The remainder of this chapter is organised as follows. A short description of the Gauss-Marquardt-Levenberg method of parameter estimation is presented in Section 3.3. Section 3.4 describes how a suite of analytical emulators can be trained to reproduce the broad-scale behaviour of a complex simulator in parameter space, and how these emulators can then be used to assist calibration of that simulator. The methodology is applied to calibration and Monte Carlo-based uncertainty analysis of a saltwater intrusion model in Section 3.5. The paper concludes with a short discussion in Section 3.6.

3.3 Gauss-Marquardt-Levenberg method

This section describes the role of the Jacobian matrix in gradient-based parameter estimation. It also explains how the Jacobian matrix, together with other readily-available by-products of gradient-based parameter estimation, can illuminate parameter and predictive uncertainty.

The Gauss-Marquardt-Levenberg (GML) method is an efficient methodology for estimation of parameters of nonlinear models. Theory and application of this method are extensively covered in texts such as *Bard* (1974), *Draper and Smith* (1998), *Aster et al.* (2013) and *Doherty* (2015b). Briefly, a weighted least squares objective function is defined using an equation such as the following:

$$\Phi = \mathbf{r}^t\mathbf{Q}_m\mathbf{r} \quad . \tag{3.1}$$

In Equation (3.1), **r** is the vector of model-to-measurement residuals. Each element of this vector is the difference between a field observation and its model-generated counterpart. $\mathbf{Q}_m$ is a weight matrix. In normal modelling practice this is a diagonal matrix with its elements proportional to the inverse of the variance of measurement noise associated with respective field observations. Such a choice can be shown to support parameter estimates which are of minimized error variance (*Koch*, 1999) (It should be pointed out, however, that in most real-world modelling circumstances model-to-measurement misfit is dominated by inadequacies of the model as a simulator of real-world behaviour, and not by measurement noise. In these circumstances, weighting schemes which afford estimated parameters some protection from the deleterious effects of so-called "structural noise" should be adopted. See Doherty and Welter 2010 and White et al. 2014 for details).

Where an inverse problem is well-posed, estimates of parameters are obtained through minimization of the objective function defined in Equation (3.1). Where a model is nonlinear, this is an iterative procedure. During each iteration of the inversion process, the model is replaced by its linearized counterpart, this being the Jacobian matrix; this is denoted as **J** in equations to follow. Each column of the Jacobian matrix is comprised of the partial derivatives of all model outputs used in the calibration process with respect to one particular parameter. The matrix therefore possesses as many columns as there are adjustable parameters. During each iteration, improvements $\Delta\mathbf{p}$ to existing parameter values **p** are calculated using the formula:

$$\Delta\mathbf{p} = (\mathbf{J}^t\mathbf{Q}_m\mathbf{J} + \lambda\mathbf{I})^{-1}\mathbf{J}^t\mathbf{Q}_m\mathbf{r} \quad . \tag{3.2}$$

The variable λ appearing in Equation (3.2) is known as the "Marquardt lambda". Its presence assists the inversion process in accommodating model nonlinearity. A large value for this variable accelerates convergence early in the inversion process. Ideally its value should fall as the minimum of the objective function is approached.

Use of Equation (3.2) assumes that the matrix inverse cited in that equation actually exists. Where it does not exist, Equation (3.2) can be modified to support use of singular value decomposition in estimation of parameter projections onto the estimable subspace of parameter space; see *Moore and Doherty* (2005) for details. Alternatively, or as well, the calibration dataset can be supplemented with information, sourced from expert knowledge that pertains directly to the parameters requiring estimation. Formulation of such a Tikhonov regularisation scheme that references preferred values for all parameters, or preferred values for relationships between them, must be accompanied by a weighting scheme that suggests to the parameter estimation process the manner in which departures from these preferred parameter conditions should arise. Normally, a global weight multiplier is then applied to all such "regularisation observations". The value of this multiplier is calculated by the inversion engine itself on an iteration-by-iteration basis such that an appropriate balance is maintained between the influence of field measurements and regularisation constraints on estimated parameter values. In PEST's implementation of Tikhonov regularisation, the weight multiplier which it calculates is related to the Lagrange multiplier that arises in solution of a constrained minimization problem. This problem is formulated as minimization of the regularisation objective function subject to the constraint that the measurement objective function is allowed to fall no lower than a user-specified value. The regularisation objective function collects residuals which quantify departures of parameters from their preferred values or conditions; this is in contrast to the measurement objective function as defined in Equation (3.1). Use of appropriately-formulated Tikhonov regularisation guides an ill- or poorly-posed inverse problem to a unique solution that satisfies conditions of minimized parameter error variance (*Tikhonov and Arsenin*, 1977; *De Groote-Hedlin and Constable*, 1990).

Following solution of the nonlinear inverse problem, the Jacobian matrix used in attainment of that solution can be used to obtain a linear approximation to the posterior parameter covariance matrix, here denoted as C´(**p**). For a well-posed inverse problem that is solved using Equation (3.2) this is calculated as:

$$C'(\mathbf{p}) = (\mathbf{J}^t\mathbf{Q}\mathbf{J})^{-1} \quad . \tag{3.3}$$

The weight matrix **Q** used in Equation (3.3) will normally be proportional to $\mathbf{Q}_m$ used in Equations (3.1) and (3.2). That is:

$$\mathbf{Q} = \sigma^{-2}\mathbf{Q}_m \tag{3.4}$$

where σ^2, the proportionality constant, is often referred to as the "reference variance". This can be calculated from model-to-measurement misfit attained through the inversion process as:

$$\sigma^2 = \frac{\Phi_{min}}{n-m} \tag{3.5}$$

where Φ_{min} is the minimized objective function, *n* is the number of observations comprising the calibration dataset and *m* is the number of parameters being estimated.

Where an inverse problem is ill- or poorly-posed, Equation (3.3) cannot be used for calculation of the posterior covariance matrix because either the matrix inverse featured in that equation does not exist, or inversion of the $\mathbf{J}^t\mathbf{Q}\mathbf{J}$ matrix leads to post-calibration parameter uncertainties that exceed prior parameter uncertainties. In such cases, a linear approximation to the posterior parameter covariance matrix can be calculated as:

$$C'(\mathbf{p}) = (\mathbf{J}^t\mathbf{Q}\mathbf{J} + C^{-1}(\mathbf{p}))^{-1} \tag{3.6}$$

where C(**p**) is the covariance matrix of prior parameter uncertainty.

Numerically, the most costly part of GML parameter estimation is calculation of the Jacobian matrix. This calculation must be repeated during each iteration of the GML process. As mentioned in the introduction to this chapter, the Jacobian matrix is usually filled by undertaking repeated model runs with the value of each parameter in

turn incremented and/or decremented. Differences in model outputs divided by differences in parameter values are then taken as approximations to partial derivatives. (Slightly more complex variants of this procedure can be employed where three or five point stencils are used for finite-difference derivatives calculation.) Finite-difference derivatives calculation thus requires that at least as many model runs be undertaken per iteration as there are parameters requiring estimation, with this number increasing where higher order stencils are employed.

Once the Jacobian matrix has been filled, parameter upgrades are calculated using Equation (3.2). Normally parameter upgrades are calculated for a few values of the Marquardt lambda. The model is then run using each set of upgraded parameters in order to monitor their effectiveness in reducing the objective function. PEST uses a trial and error procedure for selection of values of the Marquardt lambda. During any iteration, this procedure commences with the optimal lambda value inherited from the previous iteration. Selection of different lambda values for use in Equation (3.2) results in calculation of parameter upgrade vectors which point in different directions in parameter space but which are all oriented down the objective function gradient. A benefit of the trial-and-error lambda testing procedure implemented by PEST is that it allows exploration of the possibility that some directions of the parameter upgrade vector are more productive than others in terms of their ability to promulgate a reduction in the objective function. This may arise as a natural consequence of the shape of the objective function surface in parameter space. It may also be a consequence of degraded integrity of parameter upgrade calculations arising from use of a corrupted Jacobian matrix, the latter being an outcome of model output numerical granularity. The testing of parameter upgrades calculated using multiple Marquardt lambda values can therefore endow the parameter estimation process with an ability to accommodate deficiencies in the Jacobian matrix calculation process.

A benefit of the GML method is that both the filling of the Jacobian matrix and the testing of parameter upgrades are both easily parallelized as model runs required for both of these procedures are independent.

3.4 Model Emulation

3.4.1 General

In the model emulation strategy discussed in the present paper, the model emulator is used to calculate partial derivatives of model outputs with respect to adjustable parameters; hence the emulator is run instead of the real model when filling the Jacobian matrix. We refer to the emulator employed in this strategy as a "proxy model", or simply "proxy", rather than using a term such as "surrogate model" as the latter is a general term implying replacement of a (supposedly complex) simulator by a (supposedly simplified) counterpart. In contrast, the emulator described in the present study has no physical basis. It is comprised of appropriately parameterized analytical expressions that are trained to reproduce certain aspects of a simulator's behaviour.

Where a proxy model runs quickly, filling of the Jacobian matrix becomes almost instantaneous. This, of course, has the potential to reduce the numerical burden of GML-based parameter estimation enormously. However there are costs associated with this benefit. One cost is that derivatives calculated using a proxy are unlikely to have the same integrity as those calculated by running the simulator itself. A second cost is that the simulator must be run multiple times to train the proxy.

In the modelling context which is the focus of the present paper, the first cost may not be a cost at all. As was discussed in Section 3.2, a motivation for development of the methodology described herein is the often-problematical numerical behaviour of complex simulators. It is the authors' experience that corruption of finite-difference derivatives incurred by such misbehaviour can make GML-based parameter estimation impossible in some important management contexts. Furthermore, adoption of a GML strategy in which parameter upgrades calculated using a number of Marquardt lambda values are tested using the real simulator during each iteration of the inversion process, partly atones for inaccuracies in derivatives computation (whether these inaccuracies are incurred through use of the simulator or by use of a proxy model).

The second cost mentioned above, namely the simulator runs that must be expended in training the proxy, can indeed be non-trivial. As will be discussed below, this number can rise rapidly with the number of simulator parameters that require adjustment. However the authors' experience to date, and the example below, suggest that the overall number of runs required for implementation of proxy-based parameter estimation, including those required for proxy training, is often not much greater than that which would have been required if a numerically well-behaved simulator was used directly in GML-based parameter estimation. Where the numerical behaviour of a simulator is such as to corrupt finite-differenced derivatives calculation to the point where GML simply fails, investment of simulator runs in training of a proxy can enable progression of a parameter estimation process that would otherwise be impossible.

3.4.2 Formulation of an analytical proxy

Let o_j designate the j'th simulator output for which there is a corresponding measurement in the calibration dataset; the difference between this simulator-generated number and the j'th measurement comprises the j'th residual of the $\boldsymbol{r}$ vector featured in Equation (3.1). Let p_i designate the i'th parameter (optionally log-transformed for enhancement of model linearity) adjusted through the history-matching process. Suppose that there are m of these and that collectively they are denoted by the vector $\mathbf{p}$. We assume that the relationships between an individual o_j and the m elements p_i of $\mathbf{p}$ can be approximately replicated over a significant subspace of parameter space by an analytical relationship of the following form:

$$o_j(\mathbf{p}) = F_j(\mathbf{c}_j, \mathbf{p}) \tag{3.7}$$

where $\mathbf{c}_j$ is a vector of proxy model parameters chosen so that F_j emulates the behaviour of o_j as well as possible over the range of likely values collectively taken by simulator parameters $\mathbf{p}$. To avoid confusion with parameters employed by the simulator, the elements c_{ij} of $\mathbf{c}_j$ are referred to as "factors" herein. As implied by the subscript j, they are specific to an individual simulator output j and to the corresponding analytical function F_j employed for emulation of that output.

For a user-specified F_j, values of c_{ij} must be estimated through calibration against values of o_j computed by the simulator for many instances of its parameters **p**. (In a further effort to avoid confusion, the c_{ij} estimation process will be referred to herein as "training" of the proxy model pertaining to simulator output j.) Ideally, for a particular j, estimation of the c_{ij} should constitute a well-posed inverse problem. This requires an appropriate choice for F_j. It also requires that the number of instances of model parameters **p** for which values of all simulator outputs o_j are computed exceed the number of elements c_{ij} of $\boldsymbol{c}_j$. The number of proxy models (i.e. the number of instances of F_j) is equal to the number of observations employed in calibration of the simulator. Therefore as many training exercises as there are proxy models (F_j) must be undertaken.

Choice of an appropriate F_j is simulator-specific. Where different analytical proxies are employed for different o_j, it is also simulator-output-specific. Selection of an appropriate F_j must take account of the following:

- As stated above, for all likely values taken by simulator parameters **p**, values for o_j calculated using Equation (3.7) should be similar to those calculated by the simulator;
- The proxy model F_j should compute these values with minimal numerical burden;
- $F_j\left(\mathbf{c}_j, \mathbf{p}\right)$ should be differentiable with respect to the elements c_{ij} of $\boldsymbol{c}_j$, this allowing rapid estimation of c_{ij} through GML-based training of F_j; and
- $F_j\left(\mathbf{c}_j, \mathbf{p}\right)$ should be differentiable with respect to the elements p_i of **p**, this allowing use of F_j in filling the Jacobian matrix used in estimation of parameters **p** used by the simulator.

In the example presented in Section 3.5, F_j is a second order polynomial. Thus:

$$F_j(\mathbf{c}_j, \mathbf{p}) = c_0 + \sum_{i=1}^{m} c_i p_i + \sum_{i=1}^{m} \sum_{k=1}^{i} c_{ik} p_i p_k \quad . \tag{3.8}$$

In other instances of proxy model usage undertaken by the authors that are not documented herein, additional complexity has been added to F_j through imposition of bounds on its values, either through use of *max( )* and *min( )* functions applied to F_j, or in a "softer" fashion through applying the *tanh( )* function to numbers computed by F_j. Further enhancements to Equation (3.7), or completely different analytical functions, could be readily employed in different circumstances in order to promulgate better emulation of simulator behaviour by the analytical proxy.

If Equation (3.8) is used as a basis for proxy design, it is easily shown that if **p** has m elements, then the number of elements n_c which comprise $\boldsymbol{c}_j$ is given by:

$$n_c = (m+1)(m+2)/2 \tag{3.9}$$

Hence at least this many simulator runs must be undertaken to provide the instances of o_j required for training of F_j; the most numerically efficient procedure for training of all F_j proxies would be accomplished using respective o_j computed using a single set of simulator runs which employed n_c instances of **p**.

Appropriate sampling strategies through which values of **p** used in the proxy training process can be obtained are briefly discussed in Section 3.4.3. Let k designate the number of these samples, and let $\mathbf{p}_s$ designate one such sample. Once these k samples have been taken, the simulator must be run k times to obtain k instances o_{js} of each o_j. Training of all o_j-specific proxy models can then be undertaken. Ideally this can be accomplished through GML-based estimation of proxy model factors c_{ij} through solution of j o_j-specific inverse problems in which proxy model factors are adjusted to minimize the discrepancies between simulator and proxy outputs for all instances of $\mathbf{p}_s$. Even where the number of proxy models requiring training is large, GML-based proxy model training is likely to be rapid if the proxies are capable of calculating derivatives of their outputs with respect to their factors; presumably analytical expressions for these derivatives are not difficult to derive because of the analytical nature of the proxies themselves.

It is not essential for proxy model outputs to faithfully replicate simulator outputs for all $\mathbf{p}_s$; in fact this is unlikely to occur where the F_j proxies embody simple analytical expressions such as Equation (3.8). Furthermore, the greater the redundancy in parameter sample sites (that is, the greater the amount by which k exceeds the number of proxy model factors requiring estimation n_c), the less likely is a perfect match between simulator and proxy model outputs to be achieved for all parameter samples. This is not necessarily a bad thing; in fact it can be a desirable design specification for an analytical proxy model where simulator outputs are contaminated by numerical granularity induced by solver non-convergence or by some of the other simulator numerical difficulties discussed above. For this reason, proxy-to-simulator fit-limiting can comprise a design feature of the inversion process through which proxy model factors are estimated. In recognition of this proxy model training as undertaken by the PEST suite optionally employs Tikhonov regularisation. The inversion problem which is then repeatedly solved to estimate factors for all o_j-specific proxies is thereby formulated as a constrained minimization problem in which the absolute values of proxy factors c_{ij} are minimized subject to attainment of a pre-specified level of proxy-to-simulator fit. As a further option, PEST provides access to a global optimiser, (the CMAES method of *Hansen et al.*, 2003) to use as a "primer" for GML-based factor estimation. This can expedite training of highly nonlinear proxy models, at a cost of proxy training speed.

A disadvantage of the proxy model strategy discussed above is that it requires the training of hundreds, or perhaps thousands, of proxies in modelling contexts characterized by large calibration datasets. However this disadvantage is offset by the following features of this strategy:

- GML-based training of a large number of proxies is relatively fast;
- Possible inclusion of Tikhonov regularisation in the training process provides the modeller with a means of limiting proxy-to-simulator fit to the level of numerical noise associated with simulator outputs; and
- Differentiability of F_j with respect to the elements of **p** allows gradient-based methods such as the (Tikhonov regularisation enhanced) GML

methodology to be used in calibration, and/or in calibration-constrained uncertainty analysis, of complex simulators in contexts where these would otherwise be employed only with great difficulty.

3.4.3 Sampling of Parameter Space

The number k (with a lower limit of n_s) of parameter samples $\mathbf{p}_s$ used in proxy model training, and the values of these samples, can exert a strong influence on proxy model performance, and the success, or otherwise, of the simulator history-matching process which proxy model usage is designed to support. The PEST suite presently supports a number of sampling options, namely: (1) user-specified; (2) Sobol; (3) random uniform; (4) random normal and; (6) Latin hypercube.

A number of matters must be considered when designing a sampling strategy. Ideally the number of sample points should be as small as possible (though always exceeding the n_s lower limit number of proxy factors c_{ij} requiring estimation) in order to reduce simulator run requirements. However if simulator numerical behaviour is problematical, the number of parameter samples may need to be increased to accommodate the presence of "numerical noise" in the proxy training dataset, through enhanced model output redundancy. Parameter samples should provide good coverage of that part of parameter space in which parameters that are ultimately estimated through the simulator calibration process are thought to lie. Presumably initial parameter values employed in GML-based simulator calibration will reflect, at least to some extent, a modeller's anticipation of what these calibrated values may be. However calibrated parameter values may nevertheless differ widely from initial parameter values; proxy model design, and the parameter sampling strategy which supports it, must accommodate this.

Sample selection as a basis for emulator design, and in experimental design generally, has been discussed by many authors. See, for example, *Eide et al.* (1994), *Alam et al.* (2004), *Razavi et al.* (2012b) and references cited therein. So-called "space filling" strategies which attempt to maximize coverage of parameter space with a pre-determined number of samples are often used to support proxy model design. These include Latin hypercube, as well as "symmetric Latin hypercube" sampling; see

discussions of these methods in *Alam et al.* (2004), *Fen et al.* (2009), *Regis and Shoemaker* (2004) and *Mugunthan et al.* (2009). An alternative to these is that provided by *Sobol* (1979); this method is designed to ensure maximum separation of a user-specified number of samples throughout a subspace of parameter space defined by lower and upper parameter bounds.

Conceptually, there is no reason why parameter sampling should be a once-only operation. Indeed, as simulator runs are undertaken during early stages of proxy-assisted simulator calibration, the parameters used in these runs (and the simulator outputs that correspond to them) can be added to the parameter sample set and then used for proxy factor refinement. Alternatively, or as well, a multi-stage parameter sampling strategy can be pursued. A space filling strategy can be adopted prior to commencement of the simulator calibration process. Because further sampling will follow, there may be no need for initial sampling to be too dense; hence simulator runs can be conserved. After a number of proxy-assisted GML iterations have been undertaken, an approximation to the posterior parameter covariance matrix can then be calculated using Equations (3.3) or (3.6). This can then form the basis for a second sampling stage comprised of random, sobol or Latin hypercube sampling to support proxy factor re-estimation or refinement; presumably samples thus obtained will span the space in which calibrated simulator parameter values are most likely to reside. Strategic, multi-stage sampling of this kind may reduce overall simulator run requirements because the targeted nature of sampling reduces the need for high sampling density. Though not documented herein, the authors have used such a multi-stage sampling strategy successfully in calibration of complex simulators.

3.5 A Proxy-Assisted Example

3.5.1 Problem Description

Figure 3.1 depicts two confined aquifers separated by an aquitard. These comprise the domain of a 2-dimensional cross-sectional model that simulates saltwater intrusion into these units. The model domain is 12 m in length and 2 m in height,

these dimensions being similar to those employed in the example presented by *Burrows and Doherty* (2014) which, in turn, was based on those used by *Henry* (1964). The sea is on the right (i.e. the east) of the model domain, while freshwater flows into the model domain from the west. Of interest is the location of the "saltwater interface"; for present purposes this is defined as the location of the 10% saltwater isohaline. First, observations of the location of this interface under two different conditions (one steady state and one transient) are used to calibrate the model. Following calculation of the steady state interface position, freshwater inflow from the west is reduced so that the interface moves to the left; steady state observations of the interface position are then supplemented by observations of the interface location after 0.5 days of reduced western freshwater inflow to comprise the total calibration dataset. The model is then used to make a prediction, this being the location of the saltwater interface after a further 1.0 days of reduced western freshwater inflow.

Proxy-assisted estimation of stratigraphic unit hydraulic conductivities and model-wide dispersivities is now described. As a sequel to calibration, proxy-assisted calibration-constrained Monte Carlo analysis is undertaken in order to explore the uncertainty associated with the prediction of the future interface position.

Density-dependent flow and transport is simulated using SEAWAT version 4 (*Langevin et al.*, 2008). The vertical cross-sectional model domain is endowed with 600 cells in each of 100 layers, this resulting in a total of 60,000 cells. The time variation diminishing (TVD) scheme is employed in simulation of advective movement of saltwater. The time required for solution of the steady state position of the interface and movement of this interface over the first transient stress period (in which western freshwater inflow is reduced) is about 45 minutes on a i7 cpu PC clocking at 1.60 Ghz. This constitutes the simulator run time under calibration conditions.

To represent the ocean, a SEAWAT "general head" boundary condition is ascribed to all cells comprising the eastern vertical boundary of the model domain. All cells comprising this boundary are ascribed a hydraulic head of 0.0 m and a salinity of 35 kg/m^3; both of these remain fixed during the simulation. Meanwhile, inflow of water

with a salt concentration of 0 kg/m^3 into the west of the model domain is simulated using injection wells placed along this boundary. Inflow is vertically uniform; it totals 1.0 m^3/d under steady state conditions and 0 m^3/d thereafter. Flow across the top and bottom boundaries of the model domain is zero.

Porosity is uniformly 0.35 while molecular diffusion is set to a uniform value of 0.01 m^2/d. Other hydraulic properties employed by the model are listed in the first column of Table 3.1. Horizontal and vertical hydraulic conductivites are layer-specific and layer-uniform; longitudinal dispersivity and the ratio of vertical to longitudinal dispersivity are uniform throughout the model domain. All parameters featured in Table 3.1 are subject to calibration adjustment; this totalling 8 parameters in all.

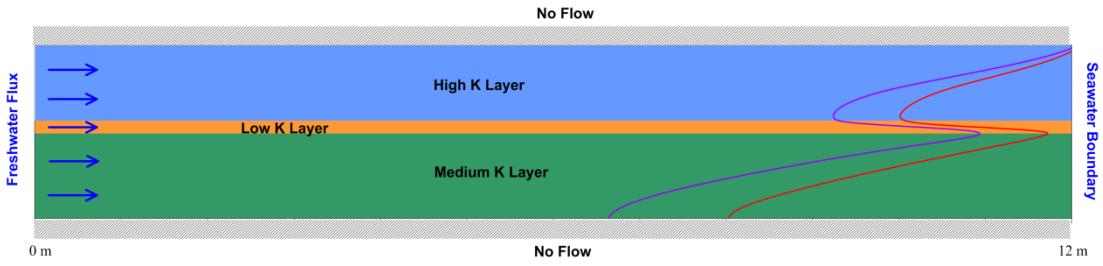

Figure 3.1: The example problem. The two saltwater interface locations comprising the calibration dataset are shown in red (steady state) and blue (0.5 days after western freshwater inflow reduction). Interface positions shown in the figure were calculated using the "reality" parameter set listed in the second column of Table 3.1.

Table 3.1: Results of calibration exercises.

Parameter identifier	Parameter description	Actual value	Initial value	Estimated parameters			
				Baseline	Proxy-assisted	Proxy	CMAES_P
kh1	Horizontal K upper aquifer (m/d)	499.4	600	397.2	521.6	500.7	539.3
kv1	Vertical K upper aquifer (m/d)	152.2	100	97.2	90.8	80.7	125.9
kh2	Horizontal K aquitard (m/d)	6.12	3	0.6	3.08	5.92	1.99
kv2	Vertical K aquitard (m/d)	0.76	0.2	0.35	0.67	0.67	0.81
kh3	Horizontal K lower aquifer (m/d)	241	300	386	267.6	251.8	279.7
kv3	Vertical K lower aquifer (m/d)	58.72	40	45.9	55.27	84.08	28.17
dspl	Longitudinal dispersivity (m)	0.0046	0.0055	0.0056	0.0081	0.0047	0.0076
tprv	Transverse/Vertical Dispersivity (-)	0.76	0.6	0.84	0.69	0.68	0.56
Number of simulator runs				567	126	120	872
Final objective function			73092	5993	672	1610	670
Parameter objective function (Equation (3.10))				1.24	0.21	0.11	0.42

The distance of the saltwater interface from the western boundary of the model domain in each of the 100 layers into which the model domain is discretised was calculated by the model (comprised of SEAWAT and appropriate post-processing software) using the parameters listed in the third column of Table 3.1. These parameter values comprise the "reality parameters" which the model calibration process seeks to estimate; they were chosen from a suite of random parameter realizations as those which lead to a prediction of the future interface position which departs significantly from the mean of these predictions. This was done to assist in assessing the integrity of the uncertainty analysis process discussed in Section 3.5.4. Random, normally distributed, "measurement noise" with a mean of zero and a standard deviation of 0.04 m was added to steady-state interface distances (from the western boundary) and those pertaining to 0.5 days of reduced western boundary inflow to form a 200-member measurement-noise-contaminated calibration dataset.

Note that in any model layer the location of the 10% (i.e. 3.5 kg/m^3 salt concentration) isohaline is obtained through linear interpolation between cell centres for which calculated concentrations bracket this marker concentration. (Salt-water interface locations rather than concentrations constitute the calibration dataset in the present example because (a) they are easily "corrupted" with structural noise born of model imperfections – see below, (b) in many real-world modelling contexts the position of the saltwater interface forms the focus of modelling interest, and (c) especially where borehole data is supplemented with geophysical data, saltwater interface locations may be approximately known on a semi-regional scale.)

As has already been discussed, a motivation for development of the proxy model strategy described in Section 3.4 is the accommodation of numerical "malperformance" of a complex simulator. "Malperfomance" was introduced to the simulator used in the present example through purposeful corruption of the linear interpolation process used to determine the location of the 10% isohaline. (It is important to note that this was not done in calculating those isohaline locations which comprise the calibration dataset. However it was implemented in all simulator runs used in proxy model training and in calibration and uncertainty analysis.) This corruption was introduced by interchanging distances from the interface to bracketing cell centres on either side of the interface. If, for example, concentrations at neighbouring cell centres are calculated to be 3.4 kg/m^3 and 3.9 kg/m^3, then the saltwater interface (i.e. the 3.5 kg/m^3 isohaline) lies 20% of the way between the first and second cell centre. In introducing numerical corruption to model outputs, the interface was instead located at 80% of the distance between the first and second cell centres. Implementation of this strategy results in an effective sawtooth pattern of concentrations along any model layer, with a significant discontinuity at each cell centre.

Figure 3.2 (red graph) shows a plot of interface location vs. the value of an adjustable parameter as calculated by the corrupted simulator, at the location of one particular observation. Parameter increments between symbols depicted in this graph are typical of those used in finite-difference derivatives calculation. Of particular note are the many instances of locally reversed gradient. In the authors' experience plots

such as this are commonly encountered when attempting to discover the reasons for poor performance of GML-based parameter estimation of numerically problematical models. Also depicted in Figure 3.2 (green graph) are proxy-model-calculated values of the same observation. Though saltwater interface positions calculated by the proxy appear to depart to some extent from those expected of the non-corrupted simulator, they are nevertheless approximately correct (as are their derivatives with respect to the adjustable parameter which is featured in this figure); furthermore they are free of gradient reversals. (Note that the offset between simulator- and proxy-calculated interface locations that is evident in this figure is local; its value and sign are a function of the local value of the parameter represented in the figure, as well as those of other parameters whose values were held fixed while this parameter was varied.)

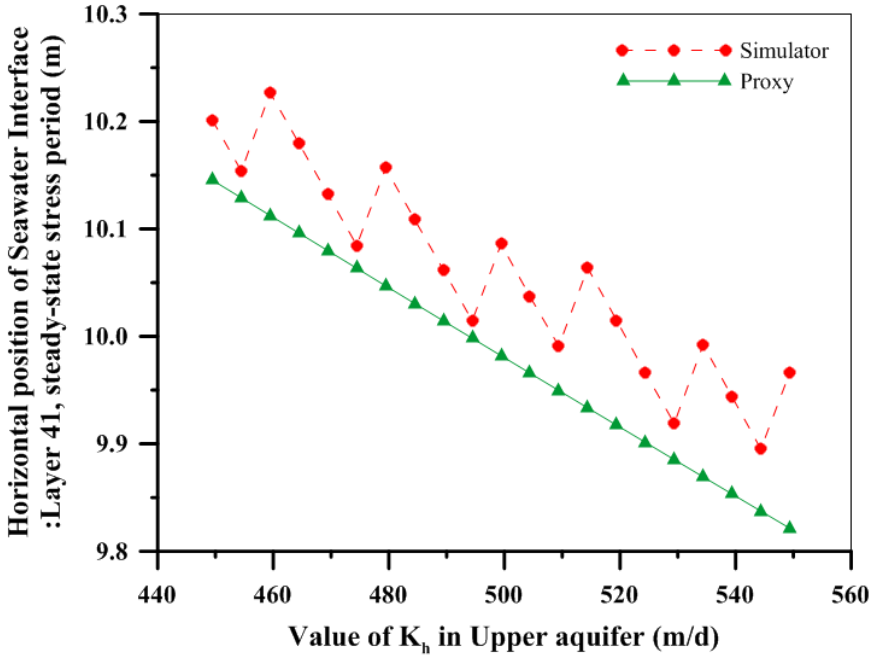

Figure 3.2: Testing of derivatives. The red dots show the location of the interface in model layer 41 calculated by the corrupted simulator using different values of upper aquifer K_h while all other parameters are held at a constant value. The green triangles show interface positions calculated by the proxy model for the same parameter increments.

Non-proxy-assisted calibration of the defective simulator was attempted using PEST. No regularisation was employed as estimation of the 8 parameters listed in Table 3.1

constitutes a well-posed inverse problem. Starting values for this (and all other calibration attempts described in the present section) are listed in the fourth column of Table 3.1; note that PEST was asked to estimate the logs of these parameters in implementing the parameter estimation process. With the diagonal elements of the $\mathbf{Q}_m$ matrix of Equation (3.1) all set to 620 (this being the inverse of the variance of random measurement noise added to the synthetic calibration dataset) an objective function of 5993 was achieved. This is well above a value of 192 (calculated as *n-m*, where *n* is the number of observations and *m* is the number of adjustable parameters, as per *Koch*, 1999) that is anticipated on the basis of measurement noise alone. Furthermore, PEST's behaviour in attempting this calibration exercise was unstable, with the objective function rising on some iterations and then recovering ground on later iterations only to rise again thereafter. The attempted calibration process required 567 runs of the complex model over a total of 26 iterations of the GML process. (Finite-difference derivatives were computed using a three-point stencil in order to mitigate the deleterious effects of model defects on these calculations. A total of 416 simulator runs were required for computation of derivatives and a total of 151 simulator runs were required for testing and refinement of parameter upgrades using different values of the Marquardt lambda.)

PEST's failure to lower the objective function to a value approaching its expected value should be seen in context. The objective function calculated using initial parameter values is 73092; PEST was able to reduce this substantially. Furthermore, the fit between the real saltwater interface position and that calculated by the model using PEST-estimated parameters isn't too bad visually; see Figure 3.3. This PEST-attempted calibration process is referred to as the "baseline" process in the discussion that follows.

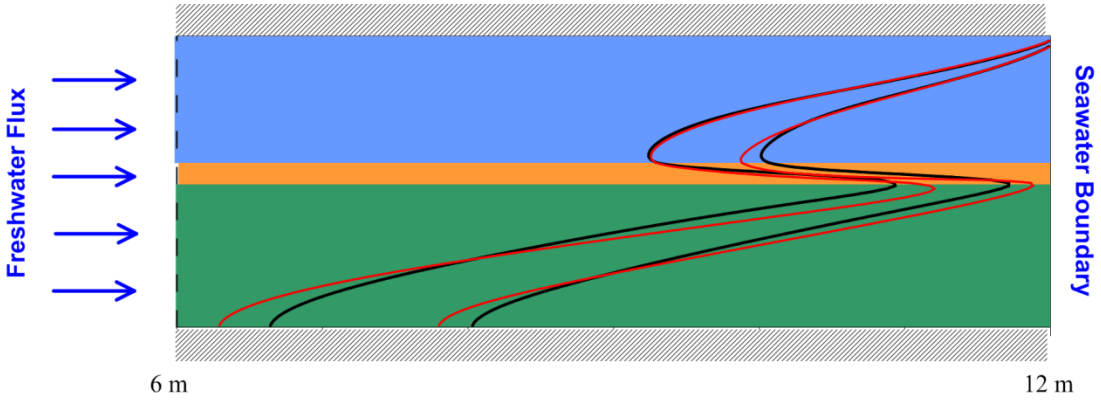

Figure 3.3: Baseline calibration. Saltwater interface positions (red) calculated using "baseline" estimated parameters. Real saltwater interface positions are shown as black lines. Interface positions are shown for steady-state conditions and after 0.5 days of reduced freshwater inflow.

To assist in comparing the outcomes of this, and a number of other calibration exercises described in this section, an objective function Φ_p is defined to measure the integrity of estimated parameters:

$$\Phi_p = \sum_{i=1}^{m} \left[\log(p_{ie}) - \log(p_{i0}) \right]^2 \quad . \tag{3.10}$$

In Equation (3.10) p_{ie} is the estimated value of parameter i and p_{i0} is its true value. As is apparent from Table 3.1, in the case of the "baseline" estimation process, while estimated values for some parameters are reasonably good, estimates for other parameters depart markedly from their true values.

3.5.2 Proxy model construction

For this example, a proxy model comprised of a second order polynomial (see Equation (3.8) was employed to emulate the relationship between each of the 200 model outputs used in the history-matching process and the 8 parameters that were estimated through that process. While a more complex proxy may have sustained a superior fit between simulator and proxy model outputs over a broader range of parameter values, accommodation of simulator "malperformance" required that each proxy emulate the general nature of the relationship between model outputs and

parameters, while eschewing emulation of corrupted simulator output details. A second order polynomial sufficed for this purpose.

The method of *Sobol* (1979) was used to generate 120 parameter samples over an 8-dimensional log-transformed subspace of parameter space defined by the bounds listed in Table 3.2. (As mentioned in Section 3.4.3, the philosophy behind Sobol sampling is to maximize the distance in parameter space between samples of parameters taken between user-specified upper and lower bounds. Thus maximum coverage of bounded parameter space is gained with the fewest number of samples.) The simulator was then run using each of these samples in order to obtain simulator-generated counterparts to observations comprising the calibration dataset. In theory 45 simulator runs are required to build a second order polynomial proxy model (see Equation (3.9) for each calibration-pertinent output of an 8 parameter simulator. The decision to generate 120 samples was somewhat arbitrary; it rested on the notion that a greater number of samples may lead to more faithful reproduction of simulator behaviour by the proxy over a broader range of adjustable parameter space.

Table 3.2: Specifications for Sobol sampling used for proxy training

Parameter Identifier	Log10 transformed parameter values			
	Actual value	Initial value	Sobol lower bound	Sobol upper bound
kh1	2.70	2.78	2.30	3.18
kv1	2.18	2.00	1.70	2.30
kh2	0.79	0.48	-0.30	1.18
kv2	-0.12	-0.70	-1.30	0.00
kh3	2.38	2.48	2.00	2.70
kv3	1.77	1.60	1.00	2.00
dspl	-2.34	-2.26	-3.00	-2.00
tprv	-0.12	-0.22	-1.00	0.00

It is worth noting that, apart from providing datasets for proxy model training, a suite of pre-calibration simulator runs can yield other useful information at the same time. For example they could form the basis for qualitative sensitivity analysis. Furthermore, the parameter sample responsible for the lowest objective function could be used as a starting point for ensuing proxy-assisted GML calibration; this was not done in the present case in order to facilitate comparison between the performances of proxy-assisted and non-proxy-assisted simulator calibration, both of which commenced from initial parameter values listed in Table 3.1.

Following completion of the 120 simulator runs, proxy models specific to the 200 observations comprising the calibration dataset were constructed. For a second order polynomial proxy model, the c_0, c_i and c_{ik} factors of Equation (3.9) require estimation. Estimation of all of these factors for proxy models associated with all observations comprising the calibration dataset was implemented in an automatic procedure that required less than 5 minutes to complete once the initial simulator runs had been carried out.

Once trained, proxy model run times were trivially small. Less than a second was required to run all observation-specific proxies to calculate a suite of 200 counterparts to measurements comprising the calibration dataset. Calculation of a Jacobian matrix based on a three point finite-difference stencil took less than four seconds. (This could be made even faster through use of analytical proxy derivatives; however to maintain flexibility in later enhancements to proxy model design, finite differences are presently employed by PEST for proxy-based derivatives calculation.) Use of the same stencil in conjunction with the simulator would have required 16 simulator runs; these would have taken approximately 11.3 cpu hrs to complete on the same computer. Furthermore, as is apparent from Figure 3.2, the integrity of these derivatives would have been questionable.

3.5.3 Proxy-assisted parameter estimation

Proxy-model assisted calibration, undertaken in the manner described in Section 3.3, reduced the objective function from an initial value of 73092 to a minimized value of 672 in 3 iterations. The inversion process required 6 simulator runs and 48 proxy

model runs. When simulator runs required for proxy model construction are taken into account, a total of 126 simulator runs were therefore required for calibration of the simulator; this value is recorded in Table 3.1. (As stated above, as few as 45 simulator runs could potentially have been used in proxy model construction, so simulator run requirements for this particular calibration example may be over-stated.)

Though the objective function is above its expected value of 192, it is greatly reduced from its initial value, and considerably lower than that achieved through simulator-only estimation (as is represented by the baseline calibration attempt). Figure 3.4 demonstrates that the fit between the observed and calculated saltwater interface is visibly very good and much better than that achieved using the defective simulator alone. Table 3.1 indicates that estimated parameters are closer to their true values than when calibration was undertaken using the simulator alone. Note that some departure from true parameter values is always to be expected where parameter estimation is based on a noisy calibration dataset. Equations (3.3) and (3.6) show that estimated parameters inherit uncertainty from this noise. In the present example the "noise" associated with the calibration dataset is enhanced by "structural noise" associated with the defective simulator as is demonstrated in Figure 3.2.

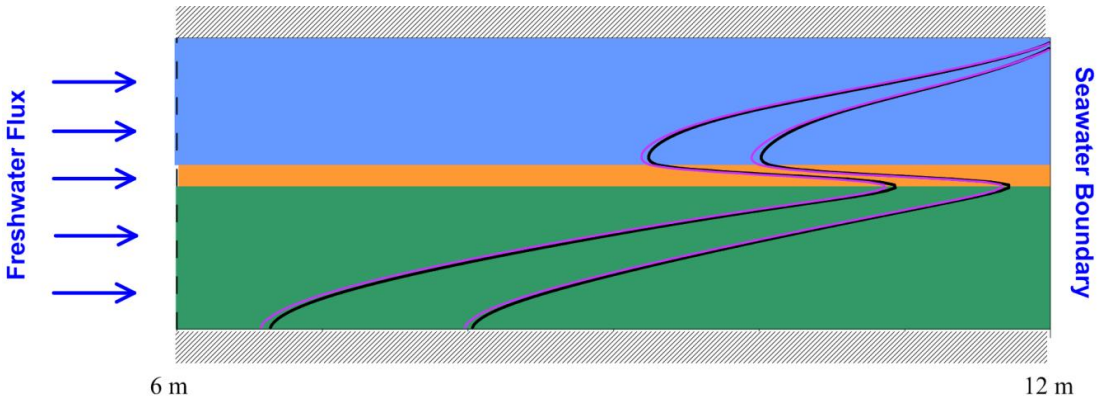

Figure 3.4: Proxy-assisted calibration. Saltwater interface positions (purple) are calculated using parameter values estimated through proxy-assisted calibration. Real saltwater interface positions are depicted as black lines.

Two more calibration exercises were undertaken. In the first of these exercises the proxy model was employed for both Jacobian matrix calculation and for testing parameter upgrades calculated using different values of the Marquardt lambda (which appears in Equation 3.2). Hence the simulator was not used at all in the parameter estimation process. As shown in Table 3.1, an objective function of 1610 was achieved; while this is greater than that achieved through proxy-assisted calibration, the value for Φ_p (see Equation 3.10) attained through this calibration exercise indicates that parameter estimates are superior. Hence while intermittent use of the simulator in proxy-assisted calibration promulgated greater reduction of the objective function, for this particular example at least, structural noise associated with simulator outputs enhanced the potential for error in estimated parameters. Ideally, such model-generated error potential should be accommodated for in post-calibration uncertainty analysis; this is done in Section 3.5.4.

Finally, calibration was again undertaken using the simulator alone without assistance from the proxy model. However, this time parameters were estimated using the CMAES global optimization algorithm (*Hansen et al.*, 2003). In that exercise 872 simulator runs were required to reduce the objective function to about the same value as that achieved through proxy-assisted GML parameter estimation. Table 3.1 demonstrates that estimated parameter values showed greater departures from their real values than those achieved through proxy-assisted GML calibration, despite attaining a similar level of model to measurement misfit as that attained using the proxy-assisted estimation process. Obviously, the computational burden was also very much higher than that of proxy-assisted GML calibration.

3.5.4 Proxy-assisted uncertainty analysis

A linear approximation to the posterior parameter covariance matrix can be obtained using Equation (3.6). Ideally, to minimize the error incurred through linear approximation, the Jacobian matrix featured in this equation should be computed using estimated, rather than prior, parameter values. In the present case this was done at minimal cost using the proxy model; and perhaps with improved integrity. The prior parameter covariance matrix C(**p**) appearing on the right side of Equation (3.6)

was assumed to be diagonal; prior parameter distributions were assumed to be independently log-normal. Prior standard deviations were obtained by dividing differences between log-transformed parameter upper and lower bounds (as presented in Table 3.2) by 4; these bounds are therefore assumed to define 95% parameter confidence intervals.

As the model is nonlinear, the posterior parameter probability distribution is not multinormal. Because of this, and the approximations which are associated with use of Equation (3.6), samples drawn from a log-multinormal distribution employing the covariance matrix of Equation (3.6) and centred on calibration-estimated parameter values do not always lead to model-calculated objective functions that are low enough to be considered as respecting calibration constraints. Theoretically, objective function thresholds associated with various posterior parameter confidence intervals can be calculated using formulas such as those provided in *Vecchia and Cooley* (1987); naturally, these thresholds exceed the minimized objective function. Sampled parameter values must be such that the objective functions with which they are associated are below these thresholds if they are to be considered to respect calibration constraints at respective confidence levels. In the present instance however we adopt a more qualitative approach to selection of an objective function threshold. We define an objective function value of 1000 as being that above which a parameter set is deemed to fail to respect calibration constraints, and is therefore very unlikely to be the real parameter set. This subjective choice of objective function threshold is compatible with a visibly good fit between measured and modelled saltwater interfaces under calibration conditions. Meanwhile the subjective nature of this choice recognizes the contribution made to model-to-measurement misfit by model-generated structural noise whose covariance structure is unknown.

The following procedure was adopted for calculation of a suite of parameter values that respect the above objective function constraint at the same time as they respect expert knowledge constraints. These values are therefore considered to be samples of an approximation to the posterior parameter distribution. As such, they can be collectively used to make predictions of future system behaviour in order to associate ranges of uncertainty with those predictions.

1. Generate a random parameter set centred on estimated parameter values (column 6 of Table 3.1) using the covariance matrix of Equation (3.6) and an assumption of posterior parameter multinormality.
2. If the objective function associated with this parameter set sample is below the threshold specified above, accept that parameter set as a sample from the approximate posterior parameter probability distribution.
3. If the objective function is above the threshold, use proxy-assisted GML parameter estimation to adjust parameter values in order to lower the objective function below the threshold. Include Tikhonov regularisation constraints in the parameter adjustment process, these being formulated in a way that minimizes parameter departure from their initial (random) values while meeting the objective function target.
4. If, after three iterations of proxy-assisted inversion, the objective function has not been reduced to the specified threshold, discard the parameter set.

Three hundred samples of the linear approximation to the posterior parameter distribution were generated; the above procedure was then applied to each of them. The same proxy model as that employed earlier for simulator calibration was used in post-calibration, proxy-assisted parameter adjustment. (In more difficult modelling contexts it may be advisable to re-train the proxy using parameter sets sampled from the linear posterior covariance matrix in order to enhance the ability of the proxy to emulate simulator response in that part of parameter space in which samples of the true posterior parameter distribution are likely to lie.) The outcome of this process was 266 parameter sets that respect calibration constraints on the one hand, and are compatible with prior knowledge of parameter values on the other hand. As such they can be considered to be samples of the posterior parameter probability distribution. These were obtained at an average cost of 1.92 simulator runs per sample.

The use of Tikhonov constraints in random parameter sample re-adjustment is an important feature of the above process. Without these constraints it would be too easy for a low objective function to be achieved without maintaining respect for the random sample from which each parameter adjustment process was initiated. While

the density of parameter samples that are forthcoming from the above random parameter adjustment procedure cannot be construed as defining the "true" posterior parameter probability density function, in contrast to those forthcoming from a procedure such as Markov chain Monte Carlo (MCMC), maintenance of respect for initial random parameter values does afford some protection against failure to sample important parts of posterior parameter space.

Disadvantages of the above procedure when compared to more exact methodologies such as MCMC must be seen in context. In particular:

- The method described herein is very efficient when considered in terms of simulator run requirements;
- Because model-to-measurement misfit is dominated by model structural noise of unknown (and probably unquantifiable) statistical properties, any method of post-calibration uncertainty analysis, whether or not it is MCMC-based, will be approximate; and
- Without a considerable investment in simulator runs, this often requiring the use of multiple Markov chains, MCMC methods are not immune to failure in sampling important parts of a posterior parameter distribution (*Gelman et al.*, 2013).

An advantage of the method documented herein, particularly its use of Tikhonov constraints to promulgate maximum adherence to initial samples, is that it could be readily extended to complement sparser, and more efficient, sampling methods such as Latin Hypercube. This would be useful where simulator run times are so long as to present an impediment to even approximate definition of a posterior parameter probability density function.

Parameter histograms resulting from the above analysis are depicted in Figure 3.5. The uncertainty associated with the prediction which is the subject of the present example (i.e. the location of the saltwater interface after 1.5 days of reduced freshwater inflow from the west of the model domain) can be assessed by employing the simulator to make that prediction using each of the 266 parameter sets obtained

as above. Predictive variability, together with the true prediction, are depicted in Figure 3.6. Importantly, in Figure 3.5 and Figure 3.6 true parameter values (in the former case), and the true value of the prediction (in the latter case) are encompassed by parameter and predictive samples respectively.

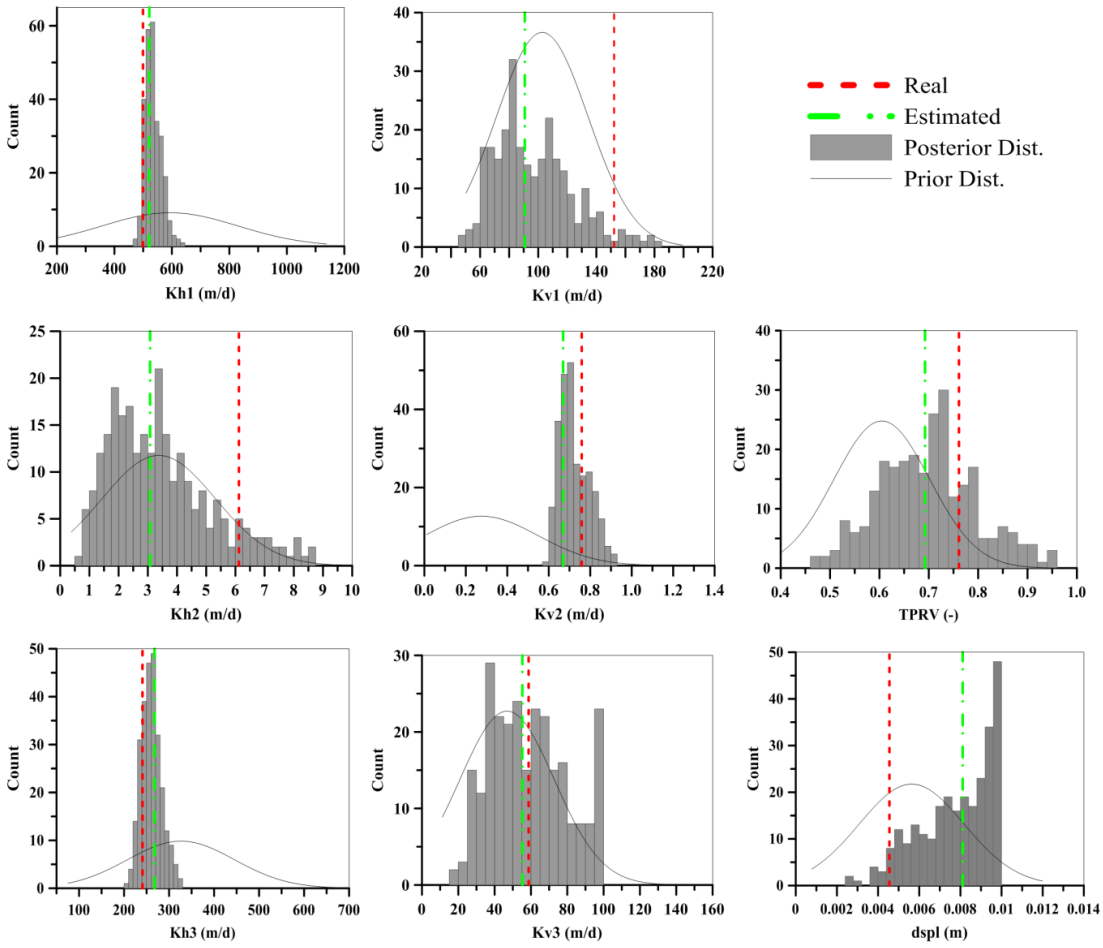

Figure 3.5: Posterior parameter distributions. Parameter histograms forthcoming from proxy-assisted sampling of the approximate posterior parameter probability distribution. Prior parameter distributions are shown as a single continuous grey line in each figure. Parameter values estimated through proxy-assisted GML calibration are shown as green dash-dot lines while true parameter values are shown as red dashed lines.

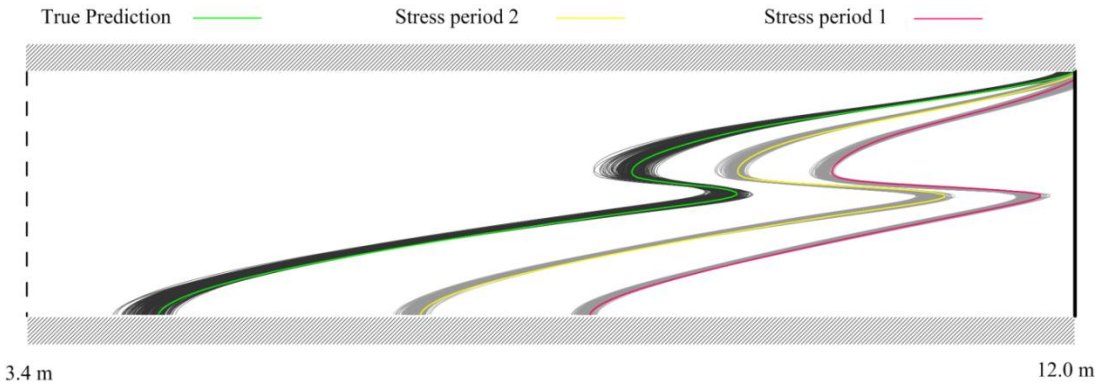

Figure 3.6: Saltwater interface after 1.5 days of reduced freshwater inflow calculated by the simulator using 266 parameter sets obtained through proxy-assisted posterior sampling (black) together with the true interface position (green). Also shown are interface positions used for model calibration (red and yellow) together with model-calculated interface positions (grey).

3.6 Discussion

A methodology has been presented for use of a proxy model in conjunction with a complex simulator to assist in calibration of the simulator, and to assist in imposition of calibration constraints on parameters employed by the simulator when analysing the posterior uncertainties of those parameters. The method complements a previous methodology presented by the authors (*Burrows and Doherty*, 2014) wherein a coarsely-gridded version of the simulator was used instead of a model proxy.

In the present paper, the term "proxy" has been used to describe a series of non-physically-based analytical functions that individually approximate the dependence of a single model output on parameters employed by a simulator. The model output that is thus emulated is matched to a measurement of system state that forms part of the simulator's calibration dataset. A different proxy is built for each such model output. The suite of proxies are individually trained through calibration against realizations of the model output that they purport to emulate, in order to optimize their capacity to represent derivatives of model outputs with respect to adjustable

parameters over a usable region of parameter space. Collectively, the suite of proxies then replaces the simulator when calculating a Jacobian matrix in support of Gauss-Marquardt-Levenberg (GML) parameter estimation. Meanwhile, the simulator is employed for the testing and refinement of parameter upgrades calculated using the proxy-derived Jacobian matrix. As implemented in PEST, the process of testing parameter upgrades can optionally be parallelized; this further adds to the efficiency of the overall proxy-assisted, simulator history-matching process.

While an analytical proxy model is unlikely to calculate environmental system states with the same integrity as a complex numerical simulator, the example presented in this paper demonstrates that the derivatives of proxy model outputs with respect to model parameters (if the analytical proxy model is chosen appropriately) have enough integrity to fill a Jacobian matrix which can then be used as a basis for calculating improved parameter sets for use by the complex simulator. Errors incurred through strategic use of a proxy model in place of a simulator in the parameter estimation process are accommodated through conjunctive use of the simulator in that process. In particular the simulator is used to test parameter upgrades, and to enhance their refinement during that part of each GML iteration wherein different values of the Marquardt lambda are employed in conjunction with a proxy-calculated Jacobian to calculate parameter value improvements. The use of different Marquardt lambdas effectively rotates the direction in parameter space along which an improved set of parameters is sought, while maintaining a search direction which is down-gradient as far as the objective function is concerned. It is therefore possible (but by no means certain) that the practical exercise of probing directions in parameter space which are close to, but not the same as, those calculated as optimal on the basis of a proxy-filled Jacobian, will find fruitful directions of parameter improvement.

Construction and use of proxy models in the manner described in this paper brings with it a number of benefits. Proxy-enabled GML-based parameter adjustment can readily accommodate numerical granularity that often accompanies quantities calculated by complex, physically-based simulators. Furthermore, it can take advantage of GML-supported schema such as singular value decomposition and

Tikhonov regularisation that can achieve solution to a non-unique inverse problem that is of minimum error variance. After a solution of the inverse problem has been found, the Jacobian matrix that is filled by the proxy model emulator can then be used to assess parameter uncertainty. If desired, a linear approximation to the posterior parameter probability distribution can then be sampled, and its samples rapidly adjusted to respect calibration constraints; samples of a nonlinear approximation to the posterior parameter probability distribution are thereby attained. Use of Tikhonov regularisation in the parameter adjustment process through which calibration constraints are enforced on random parameter samples can help to prevent gaps in the sampling of the approximate posterior probability distribution from arising in parts of posterior parameter space. At the same time it can support the use of efficient sampling strategies such as Latin Hypercube, despite the fact that the true posterior probability distribution may have no analytical characterization.

In the example presented herein, use of a proxy model enabled calibration of a numerically troublesome model has been demonstrated. Calibration of this model would have been possible, but non-optimal, without use of the proxy. For this particular example, if the simulator was not, in fact, numerically troublesome, then it could have been calibrated without the help of a proxy model with a numerical burden that is commensurate with that required for training and use of the proxy. However the same cannot be said for post-calibration uncertainty analysis. Use of a proxy model allowed the calculation of many different samples of the posterior parameter distribution with minimal simulator run cost. Even if simulator outputs were numerically uncorrupted, the computation cost of obtaining these samples would have been considerably greater without the use of a proxy.

As for any method that uses approximation to achieve computational feasibility, use of a proxy model in the manner described herein may incur some problems in some modelling contexts. It is not a foregone conclusion that outputs of any complex, physically-based model can indeed be emulated by an analytical function. While the analytical expressions presented in Equations (3.7) and (3.8) could be readily expanded to accommodate more complex simulator behaviour, parameterization of these expressions may become difficult if simulator outputs show complex

dependencies on the values of simulator parameters. Furthermore the more factors that a proxy model employs to enhance its ability to emulate complex simulator outputs, the greater will be the number of pre-calibration simulator runs that are required to train it, and the more likely it is that these factors will vary over parameter space. The latter difficulty can be accommodated to some extent by re-training the proxy as the parameter estimation process progresses.

Another problem with the methodology outlined herein is that the number of parameters that can be estimated through proxy-assisted calibration is limited. Hence highly parameterized inversion, where parameters may number in the hundreds, or even thousands, is not possible because an unworkably large number of proxy model factors would then be required to express the dependence of a single model output on a large number of simulator parameters. In many circumstances it may be possible to ameliorate this problem by identifying a handful of parameters to which any single model output may be sensitive using experimental design techniques; the outcomes of these same runs could then be used to train a proxy that expresses dependence only on thus-identified parameters. Alternatively, or as well, proxy models could be trained to emulate the dependency of model outputs on so-called "super parameters", these being orthogonal combinations of parameters, calculated through singular value decomposition of a full–parameter Jacobian matrix, which collectively span the parameter calibration solution space; see *Tonkin and Doherty* (2005) for a more detailed discussion on the use of "super-parameters" in parameter estimation. However this requires that a full parameter Jacobian matrix be calculated using the original simulator, and that derivatives which populate that matrix have integrity. Filling of such a Jacobian matrix may not be possible.

It should be borne in mind, however, that highly parameterized inversion becomes an extremely difficult undertaking whenever model run times are high and/or model numerical behaviour is questionable. So an inability to conduct highly parameterized inversion is not so much the fault of the proxy model methodology described herein, as it is of the simulator requiring calibration.

Nevertheless, the rewards of high-dimensional model parameterization are considerable. As is explained by *Doherty* (2015b), an advantage of highly

parameterized inversion and highly parameterized, inversion-constrained uncertainty analysis, is that it can support exploration of the uncertainties of predictions that are sensitive to parameterization detail. This is important, not because parameterization detail is necessarily estimable on the basis of a limited calibration dataset, but precisely because it is not uniquely inestimable. It may be possible to aspire to these same goals in difficult modelling contexts where simulators are slow and numerically problematical by adopting a simple parameterization scheme for calibration, and then introducing parameterization complexity for post calibration uncertainty analysis. Thus, for example, a few large zones of assumed parameter constancy whose values are adjusted in a parameter-parsimonious calibration process can be supplemented with a denser parameterization scheme comprised of space-dependent multipliers of zonal values. Random realizations of these multiplier fields, when superimposed on zone-based parameterization, may de-calibrate a previously calibrated model. It may be possible to then re-calibrate the simulator by proxy-enabled adjustment of zonal parameters on which the more detailed parameterization scheme is superimposed.

In summary, the methodology presented herein provides support for calibration, and calibration-constrained uncertainty analysis in modelling contexts in which these two tasks would otherwise be very difficult or even impossible. These contexts are characterized by long simulator run times and by problematical simulator numerical performance. The method has its shortcomings, and may fail in some situations. However, at a cost of a relatively small number of complex model simulations (whose outputs may then be used for other purposes such as sensitivity and pre-calibration uncertainty analysis, and as part of a strategy to seek a lower objective function through implicit or explicit global optimization) the method is easily tested. At worst, it will take very few further simulator runs to demonstrate that the methodology described herein can achieve no further reduction of the objective function. At best, the methodology described herein will reduce the objective function substantially and provide useable linear and/or nonlinear estimates of parameter uncertainty.

Chapter 4

Some Other Examples of Surrogate-Assisted Calibration Using Readily Available Modelling Software

The opportunity to adopt surrogate models for the purposes of populating an approximate Jacobian matrix exists in many everyday groundwater modelling contexts. In this chapter a few more examples are presented in which calibration of an original detailed model is successfully accomplished using surrogate models developed using differing simplification strategies. The concepts presented in this chapter may enlighten prospective practitioners or inspire further extension of these techniques.

4.1 Using SWI as a surrogate for efficient calibration of a SEAWAT model

The originality and significance of the example presented here is that it demonstrates the deployment of a surrogate model, derived from a simplified algorithm compared to the original and more physically based simulator, for the purpose of populating the Jacobian matrix during parameter estimation. This is in contrast to the implementation of the methodology described in Chapter 2 whereby the surrogate model constructed is the same model only with a coarsened grid. Furthermore, the modelling context in which this example is framed is representative of many seawater intrusion modelling scenarios that confront modelling practitioners where calibration and predictive uncertainty analysis is severely hampered by large computational burdens.

4.1.1 Introduction

Seawater intrusion modelling often requires finely discretised model domains in both horizontal and vertical directions, as well as appropriately small temporal discretisation. Such fine scale discretisation is often necessitated by the numerical constraints of the code implemented for solution to the solute transport equation. For example, fine spatial discretisation may be required to reduce the ratio of the rate of advective solute transport to the rate of diffusive solute transport, to levels that promulgate model solution convergence and mitigate numerical dispersion and/or oscillatory effects (as indicated by the Peclet number). In addition, seawater intrusion modelling exercises that require detailed characterisation of the distribution of salt concentrations, and accurate representation of the effects of density contrasts on groundwater movement especially in the vicinity of the seawater-freshwater interface, may also demand model grids that are finely discretised vertically necessitating many model layers be employed. Fine scale discretisation may also be necessary to endow the model with the ability to express fine scale geological heterogeneity particularly in contexts where geological variability is known to be present and is considered likely to have a significant bearing on model predictive outcomes. In such cases localised (model cell scale) rates of dispersion, reactions and source/sink mixing relative to the scale of the grid cell, can have a significant influence on the maximum allowable transport time step that can be tolerated for a stable and accurate solution. As an outcome of these factors regional scale seawater intrusion models that employ parameterisation schemes that allow for expression of geological heterogeneity often have extremely long run times. Furthermore, model generated outputs of salinity corresponding to measurements taking in the field, may tend to be contaminated with numerical errors that can erode the integrity of model generated observation to model parameter relationships, when these are calculated from model generated outputs arising from incremental variations in model parameters. Both long model runtimes and unreliable observation to parameter relationships can make calibration of seawater intrusion models an extremely challenging task and may render that process and a rigorous assessment of prediction uncertainties intractable.

The coastal, geomorphological environments, in which seawater intrusion modelling exercises are often framed, inherently possess heterogeneities at various scales. These geological variations are also associated with variations in hydraulic properties of the porous medium. The effects of these heterogeneities are multifaceted and may particularly influence the size and shape of the so-called "mixing zone" of the freshwater-seawater interface as well as create preferential pathways for the transport of seawater born salts within the aquifer leading to freshwater-seawater interfaces that are of complex shape (see *Werner et al.*, 2013, for further discussions on these topics). Where geological heterogeneities are known to exist, observations of salinity and hydraulic head will likely be influenced by their presence. A seawater intrusion model must then be provided with the means by which to represent these geological variations if it is to be able to adequately reproduce calibration observations. The use of many parameters in a model allows for expression of hydraulic property variability thereby allowing for the possibility of attaining a good fit with observations when the model is subjected to calibration while at the same time providing some protection against the deleterious effects of calibration induced parameter bias. As *Doherty and Welter* (2010) explain, the propensity for adjustable parameters to take on compensatory roles is increased when parameters are omitted from the model during calibration, as information contained within the calibration dataset may have "no place to go". Instead those parameters that are included may be forced to take erroneous values in order to soak up some of the model to measurement misfit. If predictions required of the model are sensitive to those parameters, then bias in predictions will be the outcome. Perhaps more importantly though is that inclusion of many parameters allows for rigorous assessment of model potential predictive error that includes the effects of fine scale geological variability. Hence, the risk of underestimating the potential for predictive error is mitigated especially where model predictions are actually sensitive to this fine scale detail.

Unfortunately, for reasons already highlighted above, highly parameterised calibration and rigorous uncertainty assessments of seawater intrusion models in heterogeneous environments is rarely performed. This has been noted by *Werner et al.* (2013) as one of the future challenges in management of coastal freshwater resources. One exception is *Herckenrath et al.* (2011), who in acknowledgement of

the absence of formal analyses of potential predictive errors in many seawater intrusion problems, adopted the efficient Null-Space Monte Carlo (NSMC) methodology described by *Tonkin et al.* (2009) and implemented in the PEST (*Doherty*, 2015a) suite of software, for predictive uncertainty analysis of a model fashioned on the Henry problem *(Henry*, 1964). The NSMC method offers a much more computationally tractable alternative to posterior parameter and predictive uncertainty assessments, particularly in highly parameterised modelling contexts, than other methods that attempt to sample directly from the posterior parameter distribution such as the Markov-Chain Monte Carlo method, employed by *Mondal et al* .(2010) and *Cui et al.* (2011) for example.

While efficient, NSMC relies heavily on the so-called Jacobian matrix. This matrix contains the sensitivities of all observations to all adjustable parameters which are usually calculated through a finite-differencing process that requires at least one forward model run for each adjustable parameter in the inversion problem. When model runtimes are long and parameter numbers are large, population of the Jacobian matrix can become an extremely computationally intensive task. Throughout the NSMC process, re-population of the Jacobian matrix is required for:

1. Calculation of parameter upgrades during each iteration of the non-linear parameter estimation process of model calibration. Ideally estimates of parameters obtained are without bias. This can be achieved through use of an appropriate regularisation strategy;

2. Definition of inverse problem solution and null spaces on which basis stochastic parameter sets are generated. In delineating these orthogonal subspaces it is usual practice to use the Jacobian matrix calculated from estimated parameters; and

3. Re-adjustment of stochastically generated and null-space projected parameter fields to ensure calibration conditions are respected. This re-adjustment process involves minimal re-calibration of these parameter fields in which only estimable components of parameter sets are adjusted.

The computational burdens experienced in populating the Jacobian matrix many times can severely impede implementation of NSMC analysis of large-scale, highly parameterised seawater intrusion models. Methodologies that can reduce these computational burdens thus making its implementation more tractable, may promote a more ubiquitous reporting of predictive uncertainty in this field.

Deployment of a grid-coarsened, surrogate version of a detailed SEAWAT model, for the purposes of Jacobian matrix population during calibration and predictive uncertainty analysis of the latter model has been demonstrated in Chapter 3 of this text to dramatically reduce these computational costs while not undermining the integrity of the NSMC analyses. The strategy of simple grid-coarsening for simplification, while relatively straightforward, may not always be appropriate. For example, where irregularly spaced pilot point parameters, with variable spatial densities, are used to parameterise differing lithological units, it may be difficult to find a coarser model grid that achieves significant improvements in numerical speed while avoiding inappropriate aggregation of two or more pilot point parameters into a single grid cell. It is demonstrated in Section 2.7 that this type of parameter lumping in the surrogate model can lead to the assignment of spurious values to estimated parameters of the original model.

An alternative strategy for construction of a fast running, surrogate, variable-density flow simulator is to use the Seawater Intrusion package (SWI) for MODFLOW. SWI was developed as a computationally efficient alternative to the runtime expensive simulators of flow and solute transport such as SEAWAT (*Langevin et al.*, 2008) and SUTRA (*Voss and Provost*, 2002). However as a consequence of the assumptions made in its design SWI has inadequacies that limit its use to regional scale simulation of these processes. Also the outputs forthcoming from the SWI simulation are very different from those that are available from the use of other variable-density simulators. When more detailed, accurate simulation of the position of the seawater interface or the evolution of the mixing zone is required, simulation packages such as those just mentioned must be used in preference to SWI. Notwithstanding its limitations, SWI's capability for variable density flow simulation, combined with its numerical speed, make it a prime candidate for deployment as a surrogate in the

conjunctive surrogate/complex calibration methodology. Furthermore, as SWI is a MODFLOW specific package, it can be readily implemented from the Groundwater Flow Process input files used by an existing SEAWAT model. The example that follows explores these concepts along with the practicalities of the implementation of this strategy.

4.1.2 A Brief Description of the SWI package

The Seawater Intrusion package (SWI) for MODFLOW was developed for computationally efficient simulation of regional scale seawater intrusion problems. Documentation of the latest version of the SWI package (SWI2) can be found in *Bakker et al.* (2013). The governing equations developed for SWI are used to represent vertically integrated variable-density groundwater flow in the MODFLOW program. In development of these equations, four approximations are made. These are:

1. The Dupuit approximation is adopted which implies that resistance to flow in the vertical direction within a single aquifer is neglegible (ie. groundwater flow vectors are in the horizontal directions of the model domain only);

2. The mass balance equation is replaced by a volume balance equation in computation of the flow field, and density effects are taken into account only through Darcy's law;

3. Dispersion and diffusion effects are not taken into account; and

4. Density inversions (whereby more dense water overlays less dense water) are not allowed within the same aquifer unit.

Instead of calculating cell-scale, solute concentration distributions throughout the model domain during solution to the variable-density flow problem, as is done in other software packages such as SEAWAT and SUTRA, the SWI package calculates the elevations of the top surface of discrete zones of differing fluid density. These discrete zones of different fluid densities are pre-defined by the user. A schematic of

this scheme is shown in Figure 4.1. With vertical density discretisation achieved in this way and with adoption of the Dupuit approximation, SWI calculates the vertically integrated, horizontal, specific discharge vector below each of the density surfaces, at each horizontal location of the model domain. The vertically integrated specific discharge vector below each surface is expressed in terms of freshwater equivalent head at the top of the surface (calculated internally on the basis of freshwater head at the top of the aquifer corrected for overlying densities zones above the surface) and the transmissivity of the aquifer below the same surface. Continuity of total flow in the aquifer may then be written in terms of head at the top of the aquifer while continuity of flow below each zone surface can be written in terms of the elevations of that surface and the surfaces below. Top of aquifer freshwater head and the elevations of the individual density zone surfaces are the dependant variables at any simulation time-step and therefore constitute the outputs of the SWI package. For presentation of these equations and a more detailed description of their derivation the reader is referred to *Bakker et al.* (2013).

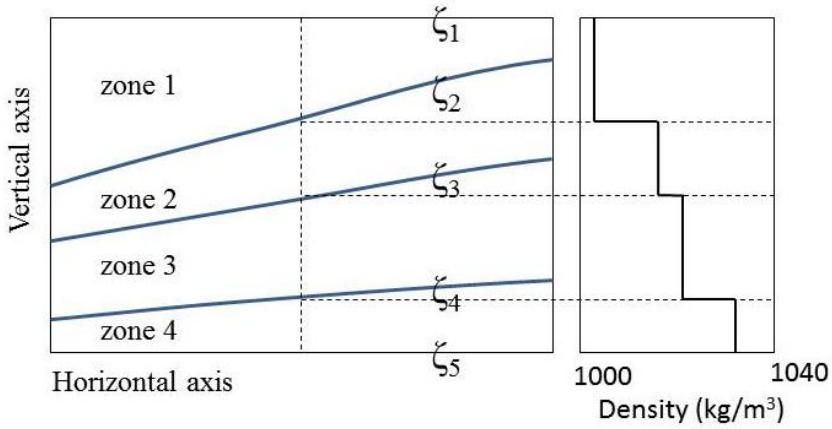

Figure 4.1: Illustration of the vertical density discretisation scheme used by the SWI package for MODFLOW. Blue lines designate the interfaces between zones of differing fluid density. The spatially defined elevations of these surfaces are known as Zetas (ζ).

This approach to the description of variable-density flow allows SWI to simulate the vertical density distribution of the aquifer without the need for many layers. In fact,

an aquifer is modelled by a single layer. Hence, the number of model grid-cells that comprise the SWI model and therefore the number of simultaneous equations requiring solution at each time-step, is vastly reduced compared to that which would be required by the approach taken in SEAWAT for example. Runtimes for the simulation of seawater intrusion problems are therefore drastically decreased by comparison. Furthermore, the SWI approach avoids the need to solve the advection-dispersion equation that proves so problematic to other simulators of variable-density flow. Hence solution of the variable-density flow problem in contexts where hydraulic properties variability is large, is far less computationally expensive and relatively free of numerical difficulties.

The assumptions used in development of SWI are only approximately correct for most coastal areas in which seawater intrusion happens. Therefore its recommended use is for regional scale seawater intrusion assessments. Where topological relief is large, where diffusive processes are not dominated by advection, and where lithological sequences vary in thickness and spatial extents within a single aquifer, the accuracy of SWI may be severely challenged. However, even in such challenging situations, SWI may still be able to calculate observation to parameter relationships that maintain integrity, in spite of its inadequacies as a simulator of more complex processes. A demonstration example follows in which SWI is used for the purpose of calculating observation to parameter sensitivities for population of the Jacobian matrix during calibration of a more physically based seawater intrusion model based on SEAWAT. As is demonstrated in this example, highly parameterised inversion of the original and more accurate SEAWAT model can be successfully achieved in this way but with much less computational cost than would be incurred if the SEAWAT model was used alone.

4.1.3 Example problem description

A synthetic test case was developed in which seawater intrusion into a heterogeneous aquifer is simulated. Figure 4.2 illustrates the example problem schematically in a plan view. The scenario assumed here is one in which we wish to calibrate a complex

flow and transport simulator to observations of the position of the seawater interface taken from a few observation wells. The purpose for calibration is so that the model may then be used as a predictor of the change in position of the interface under altered stress scenarios. As can be seen from Figure 4.2, fresh groundwater flows into the model domain from the western boundary and discharges to the sea on the eastern side. Lateral flow across the northern and southern boundaries is neglected, as are water exchanges through the top and bottom boundaries. Confined aquifer conditions are assumed. As a result of the complex interaction of advective-dispersive processes combined with aquifer heterogeneity, the freshwater-seawater interface that exists is of complex shape, and is best simulated using a distributed, physically based numerical model.

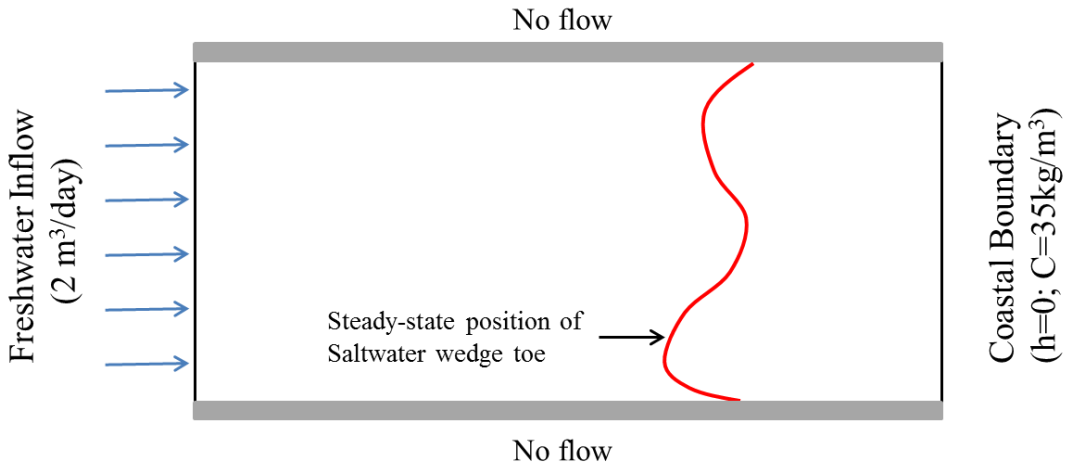

Figure 4.2: PLAN view schematic of the example problem used to demonstrate SWI assisted calibration of a SEAWAT model. The red line notionally indicates the position of the TOE of the freshwater-seawater interface. The interface is therefore inclined outwards from the page and towards the right while originating from the red line.

4.1.3.1 The complex model

SEAWAT is employed to simulate the variable density flow regime and calculate the position of the freshwater-seawater interface. For practical purposes, the freshwater-seawater interface is defined by the 10% seawater isohaline. The synthetic model

domain has dimensions 12m x 2m in the horizontal direction and has a constant thickness of 2m. The domain is discretised into 120 columns x 20 rows horizontally and 20 layers are used to partition the aquifer in the vertical direction so that the distribution of salinity can be modelled. Therefore there are 48,000 cells in the model each with dimensions 0.1m x 0.1m x 0.1m. Advection is simulated using the computationally costly time variation diminishing (TVD) scheme to provide the model with some protection against numerical dispersion and oscillatory effects. Moderate, dispersive mixing effects are introduced through specification of a molecular diffusion coefficient of 0.02 m^2/day. Longitudinal and transverse dispersivities (α^L and α^T respectively) are set to zero. Porosity is uniformly set to 0.35 in all layers. Freshwater inflow at the western boundary is simulated using injection wells with a uniformly distributed injection rate in each model cell that comprises this boundary (ie. 400 wells each injecting 0.005 m^3/day totalling 2 m^3/day inflow). The ocean boundary is simulated using general head boundary cells with reference head of 0.0 m and constant salinity of 35 kg/m^3. For the remainder of this example this model will also be referred to as the "complex" model. Solution of the variable-density flow problem under steady-state flow conditions requires on average 8 minutes model runtime when populated with pilot point parameter values that reflect moderate hydraulic conductivity variability. This runtime significantly increases however with increases in hydraulic conductivity property variability.

Initially, a few random realisations of hydraulic conductivity fields were generated and provided to the SEAWAT model for evaluation. These hydraulic conductivity fields were generated using a sequential Gaussian simulation algorithm engine based on the SGSIM code supplied with the GSLIB geostatistical library (*Deutsch and Journel*, 1998). In generation of these hydraulic conductivity fields a log exponential variogram with variance of 0.5 (in the $\log_{10}$ domain) was used. The range of the variogram is 2.5 m in the longitudinal direction and 1.25 m in the transverse direction. The mean value of log transformed hydraulic conductivity values in these fields is 2.301 m/day (200 m/day prior to $\log_{10}$ transformation). In each initial model evaluation, one of the stochastic hydraulic conductivity fields was provided for all of the 20 model layers. Hence, aquifer heterogeneity prevails in the horizontal directions of the model domain, while within any vertical column of the domain the

aquifer is assumed homogeneous. Following initial evaluations of these hydraulic conductivity fields, one was chosen to be the "reality" hydraulic conductivity field. It was chosen because it results in a complex inclination and shape of the freshwater-seawater interface, thus providing a challenge for the calibration exercise. The field chosen as the "reality" field is shown in Figure 4.3.

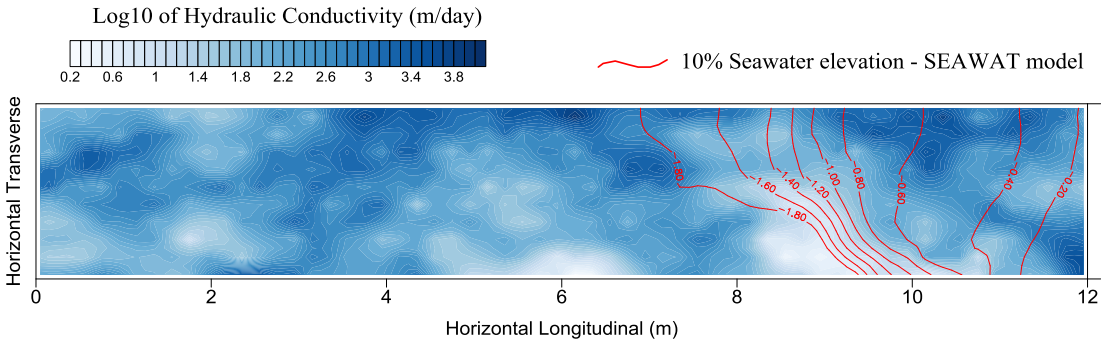

Figure 4.3: Hydraulic conductivity field chosen as the "reality" distribution. This field is used to populate properties for all 20 layers of the "complex" SEAWAT model (ie. aquifer is homogeneous vertically). Also shown is the position of the "reality" freshwater-seawater interface (indicated by the red elevation contours) calculated from SEAWAT concentration outputs. These contours indicate the calculated elevation at which the 10% seawater salinity is intercepted. Noting that the elevation of the model bottom is -2.0m the interface can be seen to rise outward from the page.

Also shown in Figure 4.3 is the position of the freshwater-seawater interface arising from the SEAWAT model when populated with the "reality" conductivity field and run in steady state mode. The lines that indicate the position of the seawater interface are elevation contours of the 10% seawater concentration delimiter that defines the interface. The elevation of the bottom of the model is set constant at -2.0 m. Therefore in Figure 4.3 the interface rises out of the page and is inclined towards the right hand side of the page where groundwater discharges from the system (recall that the coastal boundary is specified with a reference head of 0.0m). Representation

of the seawater interface in this fashion not only allows for a 2D visual depiction of the 3D surface but also provides a necessary conversion to SEAWAT concentration outputs for comparison with SWI output results. As is explained in Section 4.1.2 one of the fundamental differences between SEAWAT and SWI is the manner in which the fluid density distribution is expressed. In SEAWAT it is expressed in terms of cell-scale salinities whereas in SWI it is expressed in terms of surface elevations of predefined discrete density zones. Conversion of one to the other is necessary so that the outputs of both models can be compared to the quantities that are contained in the calibration dataset.

Conversion of SEAWAT calculated concentration outputs to contoured elevations was achieved with the aid of a post-processing software that searches progressively downward through the model layers, in each vertical model column, for a specified concentration value (the 10% seawater concentration in this case). When the specified concentration is found to lie between vertically adjacent cells, the elevation of the specified concentration is calculated by linear interpolation between the elevations of the cell midpoints, weighted by the difference between cell centre concentration and the specified search concentration. In this way a two dimensional array is computed whereby the value of each element is the calculated elevation of the specified concentration value. This can then be used within a contouring software package for presentation or to provide model generated observation equivalents to field measurements. This post-processor is executed as part of the complex (SEAWAT) model run.

Using the outputs generated by the SEAWAT model, run in steady-state flow mode and populated with the "reality" hydraulic conductivity field, synthetic observations were obtained from hypothetical observation wells. The observation wells are assumed to fully penetrate the aquifer and be slotted/perforated along the entire length of the casing below the aquifer saturation elevation. These types of well constructions are commonly employed in seawater intrusion monitoring programs. If water in the observation well is relatively stagnant, then the distribution of salinity in such a well will be the same as that which surrounds the well. Therefore acquisition of a salinity profile log from such a well of this type is representative of the local

salinity profile of the aquifer. Also measurements of groundwater head taken from such a well construction, coincides with the water table elevation. The location of these hypothetical wells are shown in Figure 4.4. It is assumed that observations of head are obtained from all 21 wells shown in Figure 4.4 and that salinity profiles are acquired from the 15 wells in the eastern half of the domain. From the profiles of salinity, a good measurement of the elevation of the 10% seawater delineator can be made. In terms of the SEAWAT model outputs, these observations are approximately equivalent to freshwater head in the top layer of the model and the elevation of the freshwater-seawater interface obtained from post-processing of concentration outputs. Random normally distributed error was added to these synthetically acquired observations and simulates the potential for error associated with acquiring these measurements. In the case of head measurements, error is characterised by a mean of zero and standard deviation of 0.5% of the total range of head measurements. For freshwater-seawater interface elevation measurements, errors with a mean of zero and standard deviation of 0.01 m were added. These 36 noise-degraded, synthetic measurements comprise the calibration dataset for this example.

For calibration purposes aquifer heterogeneity is characterised by 600 pilot point parameters. The locations of pilot points used for parameterisation of the model are also indicated in Figure 4.4. Each pilot point parameter is adjustable in the calibration exercise. Kriging is used to interpolate pilot point parameter values to the model grid cells. In doing so a variogram identical to that used in generation of initial hydraulic conductivity and "reality" fields is used.

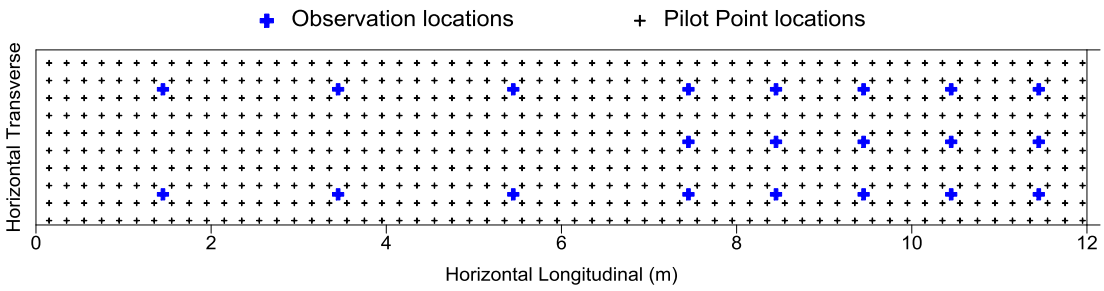

Figure 4.4: Location of observations and pilot points used in the SWI-assisted calibration example problem.

4.1.3.2 The surrogate SWI model

A simplified version of the model described above was constructed using the SWI package for MODFLOW. This model is used as a faster running surrogate for calculation of parameter sensitivities during calibration of the complex model. Construction of the surrogate model requires minimal effort as existing files required for operation of the Groundwater Flow Process of SEAWAT may simply be recycled when the SWI package is implemented. Therefore, horizontal spatial discretisation of this simplified model is identical to the complex model; however in the vertical direction the original 20 layers are amalgamated into one single layer. More precisely, the surrogate model has 20 rows, 120 columns and 1 layer (2400 model cells). A simple illustration of the intrinsic differences between the complex model and the simplified surrogate is provided in Figure 4.5. This model also shares the same pilot point parameters as those defined for the complex model with interpolation of hydraulic conductivity pilot point parameter values to the model grid performed with kriging using the identical variogram as described for the complex model in the preceding section.

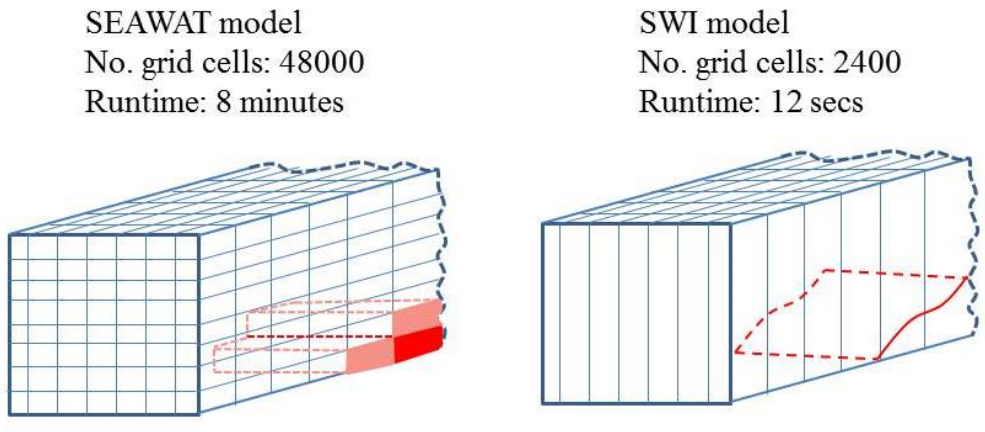

Figure 4.5: An illustration of the intrinsic differences between the complex SEAWAT model and the surrogate MODFLOW-SWI model.

The “stratified” density option for defining the vertical density distribution was selected, with 2 surfaces defining 3 constant density zones of 1000, 1002.5 and 1025 kg/m^3 in each zone; zone 1 representing freshwater, zone 2 representative of the density corresponding to our definition of the seawater interface and zone 3 representing pure seawater. Initial zone surface positions were arbitrarily defined at a position in close proximity to the seawater boundary such that the inclination of these surfaces spanned a few model cells. Tip and Toe tracking parameters were specified in accordance with recommendations in the SWI manual and the dimension of the problem. A constant number of transport time steps was specified, the number of which being heuristically determined from a few runs of SWI with uniformly distributed hydraulic conductivity values equal to the mean value of the hydraulic conductivity fields initially trialled in the complex model (ie. 200 m/d prior to log transformation). It was found that 300 transport time steps was required to ensure solution convergence for steady state simulation, starting from the initial surface positions. This number of transport time steps would have been considerably less had more representative initial surface positions been defined, thus making the simple model run much faster. This SWI based simplification of the computationally costly SEAWAT model takes on average 12 seconds to run on the same PC. This represents a 30 times speedup compared with the complex model. This speedup in runtime is a direct result of the reduced number of model grid-cells along with the simplified algorithm used by SWI in solution of the variable-density flow problem.

As can be seen from Figure 4.6 when populated with the “reality” field the SWI-based surrogate model produces a comparable result to the SEAWAT model however as a result of the assumptions which underpin it’s development some considerable differences in the shape of the freshwater-seawater interface are obvious. Differences are particularly pronounced in the southern part of the interface toe, in proximity of a region where hydraulic conductivity values are relatively low. These differences may be a result of a simplified density distribution used by the SWI model which is manifest as a few discrete zones of different density. They may also be a result of SWI’s neglect of diffusion combined with localised vertical components of flow which lead to a flushing of seawater from the aquifer. Regardless of the nature and cause of these differences, deficiencies in the SWI

model's ability to replicate the more physically based simulator are to be expected. What is important though is that the SWI-based surrogate model demonstrates comparatively good performance and with much reduced computational time.

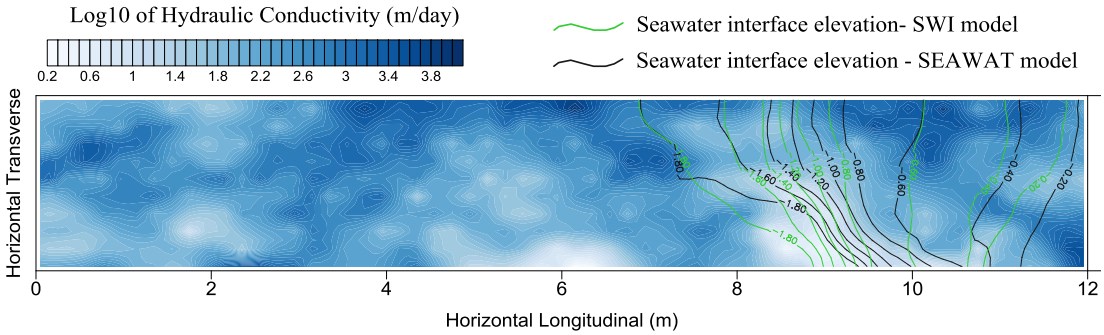

Figure 4.6: Comparison of seawater interface elevation contours arising from SEAWAT model (black lines) and from the SWI version (green lines), when populated with the "reality" hydraulic conductivity field which is shown in blue gradation.

However the inability of the SWI model to exactly replicate outputs of the SEAWAT model does not necessarily preclude its use as a proxy in calculation of derivatives for use in calibration of the latter. When derivatives are approximated using finite-differencing methods, the differences in model generated equivalents to calibration data are taken and divided by the increment of the parameter varied. Hence model structural error as shown in Figure 4.6 tends to cancel out in these calculations. The surrogate model may then at least be capable of calculation of a Jacobian matrix that supports calibration of the original complex model but with much reduced computational costs. This is now tested.

4.1.4 SWI assisted calibration of the SEAWAT model

Since many parameters have been employed for characterisation of aquifer heterogeneity (recall that 600 pilot point parameters are used for this purpose) and observations of system state are sparse (36 in total) while also contaminated with noise (as is inevitable in practice), calibration of the complex model constitutes an

ill-posed inverse problem. In spite of this, a unique solution can still be found through inclusion of regularisation in the calibration process. With use of a mathematical regularisation strategy that is appropriate for the problem, a unique solution can be found that will be without bias. Subsequent predictions made by the model when populated with this parameter field will also be without bias but may be considerably in error nonetheless. As a result of data scarcity, the estimated parameter field will necessarily lack detail. However, it can be used as a good centre-point for quantification of the model's potential for predictive error whereby calibration-constrained Monte Carlo analysis is employed as a computationally feasible proxy for posterior parameter and predictive uncertainty quantification. See *Doherty* (2015b) for discussions of these topics.

In the example presented herein calibration of the SEAWAT model is accomplished using the surrogate-assisted methodology described in Section 2.3.2 and now implemented in the PEST suite of software. As is the process in this surrogate-assisted methodology, the SWI-based surrogate model is run on those occasions when finite-difference approximations to derivatives are required, for population of the Jacobian matrix. At each iteration of the non-linear parameter estimation process, a parameter upgrade vector was calculated on the basis of the Jacobian matrix calculated by the surrogate model. A few such parameter upgrade vectors are calculated each time by adjustment of the Marquardt parameter which features in the GML method. In this way, the possibility that different directions in parameter space are better than others at that part of the estimation process is explored. This also atones to some extent for the fact that derivatives calculated by the surrogate model may lack integrity. When these parameter upgrades are tested for possible improvement to the objective function, the complex SEAWAT model is run, thereby maintaining the integrity of calibration dataset to model output equivalent residuals.

In undertaking calibration, Tikhonov constraints were applied to the 600 adjustable pilot point parameters whereby each pilot point is assigned a "preferred value" equal to 200 m/day, which is the mean of the "reality" hydraulic conductivity field; these also constitute initial values assigned to pilot point parameters. These "preferred value" observations were included in the inversion problem as a set of prior

information equations. Deviations in estimated parameters from these "preferred values" were weighted in accordance with innate hydraulic conductivity field variability described by the variogram used to generate the "reality" field. The squares of these weighted parameter deviations were collected in a so-called "regularisation objective function" whose task it is for PEST to minimise while at the same time ensuring that calibration observations to model generated equivalents misfit does not exceed a specified value. This value is known as the "target measurement objective function" and is normally set slightly higher than that which would be anticipated with regard to measurement noise. This value is often difficult to define as in most practical applications as so-called "structural noise" normally dominates model to measurement misfit; the magnitude and covariance of structural noise unknowable. In this example however, structural noise is negligible and so a value for the "target measurement objective function" was set at 40 which is approximately 10% higher than the expected value of 36 (the number of observations) assuming that measurement noise is normally distributed and that the weights applied to residuals are equal to the inverse of the standard deviation of this noise (noise introduced to the calibration dataset was described in Section 4.1.3.1).

Calibration of the SEAWAT model using the surrogate-assisted methodology in which SWI was used for derivatives calculations achieved a minimised measurement objective function of 311 in 17 iterations, requiring 55 computationally costly SEAWAT model runs and 15,670 much faster SWI runs. Following the initial 5 iterations in which model to measurement misfit rapidly reduced to a value an order of magnitude lower than the initial value calculated using initial parameter values, the estimation process proceeded much slower with model to measurement misfit rising in some iterations before falling again in other iterations. This indicates that derivatives being calculated by the surrogate model may have reduced integrity and that parameter upgrades calculated on their basis also lack integrity. In spite of this, the surrogate-assisted methodology was still able to reduce model to measurement misfit by another order of magnitude. However total runtime efficiencies were somewhat degraded by the necessity to perform several more iterations than would be expected if derivatives were of better quality.

The surrogate-assisted, calibrated hydraulic conductivity field is shown in Figure 4.7. As is to be expected, this field lacks the parameterisation detail of the reality field shown in Figure 4.6, however this is the cost of seeking a unique solution to an ill-posed inverse problem where data is limited and contaminated by noise (see *Moore and Doherty*, 2005). This hydraulic conductivity field does however represent the minimum divergence from prior mean parameter values, that is required to attain a good fit between elements of the calibration dataset and model generated equivalents. The broad scale parameterisation features that this field possesses are consistent with the broad scale features of the "reality" field depicted in Figure 4.6. Model generated freshwater-seawater interface elevation contours arising from the calibrated field are also compared with "reality" elevation contours in Figure 4.7. The match is subjectively quite good in spite of a slightly higher than anticipated measurement objective function.

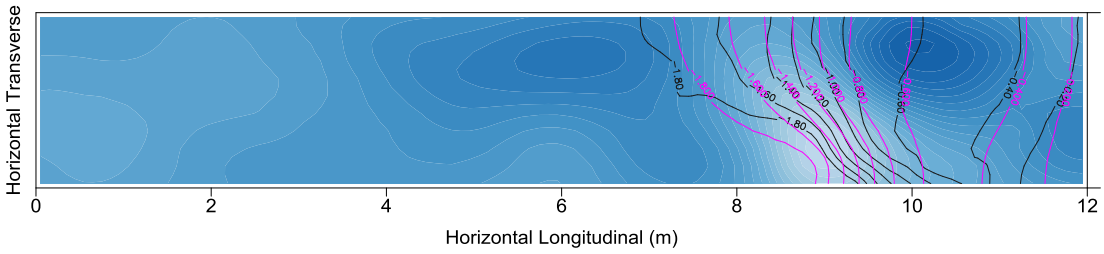

Figure 4.7: Calibrated hydraulic conductivity field arising from the "surrogate-assisted" calibration process (see Figure 4.6 for graduation scale). SEAWAT calculated freshwater-seawater interface elevation contours resulting from this calibrated field are shown as pink lines. Interface elevations from the "reality" case are shown as black lines.

For comparative purposes, a calibration exercise was conducted in which only the complex model was used. In that process, which is designated the "baseline" calibration, the complex model was used for both the runs required for Jacobian matrix population and for the testing of parameter upgrades during each iteration. That process achieved calibration in only 5 iterations but required a total of 3029

computationally costly complex model runs that results in an overall calibration time that is almost 7 times longer than that achieved with the surrogate-assisted approach. A summary of the calibration attempts model run requirements is presented in Table 4.1. The calibrated hydraulic conductivity field arising from this "baseline" calibration process is shown in Figure 4.8. It can be seen through comparison with Figure 4.7 that these calibrated hydraulic conductivity fields are not identical however they are both very similar and are reasonable expressions of the broad scale parameterisation features of the "reality" field. Also, comparatively the seawater interface elevation contours arising from both calibration attempts differ only slightly and visually depict a similar level of fit to contours of "reality" seawater interface position.

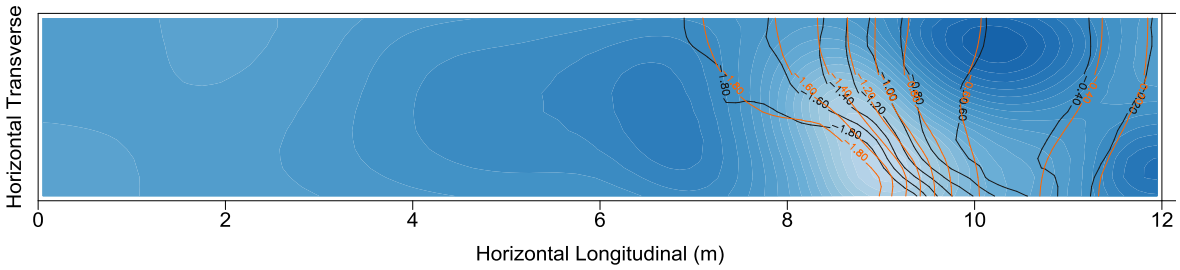

Figure 4.8: Calibrated hydraulic conductivity field arising from the "baseline" calibration process (see Figure 4.6 for graduation scale). SEAWAT calculated seawater interface elevation contours resulting from this field are shown as orange lines. Seawater interface elevations from the reality case are shown as black lines.

Importantly, both of the above calibration attempts result in estimated parameter fields that deviate from a prior expected condition by the minimum amount that is required to attain a good fit with calibration observations; this is demonstrated further in Figure 4.9. In this illustration, estimated pilot point parameter values are plotted in a similar fashion to that used in Section 2.7 to compare estimated parameters. Pilot point parameters are plotted sequentially from left to right and row by row beginning at the top (with reference to Figure 4.4). A sample of the reality parameter field was taken in which pilot point parameter values are calculated such

that the square of the differences between pilot point interpolated cell values and the reality field cell values is minimised. These parameter values represent the maximum amount of parameter detail that could be inferred if all parameters were uniquely estimable and the vastly expanded calibration dataset that would be required was without noise. Of course, in this example all parameters are not uniquely estimable and errors do attend the limited calibration dataset. Figure 4.9 shows that as expected, parameters values arising from the "baseline" calibration depict a smoothed representation of the "reality" sample with some small differences visible due to parameter estimation using a limited, noise degraded calibration dataset. Parameter values arising from the SWI-assisted calibration process are also shown in Figure 4.9 and clearly closely match "baseline" values. This demonstrates that the SWI-assisted calibration process has attained a similarly un-biased estimation of parameters as when only the complex model is employed, but with nearly an order of magnitude reduction in computational cost.

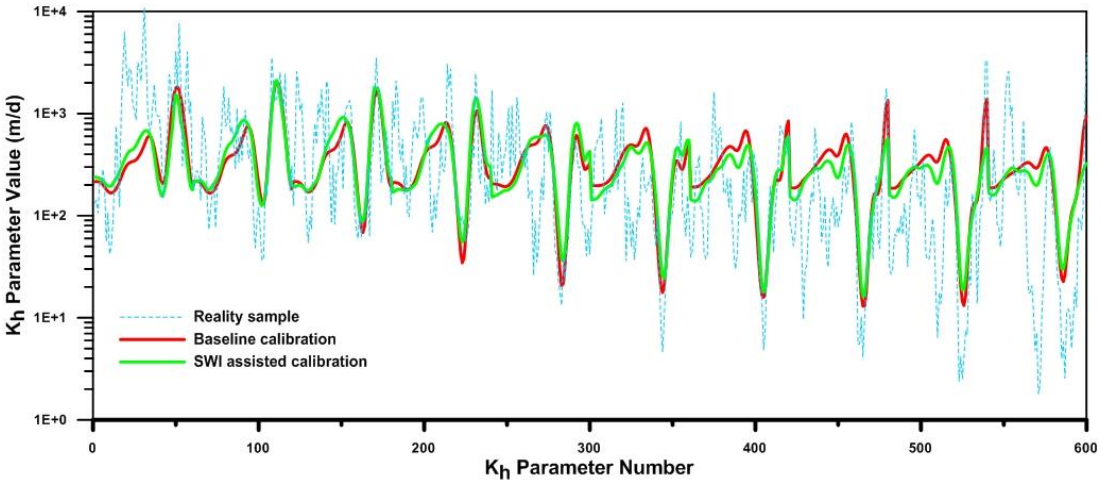

Figure 4.9: Comparison of pilot point parameter values estimated from the SWI-assisted and Baseline calibration attempts. Pilot point parameters are numbered sequentially from left to right, continuing for each row of parameters.

Table 4.1: Summary model run requirements for example SWI-based surrogate-assisted calibration problem.

Calibration process	Initial measurement objective function	Final measurement objective function	Iterations required	Complex model runs	Surrogate model runs	Equivalent total runtime on single processor (hours)	Calibration Speed-up (factor)
Baseline (SEAWAT only)	30155	44	5	3029	-	403.9	-
Surrogate-assisted	30155	311	17	55	15670	59.6	6.8

A third calibration attempt was performed wherein only the surrogate SWI model was used within the inversion process. The outcome of this calibration attempt is shown in Figure 4.10. This calibration attempt was unable to reduce model to measurement misfit to near the same level as the previous two calibration exercises. Visually the differences in the position of the freshwater-seawater interface estimated using only the SWI model compared to the "reality" position are much more obvious. What is of greater concern is that the calibrated hydraulic conductivity field achieved displays inconsistencies when compared to the "reality" hydraulic conductivity field. Notwithstanding its smoothed appearance which is an unavoidable consequence of the inversion process, bias in some estimated parameter values is becoming evident. This bias is a consequence of some parameters taking on compensatory roles in an attempt to atone for deficiencies in the simplified algorithm employed by the surrogate model. In the present case these deficiencies largely arise from SWI's neglect of diffusion/dispersive effects and the assumptions of horizontal flow which are used in its development. Parameter compensations are more obvious when compared with the "baseline" calibration values. In Figure 4.11 it can be seen that some estimable parameters are tending towards values that exceed pre-calibration variability.

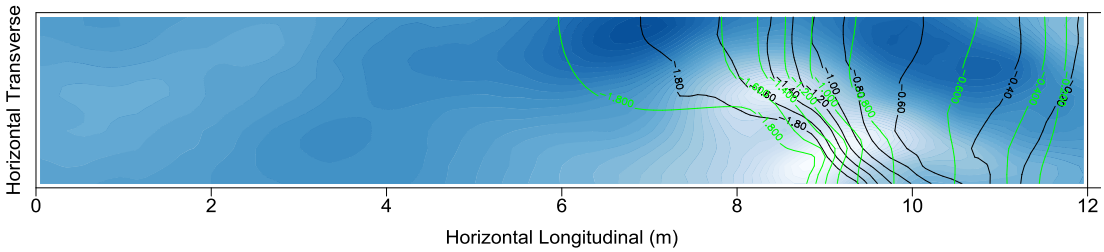

Figure 4.10: Calibrated hydraulic conductivity field arising from calibration using only the MODFLOW-SWI model (see Figure 4.6 for graduation scale). SWI calculated seawater interface elevation contours resulting from this field are shown as green lines. Seawater interface elevations from the reality case are shown as black lines.

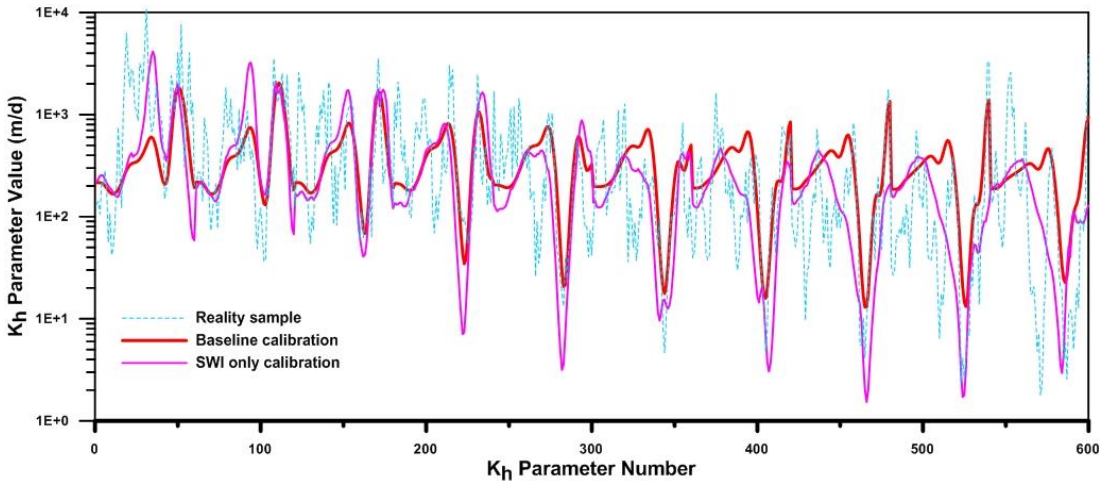

Figure 4.11: Comparison of pilot point parameter values estimated using the SWI model alone and Baseline calibration attempts. Pilot point parameters are arranged as they are in Figure 4.9.

The effects of these parameter biases become more profound when the parameter field estimated using the SWI model alone is provided to the original complex model. This outcome is shown in Figure 4.12 in which it can be seen that the ability of the complex model to reproduce the "reality" interface position is severely impaired. Consequently the complex model's ability to make predictions that are without bias

is also impaired. Furthermore use of this hydraulic conductivity field as the centre piece about which stochastic fields are generated for assessment of parameter and predictive uncertainty may compromise that analysis.

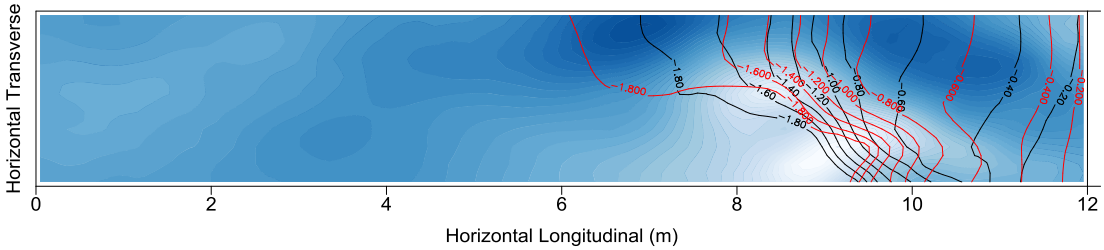

Figure 4.12: SEAWAT (complex) model generated freshwater-seawater interface (red contour lines) arising from the hydraulic conductivity field estimated using SWI model only. The "reality" interface position is shown as black contours.

4.1.5 Discussion

The example just described shows an original and unique application of the SWI package for MODFLOW. It is demonstrated that when used as a faster running surrogate model in calculation of finite-difference approximated derivatives that the Jacobian matrix populated using this surrogate model retains sufficient integrity to enable calibration of a more physically detailed, distributed parameter simulator. Almost an order of magnitude reduction in total computational costs during calibration are realised through application of this method. In this example SEAWAT was used as the accurate simulator in a seawater intrusion problem that requires accuracy in representation of the seawater interface. However the concepts presented here are just as applicable to other contexts in which sophisticated simulation software are employed to replicate physical process detail, that use many parameters in distribution of physical properties and requires matching of model outputs with experimentally acquired observations.

Although not conducted in this experiment, similar reductions in computational costs are to be expected when calibration-constrained Monte Carlo analysis is performed for quantification of predictive uncertainty using stochastic fields that are based on the calibrated field. NSMC can be used for this purpose.

It is readily acknowledged that the synthetic example presented here is relatively simple in its design when compared to more real-world applications of seawater intrusion modelling. In particular, the example problem assumes vertical homogeneity of the aquifer in spite of the presence of heterogeneity in the horizontal direction. The main reason for this is that in calculation of specific discharge vectors below the surfaces of zones defining the density distribution, SWI uses vertically integrated transmissivity calculated from the hydraulic conductivity value assigned to a particular cell and the elevation of the pertinent surface. The elevations of all surfaces are then simultaneously calculated using this in-variable hydraulic conductivity value. While it would be not too difficult to calculate a vertically averaged hydraulic conductivity from a SEAWAT model that uses many layers to represent vertical heterogeneity prior to supplying this to the SWI model, this involves a form of parameter lumping in which the relative contributions of each relevant parameter and hence the relative sensitivities of those parameters becomes blurred. As is demonstrated in Section 2.7, this situation can lead to the estimation of spurious parameter values. Perhaps a better approach would be to enlist the multiple aquifer capabilities of SWI to calculate sensitivities of adjustable parameters based on their vertical disposition within the aquifer. This would avoid parameter aggregations but likely introduce other nuances of a practical nature. In any case, in this original demonstration example, vertical homogeneity was assumed to avoid these complications.

4.2 Using MODFLOW-USG as a Surrogate for Efficient Calibration of a Finely Discretised MF2K model

This section introduces the prospect of using an unstructured grid coarsening strategy as a means of constructing a much faster running surrogate model for use in calculation of derivatives during the process of model calibration and calibration-constrained Monte Carlo analysis. The strategy used here relies on localised grid refinement, while the grid is quite coarse elsewhere. A surrogate model constructed in this way concedes numerical inadequacies over large portions of its domain, however within the region of grid refinement it performs quite well. The integrity of derivatives calculated by this surrogate model may only be reliable for observation-parameter pairs that are within the extent of grid refinement. Therefore a series of surrogate models must be used, each of which is refined in different regions of the model domain, each of which is very fast running, and collectively they are able to populate a Jacobian matrix that can support calibration of the original model. In this way dramatic reductions in the computational costs associated with the processes of highly parameterised inversion and calibration-constrained uncertainty analysis can be realised.

4.2.1 Introduction

In Chapter 2 a much faster running, simplified version of an existing complex model was constructed so that it could be used as a surrogate to the original model during the many model runs required for finite-difference calculations of derivatives during the inversion process. In that example, the surrogate model was fashioned from the original model by using a simple, uniform grid coarsening strategy that led to a reduced number of model grid cells resulting in reduced surrogate model runtimes. Surrogate model runtime savings were greatly enhanced by choice of a much faster solution scheme for the solute transport part of the problem, than the computationally costly solution scheme used by the original simulator. In large, regional scale

groundwater modelling problems that involve solution of flow only such a solution scheme substitution may not be a possible. Hence significant runtime savings may only be attainable through gross grid coarsening. It was also shown in Section 2.7 that when the surrogate-assisted inversion approach is adopted and where the surrogate model uses an overly coarsened grid, the modeller risks incurring bias in estimated parameters while also needing to accept an increased level of model to measurement misfit from the calibration exercise. An alternative grid coarsening strategy for construction of a surrogate model is now presented that offers some protection from the effects of overly coarse grid structures while still providing for a much faster running surrogate model that is capable of calculating finite-difference derivatives with the same integrity as those that would be calculated by the original model itself. Highly parameterised inversion and calibration-constrained uncertainty analysis may then proceed with greatly improved computational efficiency.

In 2013, the US Geological Survey released the initial version of MODFLOW-USG (*Panday et al.*, 2013); since then it has become popular amongst groundwater modelling practitioners. The "USG" suffix stands for UnStructured Grid. This version of MODFLOW supports a wide variety of structured and unstructured grid types that can be used to better represent the shape of features in the model domain and/or focus resolution around these features, or other areas of interest. It is based on an underlying control volume finite difference (CVFD) formulation for groundwater flow in which a model grid cell can be connected to an arbitrary number of adjacent cells. This allows for localised grid refinement and/or grid coarsening to be used in various parts of the model domain without necessarily extending these grid refinements to the model boundary as is required in standard MODFLOW. As a result groundwater flow models can be constructed with MODFLOW-USG that have greatly reduced numbers of grid cells and subsequently greatly reduced model runtimes compared to the standard version, while maintaining respect for resolution in those areas of the study site where it is most warranted.

In deciding to use an unstructured grid with variable coarsening, the modeller chooses to provide the model with maximum numerical accuracy and parameterisation detail in some areas while accepting reduced numerical accuracy

and loss of parameterisation detail in other areas. When a model is developed for decision making purposes in which numerical accuracy and variability of distributed hydraulic properties must be maximised over the entire extent of the model, this can only be supported by use of fine-scale discretisation throughout.

The unstructured grid capabilities of MODFLOW-USG may also be exploited to construct a series of much faster running versions of the original model, that are uniquely refined in small areas of the study site and very coarse elsewhere. Together, the refined areas of these models collectively span the original model's grid extent with at least nearly the same resolution. Individually these models are able to closely reproduce original model outputs in those smaller areas of the model domain where they are refined, while for the much larger proportion of the domain their outputs will likely be degraded by comparison. Theoretically, these models may then individually be used to calculate reliable derivatives for observation-parameter pairs that lie within their respective refinement areas. Derivatives of observation-parameter pairs that transgress refinement areas are expected to have reduced integrity. However this may be of little consequence to the inversion process. It is generally the case in a physical-based, distributed model that model generated observations in close proximity to a particular parameter will be more sensitive to variations in that parameter than they will be to parameters that are much farther away. When regularisation is used to obtain a unique solution to an ill-posed inverse problem, insensitive observation-parameter combinations are supressed or omitted from calculation of parameter improvements. The population of a Jacobian matrix that retains enough integrity to support highly parameterised calibration of the original model may then be achieved but at far less computational cost.

A synthetic example calibration problem follows in which an original implementation of this strategy is demonstrated.

4.2.2 Example problem description

The context in which this hypothetical, synthetic example problem is framed is a geological setting wherein two separate aquifers are being used for the production of fresh groundwater. Notionally the model has been developed to assess the potential

ramifications of increasing production rates in some of these production bores, either in terms of reductions in groundwater discharges to the natural environment or in terms of interferences with other production wells. Innate variability of distributed hydraulic properties within the aquifers is likely to play a key role in these outcomes. In order for the model to be used as a predictor of future outcomes, it must first be calibrated. With attainment of calibrated hydraulic property fields that are without bias (this being promulgated through the use of mathematical regularisation) these fields can then be used as a good starting point for quantification of uncertainties associated with predictions made by the model. However, because of the use of a large number of parameters employed to characterise geological heterogeneity and long model runtimes that often accompany transient, multi-layered, regional scale models of this type, calibration becomes a computationally demanding task. So does calibration-constrained Monte Carlo analysis. The computational burden of these processes can be greatly reduced by using several variably discretised surrogate models for calculation of the derivatives which comprise the Jacobian matrix on which those processes depend.

A schematic conceptualisation of the synthetic modelling problem is shown in Figure 4.13. An unconfined alluvial aquifer forms the top of the model domain. This is underlain by a second alluvial aquifer that has similar hydraulic properties to the top aquifer. As is the case with many alluvial deposits, heterogeneities are known to exist throughout the domain and significantly affect production rates from bores. The two aquifers are separated by a thin, relatively impermeable layer of clay that acts as a semi-confining layer. This semi-confining layer has been explicitly represented in the model as its hydraulic properties are known to also vary considerably over its spatial extent so that water exchanges between the aquifers occurs in some areas but not others. The areas where hydraulic connections between the two aquifers is present, is not well known.

Hydraulic head measurements are assumed available from a number of observation bores located in the domain. At each observation bore site there are in fact two bores, one completed in each of the two aquifers. The locations of these observation bore pairs are shown in Figure 4.14. From these measurements of hydraulic head it has

been assessed that groundwater flows into the model domain from the southern boundary and flows out of the domain at the northern boundary. There are 5 production bores in use; these are also shown in Figure 4.14. These are assumed to fully penetrate the aquifer in which they are completed. The bores labelled P1, P2 and P5 are completed in the upper aquifer while bores P3 and P4 are completed in the lower aquifer. A river traverses the middle of the model domain and is deeply incised into the upper aquifer providing for groundwater-surface water exchanges along its length. It also controls hydraulic heads within that aquifer.

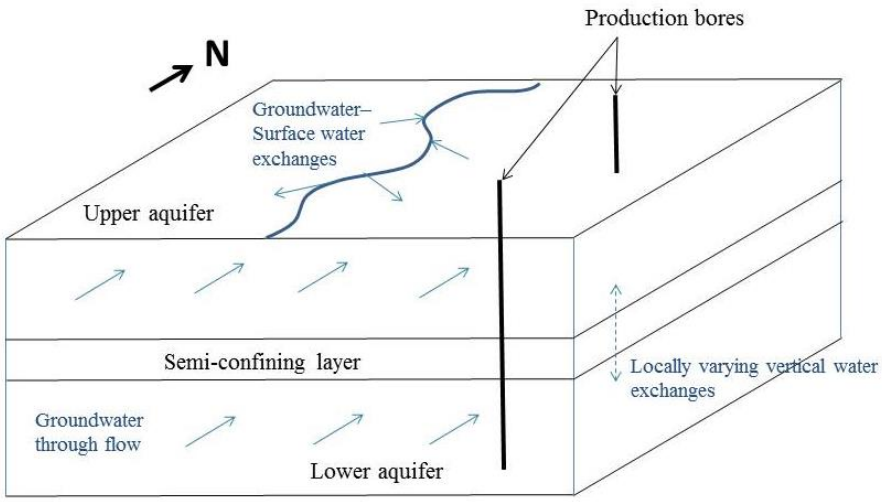

Figure 4.13: Schematic conceptualisation of the synthetic problem used in demonstration of MODFLOW-USG based surrogate-assisted calibration.

4.2.2.1 The complex model

For the purposes of this example, a finely discretised model was constructed in accordance with the above conceptualisation using MODFLOW-2000. This model is referred to henceforth as the "complex model". The model domain is square horizontally with dimensions of 2 km x 2 km discretised into 400 rows with 400 columns. Therefore grid cells have dimensions 5 m x 5 m horizontally. The model has three layers, one for each of the two aquifers and the semi-confining unit which separates them. Hence the model domain consists of 480,000 grid cells. The thickness of both aquifers is set uniformly as 20 m while the semi-confining unit is 5 m thick uniformly. Groundwater inflow at the southern boundary is uniformly

distributed in both aquifers (there is no inflow to the semi-confining unit) and is simulated using wells injecting at a rate of 15 m^3/d each. There is one such injecting well in each model grid cell along this boundary, in each aquifer layer only. Hence there are 800 injecting wells providing total inflow of 12 ML/d to the model domain. The northern boundary of the model domain is defined by constant head cells set to 8 m groundwater head in all three layers. Areal recharge rates are uniformly distributed and constant providing for the addition of 2.4 kL/d to the system via free-drainage. The river is simulated using the MODFLOW River package with the river stage elevation along its length set constant at 10 m below the model surface. It is assumed that the river is deeply incised into the upper aquifer so that the elevation of its bed is 1 m above the bottom of the aquifer. A constant river bed conductance of 2.5 m^2/d is specified along its length.

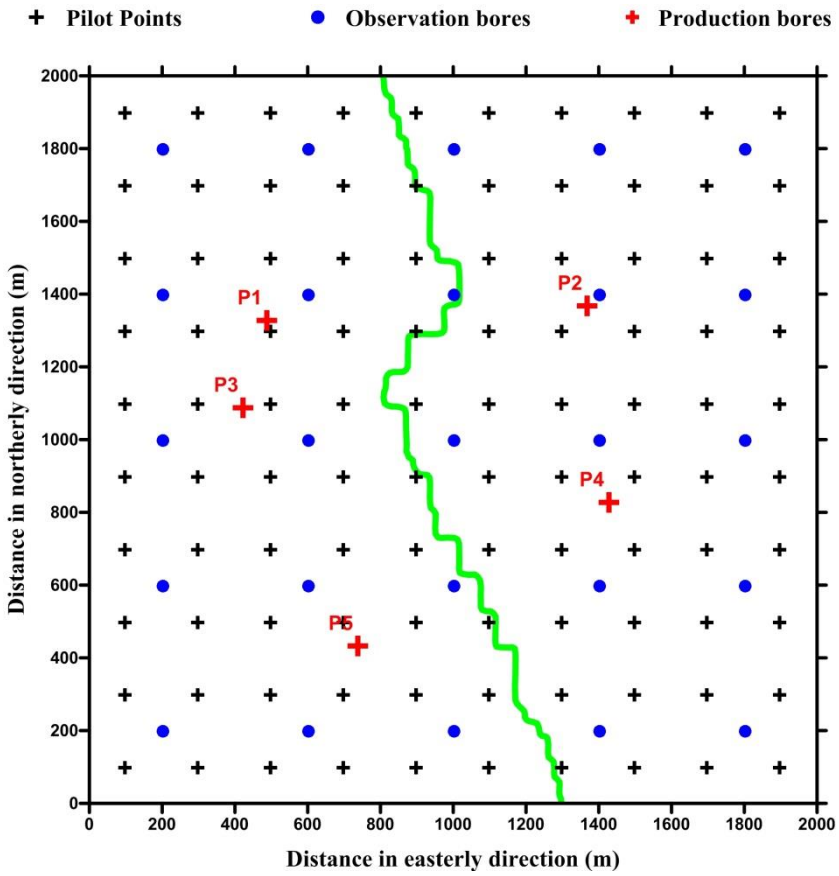

Figure 4.14: Locations of pilot point parameters, observation bores and production bores used in the USG-based surrogate-assisted calibration example. The location of the river is indicated by the green line.

Spatially distributed hydraulic conductivity within each model layer is parameterised using pilot point parameters. Values assigned to these parameters are interpolated to model grid cells using kriging; it employs a variogram that expresses geological variability to the extent that it can be ascertained from prior knowledge of the system. In this synthetic test-case, the variogram used for kriging-based interpolation is the same as that used in generation of the hydraulic conductivity fields that will be used as the "reality" case; this is described shortly. The horizontal locations of these pilot point parameters are shown in Figure 4.14. In each of the three model layers pilot points with the same horizontal coordinates have been placed. There are 100 such pilot point parameters in each model layer; hence there are 300 parameters in total that characterise distributed hydraulic conductivity. All these parameters are adjustable during the course of calibration.

Initially, a few random realisations of hydraulic conductivity fields were generated and provided to the complex model for evaluation. These hydraulic conductivity fields were generated using a sequential Gaussian simulation algorithm engine based on the SGSIM code supplied with the GSLIB geostatistical library (*Deutsch and Journel*, 1998). In generation of these hydraulic conductivity fields a log exponential variogram with variance of 0.25 (in the $\log_{10}$ domain) was used. The range of the variogram is 700 m. The fields are isotropic. The mean value of log transformed hydraulic conductivity values in these fields is 1.3 (20 m/day prior to $\log_{10}$ transformation). Three of these stochastically generated fields were chosen at random to separately populate the three model layers of the complex model and form a scenario that becomes the "reality" case. In the case of the aquifer layers these fields are applied to the model grid cells without modification. However when applied to the semi-confining layer the random field selected was divided by a factor of 100 to obtain hydraulic conductivity values that are commensurate with the reduced conductivity characteristics that this layer is intended to represent. The hydraulic conductivity fields selected as the "reality" fields are shown in Figure 4.15.

In the hypothetical scenario adopted for the purposes of this calibration example, it is assumed that demand for water from the production bores can vary from one 12 month cycle to the next and that a complete set of groundwater head measurements from observation bores is available at the end of each 12 month seasonal cycle over a three year period. In the first year of this period, all production bores operated continuously at an extraction rate of 500 m^3/d. In the subsequent two years of this three year period water demand is higher and the extraction rate is doubled in two of the better producing bores; namely production bores P3 and P4 which produce from the lower aquifer. Hence, simulation of groundwater flow at the study site which spans the timeframe of available observations consists of two stress periods. The first stress period is simulated as steady state while the second two year stress period in which production was increased is simulated as a transient stress period with constant time-stepping intervals equal to one month. A calibration dataset was constructed from model outputs that are equivalent to observation bore measurements, when the model was populated with the "reality" hydraulic conductivity fields described above. Contours of groundwater head arising from this

"reality" simulation at the end of the calibration period are also shown in Figure 4.15. Normally distributed noise with mean value zero and standard deviation equal to 1% of the range of head observation values was added to model generated observation equivalents. These noise degraded head observations comprise the calibration dataset for this exercise. As described above there are 150 observations in total.

Average runtime for the complex model over the time period described and which encompasses the calibration dataset, is 2.4 minutes. While this is not considerable in everyday terms it is not uncommon for regional scale groundwater models with many more layers which simulates time-periods of many tens of years, to have runtimes that are in the order of several hours.

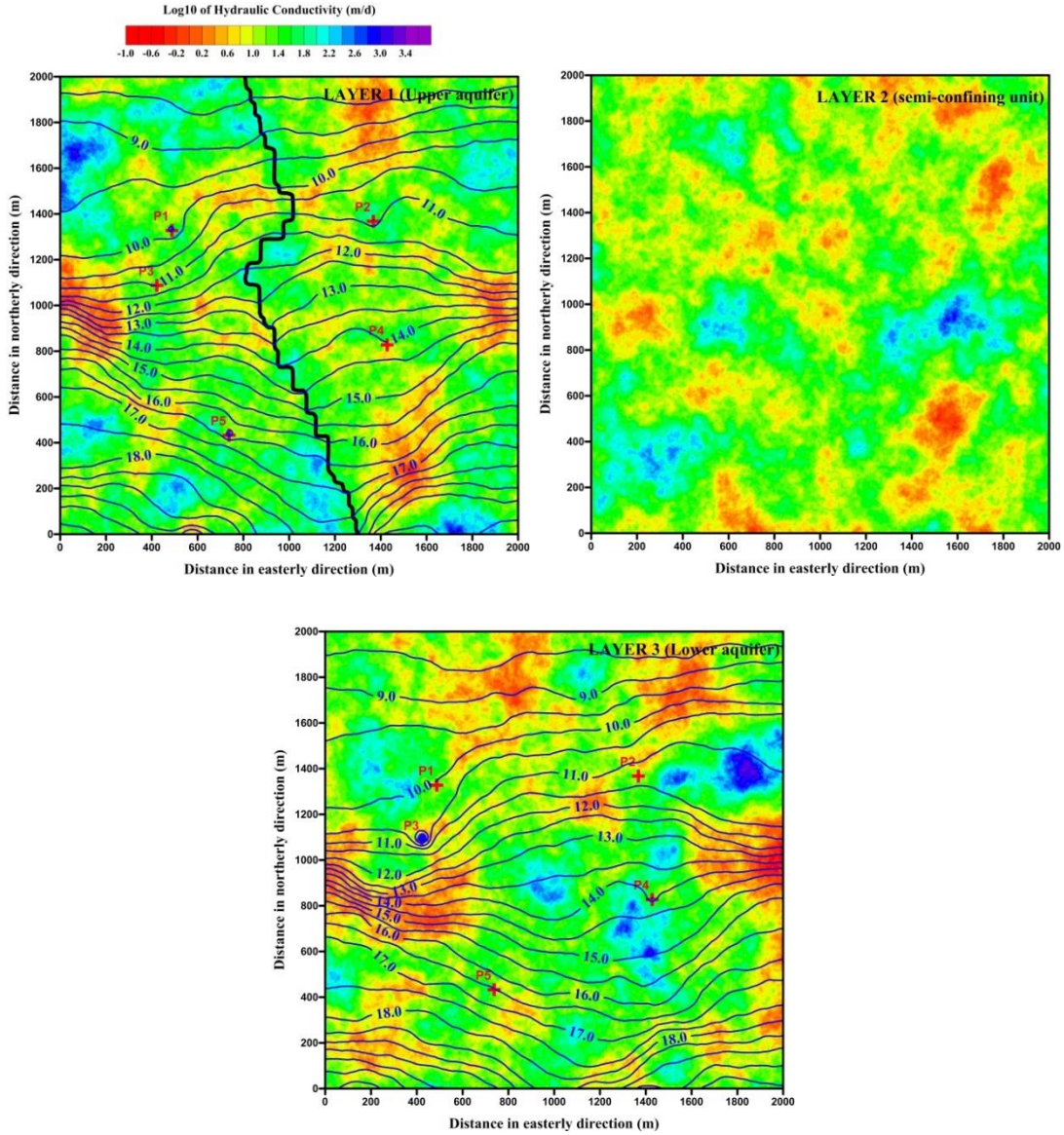

Figure 4.15: Randomly chosen hydraulic conductivity fields used to populate the "reality" model for the MODFLOW-USG-based surrogate-assisted calibration example. Note that the values shown here for layer 2 are divided by 100 when supplied to the model while the other layers are supplied without modification. Contours of hydraulic head arising from these reality fields, at the end of the calibration period, are also shown (blue lines).

4.2.2.2 The MODFLOW-USG surrogate models

Four MODFLOW-USG surrogate models were developed. These models use an unstructured grid configuration that is based on quadtree refinement. Each of these models has grid refinements that are focused on a separate quadrant of the original square model domain. In practice, regional groundwater model domains do not have square geometries. Hence partitioning of regions of focused refinement will be specific to each case and could for example be based on clusters of observation locations. All the surrogate models developed here share the same pilot point parameters that are used by the fine-scale model. These values are interpolated to the model grids using kriging with the same variogram used by the original model. Of course, different grid structures require differing cell-by-cell factors by which interpolation is undertaken. Calculation of these factors for the unstructured surrogate models grids and interpolation of parameter values is performed using the PLPROC parameter list processor (Doherty, 2012). These surrogate models are each capable of generating output equivalents to all members of the calibration dataset. They also reproduce components of water budget that are very similar to the original model.

The grid generation program GRIDGEN (*Lien et al.*, 2014) was utilised to perform the quadtree refinement of the original model grid domain. GRIDGEN begins by reading a three-dimensional base grid, which can have variable row and column widths and spatially variable cell top and bottom elevations. From this base grid, GRIDGEN continuously divides into four any cell that intersects user-provided refinement features (points, lines, and polygons) until a desired level of refinement is reached. In the example that forms the basis of this section, an initial base grid with 200 m cell row widths and 200 m column widths was employed. This is the grid dimension that prevails in the absence of the necessity to sub-divide grid cells due to intersections with defined features. This base grid dimension was chosen as it is the largest cell size that can be used without aggregating pilot point parameters into a single grid cell. GRIDGEN was instructed to perform quadtree refinements down to a level of 4 where feature intersections are present, this resulting in a minimum horizontal grid cell dimension of 12.5 m x 12.5 m. Note that this minimum cell

dimension is still 2.5 times larger (in width terms) than the original fine-scale grid cell dimensions. Refinements were performed within the quadrant to which the surrogate model pertains and around the river feature. In performing the quadtree refinements smoothing in both the horizontal and vertical directions was enforced, to ensured that no two adjacent cells differ by more than 1 level of refinement. An example of the refinement undertaken for the four quadtree grids in this process is shown in Figure 4.16.

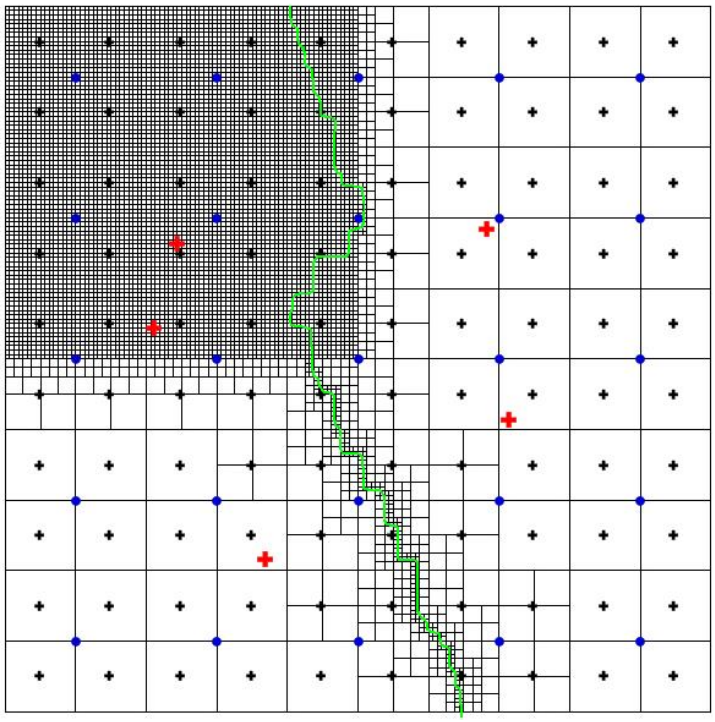

Figure 4.16: Unstructured grid generated for one of the MODFLOW-USG surrogate models. This quadtree structure and associated grid connection details were generated using GRIDGEN. Pilot point parameter locations (black crosses), observation bores (blue-filled circles), production bores (red crosses) and the river (green line) are also indicated.

In Figure 4.17 the vertical grid coarsening of these unstructured grids is illustrated. It is clear that the total number of grid cells in each of the surrogate models has been vastly reduced. Total grid cell numbers of the four coarsened grids range between 20,781 and 21,288 in comparison to the 480,000 grid cells that comprise the fine-

scale, detailed model. GRIDGEN was also used to rapidly generate grid connectivity and cell information required for construction of the MODFLOW-USG input files.

Some modifications to boundary cells of the surrogate models are required to ensure that modelled flow processes are consistent with the original model. For example, values assigned to injection wells used to simulate groundwater inflow to the model domain were modified so that original injection rates are multiplied by the number of original model grid cells that each new cell replaces along this boundary. The other significant modification that was applied to boundary cells was in defining river cell conductance values. This was a little less straight forward than was expected because of the geometry of the river and the inexactness of the overlay between the original fine-scale mesh and the quadtree refined meshes. However an appropriate value for surrogate model river conductance was found in a heuristic fashion by matching volumetric water budgets compared to the original model when homogeneous hydraulic conductivity values were used to populate the model. These settings ensure that each of the surrogate models is able to accurately reproduce original model outputs quite well, at least in their refined quadrant areas.

On average, these surrogate models take approximately 2 seconds to execute a simulation time that spans the calibration period of the original model, which represents a speedup of 72 times. Hence where population of a Jacobian matrix is performed using a two-point stencil for finite-difference approximation of derivatives, the original model requires approximately 12 hours computation time on a single processor, whereas population using the four surrogate models requires only 10 minutes to achieved the same task.

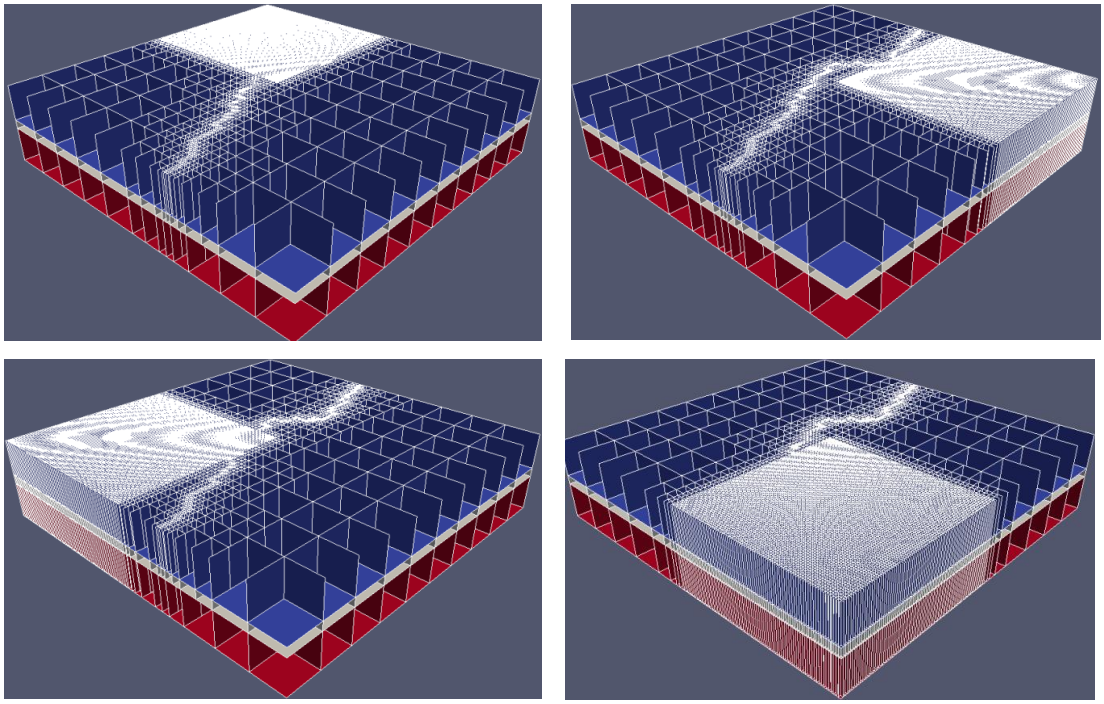

Figure 4.17: 3D images of the four unstructured grids constructed as surrogate models for use in the MODFLOW-USG-based surrogate-assisted calibration exercise. These images also demonstrate the vertical grid coarsening undertaken.

4.2.3 MODFLOW-USG assisted calibration

Surrogate-assisted calibration of the original fine-scale, detailed model is now described. In undertaking this calibration attempt, the four surrogate models developed with quadtree refinement are used selectively for population of the Jacobian matrix, during the iterative process of GML estimation. This is in contrast to all other examples presented in this document where only a single surrogate model was used for this purpose. Finite-difference approximations of derivatives are calculated using a 3 point stencil. Hopefully, this choice facilitates approximation of derivatives with greater integrity than would be obtained from use of the more simple forward-differencing method. In this process each surrogate model is initially run once with the current best estimate of parameters to obtain reference values for observations calculated by each surrogate model. Following this, each parameter is individually increased and then decreased by a pre-defined increment to obtain two further values of each model generated observation equivalent. To each set of three

values of a model generated observation equivalent, a parabola is then fit and the derivative of the observation with respect to the incrementally varied parameter is calculated from that parabola at the current best parameter value. Normally, such a method for derivatives approximations requires at least twice as many model runs than the more simple forward-differencing method. However this potential for increased computational cost is more than offset by the vastly reduced runtimes of the surrogate models.

Selection of one of the four surrogate models for finite-difference approximation of derivatives is done on the basis of the spatial location of each individual pilot point parameter. That is, when a particular pilot point parameter lies within the extent of a particular quadrant upon which quadtree refinement of MODFLOW-USG-based surrogate models have each been focused, then that surrogate model is used for the referencing model run, as well as for the two further runs with parameter increments. Of course, derivatives of all calibration observations with respect to that varied parameter are calculated by that surrogate model. Hopefully, those observations that are close to the varied parameter and therefore likely to be sensitive to that parameter will lie in the refined area of the surrogate model. Hence their derivatives will maintain integrity. For those parameters that are further away from the parameter and therefore in the less accurate part of the surrogate model domain, sensitivities will be much less; less model accuracy is therefore required.

As there are many more parameters than calibration observations in the inversion problem, regularisation is needed to obtain a unique solution to the inverse problem. With appropriate use of mathematical regularisation a unique solution can be found that is without bias. Predictions made by the model using these estimated parameter values will then also be lacking bias. In this calibration exercise, regularisation is implemented through assignment of Tikhonov constraints to each of the 300 pilot point parameters estimated through the process. These are introduced to the inversion problem as a set of prior information equations that assign a “preferred” value to each parameter. This preferred value (which is also the initial value for pilot point parameters) was set to the mean of the hydraulic conductivity fields generated to form the “reality” case; that is 50 m/d prior to log transformation. A covariance

matrix that describes prior parameter variability was constructed from the variogram used in generation of the "reality" fields. This matrix was inverted to obtain weights by which deviations of parameters from these "preferred" values are multiplied. In PEST's implementation of Tikhonov regularisation, an overall multiplication factor for this weight matrix is calculated, subject to the constraint that a user-specified "target measurement objective function" be achieved. This target was set to 150 (the number of observations) which is the expected value of the measurement objective function based on "measurement noise" accompanying the calibration dataset.

Throughout the inversion process the four surrogate models were used for all runs required for derivatives calculations. Meanwhile the original detailed model was used for calculation of the initial objective function and for testing of a few parameter upgrades, this being based on the premise that it is able to produce a better fit between original model outputs and calibration observations. The target measurement objective function was achieved in 11 iterations requiring only 36 detailed model runs while a further 6687 runs necessary for derivatives calculations was performed by the very fast running MODFLOW-USG surrogate models. Model run requirements for this calibration are summarised in Table 4.2. The estimated hydraulic conductivity fields arising from this calibration attempt are shown in Figure 4.18 together with head contours that result at the end of the calibration period. As is expected, these fields are a smoothed representation of the "reality" fields however this is the cost of seeking a unique solution to an ill-posed inverse problem. Hydraulic head contours arising from these fields also match quite well those arising from the reality case in spite of a limited calibration dataset that is also contaminated with noise.

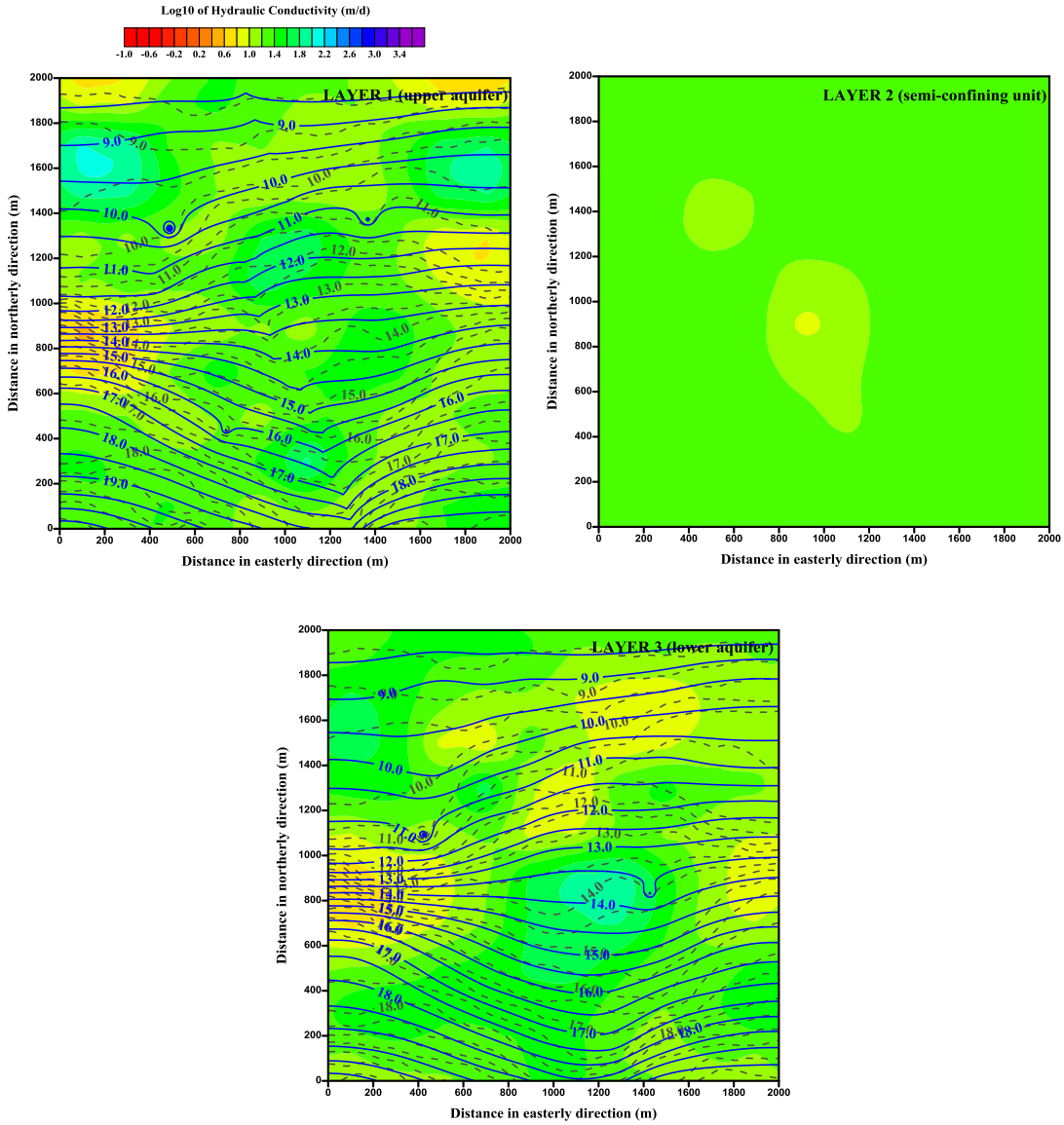

Figure 4.18: Calibrated hydraulic conductivity fields arising from MODFLOW-USG-based surrogate assisted calibration process. Heads contours resulting from these fields are shown as blue lines overlain by the "reality" head contours (dashed lines).

For comparison, a second calibration attempt was performed in which only the original detailed model was employed. That is, the original model was run for all model runs required for calculation of derivatives as well as those required for parameter upgrade testing. This calibration process denoted the "baseline" calibration process henceforth, required 9 iterations and 5461 computationally costly

original model runs to achieve the target measurement objective function. A summary of these model run requirements is also provided in Table 4.2 in which it can be seen that the total computational cost of using the original model alone for this process is approximately 42 times greater than that experienced using the surrogate models for derivatives calculations. The hydraulic conductivity fields arising from this "baseline" calibration attempt are shown in Figure 4.19. As can be seen they are very similar to those arising from calibration using the MODFLOW-USG-based surrogate-assisted methodology.

While these fields lack the parameterisation detail of the reality case, they deviate from preferred mean values only by as much as is necessary to obtain a good fit with the calibration dataset. As such they are without bias and may then be used as the centre piece for generation of many stochastic fields that also calibrate the model but express innate parameter variability. Parameter and therefore predictive uncertainty can therefore be explored, but with much less computational effort than would be required if only the original model was used for this purpose.

Table 4.2: Summary model run requirements for example MODFLOW-USG-based surrogate-assisted calibration problem.

Calibration process	Initial measurement objective function	Optimised measurement objective function	No. iterations	Detailed model runs	Surrogate model runs	Total Equivalent runtime (CPU hours)	Overall Speedup (factor)
Baseline	564,592	150	9	5461	NA	218.44	-
MODFLOW-USG-based Surrogate assisted	564,592	151	11	36	6687	5.16	42.3

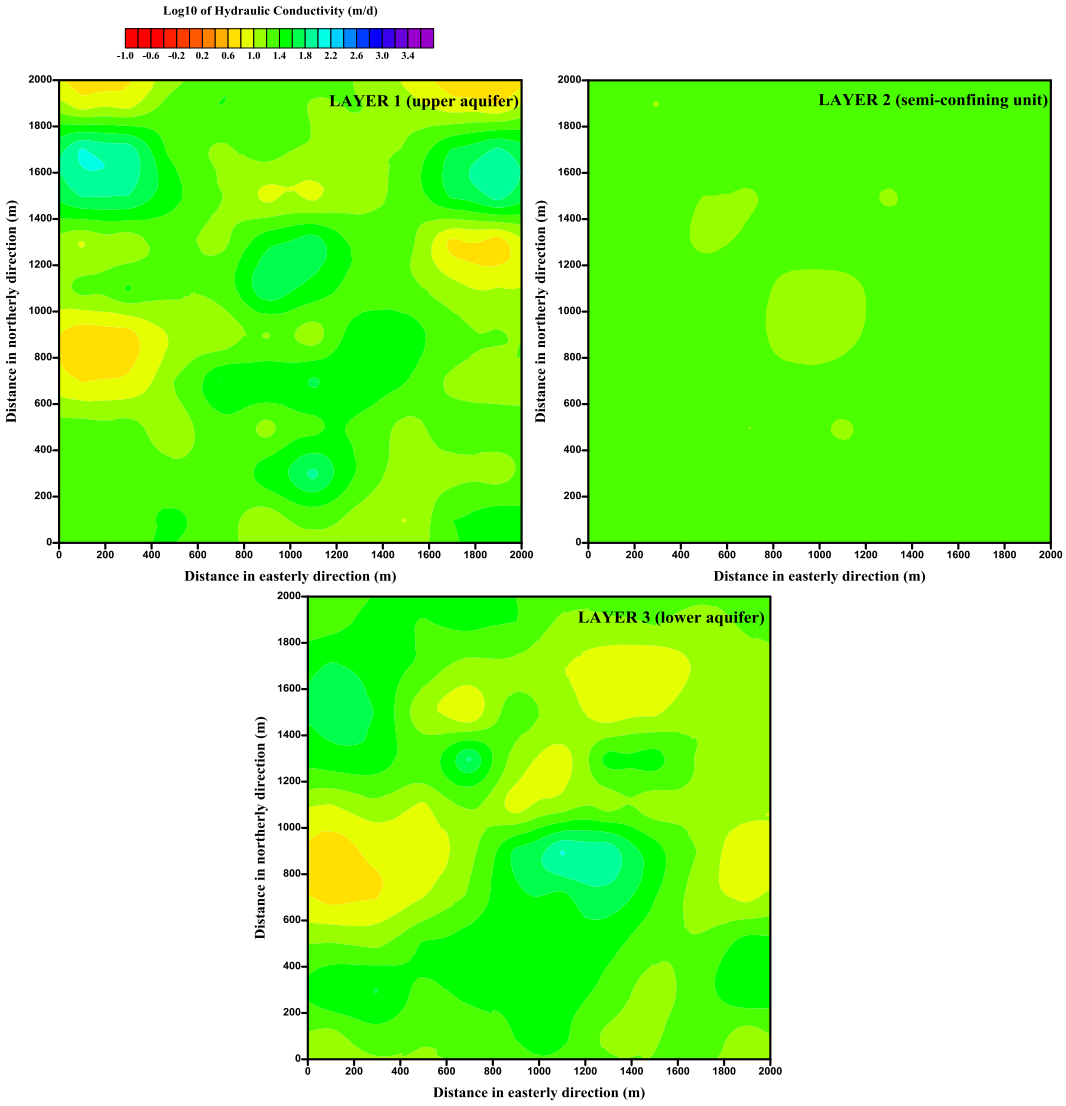

Figure 4.19: Calibrated hydraulic conductivity fields arising from "baseline" calibration process in the MODFLOW-USG surrogate-assisted calibration example problem.

Chapter 5

Conclusions

5.1 General

This thesis has explored the use of proxy and surrogate models in calibration and uncertainty analysis of complex models of subsurface flow. The author considers the work to be novel in the following respects:

- It is the first time that a proxy model has been used in a context wherein parameter adjustment is effected using gradient methods;
- The number of parameters that can be adjusted using surrogate-enabled calibration and uncertainty analysis is far greater than has been documented in any pertinent literature to date;
- The analytical proxy development strategies documented herein which involve the rapid creation of what is effectively a different proxy model for every model output used in the calibration process has not been documented elsewhere;
- The development of multiple surrogate models based on different grid refinement strategies has not been documented elsewhere;
- Nor has the use of a proxy spatial model based on a different conceptualization of three-dimensional, density dependent flow that that employed by the original model.

The advantages of using a surrogate model in the gradient concept have been extensively documented herein. They include the following:

- Gradient methods support the use of many more parameters than can be employed by so-called global methods. This supports calibration and

uncertainty analysis for a broad range of spatial models wherein complex environmental processes are simulated within heterogeneous media;

- Calculation of a Jacobian matrix (at the heart of gradient methods) allows for the use of linear methods for approximate analysis of parameter and predictive uncertainty and for calculation of data worth; and
- It also supports the use of methodologies such as Null Space Monte Carlo which provide the only viable option for exploration of parameter and predictive uncertainty where the dominant contributions to these uncertainties emerge from information deficits within the calibration dataset expressed through the null space.

As models become more complex, and as decision-makers and stakeholders demand that predictions of future environmental behaviour made by models are accompanied by estimates of the uncertainties associated with those predictions, the need for parameterization complexity to match simulation complexity will grow. So too, inexorably, will be the requirement that inversion and uncertainty analysis be based on gradient methods. Paradoxically, the rise in simulation and parameterization complexity will almost certainly be accompanied by longer model run times and by numerical instability in model solvers as they attempt to solve highly nonlinear equations pertaining to highly nonlinear systems. As the impasse between the requirements of decision-makers and the capabilities of numerical tools that can serve the decision-making process grows, it is anticipated that the role of surrogate and proxy models will expand. In fact, it is suggested that not only will their role expand, but that their use will be essential to the next generation of modelling that supports the decision-making process.

The work that is documented in this thesis suggests that there is every reason for optimism that the use of proxy and surrogate models can indeed support the next generation of decision-support simulation. It also indicates however, that the development of a suitable surrogate or proxy model will required careful consideration on each occasion of its deployment. In short, the development of context-specific surrogate and proxy modelling strategies, though essential, is likely

to be as much an art as a science. Conceptually, the use of such strategies will be essential to meeting future modelling needs. Practically, much experience needs to be gained, and the results of that experience documented.

The work documented herein has demonstrated that development and deployment of proxy models must accommodate at least the following considerations:

- A surrogate model must be capable of emulating the response of a real model over the range of parameters that represents possible system properties;
- Ideally, it should be comparatively easy to build; and
- It should run quickly and provide numerical stability.

The first point is particularly important. By definition, a proxy model will not have the ability to replicate the performance of a much more complex model. Hence derivatives that are calculated using the proxy model for the filling of the Jacobian matrix will be somewhat in error. So too, then, will be parameter upgrades calculated using that Jacobian. To some extent, these errors can be accommodated through use of the complex model for testing and selection of upgraded parameters. However the present study shows that there is a limit to the extent to which this will be possible. It also shows that, the better is the fit sought between model outputs and field measurements during the calibration process, the greater is the extent to which parameters will adopt surrogate roles to compensate for imperfect surrogate/proxy model behaviour in the calibration process. It follows, that the use of proxy models must be accompanied by reduced expectations of the level of fit sought through the history matching process. Uncertainties in estimated parameters will therefore be higher than they may otherwise be if a complex model were used for all components of the inversion process. Fortunately, however, through use of surrogate/proxy models, parameter and predictive uncertainty can be quantified but at far reduced computational cost. Though uncertainty may be inflated to some degree through the use of these history-matching-enablers, this is a far better alternative to not being able to quantify uncertainty at all. Furthermore, with the cost of their use taken into

account as part of the uncertainty assessment process, the likelihood of committing a type II statistical error (whereby unwanted events are mistakenly assigned a low probability) is mitigated.

So while the use of surrogate/proxy models comes at a cost, this cost is generally quantifiable as software which depends on their use quantifies model parameter and predictive uncertainty. At the same time, it is anticipated that in complex environmental systems that are often the focus of modelling attention, the "man-made" (and quantifiable) uncertainties induced by use of surrogate/proxy models will be considerably less than those arising from system complexities and information deficits in calibration datasets. For the latter to be quantified, representation of system complexity in a calibration-constrained setting is essential. The present study shows that this can be achieved with surrogate/proxy models in a numerical context based on gradient methods which is readily capable of assimilating such complexity.

5.2 The Future

As stated, the present study documents the first application of proxy models in the gradient context. It suggests that the future of surrogate/proxy-enabled history-matching and calibration-constrained uncertainty analysis is bright. However it is only the first such study and more studies are needed. As well as providing cause for optimism that surrogate enablement in the gradient context is not only possible but essential, the present study can suggest topics that deserve the attention of future research. A few of these are now briefly mentioned.

5.2.1 Inversion Methodologies

In common with surrogate/proxy strategies employed in conjunction with global methods, the strategy outlined herein deploys both a complex model and its surrogate/proxy in a single inversion algorithm. As has been discussed, the simplified model is used for filling the Jacobian while the complex model is used for testing upgrades. In the present context parameter upgrades are calculated using different values of the Marquardt lambda.

Modern computing takes place in highly parallelized contexts - often on a computing cloud such as Amazon of Azure. This provides the opportunity to upgrade parameters in many ways with little cost in wall-time required for testing those upgrades, all based on the same Jacobian matrix. For example, many (and not just a few) values of the Marquardt lambda can be used for calculation of parameter upgrades. Alternatively, or as well, gradient-enhanced global methodologies could be developed for use in calculating parameter upgrades. In this way it may be possible to harness the strengths of both of these approaches in a way that makes maximum use of the approximate Jacobian matrix that surrogate/proxy models can provide.

5.2.2 Surrogate and Proxy Models

The present study has exemplified the use of a number of different simplified models that are used in place of the more complex model in calculation of derivatives. These cover a broad spectrum of strategies, these being:

- Use of a single model based on a coarser grid;
- Use of multiple models based on parameter-specific coarser grids;
- Use of a model which employs an alternative simulation algorithm; and
- Use of a large suite of observation-specific analytical proxies.

Many more options are possible. All will have their strengths and weaknesses. It is possible that in many real-world modelling contexts that more than one surrogate/proxy will be required, with different proxies used for different parameter/observation combinations. Before the modelling community is ready to adopt surrogate/proxy modelling as a standard weapon in their inversion and uncertainty analysis arsenal, much work is needed in developing and testing the use of different surrogate/proxy models for different modelling occasions. Where proxy, rather than surrogate, models are employed, work must be devoted to extending the present functionality of PEST to allow rapid calibration of such proxy models through adjustment of proxy model factors so that quantities calculated by these

simple analytical equations match pertinent complex model outputs, even in highly nonlinear modelling contexts.

5.2.3 Second Stage Parameterization

As was discussed in Chapter 3 of this thesis, while the use of a suite of polynomial proxy models can overcome problems associated with numerical granularity of complex model outputs, the methodology can only support estimation of a relatively small number of parameters (when compared to the hundreds or thousands that can be accommodated with gradient methods in general). It was suggested in that chapter that this problem can be overcome by using a limited number of parameters in the calibration stage, and then introducing a much greater number of "secondary" parameters in the post-calibration uncertainty analysis stage. It was suggested that these can "ride on the back" of the broad-scale parameterization supported by proxy model adjustment; the latter can then be adjusted to maintain calibration constraints following generation of random realisations of the former.

The above strategy relies on the fact that proxy-adjustable parameters span the solution space of the inverse problem. At the same time, it relies on the fact that the secondary set of parameters spans the null space. Ideally, these two spaces are orthogonal. This suggests that it may be possible to design a secondary parameterisation set for use in parameter and predictive uncertainty analysis by using the outcomes of the primary parameter set adjustment process, and accompanying post-calibration covariance matrix. Construction of the secondary parameter set would be such as to maintain orthogonality to the primary set, while respecting the nature of geological variability that is likely to arise in a given simulation setting.

5.2.4 Objective Function Definition

Studies such as that undertaken by *White et al*. (2014) demonstrate the importance of careful formulation of an objective function for minimization during the history matching process. These authors demonstrate that direct matching of measurements with corresponding model outputs can, in many modelling circumstances, expose the history matching process to the deleterious effects of structural noise. This, in turn, can introduce bias to certain parameters, and to predictions which depend on them.

They go on to show that the situation can often be rectified by first processing observations in certain ways (e.g. through spatial, temporal, layer-differencing) and then matching these to corresponding model-generated quantities. Appropriate processing strategies can "orthogonalize out" the components of structural noise that would otherwise damage the inversion process.

As has been discussed above, the use of surrogate and proxy models is seen as essential to the next generation of decision-support modelling. However as has also been discussed above, this can introduce its own type of structural noise. However the work of *White et al*. (2014) suggest that it may not be too difficult to formulate objective functions which provide some degree of immunity from the types of structural noise incurred through surrogate/proxy model usage. An obvious strategy in the salt water intrusion modelling context is the use of seawater wedge lateral locations rather than concentrations in measurement wells; the latter are highly nonlinear and particularly prone to model error whereas the former are much less so. In addition to this, the locations of salt water interfaces are often of more direct relevance to the decision-making process than concentrations in wells.

In other calibration contexts, similar strategies should be sought. At the same time, some guidance must be sought as to appropriate levels of fit to achieve between processed measurements and similarly processed model outcomes as the two are matched through the parameter estimation process.

5.3 Final Remarks

It is hoped that the present study is the first of many. Complex simulation is here to stay. However the use of complex models in the decision making process is often compromised through their inability to deliver the only things that they can promise. These do not include predictive certainty; in contrast they include an ability to quantify lack of predictive certainty after assimilation of as much information as possible emerging from expert knowledge and the historical behaviour of an environmental system. Rectification of this situation will require strategic use not just of complex models, but of appropriately-designed complementary simplified models in conjunction with inversion software whose capabilities are expanded to embrace the use of both of these. It is hoped that the present study contributes to this pursuit.

References

Alam F.M., McNaught, K.R. and Ringrose, T.J., 2004. "A comparison of experimental designs in the development of a neural network simulation model". *Simulation Modelling Practice and Theory*, vol. 12, pp. 559-578.

Asher, M. J., Croke, B. F. W., Jakeman, A. J. and Peeters, L. J. M., 2015. "A review of surrogate models and their application to groundwater modelling". *Water Resources Research*, vol. 51, pp. 5957-5973. doi:10.1002/2015WR016967.

Aster, R.C., Borchers B. and Thurber, C.H., 2013. *Parameter estimation and inverse problems*. Second edition. New York: Academic Press.

Bakker, M., Schaars F., Hughes, J.D., Langevin C.D., and Dausman A.M., 2013. "Documentation of the seawater intrusion (SWI2) package for MODFLOW". *U.S. Geological Survey Techniques and Methods*, book 6, chap. A46, p. 47, http://pubs.usgs.gov/tm/6a46/.

Bard, Y., 1974. *Nonlinear parameter estimation*. New York: Academic Press.

Beven, K. and Binley, A., 1992. "The future of distributed models: Model calibration and uncertainty". *Hydrological Processes*, vol. 6, pp. 279-298.

Beven, K., 2005. "On the concept of structural error". *Water Science and Technology*, vol. 52, no. 6, pp. 167-175

Bevan, K., Smith, P. J. and Freer, J. E., 2008. "So just why would a modeller choose to be incoherent". *Journal of Hydrology*, vol. 354, pp. 15-32.

Bliznyuk, N., Ruppert, D., Shoemaker, C.A., Regis, R., Wild, S. and Mugunthan P., 2007. "Bayesian calibration of computationally expensive models using optimisation

and radial basis function approximation". *Journal of Computational and Graphical Statistics*, vol. 17, no. 2, pp. 270-294.

Borgonovo, E., Castaings, W. and Tarantola, S., 2012. "Model emulation and moment-independent sensitivity analysis: An application to environmental modelling". *Environmental Modelling and Software*, vol. 34, pp. 105-115.

Burrows, W.A. and Doherty, J., 2014. "Efficient Calibration/Uncertainty Analysis Using Paired Complex/Surrogate Models". *Ground Water*, vol. 53, no. 4, pp. 531-541. doi: 10.1111/gwat.12257.

Burrows, W.A. and Doherty, J., 2016. "Gradient-based model calibration with proxy-model assistance". *Journal of Hydrology*, vol. 533, pp. 114-127. doi: 10.1016/j.jhydrol.2015.11.033.

Carrera, J. and Neuman, S.P., 1986a. "Estimation of Aquifer Parameters Under Transient and Steady State Conditions: 1. Maximum Likelihood Method Incorporating Prior Information", *Water Resources Research*, vol. 22, no. 2, pp. 199-210.

Carrera, J. and Neuman, S.P. 1986b, "Estimation of aquifer parameters under transient and steady state conditions: 2. uniqueness, stability, and solution algorithms", *Water Resources Research*, vol. 22, no. 2, pp. 211-227.

Chen, Y. and Oliver, D. S., 2013. "Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification." *Computational Geosciences*, vol. 17, no. 4, pp. 689-703.

Christensen, S. and Cooley, R.L., 1999. "Evaluation of prediction intervals for expressing uncertainties in groundwater flow model predictions", *Water Resources Research*, vol. 35, no. 9, pp. 2627-2639.

Conti, S., Gosling J.P., Oakley, J.E. and O'Hagan, A., 2009. "Gaussian process emulation of dynamic computer codes". *Biometrika*, vol. 96, no. 3, pp. 663-679.

Cui, T., Fox, C. and O'Sullivan, M.J., 2011. "Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm". *Water Resources Research*, vol. 47, no. 10, W10521, DOI:10/1029/2010WR010352.

Dausman, A.M., Doherty, J., Langevin, C.D., and Sukop, M.C., 2010. "Quantifying data worth toward reducing predictive uncertainty". *Ground Water*, vol. 48, no. 5, pp. 729-740.

De Groot-Hedlin, C. and Constable, S., 1990. "Occam's inversion to generate smooth, two-dimensional models from magnetotelluric data". *Geophysics*, vol. 55, no. 12, pp. 1613-1624.

Deutsch, C. and Journel A., 1998. GSLIB: *Geostatistical Software Library and* User's Guide, 2nd Edition, Oxford University Press.

Doherty, J., 2003. "Ground Water Model Calibration Using Pilot Points and Regularisation", *Ground Water*, vol. 41, no. 2, pp. 170-177.

Doherty, J., 2012. *PLPROC : A Parameter List Processor*. Watermark Numerical Computing, Brisbane, Australia. Available at http://www.pesthomepage.org.

Doherty, J., 2015a. *PEST: Model-independent parameter estimation, user manual*. Watermark Numerical Computing. Available at :- http://www.pesthomepage.org/downloads.php.

Doherty, J., 2015b. *Calibration and Uncertainty Analysis for Complex Environmental Models*. Watermark Numerical Computing, Brisbane, Australia. ISBN: 978-0-9943786-0-6. Available at http://www.pesthomepage.org.

Doherty, J., and Christensen, S., 2011. “Use of paired simple and complex models to reduce predictive bias and quantify uncertainty”. *Water Resources Research*, vol. 47, no. 12, p. 21.

Doherty, J. and Hunt, R.J., 2009. “Two statistics for evaluating parameter identifiability and error reduction”. *Journal of Hydrology*, vol. 366, pp. 119-127.

Doherty, J. and Welter, D., 2010. “A short exploration of structural noise”. *Water Resources Research*. vol. 46, WO5525, doi:10.1029/2009WR008377.

Doherty, J.E., Fienen, M.N. and Hunt, R.J., 2010. *Approaches to highly parameterized inversion: Pilot-point theory, guidelines, and research directions*. U.S. Geological Survey Scientific Investigations Report 2010-5168, p. 36.

Doherty, J.E. and Hunt, R.J., 2010. *Approaches to highly parameterized inversion: A guide to using PEST for groundwater-model calibration*. U.S. Geological Survey Scientific Investigations Report 2010-5169, p. 59.

Doherty, J. and Simmons, C., 2013. “Groundwater modelling in decision support: reflections on a unified conceptual framework”. *Hydrogeology Journal*, vol. 21, pp. 1531-1537.

Draper, N.R. and Smith, H., 1998. *Applied regression analysis*. Third edition. New York: John Wiley and Sons, Inc.

Duan, Q.S., Sorooshian, S., and Gupta, V.K., 1992. “Effective and efficient global optimization for conceptual rainfall runoff models”. *Water Resources Research*, vol. 28, no. 4, pp. 1015-1031.

Efendiev, Y., Datta-Gupta, A., Ginting, V., Ma, X., and Mallick, B., 2005. “An efficient two-stage Markov chain Monte Carlo method for dynamic data integration”, *Water Resources Research*, vol. 41, W12423, doi:10.1029/2004WR003764.

Efendiev, Y., Datta-Gupta, A., Ma, X., and Mallick, B., 2009. “Efficient sampling techniques for uncertainty quantification in history matching using non-linear error models and ensemble level upscaling techniques”. *Water Resources Research*, vol. 45, W11414, doi:10.1029/2008WR007039.

Eide, A.L., Holden, L. and Reiso, E., 1994. "Automatic History Matching by use of Response Surfaces and Experimental Design". *4th European Conference on the Mathematics of Oil Recovery*, Røros, Norway.

Elsheikh, A.H., Hoteit, I. and Wheeler, M.F., 2014. “Efficient Bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates”. *Computer Methods in Applied Mechanics and Engineering*, vol. 269, pp. 515-537.

Fen C., Chan, C. and Cheng, H., 2009. “Assessing a response-based optimisation approach for soil vapour extraction system design”. *Journal of Water Resources Planning and Management*, vol. 135, no. 3, pp. 198-207, doi:10.1061/(ASCE)0733-9496(2009)135:3(198).

Forrester, A. I. J., Sobester A. and Keane A. J., 2007. “Multi-fidelity optimisation via surrogate modelling”. *Proceedings of the Royal Society A*, vol. 463, pp. 3251-3269.

Gallagher, M. and Doherty, J., 2007a. “Predictive error analysis for a water resource model”. *Journal of Hydrology*, vol. 334, pp. 513-533.

Gallagher, M. and Doherty, J., 2007b. “Parameter estimation and uncertainty analysis for a watershed model”. *Environmental Modelling and Software*, vol. 22, pp. 1000-1020.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B., 2013. *Baysian Data Analysis*. Chapman and Hall, CRC Texts in Statistical Science.

Gomez-Hernandez, J.J., Hendricks Franssen H.J. W. M. and Sahuquillo A., 2003. "Stochastic conditional inverse modelling of subsurface mass transport: A brief review and the self-calibrating method", *Stochastic Environmental Research and Risk Assessment*, vol. 17, no. 5, pp. 319-328.

Harbaugh, W., Edward, E.R., Hill, M.C. and McDonald, M.G., 2000. *MODFLOW-2000, The U.S. Geological Survey Modular Ground-Water Model – User Guide to Modularisation Concepts and the Ground-Water Flow Process*. Open-File Report 00-92, U.S. Geological Survey, Reston, Virginia.

Hansen, N., Muller, S.D. and Koumoutsakos, P., 2003. "Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES)". *Evolutionary Computation*, vol. 9, pp. 159-195.

Hemker, T., Fowler, K. R., Farthing M. W. and von Stryk, O., 2007. "A mixed-integer simulation-based optimisation approach with surrogate functions in water resources management". *Optimisation Engineering*, vol. 9, pp. 341-360.

Henry, H. R., 1964. Effects of dispersion on salt encroachment in coastal aquifers. *U.S. Geological Survey Water Supply Paper*. 1613-C.

Herckenrath, D., Langevin, C. D. and Doherty, J., 2011. "Predictive uncertainty analysis of a saltwater intrusion model using null-space Monte Carlo". *Water Resources Research*, vol. 47, no. 5, W05504.

Hunt, R.J., Doherty, J. and Tonkin, M.J., 2007. "Are Models Too Simple? Arguments for Increased Parameterization". *Ground Water*, vol. 45, no. 3, pp. 254-262.

James, S. C., Doherty, J. and Eddebbarh, A., 2009. "Practical Postcalibration Uncertainty Analysis: Yucca Mountain, Nevada". *Ground Water*, vol. 47, no. 6, pp. 851-869.

Johnson, J.S., Gosling, J.P. and Kennedy, M.C., 2011. “Gaussian process emulation of second-order Monte Carlo simulations”. *Journal of Statistical Planning and Inferenc*e, vol. 141, no. 5, pp. 1838-1848.

Kavetski, D., Kuczera, G. and Franks, G., 2006. “Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artifacts”. *Journal of Hydrology,* vol. 320, pp. 173-186.

Keating E.H., Doherty, J., Vrugt, J.A. and Kang, Q., 2010. “Optimisation and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality”. *Water Resources Research*, vol. 46, W10517, doi:10.1029/2009WR008584.

Kennedy, M.C. and O'Hagan A., 2000. “Predicting the output from a complex computer code when fast approximations are available”. *Biometrika,* vol. 87, no. 1, pp. 1-13.

Kennedy, M.C. and O'Hagan A., 2001. “Baysian calibration of computer models”. *Journal of the Royal Statistical Society B*, vol. 63, part 3, pp. 425-464.

Kennedy, J. and Mendes, R., 2002. “Population structure and particle swarm performance”. *In: Proceedings of the 2002 Congress on Evolutionary Computation*, CEC ’02, pp. 1671-1676.

Koch, K. R., 1999. *Parameter estimation and hypothesis testing in linear models*. Third edition. Springer-Verlag, Berlin, Hiedelberg.

Laloy, E., Rogiers, B., Vrugt, J.A., Mallants, D. and Jacques, D., 2013. “Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion”. *Water Resources Research*, vol. 49, no. 5, pp. 2662-2682.

Langevin, C.D., Thorne Jr., D.T., Dausman, A., Sukop, M.C., and Guo, W., 2008. *SEAWAT version 4: A computer program for simulation of multi-species solute and heat transport*. U.S. Geological Survey, Reston, Virginia.

Marzouk, Y. and Xiu, D., 2009. "A Stochastic Collocation Approach to Bayesian Inference in Inverse Problems". *Communications in Computational Physics*, vol. 6, no. 4, pp. 826-847.

Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory*. Revised edition. San Diego: Academic Press.

Mondal, A., Efendiev, Y., Mallick, B. and Datta-Gupta, A., 2010. "Bayesian uncertainty quantification for flows in heterogeneous porous media using reversible jump Markov chain Monte Carlo methods". *Advances in Water Research*, vol. 33, no. 3, pp. 241-256.

Moore, C., and Doherty, J., 2005. "The role of the calibration process in reducing model predictive error". *Water Resources Research*, vol. 41, no. 5, W05020, DOI:10.1029/2004WR003501.

Moore, C. and Doherty J., 2006. "The cost of uniqueness in groundwater model calibration". *Advances in Water Resources Research*, vol. 29, no. 4, pp. 605–623, DOI:10.1016/j.advwatres.2005.07.003.

Mugunthan, P., Shoemaker, C.A. and Regis, R.G., 2005. "Comparison of function approximation, heuristic, and derivative based methods for automatic calibration of computationally expensive groundwater bioremediation models". *Water Resources Research*, vol. 41, W11427, doi:10.1029/2005WR004134.

Oliver, D.S., Cunha, L.B. and Reynolds, A.C., 1997. "Markov Chain Monte Carlo Methods for Conditioning a Permeability Field to Pressure Data". Mathematical Geology, vol. 29, no. 1, pp. 61-91.

Oliver, D.S., Reynolds A. and Liu, N., 2008. *Inverse Theory for Petroleum Reservoir Characterisation and History Matching*. Cambridge University Press.

Oliver, D. S. and Chen, Y., 2011. "Recent progress on reservoir history matching: a review". *Computational Geosciences*, vol. 15, no. 1, pp. 185-221.

Panday, S., Langevin, C.D., Niswonger, R.G., Ibaraki, M. and Hughes, J.D., 2013. "MODFLOW–USG version 1: An unstructured grid version of MODFLOW for simulating groundwater flow and tightly coupled processes using a control volume finite-difference formulation". *U.S. Geological Survey Techniques and Methods*, book 6, chap. A45, 66 p.

Razavi, S., Tolson, B.A. and Burn, D.H., 2012a. "Review of surrogate modeling in water resources." *Water Resources Research*, vol. 48, W07401, doi:10.1029/2011WR011527.

Razavi, S., Tolson, B.A. and Burn, D.H., 2012b. "Numerical assessment of metamodeling strategies in computationally intensive optimization". *Environmental Modelling and Software*. vol. 34, pp. 67-86.

Regis, R.G. and Shoemaker, C.A., 2004. "Local function approximation in evolutionary algorithms for the optimisation of costly functions". *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 5, pp. 490-505.

Siade A.J., Putti, M. and Yeh, W.W.G., 2010. "Snapshot selection for groundwater model reduction using proper orthogonal decomposition". *Water Resources Research*, vol. 46, no. 8, W08539.

Siade A.J., Putti, M. and Yeh, W.W.G., 2012. "Reduced order parameterestimation using quasilinearization and quadratic programming". *Water Resources Research*, vol. 48, W06502, doi:10.1029/2011WR011471.

Sakov, P., Oliver, D. S. and Bertino L., 2012. "An Iterative EnKF for Strongly Nonlinear Systems." *Monthly Weather Review*, vol. 140, no. 6, pp. 1988-2004.

Schwede, O.A. and Cirpka, O.A., 2009. "Use of steady-state concentration measurements in geostatistical inversion". *Advances in Water Resources*, vol. 32, no. 4, pp. 607-619.

Shultz, M.T., Small, M.J., Farrow, R.S. and Fischbeck P.S., 2004. "State Water Pollution Control Policy Insights from Reduced-Form Model". *Journal of Water Resources Planning and Management*, vol. 130, no. 2, pp.150-159.

Shultz, M.T., Small, M.J., Fischbeck P.S. and Farrow, R.S. 2006. "Evaluating response surface designs for uncertainty analysis and prescriptive applications of a large-scale water quality model". *Environmental Modelling and Assessment*, vol. 11, pp. 345-359.

Sivakumar, B., 2004. "Dominant processes concept in hydrology: moving forward". *Hydrological Processes*, vol. 18, pp. 2349-2353.

Sivakumar, B., 2008, "Dominant processes concept, model simplification and classification framework in catchment hydrology". *Stoch. Environ. Res. Risk. Assess.*, vol. *22*, pp. 737–748.

Sobol, I.M., 1979. "On the systematic search in a hypercube". *SIAM Journal of Numerical Analysis*, vol. 16, pp. 790-793.

Sun, Y., C. Tong, Q. Duan, T. A. Buscheck and J. A. Blink (2012). "Combining Simulation and Emulation for Calibrating Sequentially Reactive Transport Systems." *Transport in Porous Media*, vol. 92, no. 2, pp. 509-526.

Tikhonov, A.N. and Arsenin, V.Y., 1977. *Solutions of Ill-Posed Problems*. Halsted Press, New York, p. 268.

Tonkin, M.J. and Doherty, J., 2005. "A hybrid regularized inversion methodology for highly parameterized environmental models". *Water Resources Research*, vol. 41, W10412, doi:10.1029/2005WR003995.

Tonkin, M.J., Doherty J. and Moore, C., 2007. "Efficient nonlinear predictive error variance for highly parameterized models", *Water Resources Research*, vol. 43, no. 7, W07429.

Tonkin, M.J. and Doherty, J., 2009. "Calibration-constrained Monte Carlo analysis of highly parameterized models using sub-space techniques". *Water Resources Research*, vol. 45, W00B10, doi:10.1029/2007WR006678.

Vecchia, A.V. and Cooley, R.L., 1987. "Simultaneous confidence and prediction intervals for nonlinear regression model with application to a groundwater flow model". *Water Resources Research*, Vol. 23, no. 7, pp. 1237-1250.

Vermeulen, P.T.M., Heemink, A.W, and Valstar, J.R., 2005. "Inverse modeling of groundwater flow using model reduction". *Water Resources Research,* vol. 41, no. 6, W06003.

Vermeulen, P.T.M., te Stroet, C.B.M. and Heemink, A.W., 2006. "Model inversion of transient nonlinear groundwater flow models using model reduction". *Water Resources Research,* vol. 42, no. 9, W09417.

Voss, C.I. and Provost A.M, 2002. *SUTRA, a model for saturated–unsaturated variable density ground-water flow with energy or solute transport*. Reston, USA., U.S. Geological Survey Open-File Report 02–4231, p. 250.

Vrugt, J.A., Gupta, H.V., Bouten, W. and Sorooshian, S., 2003. "A Shuffled Complex Evolution Metropolis algorithm for optimisation and uncertainty assessment of hydrologic model parameters". *Water Resources Research*, vol. 39, no. 8, pp. na-na.

Vrugt, J. and Robinson, B., 2007. “Improved evolutionary optimization from genetically adaptive multimethod search”. *Proceedings of The National Academy of Sciences of The United States of America*, vol. 104, pp. 708-711.

Vrugt, J.A., ter Braak, C.J.F, Clark, M.P., Hyman, J.M. and Robinson, B.A., 2008. “Treatment of input uncertainty in hydrologic modelling: Doing hydrology backward with Markov chain Monte Carlo simulation”. *Water Resources Research*, vol. 44, W00B09.

Watson, T., Doherty, J. and Christensen, S., 2013. “Parameter and predictive outcomes of model simplification”. *Water Resources Research*, vol. 49, no. 7, pp. 3952-3977, DOI:10.1002/wrcr.20145.

White, J.T., Doherty, J. and Hughes, J.D., 2014. “Quantifying the predictive consequences of model error with linear subspace analysis”. *Water Resources Research*, vol. 50, no. 2, pp. 1152-1173, doi:10.1002/2013WR014767.

Werner, A.D., Bakker, M., Post, V.E.A., Vandenbohede, A., Lu, C., Ataie-Ashtiani, B., Simmons, C. and Barry, D.A., 2013. “Seawater intrusion processes, investigation and management: Recent advances and future challenges”. *Advances in Water Resources*, vol. 51, pp. 3-26.