Problem Decomposition

in

Decision-Support Groundwater Modelling

by

John Doherty and Catherine Moore

gmdsi
Groundwater Modelling Decision Support Initiative

BHP
Flinders University
National Centre for Groundwater Research and Training
RioTinto

PUBLISHED BY

The National Centre for Groundwater Research and Training
C/O Flinders University
GPO Box 2100
Adelaide SA 5001
+61 8 8201 2193

DISCLAIMER

The National Centre for Groundwater Research and Training, Flinders University advises that the information in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice.

CITATION

For bibliographic purposes this report may be cited as: Doherty, J. and Moore C., (2023). Problem Decomposition in Decision-Support Groundwater Modelling. A GMDSI Monograph. National Centre for Groundwater Research and Training, Flinders University, South Australia.

ISBN:

DOI:

DOI

COPYRIGHT

© Flinders University 2023

Copyright: This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part may be reproduced by any process, nor may any other exclusive rights be exercised, without the permission of Flinders University, GPO Box 2100, Adelaide 5001, South Australia.

Glossary

Bayesian analysis

This describes methods that implement history-matching according to Bayes equation. These methods support calculation of the posterior probability distribution of one or many random variables from their prior probability distributions and a so-called "likelihood function" – a function that increases with goodness of model-to-measurement fit.

Calibration

Calibration is history-matching, implemented in a way that achieves uniqueness of estimated parameters. Uniqueness is often achieved by finding the "simplest" set of parameters that enables pertinent model outputs to match field measurements of system behaviour.

Data assimilation

This term is generally used to describe history-matching undertaken as part of Bayesian analysis. Data assimilation reduces the uncertainties of model parameters because they can only adopt values that enable model outputs to fit field measurements of system behaviour.

Data space inversion

A numerically efficient data assimilation methodology that uses a model to build statistical linkages between the past and future behaviour of an environmental system. Predictions of the latter can then be conditioned by measurements of the former.

Ensemble

A collection of realisations of random parameters.

Ensemble smoother

A software package that adjusts realisations of parameters that comprise an ensemble until all of these realisations enable pertinent model outputs to match field measurements of system behaviour.

History-matching

The adjustment of model parameters such that model outputs are able to replicate field measurements of system behaviour.

Hydraulic conductivity

The greater is the hydraulic conductivity of a porous medium, the greater is the amount of water that can flow through that medium in response to a head gradient.

Impact pathway

This is a colloquial expression that loosely describes a volume of subsurface material that enables propagation of head/ drawdown and/or contaminants to a location at which its arrival comprises an adverse environmental impact. This volume often possesses a higher hydraulic conductivity than that of media which encompass it.

Jacobian matrix

A matrix of partial derivatives (i.e. sensitivities) of model outputs (generally those that are matched with field measurements) with respect to model parameters.

Matrix

A two-dimensional array of numbers indexed by row and column.

Model structure

This term describes those aspects of the design of a numerical model that pertain to its spatial and temporal discretisation. These include cell sizes and connections, as well as its layering. The notion of "structure" can also be applied to the nature and locations of a numerical model's boundary conditions.

Null space

In the parameter estimation context, the null space is comprised of combinations of parameters that have no effect on model outputs that correspond to field measurements. These combinations of parameters are thus inestimable through history-matching.

Objective function

A measure of model-to-measurement misfit whose value is lowered as the fit between model outputs and field measurements improves. In many parameter estimation contexts the objective function is calculated as the sum of squared weighted residuals.

Parameter

In its most general sense, a parameter is any model input that can be adjusted in order to support a better fit between model outputs and corresponding field measurements. Often, but not always, these inputs represent physical or chemical properties of a simulated system. However, there is no reason why they cannot also represent water or contaminant source strengths and locations.

Pilot point

A type of spatial parameterisation device. A modeller, or a model-driver package such as PEST or PEST++, assigns values to a set of points which are distributed in two- or three-dimensional space. A model pre-processor then undertakes spatial interpolation from these points to cells comprising the model grid or mesh. This allows parameter estimation software to ascribe hydraulic property values to a model on a pilot-point-by-pilot-point basis, while a model can accept these values on a model-cell-by-model-cell basis. The number of pilot points used to parameterise a model is generally far fewer than the number of model cells.

Prior probability

The pre-history-matching probability distribution of random variables (model parameters in the present context). Prior probability distributions are informed by expert knowledge, as well as by data gathered during site characterisation.

Posterior probability

The post-history-matching probability distribution of random variables (model parameters in the present context). These probability distributions are informed by expert knowledge, site characterisation studies, and measurements of the historical behaviour of the system.

Probability density function

A function that describes how likely it is that a random variable adopts certain values, or falls between certain values.

Probability distribution

This term is often used interchangeably with “probability density function”.

Problem decomposition

A conceptual partitioning of a complex management problem into a series of carefully defined smaller problems that are individually easier to solve.

Realisation

A random set of parameters.

Regularisation

The means through which a unique solution is sought to an ill-posed inverse problem. Regularisation methodologies fall into three broad categories, namely manual, Tikhonov and subspace methods; singular value decomposition is the flagship of the last of these categories.

Residual

The difference between a model output and a corresponding field measurement.

Simulator

A computer program that calculates the states of a system using equations (often partial differential equations) that encapsulate the physical and/or chemical laws that govern its behaviour.

Singular value decomposition (SVD)

A matrix operation that creates orthogonal sets of vectors that span the range of possible vectors that a matrix can calculate on the one hand, and the range of possible vectors that a matrix can process on the other hand. When undertaken on a Jacobian matrix, SVD can subdivide parameter space into complementary, orthogonal subspaces; these are often referred to as the solution and null subspaces. Each of these subspaces is spanned by a set of orthogonal unit vectors. The null space of a Jacobian matrix is composed of combinations of parameters that have no effect on model outputs that are used in its calibration, and hence are inestimable.

Solution space

The orthogonal complement of the null space. This is obtained by subjecting a Jacobian matrix to singular value decomposition.

Spatial correlation

The propensity for the value of one random variable (such as the value assigned to a pilot point) to be influenced by that of another that is separated from it in space.

Stochastic

A stochastic variable is a random variable.

Stress

This term generally refers to those aspects of a groundwater model that cause water to move. They generally pertain to boundary conditions. User-specified heads along one side of a model domain, extraction from a well, and pervasive groundwater recharge, are all examples of groundwater stresses.

Tikhonov regularisation

An ill-posed inverse problem achieves uniqueness by finding the set of parameters that departs least from a user-specified parameter condition, often one of parameter equality and hence spatial parameter homogeneity.

Vector

A collection of numbers arranged in a column and indexed by their position in the column.

CONTENTS

1. INTRODUCTION

1.1 General

This document complements other GMDSI manuscripts on related topics. These manuscripts include Doherty and Moore (2021) and Doherty (2022). The former discusses appropriate model complexity, while the latter discusses the relationship between decision-support modelling and the scientific method. Both of these issues are closely related to that of problem decomposition, this being the subject matter of the present document.

We use the term “problem decomposition” herein to characterise an approach to environmental management that renders it amenable to the type of quantitative assistance that numerical modelling can provide. It requires that modelling goals be carefully defined, and that modelling workflows be then designed to serve these goals. As the term “decomposition” implies, it often involves the development of one or a number of conceptual simplifications which may invoke concepts such as “impact pathways”. These provide a focus for model-based processing of environmental data in ways that improve the likelihood of a management decision being “good” according to the values that system management serves.

In this manuscript we attempt to develop a framework that identifies and characterises components that are common to the many management contexts which modelling is asked to support. This framework allows the problem of “decomposing the problem” to be itself decomposed in a given management context as a precursor to developing a modelling workflow that is suitable for that context.

We also develop the notion of “modelling as inquiry”. Inquiry is at the heart of the scientific method. Ideally, the scientific method should be at the heart of modelling-assisted environmental management. Application of the scientific method in environmental investigations is discussed in texts such as Baker (2017), Caers (2018) and Doherty (2022).

What is just as important as the scientific method is the scientific mindset. This is one of humility when dealing with natural systems whose complexity is daunting, and whose details are incompletely known. It is one of perpetual scepticism of current theories, explanations and conceptual models. It is also collegiate, as it acknowledges that the path to a deeper understanding of natural systems is paved by the insights of many.

1.2 Decision-Support Modelling

We focus on decision-support groundwater modelling. It is important to distinguish this type of modelling from that which is undertaken for purposes other than direct support of groundwater management. In other contexts, modelling may be undertaken in order to better understand how complex subsurface processes may interact with each other and with the heterogeneous media that host them. However, this type of modelling does not attempt to replicate historical measurements of groundwater behaviour at a particular study site. Nor does it attempt to predict the specifics of its future behaviour at that site and associate uncertainties with these predictions. Nevertheless, “investigatory modelling” of this type may influence the design of decision-support modelling workflows which pursue the latter objectives.

Decision-support modelling is often directed towards forecasting the future behaviour of a groundwater system when that system will be subjected to human-induced stresses that it has not experienced in the past. These stresses may comprise its new management regime. The list of issues which decision-support groundwater modelling may address is long. It includes settings wherein groundwater is a resource, and settings wherein groundwater is a problem. It also includes contexts wherein groundwater engages in complex interactions with other natural systems such as surface water and biological systems. Issues which decision-support modelling commonly addresses include the following:

- design of a dewatering system for a mine;
- assessment of the effect of water extraction on a surface water body or a groundwater-dependent ecosystem;
- mitigation of subsidence induced by groundwater extraction;
- design of a contaminant remediation system;
- prevention of extraction-induced salt water intrusion;
- low or high temperature geothermal energy extraction and/or storage;
- mitigation of river pollution by groundwater-conveyed agricultural nutrients.

The term “management” implies evasion of a problem. Sometimes the problem is ongoing and management wishes to abate it. Sometimes the problem is conjectural, and management is designed to forestall its occurrence.

In many cases, a groundwater system will soon be subjected to stresses which it has not hitherto experienced. Measurements of its past behaviour are therefore only partially indicative of its future behaviour. It follows that prediction of its future behaviour requires some knowledge of the hydraulic properties of the media through which groundwater flows. This knowledge is always incomplete, but must be exploited to the extent that it is available.

Different workflows will suit different management settings, and different data availability circumstances. All of these workflows will bear the marks of those who develop them, for although groundwater modelling is a scientific (and therefore presumably objective) enterprise, its fruitful application requires the making of many compromises on which practitioners may disagree. However differences in approach may matter little if modellers and modelling stakeholders possess a clear understanding of the trade-offs that spawn the need for compromise.

1.3 Language

Words such as “model”, and words with which “model” is often associated, mean different things to different people. It is not our intention to offer precise definitions of these words. However, it may make the meaning of parts of this manuscript a little clearer if we state the meanings that we associate with them (while acknowledging that we occasionally stray into more colloquial use of these words ourselves).

We start with “simulator”. A simulator is a computer program that calculates the state of a system using equations that describe how the system works. These are often partial differential equations. Where a system has many states (for example heads at many locations), these equations may be difficult to solve.

“Simulation” is what a simulator does.

A “model” of a system is simulation applied to that system. It is therefore more than a computer program. It is a simulator, together with files that describe the geometry and properties of a particular system. When the simulator runs, it therefore calculates states for that particular system, and stores these states in other files.

“Modelling” is used to describe the act of building a model for a system of interest, often for the purpose of supporting management of it. As is discussed extensively herein, this encompasses a range of activities which extend well beyond that of selection of an appropriate simulator, and construction of a set of files that render simulation specific to a particular site. It also involves activities such as history-matching, uncertainty analysis and reporting of modelling results. It therefore requires the use of software packages whose roles extend beyond that of purely simulation.

“Modelling” is the present participle of the verb “model”. Used as a verb, the word “model” therefore describes the act of modelling. This is a source of confusion, as the act of modelling (in the decision-support context at least), should pertain to much more than the construction of a model (whose meaning is provided above). The authors opine that confusion surrounding the different meanings of “model” when used as a verb and as a noun may be a contributing factor to the sometimes disappointing performance of modelling in decision-support.

Finally, the term “modelling workflow” describes a structured set of activities that collectively comprise the act of modelling. The components of this workflow can be complex. Each component requires the use of simulator-complementary software that is specific to that component.

1.4 Target Audience

Although this manuscript addresses technical issues, it is purposefully non-technical. Its intention is to propose a philosophical and practical basis for decision-support modelling that has the potential to remove some impediments to its creative use in supporting the management of complex natural systems. While some equations are presented, this is done purely to demonstrate the logical basis of some of the arguments presented herein.

We hope, therefore, that this manuscript will be suitable for modellers, as well as for those who do not themselves engage in modelling, but would like to better understand what modelling can, and cannot, provide to the decision-making process.

2. INFORMATION AND METRICS

2.1 General

We see problem decomposition as a necessary precursor to design and implementation of an effective decision-support modelling workflow. However before attempting to demonstrate why this is so, some metrics are needed.

This section provides metrics for decision-support modelling. In doing this, it attempts to characterise the context in which decision-support modelling must operate, and the tasks that decision-support modelling must accomplish.

2.2 Management Failure

Management of a natural system is undertaken to prevent the occurrence of events and/or states that are unwanted. The nature of these events and states, and the reasons for their undesirability, are context-specific. Their occurrence can be considered as management failure.

In some contexts, management failure results in loss of money, or loss of a costly asset. This can occur if a mine dewatering system fails to maintain dryness of a pit, or does not reduce near-pit hydraulic pressures to the extent required to ensure pit wall stability. It can occur when a contaminant stubbornly remains underground after prolonged operation of a pump-and-treat system. In other contexts, loss or damage to an environmental asset on which it is difficult to place a monetary value can be decreed as management failure. For example, reduction in surface water flows, or damage to a groundwater-dependent ecosystem may be the result of excessive groundwater extraction. Algal blooms may follow groundwater transmission to rivers of nutrients applied to farms as fertiliser.

Central to the Popperian view of the scientific method is the notion of hypothesis-testing. Central to the notion of scientific hypothesis-testing is that a hypothesis can be rejected, but never accepted. This is because competing hypotheses that exists now, or may emerge in the future, may also defy rejection. A hypothesis can be rejected if it is demonstrably incompatible with system processes, system properties, and/or the measured behaviour of a system. In practice, environmental processes and properties are incompletely known. It may therefore be possible to portray a hypothesis as unlikely, while not being able to reject it completely.

Science-based environmental decision-making is presented with a ready-made hypothesis. It is the hypothesis of management failure. If management failure can occur in multiple ways, multiple hypotheses are thereby defined.

Science-based environmental management, as implemented at a particular site, can be furnished with a “scientific instrument” that can be turned to the task of testing management-salient hypotheses. This instrument is a numerical model. Ideally, a numerical model can mimic the behaviour of an environmental system (to some extent at least). Its parameters can represent the physical and chemical properties of that system (to some extent at least). Ideally, history-matching can ensure that these processes and properties are compatible with measurements of system behaviour. The ingredients for management-pertinent hypothesis-testing are therefore available. However, their existence alone does not guarantee that groundwater modelling implements the scientific method. As for any other instrument, it is the experimental setup, and the dexterity of the scientist, that defines the scientific integrity of its deployment.

The testing of management-salient hypotheses therefore defines the “problem” that decision-support modelling must address. This problem must be satisfactorily decomposed before it can be solved.

2.3 Accuracy of Numerical Simulation

2.3.1 General

Implied in the words “numerical simulation” is a capacity to replicate on a computer the complex processes that prevail in a natural system. It follows that simulation is “better” when it mimics those processes more closely. However, because the pursuit of “simulation perfection” can actually erode the decision-support potential of numerical simulation, we will shortly define alternative metrics for decision-support modelling that are more aligned with achievement of its scientific potential.

2.3.2 Upscaling

Numerical simulation is generally based on a grid or mesh. Use of this grid rests on the assumption that partial differential equations that describe processes within tiny volumes of a porous medium can also be applied to the larger volumes of material that occupy cells within the grid. It is further assumed that each model grid cell can be endowed with upscaled hydraulic properties that represent averages of real-world hydraulic properties over that cell.

The mathematics and practicalities of upscaling have received a great deal of attention in both the water and petroleum literatures. These studies demonstrate that the relationship between upscaled and point hydraulic properties is complicated at best and tenuous at worst. Rarely are they linked through simple spatial averaging. Furthermore, in an upscaled simulation environment, groundwater flow should be considered as tensorial. Hence a head difference in one direction can induce groundwater flow in orthogonal directions. This is rarely simulated in everyday groundwater modelling practice.

Highly nonlinear processes such as recharge are even more difficult to upscale – both in space and in time. At many locations, yearly recharge is dominated by a small number of intense rainfall events. During these events, surface water ponds at low points in the landscape which then become the foci of intense recharge. At the same time, macropore flow renders vertical recharge processes highly nonlinear and difficult to describe mathematically. Recharge through the beds of streams and rivers (whose footprint increases dramatically during wet events) is also highly nonlinear under these climatic circumstances.

Groundwater models often employ a single layer to represent the entire thickness of an aquifer. Flow of water is assumed to be vertically uniform within this layer. Intricate flow patterns induced by variations in the properties and dispositions of individual lithologies which comprise an aquifer are ignored.

Even if it were possible to undertake perfect upscaling of subsurface hydraulic properties so that their representation at the model grid scale does not violate the applicability of the partial differential equations that are encapsulated in a simulator, the information that is required for correct assignment of these properties to every model cell is never available. Subsurface hydraulic properties are rarely measured at more than a handful of locations. Spatial interpolation of these properties is fraught with uncertainty. Nor can they be back-calculated uniquely through history-matching.

Many model algorithmic details are approximate. “Hydrodynamic dispersion” is a surrogate for the effects of tortuous flow paths through heterogeneous media. “Secondary porosity” is a surrogate for a similar thing. Field measurements of contaminant concentration are often highly variable in both the horizontal and vertical directions – far too variable for a model to reproduce. This makes history-matching of concentration measurements difficult, notwithstanding their high information content with respect to local and medium-scale hydraulic properties.

2.3.3 Some Other Problems

In many, if not most, managed groundwater systems, important historical stresses are poorly known. Land use is constantly changing. Historical human extraction from a groundwater system may have been considerable, but unmetered.

Model boundary conditions are often approximate. A no-flow boundary is rarely impervious. The elevations of the top and bottom of riverbed sediments can vary within an individual model cell, and are rarely known precisely. The permeability of these sediments can change with season, and with the direction of groundwater flow through them. The nature of a river's connection with an aquifer can change as its head rises and falls during a river flow event, or as nearby groundwater extraction reduces groundwater pressures beneath the river.

2.3.4 Is Accuracy Really Required?

In summary, if the metric for a good numerical model is that it simulates a particular natural system "accurately", then no numerical model is very good. Fortunately, as is discussed herein, this metric is not aligned with the requirements of decision-support modelling. A metric that is more aligned with these requirements (one that will be refined later), is that the modelling process quantify the uncertainties of predictions of management interest, and that it reduce these uncertainties as much as available data allow. Compliance with this metric on the one hand, and simulation "accuracy" on the other hand, are two different things. Furthermore, while achievement of this metric may necessitate a certain degree of simulation integrity, this is never a sufficient condition for its achievement.

2.4 Information

The term "information" has many colloquial meanings. It also has a strict mathematical meaning. In this monograph we characterise "information" as "that which reduces uncertainty".

When undertaking decision-support modelling we are not interested in all aspects of system uncertainty, for these aspects are too numerous to contemplate, too burdensome to assess, and are mostly of little interest. We are generally interested only in those aspects of system uncertainty that effect predictions of management interest. These normally pertain to management failure. As is discussed above, "management failure" is a context-specific notion. In some contexts it may denote the economic cost of unpreparedness for unforeseen aspects of system behaviour. In other contexts it may denote damage to a natural asset. These are important (by definition). However, typically only a few management failure mechanisms exist in any given modelling context.

We therefore limit our concept of "information". We use this term to denote that which reduces the uncertainties of predictions that are of interest to us. This definition makes intuitive sense. After all, information (in the colloquial sense) is not information unless it pertains to something that we care about.

Actually, the notion of "care factor" can prompt further refinement of the concept of "information" as it applies to decision-support modelling. Assessment of the possibility of management failure is the primary task of decision-support modelling. It follows that the pessimistic end of the uncertainty interval of a failure-pertinent model prediction is of greater interest to us than its optimistic end, for the pessimistic end of this distribution embodies a hypotheses that decision-support modelling must test. “Information” can therefore be conceptualised as that which reduces the pessimistic extent of management-failure-pertinent predictive uncertainty intervals. Information may therefore allow rejection of a hypothesis of management failure; alternatively, it may demonstrate that failure is inevitable. Conversely, lack of information may admit the possibility of management failure, while precluding its inevitability.

Information resides in data. It is the task of decision-support modelling to harvest information from data. Sources of information can be divided into two broad categories.

Information that emerges from site characterisation defines one of these categories. This includes inferences of three-dimensional geology from geological mapping and drilling, local estimates of system hydraulic properties obtained through aquifer test analysis, indirect estimates of system properties obtained from borehole and surface geophysics, and identification of recharge mechanisms and sites from aerial and satellite imagery. This type of information is inherently probabilistic; it should therefore be expressed as such in a model. Large uncertainties often accompany two- and three-dimensional extrapolation of hydraulic properties and lithologies between their measurement points.

The other source of information is comprised of measurements of present and/or historical system states and behaviour. These include measurements of borehole water levels and/or contaminant concentrations, groundwater contributions to streamflow, etc. The information content of these data depends on both their quantity and their quality. Their quality deteriorates to the extent that measurements are contaminated by error, or reflect local conditions that cannot be represented in a model.

2.5 Parameters and Structure

A model’s components can be broadly subdivided into broad two categories, namely its parameters and its structure. Where decision-support modelling follows the principles of problem decomposition that are outlined herein, these two component types collectively embody the numerical representation of an impact pathway. However, in doing this, their roles are fundamentally different.

Groundwater model parameters often represent hydraulic properties of the subsurface, these being the coefficients that appear in the partial differential equations of groundwater flow and contaminant transport that simulators solve.

However the term “parameters” has broader connotations. A key characteristic of parameters is that they are adjustable. They can be adjusted during history-matching so that model outputs match field measurements. They can be adjusted prior to history-matching to express the fact that they are imperfectly known. Parameters can also be adjusted after history-matching to express the fact that constraints that are imposed on them by history-matching do not guarantee their uniqueness.

The notion of adjustability broadens the potential population base of parameters. If they are imperfectly known, historical pumping rates should be denoted as parameters, as should many boundary condition specifications. Ideally, any aspect of model construction that is imperfectly known should be awarded parameter status in order to acknowledge its uncertainty. Put simply, “if something is imperfectly known, then let it wiggle”. Failure to do this may result in underestimation of the uncertainties of one or more predictions of management interest.

As is discussed in the following section of this document, parameters play a unique role in decision-support modelling. They carry both information, and lack of information. Good decision-making requires access to both of these.

To distinguish them from parameters, we use the word “structure” to denote immutable aspects of a model. These include its gridding and layering, as well as the nature and locations of its boundary conditions. A model’s structure plays host to its parameters.

Unfortunately, rarely are all elements of a model’s structure completely known. Some are uncertain, while others (such as model cell size in the vicinity of known or implied structural or anthropogenic features, and the number of model layers) are somewhat arbitrary. This may affect model outcomes. Imperfections in a model’s structure can affect the values that parameters must adopt in order for a model to be capable of replicating measurements of system behaviour. This has repercussions for a model’s ability to predict future system behaviour.

2.6 Modelling Metrics

We now provide metrics for decision-support modelling. Metrics are important, for they can be used to assess the quality of modelling work. An assurance of modelling quality is important for those who pay for it, and for those who are stakeholders in model-based environmental management.

As a scientific activity, modelling is undertaken to extract information from data. Ideally, it should be a continuous activity that encourages alteration of site concepts as more data are acquired and processed. More importantly, it is an activity that relies on more than simulation alone, for a simulator must be used in conjunction with other software in order to implement the hypothesis-testing tasks that decision-support modelling demands. It follows that metrics should be applied to the modelling process as an activity, rather than to a discrete model that provides only partial support for this activity.

For reasons that are addressed later in this manuscript, decision-support modelling as an activity is replete with compromises. Perfection can never be attained; frustration is inevitable; subjectivity is a necessity. Formulation of decision-support modelling metrics that are relevant to real-world modelling must accommodate these defining characteristics of the decision-support modelling process.

Doherty and Simmons (2013) suggest two appropriate metrics.

The first of these metrics pertains to modelling failure. Decision-support modelling fails when it results in false rejection of a management failure hypothesis. The consequences of a false rejection are obvious; in some circumstances they may be disastrous. It follows that it is incumbent on a decision-support modeller to demonstrate to his/her clients and stakeholders that his/her modelling workflow satisfies this metric.

The second metric pertains to modelling utility. It may be easy for a modeller to preclude inappropriate rejection of a management failure hypotheses by endowing his/her predictions with gratuitously wide uncertainty margins. Information that is resident in data is therefore ignored rather than harvested – information that may, if properly processed, allow rejection of a management failure hypothesis. Doherty and Simmons (2013) declare a modelling enterprise to be lacking in utility (or “useless”) when it fails to subject decision-pertinent hypotheses to the scrutiny that the information content of site data allows.

2.7 Repercussions for Model Design

It is obvious from the above considerations that a model that is used to test a hypothesis of management failure must actually be capable of simulating management failure. It must therefore include any structural complexity which admits the possibility of failure. At the same time, it must not preclude failure through deployment of an artificially simplistic parameterisation scheme that prevents the probabilistic representation of impact pathways that may exist in a real system.

Just as important as flow of water is flow of information. A decision-support modelling workflow must ensure that failure-pertinent information is harvested from site data wherein it resides.

Model-harvested information is carried by its parameters. The amount by which parameter values can be varied, while still being considered acceptable, is a measure of their uncertainty (i.e. lack of information). If site characterisation studies and/or history-matching can constrain a model's parameters, they must be given the opportunity to do so. At the same time, continued lack of information must be free to express itself as post-history-matching parameter adjustability. This generally requires that spatial parameterisation density reflects the natural heterogeneity of material that comprises impact pathways. If this is not done, data may be credited with more information than it deserves as history-match-adjustment of artificially broadscale parameters (such as large zones of presumed parameter constancy) closes impact pathways. The risk of modelling failure (as defined above) is thereby increased.

It follows that design of a decision-support model should reflect the hypotheses that it must test, and therefore the hypothesis-pertinent predictions that it must make. This may require that it represent certain aspects of a natural system in detail – not because they are known but because they are unknown and are relevant to a prediction. To the extent that it is possible, this detail should be represented using parameters rather than structure. This is because parameters are adjustable. They can therefore express lack of information at the same time as their adjustability can be constrained by acquisition of information.

The first corollary of this imperative is that aspects of a decision-support model that are of secondary relevance to the testing of a particular management-salient hypothesis may be awarded simpler structure and simpler parameterisation. This may allow the model to evade slow run times and numerical complications, both of which obstruct the uncertainty quantification and uncertainty reduction tasks that are required of decision-support modelling.

A second corollary of this imperative is that different models may be required to test different management failure hypotheses, or even the same hypothesis if management failure can occur through multiple impact pathways.

The reader is reminded that the primary purpose of decision-support modelling is not to provide a numerical replica of the subsurface. It is to assess the uncertainties of predictions of management interest while reducing these uncertainties (or at least their pessimistic extremes) to the extent that the information content of site data allows. Uncertainty analysis should not therefore be considered as an "optional extra" for site numerical modelling. Instead, a model should be built from the ground up to assess and reduce the uncertainty of one or a number of management-pertinent predictions. This requires problem decomposition.

3. PROBLEM DECOMPOSITION

3.1 General

This section begins by focussing on some concepts that are salient to the notion of problem decomposition. While there is no strict “theory of problem decomposition” (because it is something that must be undertaken by a human being rather than by an algorithm), there are some theoretical considerations that manifest its necessity.

The section finishes with some practical examples of problem decomposition in action. Site-specific conclusions that are reached in this discussion are of secondary importance as they are context-specific. It is the way of thinking about these problems that matters.

We warn the reader that he/she is approaching some equations. As stated previously, it is the concepts that these equations embody that are important, and not the equations themselves. These concepts are rooted in common sense. We present equations because:

- they demonstrate that common sense has theoretical underpinnings;
- the theory on which they are based can be extended to examine issues such as predictive bias which are discussed in the next section of this manuscript.

3.2 Stochasticity

Characterisation of subsurface hydraulic property values and dispositions can only be probabilistic. It must express knowledge, at the same time as it must express incompleteness of knowledge. Thus, for example, we may know the rock types that prevail under a particular study site, and we may know the range of hydraulic properties that characterise these rock types. However we rarely know the exact dispositions of rock units within the subsurface, nor the nature and extent of heterogeneity that characterise each unit. This “lack of knowledge” problem is exacerbated where hydraulic properties are affected by local structural features such as faults. Hydraulic properties that are associated with these features may replace or interrupt those of their host rocks. At the same time, faulting may introduce uncertainty into local rock type juxtapositional relationships.

Patterns adopted by subsurface heterogeneity are of particular importance, especially where they embody connected permeability or impermeability. These can create or disrupt impact pathways.

Hydrogeological site knowledge is therefore comprised of a panoply of “might’s” and “maybe’s”. This raises many issues for decision-support modelling. The first issue is how best to represent this type of knowledge. Ideally it should be given mathematical expression so that geostatistical algorithms can create “realistic” realisations of what may exist underground, with these realisations constrained by direct measurements where they are available. However provision of such a stochastic description is difficult, as the disposition of subsurface features and properties is the outcome of complex natural processes rather than a random number generation process. Stochastic descriptors can only ever therefore be approximate. Furthermore, approximations may be greatest, and repercussions may be most serious, when attempting to characterise hydraulic property connectedness.

Another issue is that of scale. As was discussed in the previous section of this manuscript, a single model cell may encompass a volume that includes several different lithologies. The cell must therefore be endowed with suitably averaged hydraulic properties. Stochastic descriptors of subsurface heterogeneity that are formulated for use by a model must pertain to these cell-averaged hydraulic properties rather than to the properties of individual lithologies.

Actually, upscaling of hydraulic properties to the model parameter scale is generally more difficult than this (which is why it is rarely done for groundwater models). This is because groundwater model parameters often pertain to hydraulic properties that are averaged over much larger volumes than those of a single model cell. This is the case where parameterisation devices such as pilot points or zones of piecewise constancy are employed. Paradoxically, it is often simpler (but not necessarily more correct) to assign stochastic descriptors of hydraulic proper heterogeneity to these abstract parameterisation devices than to those of the real world. Furthermore, values ascribed to these parameter types are relatively easy to adjust, either during history-matching or to express their uncertainties. However their use may entail a diminution of a modeller's ability to express hydraulic property connectedness. Representation of possible impact pathways (or interruption of these pathways) may suffer as a result.

As is discussed in the previous section, any feature of a model that is incompletely known should ideally be represented as a parameter. This allows it to be awarded stochastic status. Many model parameters may therefore represent system features other than subsurface hydraulic properties. They may, for example, represent historical pumping rates, boundary condition specifications, and the spatial and temporal disposition of recharge. As for subsurface properties, pertinent parameters should be provided with stochastic descriptors that reflect both their innate variability, the degree to which they are inter-related, and their temporal and spatial connectedness. Such descriptors are likely to be approximate.

Taken together, the suite of probabilistic descriptors of parameter variability and parameter inter-relatedness that emerges from site characterisation studies embodies what is called their "prior parameter probability distribution". "Prior" refers to the fact that these descriptions are unconstrained by history-matching.

Parameters are imbued with information when they are collectively assigned a prior probability distribution. This distribution expresses their uncertainties, at the same time as it simultaneously places limits on their uncertainties. Some of these limits are expressed through stochastic linkages with other parameters that discourage totally independent behaviour of any one parameter. However if restrictions that are placed on parameter adjustability and linkages are too stringent (for example if inappropriate use is made of zone-based parameterisation, or if stochastic descriptors of upscaled parameter variability preclude the emergence of hydraulic property connectedness at a site where connectedness is a hydrogeological possibility), the prior parameter probability distribution expresses disinformation. This false information is then hosted by the model.

3.3 Bayes Equation

The role of history-matching is to constrain parameter adjustability. If (as is usually the case) it is incumbent on a model to reproduce the measured behaviour of a system before making predictions of its future behaviour, then the range of values that parameters can adopt – both individually and in concert – is reduced. This reduction in parameter variability comprises the harvesting of information from field measurements of system behaviour.

Acquisition of information from measurements of system behaviour is described using Bayes equation. It can be written as follows:

$$P(\mathbf{k}|\mathbf{h}) \propto P(\mathbf{h}|\mathbf{k})P(\mathbf{k}) \qquad (3.1)$$

In equation 3.1, the vector **k** contains model parameters. (It is a vector because it is a collection of numbers; vectors are normally represented by bold, lower-case letters.) The vector **h** comprises field measurements. The symbol $P()$ denotes probability.

P(***k***) depicts the prior probability distribution of parameters **k**. *P*(**h**|**k**) is referred to as the "likelihood function". This increases to the extent that model outputs replicate field measurements of system behaviour. However a perfect fit with these measurements is not required. The probability distribution on which the likelihood function is based is that associated with measurement error (sometimes referred to as "measurement noise"). Hence a certain degree of model-to-measurement misfit is expected.

P(**k**|**h**) denotes the posterior probability distribution of parameters **k**. The vertical bar between **k** and **h** (i.e. the "|" symbol) is normally read as "conditional on". Hence the posterior parameter probability distribution denotes the probability distribution of parameters, conditional on pertinent model outputs respecting field measurements.

The posterior variability of some parameters can be reduced significantly through history-matching. That of others may be reduced very little. Permissible spatial and temporal relationships between parameters may undergo change through history-matching. Some potential impact pathways may be closed; others may be unaffected.

Suppose that the following conditions are met.

1. The relationship between model outputs **h** and parameters **k** under history-matching conditions is linear, and hence can be represented by a sensitivity matrix. We denote this matrix as **Z**.
2. The relationship between a single predictive model output *s* and parameters **k** is linear, and can therefore be represented by a sensitivity vector. We denote this vector as **y**.
3. The prior probability distribution of model parameters is multiGaussian (i.e. multinormal). Encapsulated in this probability distribution is the prior covariance matrix of parameters, C(**k**).
4. The probability distribution of measurement noise **ε** is also multiGaussian. Its covariance matrix is C(**ε**).

The diagonal elements of a covariance matrix express variability of individual elements of a random vector. Off-diagonal elements express interdependency between these elements, that is, the tendency for them to vary together either in the same or opposite directions.

For those interested, the above four conditions can be expressed mathematically as follows. For simplicity (and without loss of generality) we assume that the prior mean of **k** is **0**. The following equations are therefore written in terms of departures of parameters from their prior mean values.

Under calibration conditions:

$$\mathbf{h} = \mathbf{Zk} + \boldsymbol{\varepsilon} \tag{3.2}$$

That is, what we measure in the field is the action of the model **Z** on its parameters plus measurement noise. Under predictive conditions:

$$s = \mathbf{y}^{T}\mathbf{k} \tag{3.3}$$

where the "T" superscript denotes matrix transpose. That is, the value of a prediction is the sum of the sensitivity of that prediction to each model parameter times the value of the respective parameter.

$$\mathbf{k} \sim N(\mathbf{0},C(\mathbf{k})) \tag{3.4}$$

That is, the prior probability distribution of parameters **k** is multiGaussian (i.e. multinormal), with a mean of zero and covariance matrix C(**k**).

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0},C(\boldsymbol{\varepsilon})) \tag{3.5}$$

That is, measurement noise has a mean of zero; its distribution is multiGaussian with covariance matrix C(**ε**). C(**ε**) is often assumed to be diagonal (and to therefore possess zero-valued off-diagonal elements). This is because the error associated with one measurement is presumed to be independent of that associated with any other measurement. (This is often a questionable assumption, especially where model-to-measurement misfit is an outcome of model imperfections.)

Under the above conditions, it can be shown that the posterior predictive probability distribution of the model prediction is also normal. Let σ^2_s denote its variance. (Variance is the square of standard deviation). This can be calculated using the following equation:

$$\sigma_s^2 = \mathbf{y}^{\mathrm{T}}\mathrm{C}(\mathbf{k})\mathbf{y} - \mathbf{y}^{\mathrm{T}}\mathrm{C}(\mathbf{k})\mathbf{Z}^{\mathrm{T}}[\mathbf{Z}\mathrm{C}(\mathbf{k})\mathbf{Z}^{\mathrm{T}} + \mathrm{C}(\boldsymbol{\varepsilon})]^{-1}\mathbf{Z}\mathrm{C}(\mathbf{k})\mathbf{y} \tag{3.6}$$

The details of equation 3.6 are of secondary importance. The form of this equation matters most. The first term on the right of equation 3.6 expresses the prior uncertainty of the prediction. The second term expresses reduction in its uncertainty accrued through history-matching. It is thus a direct measure of the information content of a calibration dataset as it pertains to the prediction.

It can be easily shown using equation 3.6 that history-matching may reduce the uncertainties of some predictions considerably, while having little impact on those of other predictions. It follows that there may be decision-support circumstances where it is not necessary for a model to harvest all of the information that is resident in a history-matching dataset. Harvesting of information that reduces the uncertainties of predictions that do not pertain to management failure may be nice, but is not critical. Identification of what information is worth harvesting, and construction of a decision-support modelling workflow that is capable of harvesting this information, are central to the notion of problem decomposition.

The role of history-matching in reducing uncertainty is illustrated graphically in figure 3.1. This figure depicts the prior probability distribution of a prediction alongside the posterior probability distribution of the same prediction. The posterior probability distribution is narrower and taller than the prior probability distribution. It is narrower because its predictive uncertainty is smaller than its prior uncertainty. It is taller because the area under any probability distribution curve must integrate to 1.0.

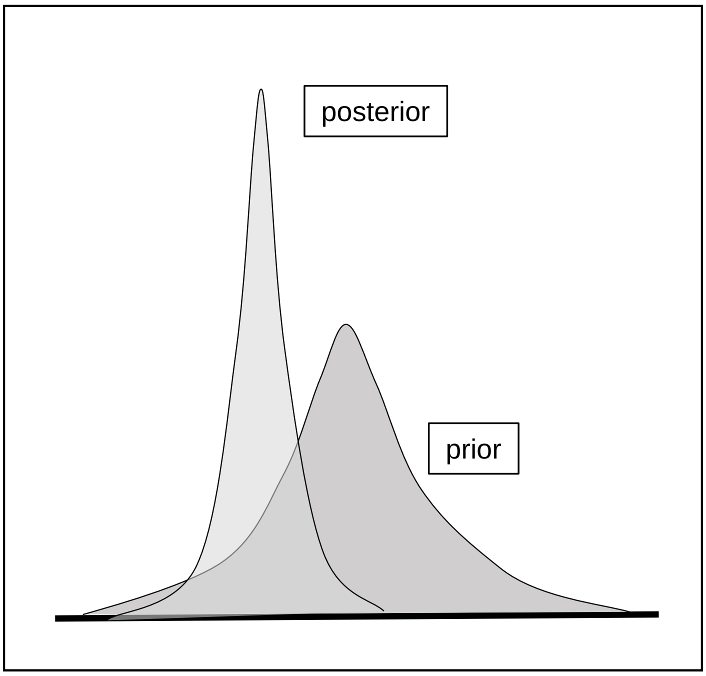

Figure 3.1. Prior and posterior probability distributions of a prediction of interest.

3.4 Model Calibration

Bayes equation expresses the important principle that acquisition of knowledge through processing of data must be expressed in terms of probabilities. It gives no reason to imply that history-matching should result in parameter or predictive uniqueness.

In spite of this, groundwater models are often "calibrated". This implies a unique solution to a nonunique inverse problem, namely that of back-calculating parameter values from field measurements of system behaviour. Bayes equation provides no justification for this endeavour. Nevertheless, justification can be sought through other means, as long as calibration is undertaken as a precursor to posterior uncertainty analysis, and as long as the conditions under which parameter uniqueness is sought are well understood.

It can be shown that parameter uniqueness can be achieved by stipulating that all predictions made by a calibrated model must lie near the centres of their respective posterior probability distributions. It can also be shown that the unique parameter field that satisfies this special (and useful) constraint is often the simplest parameter field that enables model outputs to fit a measurement dataset.

Model calibration can be viewed as a form of hypothesis-testing. If calibration cannot achieve a good fit with a history-matching dataset, this indicates a need for model refinement. Alternatively, if the simplest parameter field that allows model outputs to fit field measurements of system behaviour is more complex than expected, or requires parameters to adopt unrealistic values, this has important didactic implications. It can indicate that subsurface hydrogeology is more complex than was hitherto imagined. The prior parameter probability distribution may therefore need revision. Alternatively, it may indicate that parameters are compensating for model structural errors as they are adjusted to fit the measurement dataset. As is discussed elsewhere in this manuscript and in White et al (2014), this may, or may not, be a problem. If it is a problem, calibration evinces the problem, and may provide guidance on how to rectify it. Perhaps the conceptual model on which the numerical model is based is in need of refinement.

The concept of model calibration also provides some important insights into problem decomposition. As we show below, it demonstrates how a model prediction can be decomposed into two components. One of these components is informed by field measurements while the other is uninformed by these measurements. In doing this, calibration exposes information trajectories that can have important consequences for model design.

"Regularisation" is the mathematical term that is used to describe the quest for uniqueness. Parameter uniqueness implies parameter simplification. Ways in which regularisation is implemented fall into three broad categories. They are as follows.

1. Manual regularisation is achieved through use of a parsimonious parameterisation scheme such as zones of piecewise constancy. This type of regularisation is far from optimal, as it is unlikely to yield parameter sets for which all model predictions are close to their posterior means.
2. Subspace methods, of which singular value decomposition is the flagship.
3. Tikhonov methods which, if formulated in certain ways, are closely related to Bayesian methods.

If the action of a model on its parameters is linear (see equation 3.2), regularisation yields a matrix **G** which acts on a measurement dataset **h** to yield calibrated parameters $\underline{\mathbf{k}}$. This is described conceptually by the following matrix equation:

$$\underline{\mathbf{k}} = \mathbf{G}\mathbf{h} \tag{3.7}$$

See Doherty (2015) for how to calculate **G** using the above three regularisation schemes. If the calibrated model is now employed to make a prediction *s*, it is easily shown that the error variance $\sigma^2_{s-\underline{s}}$ of this prediction is calculable using the following equation:

$$\sigma^2_{s-\underline{s}} = \mathbf{y}^{\mathrm{T}}(\mathrm{I}-\mathbf{R})\mathrm{C}(\mathbf{k})(\mathbf{I}-\mathbf{R})^{\mathrm{T}}\mathbf{y} + \mathbf{y}^{\mathrm{T}}\mathbf{G}\mathrm{C}(\boldsymbol{\varepsilon})\mathbf{G}^{\mathrm{T}}\mathbf{y} \tag{3.8}$$

where:

$$\mathbf{R} = \mathbf{GZ} \tag{3.9}$$

The square root of error variance is error standard deviation. Hence equation 3.9 expresses the potential for wrongness of predictions made by a calibrated model. This is generally similar in magnitude to (but slightly larger than) the posterior uncertainty of a prediction.

3.5 Decomposing the Problem

Equation 3.8 looks complicated. However it is conceptually simple, and has much to say about problem decomposition.

The equation has two terms. Each of them is a scalar (i.e. a single number). This is despite the fact that the number of elements which comprise the vector **k** (that is, the number of model parameters) may be high, and the model itself may be complex. What turns a complex parameter vector into a scalar is its prediction-specificity; this is achieved through pre-multiplication by $\mathbf{y}^{\mathrm{T}}$. (Recall that **y** expresses the sensitivity of a prediction *s* to model parameters **k**.)

The second term of equation 3.8 can be thought of as "the machine learning term". This embodies what can be learned about the future behaviour of a system from measurements of its past behaviour. This term makes no reference to prior parameter uncertainty – only to uncertainties in field measurements.

The first term of equation 3.8 is the expert knowledge term. At its heart is the prior parameter covariance matrix C(**k**). It expresses the fact that where past system behaviour does not fully inform future system behaviour, reliance must be placed on knowledge of site properties that emerges from site characterisation. As has been discussed, this type of knowledge requires stochastic expression.

If regularisation is properly implemented as a model is calibrated, the two terms of equation 3.8 are independent of each other. (This is an outcome of the fact that the **G** and (**I**-**R**) matrices should be orthogonal to each other – or nearly so.) Hence any model prediction, together with its associated uncertainty, can be neatly subdivided into two separate components - the machine learning component and the prior knowledge component. The latter is only required to the extent that the former is limited. For some model predictions, the machine learning term may be dominant. Meanwhile, other predictions made by the same model may be prior knowledge dominated. This has important implications for design of a decision-support model. In particular, it suggests that a model that is designed to make one type of prediction may be history-matching-orientated, while that which is designed to make another type of prediction may focus on stochastic expression of expert knowledge. This, and the fact that each of the terms in equation 3.8 is a scalar, accentuates the advantages of prediction-specificity in model design.

The principles that are embodied in equation 3.8 extend beyond the equation itself. It can be shown that if a model prediction is machine-learning-dominated, then its post-calibration integrity is independent of those of the parameters to which it is sensitive. Values ascribed to the latter during model calibration do not need to be hydrogeologically plausible according to metrics that arise from expert knowledge, i.e. that arise from C(**k**); recall that C(**k**) is absent from the second term of equation 3.8. All that is required is that model parameters allow the model to replicate the past. The uncertainty of the prediction then depends only on the extent of model-to-measurement misfit and the measurement noise that this evinces. The decision-support model then becomes a machine learning tool. The role of its "machine-learning parameters" is to convey information directly from field data to a model prediction; there is no requirement for them to express expert knowledge. At the same time, model-based machine learning is computationally efficient because a model's parameters provide ready receptacles for prediction-pertinent information. A machine-learning-dominated modelling process therefore requires construction and deployment of a structurally simple model that hosts enough parameters to enable attainment of a good fit between pertinent model outputs and a history-matching dataset.

At the other extreme are prior-knowledge-dominated predictions. There is little need for a model that must make a prediction of this type to be history-matched, for history-matching will reduce the uncertainty of this prediction only minimally. Adaption of the modelling process to its required prediction may then take one of two different forms. One option is to make the model, and its parameterisation, as "realistic" as possible so that associations between parameters and real-world hydraulic properties are clear; ideally, this can facilitate construction of an appropriate prior parameter probability distribution. History-match-constrained adjustment of these parameters is not required. Alternatively, a modeller may decide to focus on the pessimistic end of the posterior predictive probability distribution by undertaking worst case scenario analysis based on a particular impact pathway. As explained by Doherty (2015) and Doherty and Moore (2022), it is dangerous to history-match such a model, as its simplistic structure may induce predictive bias. However if history-matching is foregone because it contributes little to predictive uncertainty reduction, then this danger is eliminated.

Most predictions of decision-support interest lie between these two extremes. It is then incumbent on a modeller to respect both of the terms of equation 3.8, while maintaining their independence. This is where skills in problem decomposition are warranted.

3.6 Singular Value Decomposition

Singular value decomposition is a numerical device that decomposes a matrix into other matrices that have special properties, of which orthogonality is one. We mention it in this manuscript because of the insights that it provides. The fact that it shares the term “decomposition” with the subject matter of this manuscript is no coincidence.

We present a few equations for the benefit of those who are interested. Details of these equations are unimportant; only the insights that they provide are relevant.

Consider the matrix **Z** of equation 3.2. It can be shown that for any matrix **Z**:

$$\mathbf{Z} = \mathbf{USV}^{\mathrm{T}} \tag{3.10}$$

where **U** and **V** are orthonormal matrices. This means that the columns of these matrices are comprised of orthogonal unit vectors. Collectively, these vectors define new coordinate systems in model output space and parameter space respectively. Meanwhile **S** is a matrix whose non-diagonal elements are all zero. Its diagonal elements are positive or zero; these elements are referred to as “singular values”.

The matrix **V** can be partitioned into two orthogonal submatrices $\mathbf{V}_1$ and $\mathbf{V}_2$. Their respective columns subdivide parameter space into two orthogonal subspaces. The first of these subspaces is informed by field measurements of system behaviour. The second is entirely uninformed by these measurements. The former subspace is referred to as the “solution space”, while the latter subspace is referred to as the null space. Parameter combinations which lie within the null space cannot be back-calculated through model history-matching; conversely, those which lie within the solution space can.

In recognition of SVD-enabled parameter space decomposition, equation (3.10) can be re-written as:

$$\mathbf{Z} = \begin{bmatrix}\mathbf{S}_1 & \mathbf{S}_2\end{bmatrix}\begin{bmatrix}\mathbf{V}_1^{\mathrm{T}} \\ \mathbf{V}_2^{\mathrm{T}}\end{bmatrix} \tag{3.11}$$

Where calibration is achieved through singular value decomposition, equation (3.8) becomes particular simple:

$$\sigma^2_{s-\underline{s}} = \mathbf{y}^{\mathrm{T}}\mathbf{V}_2\mathbf{V}_2^{\mathrm{T}}\mathbf{y} + \mathbf{y}^{\mathrm{T}}\mathbf{V}_1\mathbf{S}_1^{-2}\mathbf{V}_1^{\mathrm{T}}\mathbf{y} \tag{3.12}$$

Matrices that comprise the two terms of equation (3.12) are orthogonal; the terms are therefore independent. As stated above, the second term depicts what can be learned about the prediction from measurements of system behaviour, while the first term depicts what cannot be learned. For most predictions made by groundwater models, the second term is much larger than the first.

Equations such as 3.12, and others that are similar to it, provide powerful insights into what model calibration actually achieves. The most important of these insights is illustrated graphically in figure 3.2.

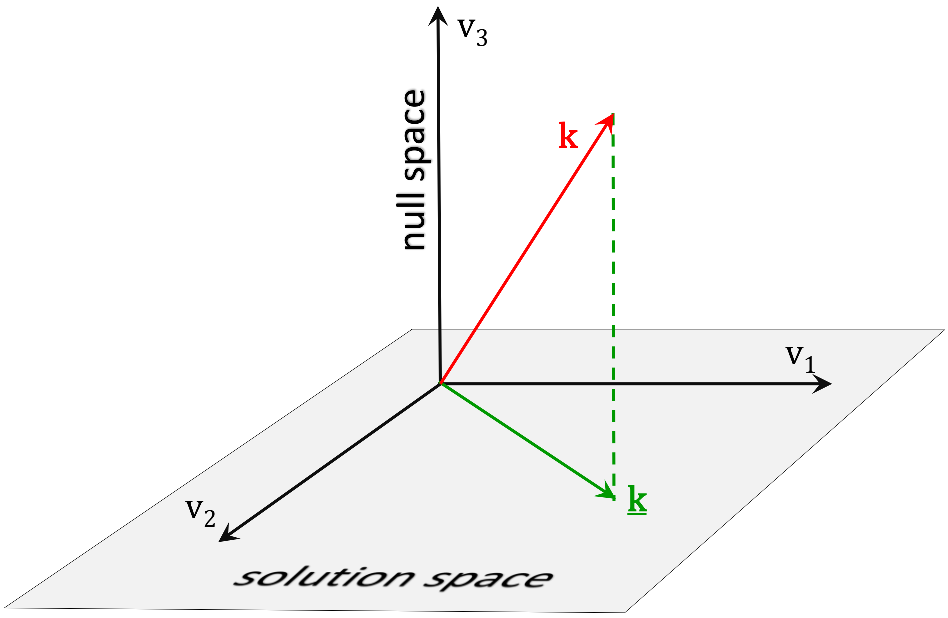

Figure 3.2. Model calibration achieved using singular value decomposition.

The large red **k** vector in figure 3.2 (drawn as an arrow), represents the real, but unknown values of a model's parameters. The green $\underline{\mathbf{k}}$ vector denotes the calibrated parameter set. This is the projection of reality (i.e. **k**) onto a generally low-dimensional parameter subspace – the subspace that is spanned by vectors comprising the columns of $\mathbf{V}_1$ (i.e. the unit vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in this three-dimensional example). This figure, and the equations of which this figure is a representation, therefore state that the most that can be inferred through model history-matching is the shadow of a far more complex reality. In practice, this means that we can never know more than broadscale, spatially averaged, hydraulic properties of a real world system. Hydraulic property detail is not ours to know; it can only be expressed stochastically, even after a model has been calibrated.

A challenge that is faced by decision-support modelling (and by problem decomposition that should support its implementation) is separation of that which can be known from that which cannot be known. The former can be back-calculated from measurements of system behaviour, while the latter must be expressed probabilistically. Gratuitous stochastic expression of that which is inferable from field measurements engenders model-to-measurement misfit. Information which is otherwise available for harvest is thereby rejected. In contrast, failure to provide adequate stochastic expression of that which is uninformed by field measurements can result in under-representation of predictive uncertainty.

We conclude this subsection with an important observation. The neat orthogonality provided by singular value decomposition that is pictured in figure 3.2 and explained in equation 3.12 relies on an important assumption. Mathematically, this is written as:

$$C(\mathbf{k}) \propto \mathbf{I} \qquad (3.13)$$

That is, C(**k**) is proportional to the identity matrix. This implies that parameters (which include system hydraulic properties) have no spatial correlation. Obviously, this is not the case. However, in theory, this is not a problem because the parameters **k** of a model can undergo mathematical transformation in order to satisfy equation 3.13. The above concepts can then be applied to transformed parameters.

Unfortunately, transformation of a model's parameters to respect equation 3.13 requires that two important conditions be met. The first is that C(**k**) is actually known, for the orthogonality-enabling transformation is derived from C(**k**). The second is that real-world complexity can be described by a C(**k**) matrix at all. In most hydrogeological contexts neither of these conditions is met. Violation of either of these conditions erodes orthogonality. This can inculcate predictive bias. Bias is the subject of the next section. Unfortunately, bias can rarely, if ever, be completely eliminated; hence it is important that it be managed.

Ideally, mental decomposition of a management problem can lay the conceptual groundwork for a decision-support modelling workflow that achieves qualitatively similar outcomes to those of singular value decomposition. A few examples of practical problem decomposition follow.

3.7 Three Examples

3.7.1 Simulation of Recharge

Software to which groundwater modellers have access provides a range of options for simulating unsaturated zone processes, including groundwater recharge. These options range from very complex to very simple. At the complex end of the spectrum are packages such as HYDRUS that employ Richards equation to simulate movement of water through the vadose zone. At the simple end of the spectrum are lumped parameter soil moisture store simulators such as LUMPREM (downloadable from the PEST web pages). Figure 3.3 schematises both of these simulators. We refer to them as "model A" and "model B" respectively in the discussion that follows.

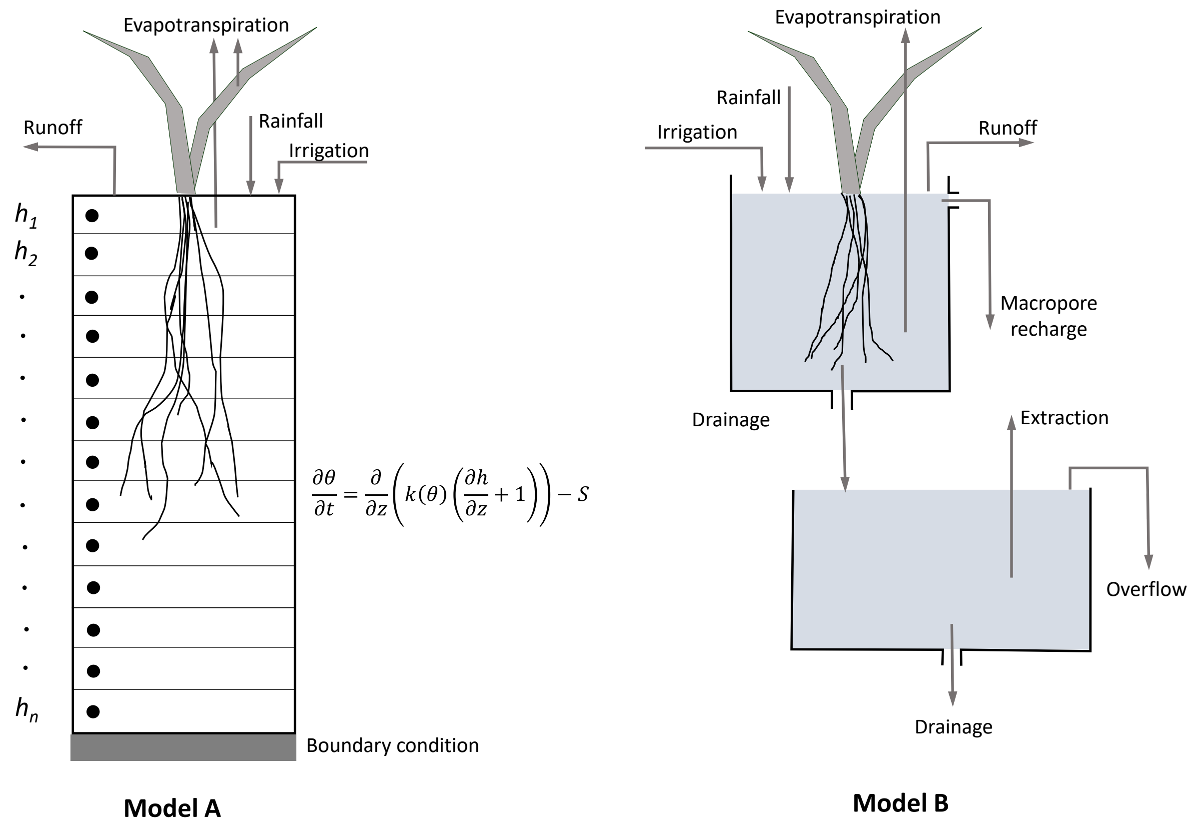

Figure 3.3. Schematic of a Richards equation model on the left (model A), and a lumped parameter soil moisture store model on the right (model B).

The Richards equation simulator is structurally complex. Because of this, it has the capacity to host many parameters. These parameters characterise soil properties such as saturated hydraulic conductivity, van Genuchten soil water retention characteristics, and relationships between water saturation and hydraulic conductivity. A suite of these properties can be associated with every node of a model's grid; however simplifying assumptions such as vertical uniformity, or limited vertical zonation, are often made. In addition to this, model A hosts parameters that govern interchange of water between soil and plant roots, between vegetation and the atmosphere and between soil and the atmosphere. Model A may run reasonably quickly if it is one-dimensional. However, two- and three-dimensional variants of this model have much greater run times. Severe nonlinearities associated with simulation of unsaturated flow may sometimes inflict solution convergence difficulties on models of this type. These can often be overcome using numerical strategies such as adaptive time-stepping. Unfortunately, these strategies may threaten continuity between model parameters and model outputs, this degrading the history-matching performance of model A.

Model B runs in a fraction of a second over arbitrarily long time periods. It has between 6 and 10 parameters depending on whether, or not, the lower moisture store is activated. Though it simulates nonlinear system behaviour, it rarely encounters numerical difficulties.

Defining the problem

Let us assume that we need to select one of these models to calculate recharge for a regional groundwater model. Let us further suppose that the regional groundwater model is tasked with supporting irrigation sustainability. Therefore different incidences of either model A or model B will be used to calculate irrigation demand and recharge under different hydrogeological units (HGUs); a HGU normally specifies a unique combination of land use and soil type. Therefore, each of these models must simulate processes that are presumed to be reasonably spatially uniform over considerable areas.

Sources of information

Suppose that the dataset that is available for history-matching of the regional model includes the following (as is fairly typical):

- time series of groundwater heads in a number of observation wells;
- spot estimates of steady-state recharge derived using methods such as salt mass balance;
- historical records of irrigation water delivery from surface water sources to different sectors of the irrigation area.

So which of the above unsaturated zone simulators should be used?

Flow of information

Model B can be considered as a simplification (i.e. a decomposition) of model A. It operates in a much smaller dimensional parameter space than does model A. Its components, and the parameters that are associated with these components, embody undefined combinations of model A components and parameters. (These combinations can actually be identified; see Watson et al, 2013). Model B is easily history-matched, either on its own or in combination with the regional groundwater model for which it provides inputs. As has been discussed, history-matching comprises information harvesting. This reduces the uncertainties of some parameters, together with the uncertainties of predictions that are sensitive to these parameters.

However, it is possible that history-matching of model B may introduce bias to some model predictions. This is because parameter and process decomposition that is implied in its simplification of model A may not be exactly aligned with singular value decomposition applied to the parameters of model A (see the next section of this document). At the same time, the limited number of parameters that model B possesses may limit its ability to express uncertainty arising from information insufficiency. Whether this matters or not depends on the prediction. In general, predictive extremes are likely to be more susceptible to bias and uncertainty underestimation than are predictions of system behaviour under more benign conditions.

Field measurements of soil properties can be more directly linked to the parameters of model A than they can to the parameters of model B. These linkages admit entry into the modelling process of information that is resident in soil mapping and soil property inferences. Like all other information, this has the capacity to reduce the uncertainties of some of its parameters. In doing so, it may reduce the uncertainties of some of its management-salient predictions. Whether it does this or not depends on the sensitivities of these predictions to parameters that are informed by these data. It also depends on the availability of other information that also has the capacity to reduce the uncertainties of these predictions.

The advantages of expert-knowledge-informed parameterisation of model A may, however, be outweighed by some of the numerical disadvantages associated with its use. These include longer run times – especially when imbued with random realisations of those of its parameters which are uninformed by expert knowledge for the purpose of predictive uncertainty evaluation.

Model B does not suffer from these numerical difficulties. However, its use must rest on a demonstrable ability to represent the full range of recharge and irrigation demand contingencies that can be simulated by model A. These include:

- recharge under high rainfall;
- recharge under conditions of both high and low antecedent moisture;
- loss of soil moisture through evapotranspiration under a variety of climatic conditions;
- recharge recession with increasing time since the last rainfall or irrigation event.

Other considerations also deserve attention. The selected model must be used to simulate spatially-averaged recharge over an area that is much larger than that of a soil column. Soil is notoriously heterogeneous. It is natural to ask whether Richards equation applies to a very large number of soil columns operating in parallel, each of which has different soil properties. Furthermore, what "average" hydraulic properties should be awarded to a single simulator that represents recharge through all of these soil columns. A nonlinear upscaling problem therefore presents itself. The benefits of soil property measurements at individual sites begin to fade under these circumstances.

The composition of the history-matching dataset must also be considered when choosing between the two models. Intuitively, records of historical water delivery to different parts of an irrigation area are information-rich with respect to future demand for irrigation (perhaps under altered cropping and/or altered climate). Which of the two models is able to better harvest this information?

Unfortunately, history-matching against regional water delivery data is problematical, for these data pertain to the average actions of many individual irrigators who apply water on an as-needed basis to their individual farms at individual times. Hence a model that is capable of retrieving information from regional water delivery records must simulate average conditions rather than evaporative demand and irrigation events at a single site. This requires some abstraction in simulating unsaturated zone processes. This, in turn, requires temporal as well as spatial upscaling of soil processes and properties while preserving the model's ability to simulate spatially and temporally averaged recharge. Model A may be too "realistic" to handle this situation. In contrast, because model B is already abstract, it can be more easily turned to this task; see Doherty (2021).

It is apparent that the choice between models A and B is multi-faceted and, ultimately, subjective. Because it runs fast, and can capture prediction-pertinent information from historical measurements of system behaviour, model B is attractive. However the decision-pertinent predictions which the overall model (i.e. the groundwater model together with its "feeder" recharge models) must make must also be born in mind. If conditions in the unsaturated zone are salient to management failure, then model A may be the better choice. If this is the case, then consideration may then be given to use of a model such as HydroGeoSphere that implements full numerical coupling between the unsaturated and saturated zones. However with every advantage comes a disadvantage. Long run times which accompany this choice would then make harvesting of information from historical system behaviour a very difficult undertaking indeed.

Choice of an appropriate simulation strategy may therefore be complicated. In making these choices it is important that a modeller bases them on the ability (or otherwise) of a particular modelling strategy to quantify and reduce the uncertainties of management-salient predictions. In doing so, he/she must recognise that the context in which he/she operates is defined as much by flow of information as it is by flow of water.

3.7.2 Steady State or Transient

It is often presumed that a transient groundwater model provides superior decision support to that which a steady state model provides. The premise is that "simulation" and "steady state" are incompatible concepts as most natural systems are in a state of perpetual change.

The problem

As always, considerations pertaining to decision-support modelling appropriateness must begin with the management-salient prediction that a model must make and sources of prediction-pertinent information that the model must harvest.

If a study site will soon undergo large and permanent alterations to its boundary conditions or stresses, then predictions of great importance are likely to be associated with the spatial distribution of the new equilibrium water level. The time required to achieve this new equilibrium, and amplitudes of seasonal oscillations about this new equilibrium, may also be of some importance. However the integrity of these latter predictions will depend on the integrity of the equilibrium prediction.

Equilibrium water levels are determined by the conductance properties of a system. It follows that, to the extent that the uncertainties of conductance parameters can be reduced, the uncertainties of new equilibrium water level predictions can be reduced.

Sources of information

Steady state history-matching allows a calibration dataset to directly inform system conductance parameters without having to simultaneously inform its storage parameters and parameters that govern seasonal variability of recharge. Obviously, a steady state dataset hosts less information than a transient dataset. However, a steady state dataset is asked to inform fewer parameters. Furthermore, the parameters that a steady state dataset does not inform are of secondary importance to evaluation of new equilibrium water levels.

Flow of information

Problems with storage of harvested information within a model are also reduced when undertaking steady state history-matching. Appropriate dissemination among different parameter types of information that enters a groundwater model through history-matching depends on prior probability distributions that are ascribed to these parameter types, as well as on their joint prior probability distribution. If prior parameter probability distributions are unsuitable, information is misdirected. This can artificially constrain parameters, or combinations of parameters, that are decreed to be highly uncertain instead of illuminating parameters, and combinations of parameters, that have wrongly been described as relatively certain. As will be discussed in the next section of this document, this can bias some important model predictions.

It may be concluded, therefore, that steady state history-matching is appropriate for models that are built to address some management issues. Large transient model run times, and the difficulties of building a transient model, may contribute to this conclusion at some study sites. A small amount of management-salient information is thereby foregone. The uncertainties of some predictions may rise as a consequence. Hopefully, however, the amount by which they rise is smaller than the predictive bias that may be incurred by attempts to harvest the foregone information.

Nevertheless, depending on the site, other considerations may come into play. “Steady state” is more of an artificiality at some sites than it is at others. This has implications for definition of “steady state” heads that comprise a history-matching dataset. One modeller may select average heads, while another may select heads that were measured during a time period over which they were relatively stable. Both of these choices may be justifiable. However, an inability to directly measure steady state heads suggests that surrogates for these heads will be accompanied by “measurement noise”. A modeller must therefore accept a level of model-to-measurement misfit that is commensurate with this noise. Information is thereby rejected. At the same time, the statistics of model-to-measurement misfit that is incurred in this manner are unknown; furthermore, the pertinent $C(\boldsymbol{\varepsilon})$ matrix (see equation 3.5) may not be diagonal. The assumption that it is, indeed, diagonal may bias estimated model parameters, and with them some model predictions.

Trade-offs in model design, and in how a model is history-matched, are therefore inevitable. A propensity for predictive bias is traded off against the ability of harvested information to reduce uncertainty. While uncertainty can (in theory) be calculated by a model, bias cannot. Choices will therefore be personal. It cannot be otherwise, so this is not a problem. Decision-support modelling only becomes problematic if the need for trade-offs is unrecognised.

See Moore and Doherty (2021) for a further discussion of trade-offs between steady state and transient model calibration.

3.7.3 Separation of Impact Pathways

Figure 3.4 schematises a particular regulatory context.

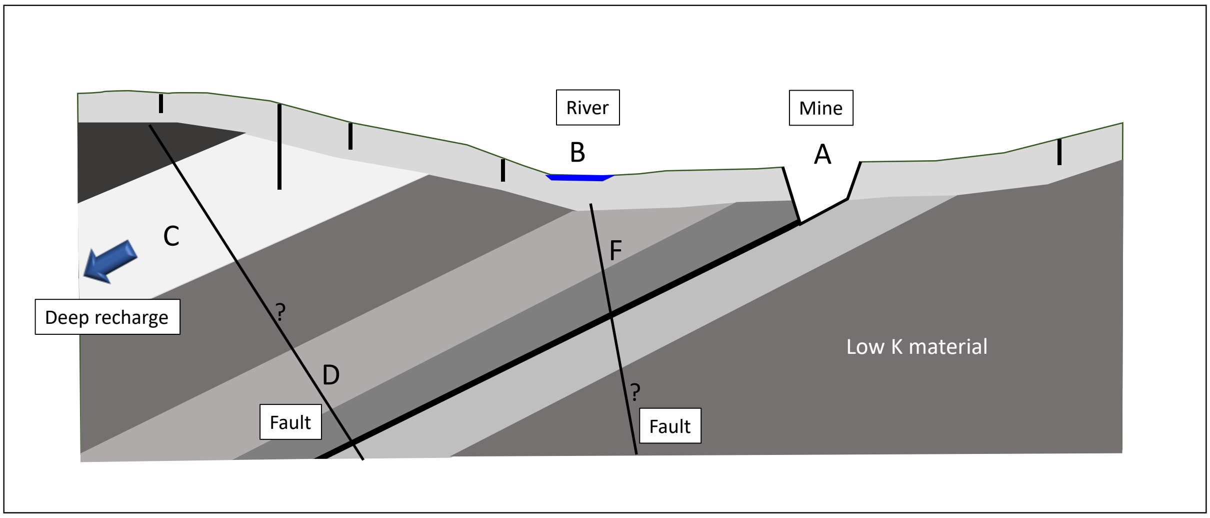

Figure 3.4. Schematisation of the impact of a mine.

The problem

A mine is proposed at location A. Excavation will proceed below the water table, so the mine will lower the phreatic surface. Of interest is its impact on the river at location B. Also of interest is its impact on aquifer C that supplies recharge to a deep regional artesian system. It is conjectured that drawdown that propagates along the hydraulically conductive coal seam may then propagate up a possible fault at location D, thereby withdrawing water from aquifer C. The local system may contain other faults; fault F may transmit coal seam drawdown to the river. However the existence, locations, and hydraulic properties of faults are only vaguely known.

It follows from the above that there are multiple impact pathways that connect the proposed mine to multiple impact points. Each of these pathways possesses different hydraulic characteristics.

Sources of information

We assume that data are scarce (a common occurrence in regulatory contexts such as this). We assume that a limited number of wells penetrates the surficial system, often near creeks. Fewer wells probe the deeper system. Water levels have been measured in all of these wells.

Flow of information

The conventional approach to regulatory modelling in a hydrogeological context such as this is to construct a model that reflects the complexity of the site. This model is then used to examine all impact pathways. The complexity (and expense) of the model serves a social purpose as well as a scientific purpose. It demonstrates to the public the seriousness with which potential impact is being treated by regulatory authorities.

Unfortunately, complex models that address issues such as those that are depicted in figure 3.4 often fail to provide the regulatory support that is required of them. As often as not, they end up in court. However here we focus only on their practical difficulties. They run slowly. This obstructs data assimilation and uncertainty quantification. At the same time, they are cumbersome and inflexible, and hence difficult to use in a hypothesis-testing capacity.

A critical surface in the geological model that is schematised in figure 3.4 is that which defines the base of the weathered zone. This surface separates the shallow regolith groundwater regime that is characterised by moderate to high hydraulic conductivities from the deeper system that is characterised by low hydraulic conductivities. Problems in representing this boundary in a complex model are exacerbated by the nature of numerical linkages that a typical simulator provides; see figure 3.5. In a model, layers that subcrop against the regolith unconformity are linked to the regolith layer through vertical model cell hydraulic conductivities; in reality the horizontal conductivity of sub-weathering strata probably dominates water exchange between the deep and shallow systems. This problem can be overcome by directly ascribing conductances to cell connections (where a simulator permits this). However this can get complicated.

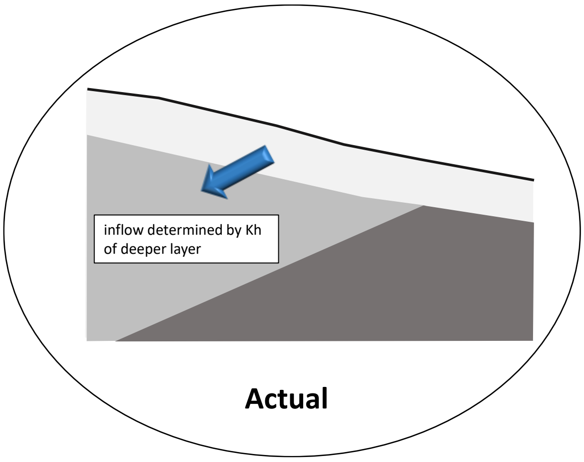

Figure 3.5. Schematic and numerical representation of the regolith unconformity.

In figure 3.6 we decompose the problem. Groundwater flow in the regolith is simulated separately from that in the deeper system using a one-layer, fast-running model. Simulation errors that are incurred by this partitioning are small if the amount of water that flows in the shallow system is much greater than that which enters the deeper system (which is generally the case).

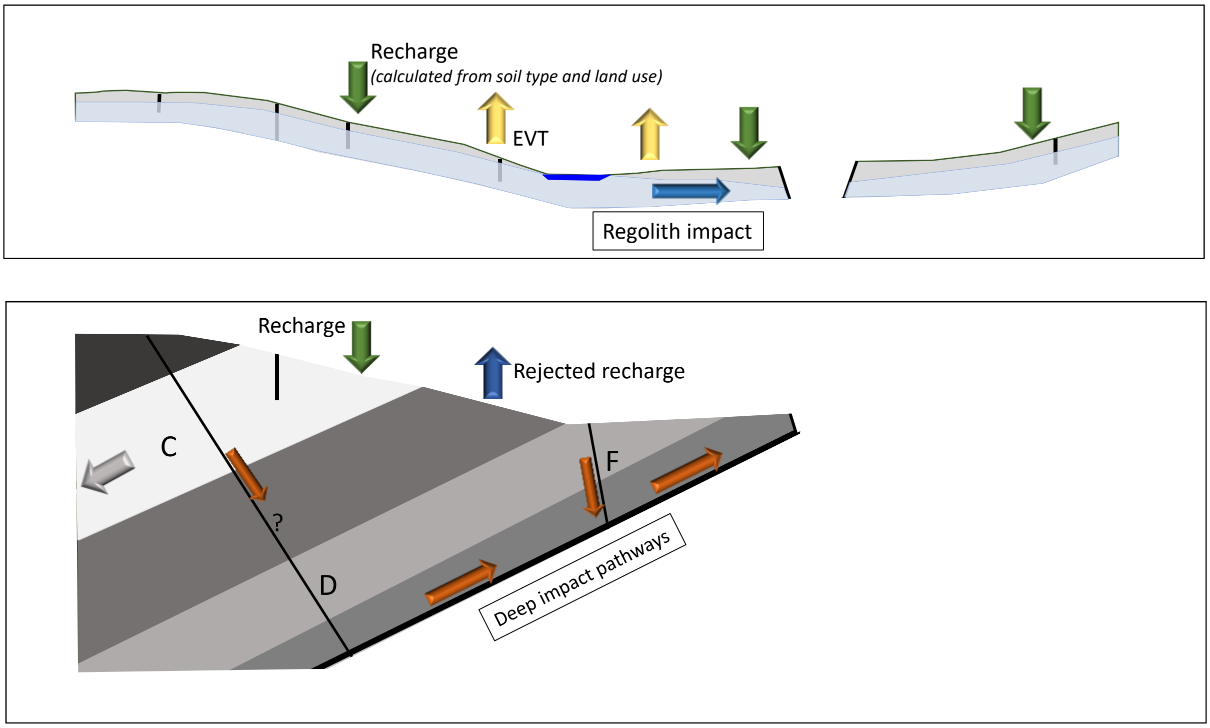

Figure 3.6. Problem decomposition according to impact pathways.

In constructing the regolith model, we assume that:

- steady state conditions prevail; hence there is no need to estimate storage parameters, and the model can be parameterised using transmissivity instead of hydraulic conductivity;
- steady state recharge can be reasonably well established from soil type and land use;
- shallow water tables under low-lying flat terrain are controlled by ET and surface seepage;
- groundwater tends to follow the topographic surface, being 30 m deep at the most.

The last point is important. It allows us to calibrate this fast-running model against heads that can be approximately inferred everywhere. Values of regolith transmissivity can thereby be inferred throughout the model domain. To be sure, these values will be uncertain, especially in flat-lying areas where groundwater levels are controlled by seepage and ET. Nevertheless, this easily-history-matched model provides access to the significant amount of information that resides in landscape attributes; the usefulness of this information is amplified by the lack of alternative sources of information.

Once it has been history-matched, this regolith model can be used to calculate the impact of the mine on the river through the regolith impact pathway. At the same time, the simplifying assumption of invariant transmissivity with groundwater elevation prevents underestimation of the ability of the mine to drain the regolith aquifer. Predictive uncertainties are easily evaluated using this model, and are more likely to be overestimated than underestimated.

Meanwhile another model can focus on the deep system. Here, uncertainties are dominated by the hydraulic conductivity of the coal seam, by the existence (or otherwise) of a small number of faults, and by the hydraulic properties of these faults. The effectiveness of these faults as impact pathways can be easily calculated using relatively simple models, once estimates have been made of mining-induced coal seam drawdown. The latter is easily evaluated, as is its uncertainty.

Design of the deep system model is facilitated through recognition of the fact that uncertainties in deep system impact pathways to the river and to aquifer C are dominated by faults. Consequently they are high. This reduces the burden of numerical simulation. Errors that are incurred through use of a relatively simple model to evaluate the possibility of impact are likely to be small in comparison with the uncertainties of these impacts.

4. PREDICTIVE BIAS

4.1 General

A clear and present danger that is posed by just about any type of model simplification (including structural and parameterisation simplifications arising from problem decomposition) is the unwitting introduction of predictive bias. This section discusses how predictive bias can arise, and what can be done about it.

A few mathematical concepts follow. Do not worry. They are harmless.

4.2 Bias in Groundwater Modelling

The previous section makes it clear that a groundwater model should not be asked to make a prediction using a single parameter set. Instead, it suggests that a groundwater model should be asked to make a prediction many times, using a parameter set that samples a probability distribution on each occasion. If history-matching has not been undertaken, the prior parameter probability distribution should be sampled. If parameters are subject to history-matching constraints, the posterior parameter probability distribution should be sampled. In either case, it is incumbent on a modeller to ensure that the predictive probability distribution which is sampled using the model is wide enough to span the "true" prediction, while being as narrow as information harvested from site characterisation data and measurements of historical system behaviour allows.

For reasons that are discussed in the present section, it is possible for the predictive probability distribution that is sampled in this manner to be in error. Sometimes this sampled distribution may be artificially shifted towards high or low values. This shift is referred to as "bias". The amount of bias is rarely known. To the extent that it can be guessed, the sampled predictive probability distribution should be widened. If the direction of bias is unknown, then it should be widened in both directions in order to ensure that it spans the true value of the prediction.

The existence of predictive bias implies that there is something wrong with the model that is making the prediction - something that endows it with a propensity for predictive error. It also implies that this propensity is unaccounted for in the model-evaluated predictive uncertainty interval.

We will loosely characterise predictive bias as the potential for model predictive error that arises from sources other than information paucity. As such, it is generally symptomatic of model defects. Obviously, because all numerical models are simplifications of reality, they all possess a propensity for predictive bias. Good model design should attempt to minimise bias in those of its predictions that it is required to make. At the same time, model usage should attempt to ensure that the potential for bias is included in model-quantified predictive uncertainty intervals.

This may be more difficult to achieve than it sounds. If a model is inclined to wrongness in making a particular prediction, it is likely to possess a proclivity for wrongness in evaluating the uncertainty of that prediction. This component of predictive uncertainty may therefore need to be evaluated outside of the modelling process, or simply guessed. However the imperative that modelling failure be avoided (see section 2.6), demands that an attempt be made to quantify the potential for predictive bias. Obviously, the smaller is this potential, the easier it is to accommodate.

4.3 Repercussions for Model Design

Ensuing subsections of the present section examine why a model prediction may be accompanied by bias. Identification of sources of model bias may lead to its reduction. However, while it may be possible to reduce bias for some predictions, it will rarely be possible to reduce it for all predictions made by the same model. A model must therefore be designed with minimisation of the bias associated with one or a small number of predictions in mind.

4.4 Prior Model Bias

Prior model bias is "inbuilt" model bias. This can have various degrees of severity. In the decision-support context, bias is severe if a model does not possess the ability to predict management failure. This can occur if it does not simulate failure-pertinent processes. It can also occur if a model's structure is too simple to represent impact pathways that induce management failure (for example if it has insufficient model layers, or does not represent certain structural features such as faults).

Prior bias may also result from incapacity of a model's parameterisation scheme to represent aspects of hydraulic property heterogeneity that can induce management failure. Of particular importance is hydraulic conductivity connectedness. Avoidance of parameterisation-induced predictive bias requires two things. Firstly, as has already been discussed, parameters must be represented at a level of spatial complexity that allows impact pathways to develop. (Beware of inappropriate use of zone-based parameterisation.) Secondly, the prior probability distribution of parameters (including their spatial correlations) must not impede the formation of these pathways.

It is apparent, therefore, that a suitably complex parameterisation scheme must be complemented by a suitably flexible prior parameter probability distribution. Though easy to state, this may be difficult to achieve in practice. In some circumstances it may require a three-dimensional, direction-dependent, spatially-variable, stochastic depiction of complex parameter interrelationships. Simplifying assumptions such as Gaussianality and stationarity (the latter implying that descriptors of parameter stochasticity prevail over large distances) may therefore be inappropriate.

In some modelling circumstances, inappropriate simplicity of a model's processes, structure and/or parameterisation may become apparent as a model is history-matched. This is because they may obstruct attainment of a good fit between model outputs and field measurements of system behaviour. In other cases, these simplifications may not erode a model's history-matching performance. However history-matching may require adoption by the model of strange parameter values. As we now discuss, this may, or may not, induce predictive bias.

4.5 Bias Induced by History-Matching

All models are defective. Sometimes their defects can be "calibrated out". Other times history-matching can exacerbate their potential for predictive error. This matter is discussed in detail by White et al (2014) and Doherty (2015).

As has been demonstrated in the previous section of this document, singular value decomposition offers insights into what happens when a model is history-matched. It can also offer insights into what happens when an inappropriately simplified model is history-matched.

Ideally, model simplification should mimic that which is implicitly undertaken by singular value decomposition acting on a perfect model. We use the matrix **W** instead of the matrix **Z** to specify the linearised form of this perfect model (see equation 3.2). Obviously we do not know **W**.

As has been discussed, singular value decomposition of any sensitivity matrix (including **W**) implements a kind of parameter simplification. This can be considered as a surrogate for model simplification. It does this by removing from adjustable parameter space those combinations of parameters (and, by inference, the structures and processes associated with them) that lie within the null space of the matrix. Solution space parameter combinations (which are generally smaller in number than null space parameter combinations) can then be estimated without bias because they are orthogonal to null space parameter combinations; the latter are not featured in the history-matching process. Meanwhile subspace orthogonality guarantees that adjustment of solution space parameter components does not incur intrusions into the information-unsubstantiated null space.

It is important to note that loss of all or part of the null space may compromise a model's ability to evaluate the uncertainties of some predictions. It is therefore incumbent on a modeller to simplify a model in a way that removes only prediction-irrelevant null space components from its parameterisation scheme and, by implication, from its design.

Figure 4.1 illustrates what can happen if model simplification is inappropriate for the data context in which simulation operates. In this figure, parameter space has three dimensions. It is spanned by three vectors $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ that form the columns of the **V** matrix that is obtained through singular value decomposition of **W**. We assume that the solution space has two dimensions; these are spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$. The null space has a single dimension; this is spanned by $\mathbf{v}_3$.

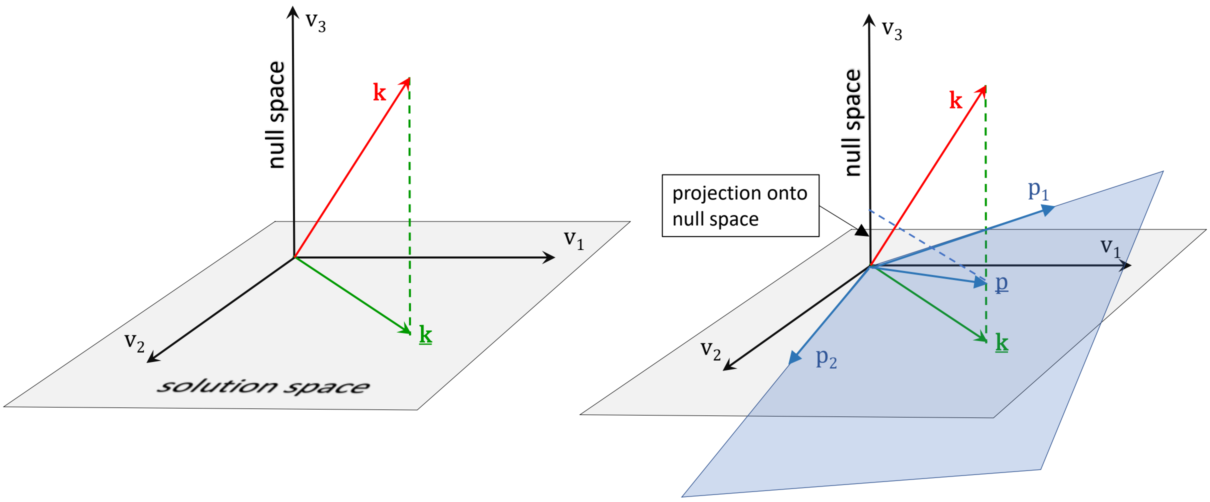

Figure 4.1. The right figure depicts how inappropriate model simplification can incur history-matching-induced parameter bias.

Let us suppose that a simplified model possesses just two parameters; these are $\mathbf{p}_1$ and $\mathbf{p}_2$. Because the dimensionality of the solution space is only 2, this is enough parameters to ensure a good match between simplified model outputs and field data. However, because design of the simplified model is imperfect, these simple model parameters are not quite aligned with the solution space of "reality".

Calibration of the simplified model yields the vector $\underline{\mathbf{p}}$. This has the correct projection onto the solution space of reality; this is why the simplified model fits the measurement dataset. This projection is correct because it coincides with the vector **k** that is the projection of the real-world parameter set onto the real-world solution space. However **p** has a non-zero projection onto the null space axis $\mathbf{v}_3$. History-matching of the simplified model therefore requires that at least some of the real world null space must adopt a value which is unsupported by data. The result is parameter bias.

Imagine a simple groundwater model with a fixed head boundary condition at its lower end and a measured head at its upper end. If the lower boundary's fixed head is incorrectly specified, the calibrated value of transmissivity must compensate for this in order for the model to replicate the head at the top of the system. Calibration therefore induces bias in estimated transmissivity.

While figure 4.1 seems to comprise an unnecessarily complicated way to explain the fact that some parameters may need to adopt compensatory roles as they are adjusted during history-matching of a simplified model, concepts that are expressed by this figure can be extended to examine some of the ramifications of this phenomenon. We now discuss them briefly. See Doherty (2015) for details.

Firstly, does calibration-induced parameter bias matter? It can be shown that this depends on the prediction that the model makes.

Suppose that a prediction of management interest is insensitive to parameters that have suffered calibration-induced bias. This prediction therefore escapes calibration-induced bias. It follows that if a structurally simple model possesses many parameters, some of these parameters may have the capacity to "soak up" prediction-irrelevant information contained within a measurement dataset. This same information would induce model-to-measurement misfit with unknown stochastic attributes if the model were endowed with a simpler parameterisation scheme. Sensitivity of a prediction to these simplified parameters would induce predictive bias.

More complex situations can (and usually do) arise. Suppose that a prediction of interest is sensitive only to parameter combinations that comprise the solution space of the real world sensitivity matrix **W**. This occurs when the future of a managed system is entirely informed by its measured past. It can be shown that the prediction suffers no bias even if it is made by a well-calibrated simplified model whose parameters suffer history-match-induced bias. As has already been discussed, a model such as this can be considered as a machine learning tool whose learning efficiency is enhanced by endowment with an information-ready parameter set.

In many cases, however, calibration-induced predictive bias cannot be avoided. The potential for this kind of bias must therefore be included in the posterior uncertainty margin of a prediction of management interest. This can be done in a number of ways. The simplest is to add an "engineering safety margin" to the prediction that extends its uncertainty interval beyond that which can be evaluated using the model itself. Alternatively, as Mathews and Vial (2017) show, it can also be accommodated by magnifying the prior probability distributions of affected parameters to accommodate variability incurred by the compensatory roles that they play during history-matching. Unfortunately, application of either of these approaches requires a certain amount of guesswork.

White et al (2014) and Doherty (2015) discuss ways in which it may be possible to shield management-salient predictions from calibration-induced bias through strategic formulation of the objective function that is minimised through history-matching. This formulation may employ spatial and temporal observation differences instead of (or as well as) observations themselves. Data-hosted information for which a model does not possess apposite receptacles can thereby be filtered out and disregarded. While this may decrease the proclivity for predictive bias, it may also increase predictive uncertainty. If bias diminution exceeds uncertainty expansion, the decision-support modelling process is well-served by this strategy.

It is apparent from this discussion that accommodation of model defects requires the making of subjective decisions. It is also apparent that the effects of these defects are prediction-specific, and that they are also specific to the data context in which modelling takes place.

4.6 Incorrect Prior Parameter Probabilities

It follows from Bayes equation (i.e. equation 3.1) that unless a model's parameters are completely informed by measurements of system behaviour, incorrect specification of the prior parameter probability distribution will incur errors in evaluation of the posterior parameter probability distribution. Depending on the sensitivity of a management-salient prediction to model parameters, this may impair calculation of the posterior predictive probability distribution of that prediction. This, in turn, may lead to over-estimation or under-estimation of the possibility of management failure.

The integrity of regularisation that is used to achieve calibration uniqueness also relies on the integrity of the prior parameter probability distribution. If singular value decomposition is used for this purpose, maintenance of orthogonality of solution and null spaces requires that model parameters undergo pre-decomposition transformation. This transformation depends on the prior probability distribution; see section 3.6. Non-orthogonal decomposition of parameter space can result from parameter transformation that is based on an inappropriate prior parameter probability distribution. This can precipitate calibration-induced predictive bias.

Assignment of a valid prior probability distribution to real-world hydraulic properties that encompasses the complex spatial relationships of heterogeneous materials that occupy the subsurface is an almost impossible undertaking, for the real world is not a realisation of a statistical process. However, it may be easier to assign a prior probability distribution to the upscaled parameter set employed by a groundwater model. As has already been discussed, model parameterisation devices include (but are not restricted to) model cells, pilot points and zones of piecewise parameter constancy. All of these parameterisation devices pertain to spatially averaged hydraulic properties. Unfortunately, however, the spatial averaging that is implied in use of these parameterisation devices is generally unknown unless special measures are adopted to pursue a particular type of spatial averaging through the history-matching process itself; see, for example, Cooley (2004) and Cooley and Christensen (2006).

Moore and Doherty (2006) demonstrate that it cannot be assumed that the hydraulic property that is estimated for a broadscale model zone of designated parameter uniformity is representative of the average hydraulic property within that zone. Instead, the averaging process incurred by history-matching must be discovered using a so-called "resolution matrix". This is too numerically burdensome to calculate in everyday modelling practice. Hence calculation of upscaled prior parameter probability distributions from those which prevail at a finer scale (even if these are known) is generally impossible. So it must be guessed. In doing this, a multiGaussian assumption is generally invoked for the sake of convenience. Repercussions for model predictive bias are generally ignored.

4.7 Complex or Simple?

A complex numerical model has a certain allure. It is often presumed that, because a model is complex, it can represent real-world hydraulic processes with integrity. At the same time, it is presumed that its representation of subsurface hydraulic properties has greater integrity than that of a simple model because less upscaling is required to represent them. The prior probability distribution of complex model parameters can therefore resemble that of real-world hydraulic properties to the extent that the latter can be surmised. (This presumes that a modeller decides to complement model structural and process complexity with a commensurate level of parameterisation complexity. It makes little sense to do otherwise.) It is therefore concluded that predictions made by a complex model are likely to be relatively free of bias.

Arguments such as these should be treated with extreme caution. Accompanying the decision to represent hydraulic processes, and the hydraulic properties that govern them, in a “realistic” manner comes the responsibility to represent their stochasticity correctly. This presents problems that are rarely addressed in real-world modelling practice.

Complex models are complex because they explicitly represent system detail. Unfortunately, with representation of greater detail comes a greater propensity for representational error, for the details of system details are uncertain. These details should therefore be represented stochastically. However if details are hardwired into a complex model’s structure, their stochastic representation becomes impossible, for only parameters can be stochastic. Hence the structure of a complex model represents just one realisation from an implied probability distribution of model structures. If this realisation is incorrect, predictive bias may arise as parameters compensate for its incorrectness during history-matching.

For reasons that have already been discussed, stochastic representation of parametrically-expressed hydraulic property detail is problematic. This applies particularly to stochastic representation of connected permeability.

Repercussions of assignment of an incorrect prior probability distribution to complex model parameters may be experienced when the model undergoes history-matching. Suppose, for example, that history-matching of a multi-layer model suggests that higher values of hydraulic conductivity should be introduced to a certain part of the model domain. The information content of the history-matching dataset may not be sufficient to determine whether elevated hydraulic conductivities should be introduced to a single model layer or to multiple model layers, or whether they should be ascribed predefined patterns or distributed over larger areas. Ultimately, data-resident information is directed to its model parameter receptacles by the prior parameter probability distribution, particularly that aspect of this distribution that describes spatial parameter correlation. Incorrectness of the prior parameter probability distribution may therefore induce incorrect model hosting of prediction-relevant information. This, in turn, may induce predictive bias.

It follows that the often unstated (but widely believed) model design metric that “more complex is better” is not necessarily justifiable. With greater model complexity comes a greater possibility of incorrect specification of a necessarily complex, complimentary, prior parameter probability distribution, and hence a greater possibility that history-matching will induce predictive bias. As in most aspects of model construction, trade-offs are required. Different modellers may decide on different trade-offs. As has already been stated, given the subjective nature of modelling, this is not a problem. Failure to recognise the requirement for trade-offs is, however, a problem.

Decision-support deployment of complex models is also beset by practical problems. Complex models often have long run times and a penchant for numerical instability. Both of these render history-matching and uncertainty analysis difficult. This should also be taken into account when deciding on an appropriate level of model complexity. Failure to attain a good fit with a history-matching dataset is tantamount to denying entry of information to the decision-support modelling process. Excessive reliance is thereby placed on the prior parameter probability distribution. If parameterisation is too simple, and/or if the prior parameter probability distribution fails to represent the possibility of hydraulic property connectedness, management-salient predictions may be biased and/or their uncertainties may be underestimated.

4.8 Using a Model to Infer the Prior

History-matching can often inform a modeller of locations and patterns of subsurface hydraulic property heterogeneity of which he/she was previously unaware. See, for example, the following GMDSI worked example:

[Worked Example: Simultaneous Interpretation of Six Pumping Tests - Groundwater Modelling Decision Support Initiative (gmdsi.org)]

Patterns of heterogeneity that emerged through highly-parameterised history-matching of the model that is described in this report revealed strong structural influences on groundwater flow. They also revealed the possibility of connections to an upper aquifer. These insights into the groundwater system could not have been achieved without a workflow that was dedicated to pursuit of these insights. The structural controls which history-matching evinced are difficult to describe stochastically. Too great a reliance on a prior parameter probability distribution during the history-matching process, and/or use of a simplistic parameter field during that process, may have suppressed their emergence. Underestimation of predictive uncertainty and the inculcation of predictive bias would have been the inevitable result.

This mode of model deployment suggests that definition of a prior parameter probability distribution on the one hand, and history-matching on the other hand, do not need to be seen as separate processes. It acknowledges that in many modelling contexts, much can be learned about the nature and disposition of subsurface heterogeneity by model-based processing of measurements of groundwater behaviour. This illustrates the importance of so-called "abductive inference", an often-overlooked aspect of implementation of the scientific method. This is seen by some writers to be just as important as deductive inference as embodied in the Popperian, hypothesis-testing view of the scientific method. This type of inference does not require a hypothesis. Its underlying philosophy is to process all available data and be prepared for surprises.

Of course, caution must be exercised. Overfitting should be avoided. Furthermore, model-based abductive inference generally requires a fast running and numerically stable model. Such a model may be purposefully somewhat "abstract"; it may be structurally simple but parametrically complex. The prior probability distribution that is ascribed to model parameters may be intentionally uninformative. Parameter patterns that emerge from the history-matching process may therefore require human interpretation. This does not diminish their didactic worth, as aspects of the subsurface that are unveiled in this manner may have a significant effect on the future behaviour of a groundwater system when it is subjected to a new stress regime under altered management practices.

Inference of some aspects of the prior parameter probability distribution through dedicated history-matching in this manner may serve a secondary purpose. History-matching-emergent parameter patterns may suggest that parameters must adopt compensatory roles to accommodate model defects. In some modelling contexts the nature of these defects may be of secondary importance, for their existence may not undermine a model's decision-support potential. In fact, they may be an outcome of simplifications that render a model's decision-support usage tractable. However, the prior parameter probability distribution may need to be adjusted in order to reflect the compensatory roles that parameters must play when predicting the future, as evinced by their need to play these same roles when replicating the past. This may be of particular relevance when using a model to quantify the uncertainties of management-salient predictions – a matter that is now discussed.

5. TESTING THE FUTURE

5.1 General

This section briefly discusses three approaches to implementing the imperatives that decision-support modelling must serve. In doing this, they can be used to inquire whether it is possible that a particular course of action will result in management failure. None of these approaches can be guaranteed to work under all conditions, for environmental systems have their own rules. Nor may it often be possible to categorically reject a management failure hypothesis. Nevertheless, model-based processing of environmental data, undertaken in ways described herein, can provide managers with a solid ground to hope for the best, while endowing them with the capacity to prepare for the worst.

Some of the methods that are described herein are standard. Others are less standard, but are aligned with problem decomposition concepts that are addressed in this document. All are more easily and more effectively implemented if a decision-support modelling workflow is founded on problem decomposition principles that are outlined herein.

5.2 The Stochastics of Pessimism

Previous sections of this document explain why formulation of a prior parameter probability distribution is difficult. It is partly an outcome of the fact that model parameters are hosted by imperfect simulators of natural system behaviour. It is also an outcome of the fact that natural systems (like social and economic systems) defy neat stochastic characterisation. As explained by Kay and King (2020), this erodes the capacity of human beings to base decision-making on mathematically exact optimisation of a utility function. Instead, informed intuition, supported by a decision-making culture that recognises that natural and social systems are beyond stochastic portrayal, often fosters a better decision.

The need for decision-support modelling to focus on the pessimistic end of posterior predictive probability distributions has also been discussed. A posterior predictive probability distribution is caricatured in figure 3.1. Unfortunately, this predictive probability distribution bears very little resemblance to those that are likely to be encountered in real world decision-support modelling practice. This is because geological media are complex beyond measure. The presence, or otherwise, of features that may establish unexpected impact pathways (such as faults, buried alluvial channels and aquitard perforations) does not lend itself to neat probabilistic characterisation. Furthermore, if such a geological feature does exist at a study site, its geometry, properties, and relationships with other geological features are likely to be complex. For example, a fault (if it exists) may (or may not) juxtapose aquifers on either side of an aquitard; if it does, then the inter-aquifer contact length will depend on many things, including the (possibly variable) thickness of the aquitard, the length of the fault, and the distribution of throw along the fault.

Predictive repercussions of the discrete and categorical nature of hydrogeological nuances that may create or disrupt impact pathways may therefore be better caricatured by figure 5.1 than by figure 3.1.

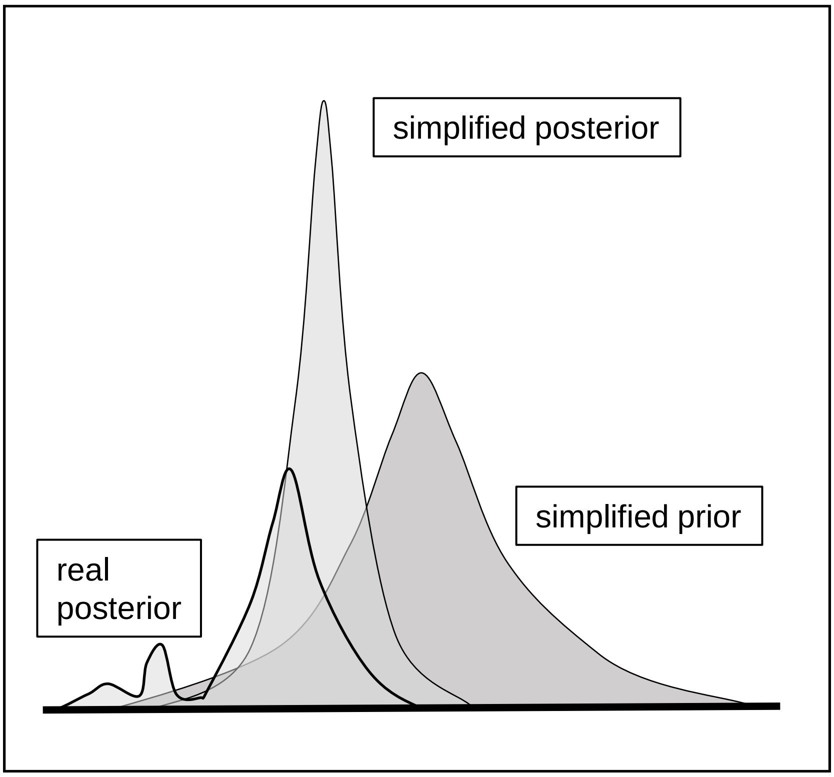

Figure 5.1. A more realistic characterisation of a posterior predictive probability distribution than that depicted in figure 3.1.

It follows that a decision-maker (whether working in the regulatory or operational framework), rarely has the luxury of ascribing an exact probability to an unwanted event, and then using that probability to evaluate the risk that he/she is taking. (This may be a blessing for regulatory decision-makers.) Instead, the best that he/she can do is to make a semi-quantitative assessment of management failure based on information that has been extracted from site data. Rarely, however, will this information be sufficient to relegate the probability of management failure to zero. Nevertheless, analyses that implement problem decomposition concepts that are discussed herein will probably have exposed mechanisms for possible management failure. They will also have identified potential sources of new information that may better illuminate these mechanisms. Acquisition of further data, either now, or as the system undergoes new management, can be based on these insights. Presumably, monitoring data will be subjected to continuous model-based processing as management proceeds so that information that is garnered from them can precipitate alterations to site management if this is warranted.

5.3 Opportunities Afforded by System Complexity

Figure 5.1 characterises decision-making contexts in which the uncertainty of a decision-critical model prediction can be at least partially attributed to the existence or otherwise of a small number of high impact pathways.

As is discussed in section 2, decision-support modelling is best implemented when it is focussed on a carefully defined hypothesis that pertains to a particular impact pathway. Where a single hypothesised impact pathway dominates predictive uncertainty, decision-support modelling can focus on that pathway. Perhaps other models can focus on other pathways; see the conceptual example that is discussed in section 3.7.3. Where currently available data yields little information about the existence, or otherwise, of a high risk pathway, decision-support modelling may become a relatively easy task, embodying worst case analysis undertaken using a simple simulator. Alternatively, other contexts may require development and deployment of more complex models to investigate how strategic acquisition of new data can shed light on the existence, or otherwise, of hypothesised impact pathways.

5.4 History-Match-Constrained Monte Carlo Analysis

Previous sections of this document explain why it is often necessary to endow a decision-support groundwater model with many parameters. These can support attainment of a good fit between pertinent model outputs and measurements of system state, thereby enabling extraction of information from those measurements. Just as importantly, they can ensure that the uncertainties of model predictions that are sensitive to hydrogeological detail are not underestimated. One means through which the uncertainties of these predictions can be explored is through Monte-Carlo analysis. However this analysis must ensure that information that is harvested from measurements of system state is respected.

Where parameter numbers are large, history-match-constrained Monte Carlo analysis is most efficiently implemented using ensemble methods, of which the PESTPP-IES iterative ensemble smoother is the most widely used in the groundwater industry. See Chen and Oliver (2013) and White (2018) for algorithmic details. Use of PESTPP-IES with the MODFLOW suite of simulators is facilitated by the PyEMU library.

Ensemble methods adjust an ensemble of random realisations (i.e. samples) of the prior parameter probability distribution until they become samples of the posterior parameter probability distribution. A parameter set achieves this status when it enables model outputs to replicate field measurements of system behaviour. The model can then be used to make a prediction of management interest using all of the history-match-constrained parameter realisations which comprise the ensemble. The posterior probability distribution of that prediction is thereby sampled.

Ensemble methods are extremely efficient. They require only as many model runs per history-match iteration as the number of parameter realisations that comprise an ensemble. Meanwhile the number of parameters that comprise a realisation can number in the tens or hundreds of thousands.

Caution must be exercised when inspecting a posterior predictive probability distribution that is sampled in this manner. Reasons for caution include the following.

- The theory on which ensemble-based parameter adjustment is based relies on a multiGaussian assumption of prior parameter variability. History-match-constrained adjustment of categorical parameter fields, and parameter fields that embody a high degree of connected permeability, is difficult to achieve with present ensemble technology.
- For reasons that have already been explained, it is often difficult to ascribe a prior probability distribution to either real world hydraulic properties or to their upscaled counterparts. Characterisation of posterior parameter (and hence predictive) predictive uncertainty is therefore itself uncertain.
- Generally, for reasons of numerical efficiency, ensemble sizes are restricted to only a few hundred parameter realisations. This is insufficient to probe the extremes of a posterior predictive probability distribution.
- Ensembles can sometimes “collapse”, as history-matching progresses. When this occurs, they sample only a part of posterior parameter space. Ensemble collapse can be ameliorated by increasing the number of parameter realisations that comprise an

ensemble, and/or by introducing “localisation” to the ensemble-based inversion process.
- On other occasions, particularly where a model is highly nonlinear, ensemble methods may fail to fit a history-matching dataset as well as is desired. The uncertainties of some management-salient model predictions may then be overstated.
- Attempts to solve a nonunique inverse problem using a limited number of parameter realisations results in computation of a rank-deficient approximation to the Jacobian matrix on which parameter adjustment is based. In some contexts, this can bias some model predictions.

Many of the above problems are the focus of current research and may diminish over time. Others may remain endemic to use of ensemble methods. Still others arise from the need to sample a prior probability distribution which is uncertain at best, and unrepresentative of reality at worst.

5.5 Data Space Inversion

Data space inversion (DSI) is described by Sun and Durlofsky (2017) and Lima et al (2020); it is also described in the PEST manual (Doherty, 2023). Its primary attraction is its extreme model run efficiency. Another advantage is that it can be used in conjunction with complex prior parameter probability distributions that can express intricate patterns of subsurface hydraulic property connectedness.

The DSI process begins by sampling the prior parameter probability distribution. The model is then run using each of these samples. It is run over the past (where field measurements of system behaviour are available), and over the future (where predictions of system behaviour under altered management conditions are required). Based on these model runs, a statistical model is developed that links past system behaviour to future system behaviour. This surrogate model is then history-matched against field measurements. Probabilistic predictions of future system behaviour are then made, and posterior predictive histograms are thereby constructed. Predictive extremes can be explored using calibration-constrained prediction maximisation/minimization. Data worth analysis can be undertaken using linear or nonlinear methods.

DSI does not adjust parameters. In fact, model “parameters” do not even need to be defined. Hence representation of geological complexity can be as complex is it needs to be in order to provide a valid representation of those aspects of the subsurface that may comprise impact pathways. However, in common with other ensemble-based methods, predictive probabilities that are based on a limited number of parameter samples cannot be guaranteed to expose the existence of all impact pathways that may threaten future system management. In general, the greater is the number of samples of subsurface hydraulic properties that is used for surrogate model construction, the greater will be the integrity of analyses that are based on these samples.

Despite its extreme numerically efficiency, use of DSI can sometimes be hydrogeologically frustrating. While parameters of the DSI surrogate model are adjusted, those of the numerical model on which this surrogate model is based are not. Hence linkages between failure probabilities and impact pathway geometries and properties may be difficult to determine. This problem is exacerbated by the multiGaussian nature of the DSI surrogate model. While the assumption of a multiGaussian relationship between some model outputs and other model outputs is much more benign than the assumption of a multiGaussian relationship between model outputs and model parameters, it may nevertheless compromise evaluation of extreme predictive probabilities, especially in highly nonlinear contexts such as exploration of contaminant fate.

Despite its shortcomings, the DSI methodology deserves consideration in any real-world decision-support modelling context – either on its own, or in conjunction with other methods. DSI is supported by both the PEST and PEST++ suites.

5.6 Direct Predictive Hypothesis-Testing

In contrast to DSI, direct predictive hypothesis-testing (DPHT) requires the adjustment of model parameters. The relationship between parameters and model outputs must therefore be continuous; parameters cannot, therefore, be categorical in nature. While this may sometimes affect how "real" they look, this does not detract from the ability of DPHT to expose complex impact pathways.

In contrast to ensemble methods and DSI, the DPHT methodology does not require that a modeller define a prior parameter probability distribution in advance of the decision-support modelling process. This is because the DPHT methodology requires no samples from this distribution. Instead, it requires that a modeller decide whether a parameter set that emerges from the DPHT process is compatible with what is known about the subsurface. A prior probability distribution is therefore implied, but does not need to be explicitly defined.

Use of DPHT is based on the premise that a hypothesis pertaining to future system behaviour (for example management failure) can be rejected if its occurrence is demonstrably incompatible with any or all of the following:

- system processes;
- system properties;
- the observed behaviour of the system.

The first of the above is encapsulated in the simulator, while the second is encapsulated in the simulator's parameters. Lack of compatibility with the third of the above constraints is exposed through either of the following conditions.

- The model is incapable of simulating an unwanted future using parameters that also allow it to replicate the measured past; or
- Parameter fields which allow simultaneous fitting of an unwanted future and the measured past are incompatible with expected system hydraulic properties and patterns of heterogeneity.

The DPHT process begins by history-matching the model against its measured behaviour. Tikhonov regularisation can be used to seek the "simplest" parameter field that is compatible with this behaviour. If a satisfactory fit cannot be attained with field measurements, this suggests to the modeller that his/her model may be defective, and may therefore require alteration. On the other hand, if the fit is good, a modeller may learn something from the calibrated parameter field about hydraulic property variability that he/she did not know before. As has already been discussed, this may affect his/her characterisation of the prior parameter probability distribution.

Tikhonov regularisation subdues the emergence of parameter heterogeneity unless a satisfactory level of model-to-measurement fit cannot be achieved without it. Therefore, if a calibration-emergent parameter field exhibits somewhat anomalous heterogeneity, this may indicate that some parameters must adopt history-matching roles that allow them to compensate for mild model inadequacies. In spite of this, a modeller may decide to retain the model, and the parameter field, while also deciding to revise his/her conceptualisation of the implied prior parameter probability distribution in appreciation of the compensatory roles that parameters must adopt during history-matching.

As well as learning something about the prior parameter probability distribution, a modeller also learns the degree of model-to-measurement misfit that he/she is prepared to tolerate during history-matching. Misfit can arise from measurement noise; more commonly it is indicative of model defects.

The modeller's calibration-acquired knowledge of tolerable parameter heterogeneity and tolerable model-to-measurement misfit accompany him/her as he/she then embarks on the next stage of the DPHT process. During this next stage, the calibration dataset is supplement with a single "measurement" of an unwanted, but hypothesised, future. During history-matching of this prediction-enhanced calibration dataset, Tikhonov regularisation is used to ensure minimal departure of parameters from the previously estimated parameter set. The modeller then assesses the emergent parameter field, as well as the level of fit that is achieved with historical measurements of system behaviour when the model is asked to simulate both the measured past and the hypothesised future. If either of these is unacceptable, the hypothesis of future system behaviour that the DPHT process was designed to test can be rejected.

From the above description it is apparent that DPHT does not require that a modeller confine his/her investigation of management failure to samples from a prior parameter probability distribution that is itself uncertain. Instead, he/she decides whether to reject a history-match-emergent parameter field on the basis of incredulity of that field. This decision will be subjective; it does not require mathematical characterisation of the prior. At the same time, this decision can take account of the fact that parameters may need to compensate to some extent for model inadequacies as they are history-match-adjusted. As is discussed above, this is something that can never be completely avoided.

Parameter patterns that emerge from DPHT may illuminate complex impact pathways. They can therefore guide acquisition of further data, and/or the design of a monitoring network that provides early warning of incipient management failure.

Success of the DPHT process requires that advice provided in previous sections of this document be heeded. In particular, processes and parameters employed by the simulator must not artificially preclude the occurrence of an unwanted system condition. Parameter parsimony therefore needs to be avoided unless it reflects a known system condition.

A disadvantage of the DPHT methodology is that a model that is subjected to history-matching must be run over both the past and the future. If predictions of long-term system behaviour under an altered management regime are of interest, model run times may therefore be long. Under these circumstances, the run-time burden of history-matching can be alleviated by using ensemble methods such as the PESTPP-IES ensemble smoother to fit pertinent model outputs to both the measured past and an unwanted future. The benefits of Tikhonov regularisation are thereby lost. On the other hand, if predictions of unwanted system behaviour are predominantly null-space-dependent, ensemble-based parameter adjustment may expose not just one, but multiple failure-pertinent impact pathways.

5.7 Appropriate Model Parameterisation

As discussed above, it is important to ensure that inappropriate parameterisation of a decision-support model does not artificially trigger rejection of a failure-pertinent hypothesis. This can occur if parameters are too few in number, and/or if they are awarded prior probability distributions which inhibit hydraulic property connectedness.

Efficacy of the DPHT methodology can be enhanced if model parameterisation reflects the failure hypothesis that DPHT is designed to test. Numerical adjustment of failure-pertinent parameters is therefore facilitated. The chances of false hypothesis rejection are thereby diminished, while the didactic outcomes of DPHT may be increased.

Special attention should be given to the importance of subsurface structural features such as faults, or alluvial features such as buried channels, as purveyors of connected permeability.

To facilitate the emergence of management-failure-pertinent connected permeability during history-matching, PLPROC (the generalized parameter preprocessor that accompanies the PEST suite) supports the use of so-called “structural overlay parameters”. These embody polylinear and polygonal features that can be superimposed on background model parameterisation. The shapes, locations and hydraulic properties of these features are all history-match-adjustable. A model can therefore attempt to reproduce past and/or hypothesised future system behaviour through adjustment of a reduced number of parameters. Meanwhile prior probabilities that are ascribed to these parameters can ensure that emergent connected permeability is realistic in both nature and location.

6. CONCLUSIONS

6.1 General

We close this manuscript by summarizing its contents in two different ways. First we present eight "principles of problem decomposition" as they pertain to decision-support modelling. It is hoped that implementation of these principles can motivate and guide the design of decision-support modelling strategies whose technical efficacy and societal acceptability are based on their demonstrable ability to implement the scientific method. Then we summarise implications of these problem decomposition principles for a decision-support modelling workflow.

6.2 Principles of Problem Decomposition

1. "Model" should be considered as a verb and not a noun.

Decision-support modelling is an activity. This activity is dedicated to the processing of data that have been acquired at a site in ways that support management of that site. This activity cannot be undertaken using a simulator on its own. It requires concomitant use of simulator-partner software such as programs of the PEST and PEST++ suites, to conduct tasks such as data assimilation and predictive uncertainty quantification.

2. Implied in any decision-making process are one or a number of hypotheses that pertain to management failure. It is the responsibility of decision-support modelling to test these hypotheses.

This principle is self-evident, but often forgotten in everyday decision-support modelling practice. It bestows on a modeller the responsibility to create a decision-support modelling workflow that is capable of testing management failure hypotheses. It does this by processing site data in ways that allow site personnel to establish whether, or not, the hypothesis is compatible with these data.

3. Information is not information unless it is relevant. Decision-support modelling is not decision-support modelling unless it is prediction-specific. A failure-pertinent prediction bestows relevance on information that decision-support modelling should harvest, at the same time as it suggests workflows for its harvesting.

The task of decision-support modelling is to assess the uncertainties of predictions that are salient to management of a natural system. Of particular relevance are the pessimistic ends of management-salient predictive uncertainty intervals.

To the extent that it can, decision-support modelling should also reduce these uncertainties. This requires the harvesting of decision-pertinent information from data, and the transmission of that information to decision-makers in a form that they can use.

Simulation is imperfect. Compromises must be made. Compromise implies a ledger wherein the benefits and drawbacks of different compromises can be compared.

No model can harvest all available information. However it's structure and parameterisation can tune it to the harvesting of information that pertains to a particular prediction and a particular impact pathway. These provide a point of reference for the many decisions that attend design of a decision-support modelling workflow.

Harvesting of information requires:

- reasonably rapid simulator run times;
- simulator numerical stability;
- adjustability of model parameters;
- minimisation of hardwired disinformation (as it pertains to a particular model prediction) in a model's structure and in the prior probability distribution of parameters that are hosted by this structure;
- formulation of a history-match objective function that filters out information for which a model has improper receptacles.

4. Where a decision-support modelling workflow incurs predictive bias, and/or artificially diminishes the size of an evaluated predictive uncertainty interval, the width of the latter must be increased to accommodate this.

This reduces the chances of false rejection of a failure-pertinent hypothesis. Unfortunately, the extent to which a particular predictive probability interval must be expanded is generally a matter of informed subjectivity.

5. The primary role of simulators that are used in decision-support is the hosting of parameters. Parameters have two roles. They store and transmit information to decision makers. They also express the predictive consequences of information insufficiency.

This is not to say that integrity of numerical simulation is unimportant. However it is important only to the extent that it supports creation of receptacles whose design is optimal for the information that they must carry.

6. Pessimism is often decomposable.

Management failure implies the existence of one or more impact pathways. It is often possible to consider these pathways separately, (possibly using different models) when testing hypotheses pertaining to their existence or effectiveness in instigating management failure. This is particularly the case for failure mechanisms induced by structural features whose impacts may be significant, but whose existence and properties are a matter of conjecture.

7. The optimal level of model complexity depends on the hypothesis that the modelling process is designed to test.

Many aspects of the decision-support modelling process become increasingly difficult as the level of model complexity rises. Simulator run times become larger, and the propensity for numerical instability becomes higher. These make data assimilation and uncertainty analysis difficult.

Theoretical problems accompany both model simplicity and model complexity. Those that accompany simplicity are well known. Those that accompany complexity are less well known, so we restate them here.

Explicit representation of a high level of subsurface detail in a complex model risks misrepresentation of this detail in hardwired model structure. Explicit representation of accompanying parameterisation detail risks misrepresentation of the prior parameter probability distribution. Both of these may introduce unquantifiable bias to management-salient model predictions.

8. Rather than starting with a "geological fantasy" enshrined in an uncertain and simplified prior parameter probability distribution and using Monte Carlo methods to work forwards, it may be safer to start with a potential management problem and work backwards.

Ascribing prior probability distributions to model parameters is problematic. This applies to complex, real-world hydraulic property fields for which stochastic characterisation is difficult or impossible. It also applies to upscaled model parameter fields wherein the existence of connected permeability is often suppressed.

When working backwards, with hypothesised management failure included in the history-matching dataset, parameter field connectedness can arise naturally. This provides a modeller with the opportunity to subjectively assess the hydrogeological credibility of parameter patterns that this numerical experiment yields. This reduces the possibility of false hypothesis rejection, and with it the chances of decision-support modelling failure.

6.3 The Decision-Support Modelling Workflow

At the risk of being prescriptive, we provide a high-level overview of steps that comprise a decision-support modelling workflow that is guided by the above principles.

- Identify management failure criteria.
- Identify potential impact pathways.
- For a particular failure criterion and corresponding impact pathway, establish the following:
    - subsurface processes and properties that comprise that impact pathway;
    - information that can illuminate the existence, or otherwise, of the pathway.
- Build a model that is capable of:
    - simulating system failure as it pertains to a particular impact pathway;
    - harvesting and storing information that is pertinent to this impact pathway;
    - displaying the outcomes of information insufficiency as they pertain to this impact pathway.

Modelling is a continuous process that does not necessarily reach its conclusion once a management decision has been made. Therefore the above steps may not comprise the entirety of a decision-support modelling workflow. Ancillary tasks that may be required of a workflow include the following.

- Identification of data types and locations that may reduce the uncertainty of a particular management-salient prediction, and may therefore render a failure hypothesis more or less difficult to reject;
- Design of a monitoring network that will provide early warning of management failure;
- Processing of data that emerge from such a network.

7. REFERENCES

Baker, V.R., 2017. Debates — Hypothesis testing in hydrology: Pursuing certainty versus pursuing uberty. *Water Resour. Res*., 53, 1770– 1778, doi:10.1002/2016WR020078

Caers, J., 2018. Bayesianism in the Geosciences. in B. S. Daya Sagar et al. (eds.). Handbook of Mathematical Geosciences, https://doi.org/10.1007/978-3-319-78999-6_27

Chen, Y. and Oliver, D. S., 2013. Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Computational Geosciences*. 17(4): 689–703.

Cooley, R. L., 2004. A theory for modeling ground-water flow in heterogeneous media. *U.S. Geol. Surv. Prof. Pap. 1679*, 220220 pp.

Cooley, R. L. and Christensen, S., 2006. Bias and uncertainty in regression-calibrated models of groundwater flow in heterogeneous media. *Adv. Water Resour*. 29(5), 639–656.

Doherty, J. and Simmons, C.T., 2013. Groundwater modelling in decision support: reflections on a unified conceptual framework. *Hydrogeology Journal.* 21: 1531–1537.

Doherty, J. and Moore, C., 2021. Decision-Support Modelling viewed through the Lens of Model Complexity. Downloadable from: Monograph: Decision-Support Modelling Viewed through the Lens of Model Complexity - Groundwater Modelling Decision Support Initiative (gmdsi.org)

Doherty, J., 2022. Manual for LUMPREM. Watermark Numerical Computing. Downloadable from Programs | PEST (pesthomepage.org).

Doherty, J., 2022. Conceptualisation of Scientifically-Based Groundwater Modelling. GMDSI monograph. Downloadable from: Scientifically-Based, Decision-Support Groundwater Modelling - Groundwater Modelling Decision Support Initiative (gmdsi.org)

Kay, J. and King, M., 2020. Radical Uncertainty: Decision-Making for an Unknowable Future. The Bridge Street Press.

Lima, M.M., Emerick, A.A. and Ortiz, C.E.P., 2020. Data-space inversion with ensemble smoother. *Comput. Geosci*. 24: 1179-1200.

Matthews, G.M. and Vial, J., 2017. Overcoming model simplifications when quantifying predictive uncertainty. Downloadable from https://arxiv.org/abs/1703.07198

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

Moore, C. and Doherty, J., 2021. Exploring the adequacy of steady-state-only calibration. *Front. Earth Sci.* 9:692671. doi: 10.3389/feart.2021.692671

Sun, W. and Durlofsky, L.J., 2017. A new data-space inversion procedure for efficient uncertainty quantification in subsurface flow problems. *Math. Geosci.* 49:679-715.

Watson, T.A., Doherty, J.E. and Christensen, S., 2013. Parameter and predictive outcomes of model simplification. *Water Resourc. Res.* 49 (7): 3952-3977. doi: 10.1002/wrcr.20145

White, J.T., Doherty, J.E. and Hughes, J.D., 2014. Quantifying the predictive consequences of model error with linear subspace analysis. *Water Resour Res,* 50(2): 1152-1173. doi: 10.1002/2013WR014767.

White, J.T., 2018. A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in high dimensions. *Environmental Modelling and Software.* 109: 191-201.

gmdsi

Groundwater Modelling
Decision Support Initiative

CRICOS NO 00114A

BHP
Flinders University
National Centre for Groundwater Research and Training
RioTinto