Decision-Support Modelling

viewed through the lens of

Model Complexity

by

John Doherty and Catherine Moore

Groundwater Modelling Decision Support Initiative

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

CITATION

For bibliographic purposes this report may be cited as: Doherty, J. and Moore C., (2021). Decision Support Modelling Viewed through the Lens of Model Complexity. A GMDSI Monograph. National Centre for Groundwater Research and Training, Flinders University, South Australia.

ISBN: 9781925562606

DOI: 10.25957/p25g-0f58

DOI as a weblink: https://doi.org/10.25957/p25g-0f58

COPYRIGHT

Glossary

Bayesian analysis

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.

Boundary condition

A condition within, or at the edge of, a model domain that allows water or solutes to enter or leave a simulated system.

Boundary conductance

The constant of proportionality that governs the rate of water movement across a model boundary in response to a head gradient imposed across it.

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 states and/or fluxes.

Capture zone

The three-dimensional volumetric portion of a groundwater flow field that discharges water to a well.

Contributing area

The two-dimensional areal extent of that portion of a capture zone that intersects the water table and surface water features where water entering the groundwater flow system is discharged by a well. (This is also referred to as the *area contributing recharge* and sometimes as the *capture zone*.)

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 states and/or fluxes.

Ensemble

A collection of realisations of random parameters.

Ensemble smoother

A software package that adjusts realisations of parameters that comprise an ensemble until they all enable pertinent model outputs to match field measurements of system states and/or fluxes.

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.

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 index by row and column

Model structure

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 are matched 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 is adjusted in order to promulgate a better fit between model outputs and corresponding field measurements. Often, but not always, these inputs represent physical or chemical properties of the system that a model simulates. However, there is no reason why they cannot also represent water or contaminant source strengths and locations.

Phreatic surface

The water table.

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 a system.

Probability density function

A function that describes how likely it is that a random variable adopts different ranges of values.

Probability distribution

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

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 singular value decomposition.

Residual

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

Singular value decomposition (SVD)

A matrix operation that creates orthogonal sets of vectors that span the input and output spaces of a matrix. 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 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 defined by undertaking singular value decomposition of a Jacobian matrix.

Spatial correlation

The propensity for one spatial parameter (such as the value assigned to a pilot point) to vary with another that is separated from it in space.

Specific storage

The amount of water that is stored elastically in a cubic metre of porous medium when the head of water in which that medium is immersed rises by 1 metre.

Specific yield

The amount of accessible water that is stored in the pores of a porous medium per volume of that medium.

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.

Stress period

The MODFLOW family of models employs this terminology to describe each member of a series of contiguous time intervals that collectively comprise the simulation time of a model.

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 homogeneity.

Vector

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

.

Acknowledgement

The contents of this manuscript are the outcome of many years of collaboration between its co-authors. Some of this collaboration took place while Catherine Moore was employed by GNS Science, New Zealand. A project in which she was engaged during part of that time was titled “Smart Models for Aquifer Management”. This project was funded by the New Zealand Ministry of Business Innovation and Employment (grant no. C05X1508). Much of the focus of that project was similar to that of the present document.

Executive Summary

Many groundwater models are too complex to provide the decision-support services that are required of them. They are often built under the premise that the goal of decision-support modelling is construction of a numerical replica of the subsurface. Once this replica has been built, decision-makers can test a variety of management options on a computer before they are implemented in the real world. Reality is complex. Hence a decision-support numerical model must also be complex.

The authors of this manuscript maintain that this premise is flawed. As a result, decision-support services that are rendered by numerical modelling are often mediocre. Construction of a digital replica of the subsurface is impossible. Our knowledge of subsurface processes, and of the hydraulic properties that govern them, is vague. Predictions of future groundwater behaviour are therefore uncertain.

This does not mean that groundwater modelling cannot provide valuable assistance in groundwater management. However, it must be designed to do so. This requires that its construction and deployment acknowledge the uncertainties of any predictions that it makes – uncertainties of which decision-makers should be aware so that they can evaluate the risks associated with management actions that they propose. Decision-support groundwater modelling must quantify the uncertainties of decision-critical model predictions while reducing them to the extent that available information allows. These are the principles on which decision-support modelling should rest. Simulation of subsurface processes, to the extent that this is possible, must serve these principles.

It follows that the primary task of decision-support groundwater modelling is the harvesting of information, for it is information that reduces uncertainty. Once harvested, this information should be directed to predictions that matter.

Information that supports groundwater management falls into two broad categories. The first is that which resides in expert knowledge and the outcomes of studies that are devoted to site characterisation. The second category of information is that which is hosted by measurements of the current and historical behaviour of a groundwater system. Information of the first type is stochastic in nature, for it is insufficient to allow hydraulic property values to be assigned with surety to all parts of a model domain. Information of the second type is accessed through history-matching. This type of information reduces the uncertainties of hydraulic properties that are represented in a model, and hence of predictions that are sensitive to them. The mixing of these two types of information, and the reduction in predictive uncertainty that is accrued through this mixing process, is described by Bayes equation. Decision-support modelling must therefore be designed to implement Bayes equation – in spirit or in fact.

The components from which a numerical groundwater model is assembled can also be divided into two broad categories. These are its structure on the one hand, and its parameters on the other hand. The primary distinction between these two components is that structure is rigid while parameters are adjustable. Parameters are of particular importance. They provide receptacles for information which the decision-support modelling process is designed to harvest. Because they are adjustable, they can express stochasticity (i.e. randomness) that reflects a deficit of information. They can also express a reduction of randomness that accompanies assimilation of information through history-matching.

The task of model structure is to host parameters. Its success in playing this role can be judged by how well parameters play their roles. If a model structure is too simple, it can distort the information-bearing capacity of the parameters that it hosts; this can introduce bias to model predictions. Hence model structure and model parameters are inter-related. However, they

are separate enough to warrant individual consideration when designing a numerical model for use in decision-support. In general, the design of a numerical model should be contingent on the prediction that it is required to make. Where groundwater management relies on a number of different predictions, it may be necessary to devise a different model for the making of each.

Predictions can be classified according to the sources of information that must be accessed in order to reduce their uncertainties. A prediction of management interest can often be classified in this way before design and implementation of a decision-support modelling strategy. Design of this strategy can therefore be tuned to quantifying and reducing the uncertainty of that prediction. Prediction types occupy positions along a continuum. However the endpoints of this continuum are important in their own right, as they are representative of many common groundwater management contexts.

At one end of the prediction spectrum are data-driven predictions. These are almost entirely informed by field measurements of system behaviour. These types of prediction are generally required at similar locations, and under similar conditions, to those which prevailed when measurements of system behaviour were made. For predictions of this type, the past is the key to the future. Hence a model must possess sufficient parameterisation complexity for adjustment of its parameters to allow model replication of the past. However its structure need not be very complex, even if this causes some parameters to adopt roles that compensate for deficiencies in its structure as they are adjusted during history-matching. It can be shown that any bias that compensatory parameters incur through this process is not inherited by data-driven predictions that are sensitive to them.

At the other end of the prediction spectrum are predictions that are uninformed (or nearly so) by field measurements of system behaviour. History-matching is not a requirement for the making of these types of prediction. Nevertheless history-matching, in a stochastic sense, can be useful for exploring prior-data conflict. In order to make predictions of this type, and in order to quantify their uncertainties, a modeller may populate a structurally complex model with complex, geostatistically-generated hydraulic property fields; geostatistical field generation is informed by expert knowledge, as well as by data that emerges from site characterisation. Alternatively, a modeller may implement strategic worst-case scenario analysis using a model that is simple in both its structure and its parameterisation.

Mixed predictions occupy the middle part of the prediction spectrum. These are the types of prediction that the decision-support modelling process is most often required to make. Information emerging from expert knowledge and site characterisation on the one hand, and from measurements of system behaviour on the other hand, must be blended to reduce the uncertainties of these types of prediction. Unfortunately, assimilation of information from one of these sources can interfere with assimilation of information from the other. This is because assimilation of the former type of information is best served by a complex model structure, whereas assimilation of the latter type of information is best served by a simpler model structure. So compromises must be made. We show that the groundwater industry's proclivity for adoption of complex model structures may actually serve the imperatives of decision-support modelling rather poorly, as it may deny measurements of system state opportunities to question the many assumptions on which assembly of a complex model structure is based. Therefore, in many decision-support contexts, pursuit of a strategy which we describe as "fit first and ask questions later" may be more productive.

We summarise our perspectives on decision-support modelling pictorially using a "roadmap". This is intended to provide modellers with scientifically-based guidance for selection of a level of structural and parameterisation complexity that is appropriate for the decision-support context in which they are working. It may also assist modelling stakeholders to understand how groundwater modelling can best respond to the decision support imperatives that it is meant to serve.

CONTENTS

1. Introduction ... 1
1.1 General ... 1
1.2 Modelling as a Scientific Instrument ... 1
1.3 Model Design and Flow of Information ... 2
1.4 Bias and Uncertainty ... 3
1.5 This Manuscript ... 3
2. A Modeller's Job Statement ... 6
2.1 Risk and Decision-Making ... 6
2.2 What Decision-Support Groundwater Modelling is Not ... 8
2.3 The Scientific Method ... 9
2.3.1 Hypothesis-Testing ... 9
2.3.2 Scientific Instrument ... 10
2.4 Uncertainty Quantification and Reduction ... 11
2.5 Metrics for Decision-Support Groundwater Modelling ... 12
3. Flow of Information ... 14
3.1 Introduction ... 14
3.2 The Conceptual Model ... 14
3.2.1 General Considerations ... 14
3.2.2 The Point of Reference ... 15
3.2.3 Where Information Resides in a Model ... 16
3.3 Sources of Information ... 16
3.3.1 Information Source Categories ... 16
3.3.2 Site Characterisation ... 16
3.3.3 Measurements of System Behaviour ... 17
3.3.4 Other Types of Measurements ... 17
3.4 Model Receptacles for Information ... 18
3.4.1 Model Design Categories ... 18
3.4.2 Parameters as Receptacles for Information ... 18
3.5 Transfer of Information ... 19
3.5.1 Bayes Equation ... 19
3.5.2 Subspace Analysis ... 21
3.5.3 Insights from SVD ... 22
4. History-Matching ... 25
4.1 Introduction ... 25
4.2 The Benefits of Many Parameters ... 25
4.3 The Cost of Many Parameters ... 26
4.4 Parameter Value Distortion ... 27
4.4.1 An Example ... 27
4.4.2 Parameter Surrogacy ... 28
4.4.3 Prediction Specificity ... 29
4.4.4 The Art of History-Matching ... 29
4.4.5 The Art of Predicting ... 30
4.5 The Dimensions of Complexity ... 31
5. Traditional Decision-Support Modelling ... 32
5.1 General ... 32
5.2 Some Practical Considerations ... 32
5.3 Physically-Based Models ... 33
5.3.1 Model Structural Complexity ... 33

5.3.2 Hydraulic Property Upscaling 33
5.3.3 Geological Structure 34
5.3.4 A Common (and Flawed) Approach to Decision-Support Modelling 35
5.4 Benefits of Abstraction 36
5.5 Problems with Abstraction 37
5.6 Starting from the Problem and Working Backwards 38
6. The Prediction Spectrum 39
6.1 SVD Revisited 39
6.2 Predictions 39
6.3 Two Models 40
6.4 A Roadmap 40
6.5 Model Assessment 42
7 Data-Driven Predictions 43
7.1 Examples 43
7.2 Strategies 44
7.2.1 Structural Complexity 44
7.2.2 Parameterisation Complexity 45
7.2.3 Other Considerations 45
7.3 The Roadmap Lane 46
7.4 Model Assessment 47
7.4.1 Goodness of Fit 47
7.4.2 Parameter Credibility 47
8. Expert-Knowledge-Driven Predictions 48
8.1 Examples 48
8.2 Discussion 48
8.2.1 Costs and Benefits of History-Matching 48
8.2.2 A Note on History-Matching Technology 49
8.2.3 Options 49
8.3 Strategy 1: Geostatistical Methods 49
8.3.1 Description 49
8.3.2 Prior-Data Conflict 51
8.3.3 Primitive Rejection Sampling 51
8.3.4 Data Space Inversion 51
8.3.5 Some Problems 51
8.3.6 Roadmap Lane 52
8.3.7 Assessment: Goodness of Fit 52
8.3.8 Assessment: Parameter Credibility 53
8.4 Strategy 2: Worst Case Analysis 53
8.4.1 Description 53
8.4.2 Roadmap Lane 55
8.4.3 Assessment: Goodness of Fit 55
8.4.4 Assessment: Parameter Credibility 55
9. Mixed Predictions 56
9.1 General 56
9.2 Examples 56
9.3 Hydrogeologically-Based Modelling 57
9.3.1 Model Structure 57
9.3.2 Model Parameters 58
9.3.3 Resolving the Paradox 59
9.3.4 A Conceptual Example 59
9.3.5 Roadmap Lane 61

9.3.6 Assessment: Goodness of Fit.......... 61
9.3.7 Assessment: Parameter Credibility.......... 63
9.4 Fit First and Ask Questions Later.......... 63
9.4.1 Discussion.......... 63
9.4.2 An Example.......... 64
9.4.3 Roadmap Lane.......... 70
9.4.4 Assessment: Goodness of Fit.......... 71
9.4.5 Assessment: Parameter Credibility.......... 71
10. Conclusions.......... 72
10.1 Decision-Support Modelling Metrics.......... 72
10.2 New Perspectives.......... 72
10.3 Modelling as a Compromise.......... 73
10.4 Parameter Subspaces.......... 73
10.5 Decision-Support Modelling Strategies.......... 74
10.6 The Roadmap.......... 75
11. References.......... 76

1. INTRODUCTION

1.1 General

The subject matter of this document is decision-support groundwater modelling. This is the kind of modelling that accompanies an environmental impact statement, that is used to evaluate next year's allocation of water to irrigators, or that forms the basis for design of a contaminant remediation system. We are not concerned with simulation that is undertaken for research purposes. This is another type of modelling altogether.

We focus on groundwater model complexity. However a prerequisite for a discussion of model complexity is an examination of decision-support modelling in general. Complex models are complex by design; simple models are simple by design. So a discussion about model complexity is a discussion about decision-support model design. Any discussion about model design must begin with an appraisal of the metrics by which decision-support modelling should be judged.

This is not the first time that the authors of this document have written (or spoken) on this subject. Much of what is written herein echoes sentiments that have been expressed elsewhere. Our primary purpose in writing the current manuscript is to present them in a way that can be understood by a broad cross-section of the groundwater community.

To be more particular, we hope that concepts that are enumerated in this monograph can serve as reference points for those who undertake decision-support modelling on a regular basis. At the same time, we hope that this manuscript is also readable by those who do not build models, but who require some guidance in assessing models that have been built by others. With the latter audience in mind, we keep mathematical notation to a minimum. Nevertheless, we attempt to present concepts with a rigor to which mathematics gives voice; references to pertinent textbooks and journal articles are provided where necessary. We also make liberal use of diagrams and pictures where these can illuminate concepts in ways that save a thousand words.

1.2 Modelling as a Scientific Instrument

A groundwater model is the outcome of many subjective decisions. Conclusions that are drawn from groundwater modelling are often hotly debated. Someone who challenges a particular groundwater model is often armed with his/her own groundwater model that makes very different predictions. This seems to contradict the notion that groundwater modelling can be a scientific enterprise. Nevertheless, stakeholders in groundwater management are often asked to respect decisions that are supported by groundwater modelling on the basis that these decisions are "scientifically-based".

The public has a right to be sceptical about predictions made by a groundwater model. This should not be seen as a bad thing; after all, scepticism is the cradle of science. Groundwater modelling can characterise itself as a scientific enterprise only if it can justify its methods and practices using logic that is beyond refutation by its critics. We contend that this is possible if a groundwater model is fashioned as a component of a scientific instrument that is deployed by individuals whose intellects are humbled by the complexities of natural systems, and whose minds are open to what may be revealed by the scientific inquiries that are facilitated by their instrument.

Subjectivities in various aspects of model design are an outcome of the fact that scientific instruments are built by human beings. Ideally, however, conclusions that are drawn from judicious use of a scientific instrument are independent of the design of that instrument. This does not mean that these conclusions are exact, particularly when they pertain to natural systems. However, this should not diminish their importance. Conclusions will often be stated in terms such as this. “We cannot be sure of what the outcome of a proposed management action will be; however the assertion that a future system state will exceed a critical management threshold can be deemed as unlikely, as this assertion is incompatible with the measured properties of the system, and with its measured behaviour.” The decision-support modelling process should be designed to enable the drawing of conclusions such as these.

The task of decision-support modelling can therefore be viewed as that of testing assertions. These assertions often pertain to one or a number of unwanted outcomes of a proposed management strategy. These are called “null hypotheses” in the scientific jargon. Because the outcomes of scientific inquiry into natural systems are rarely certain, predictions of future system behaviour, especially if the system is subjected to a new set of stresses, normally embody a range of possibilities. If undertaken in the spirit of scientific inquiry, decision support modelling can characterise this range, while simultaneously reducing it. A null hypothesis regarding an unwanted management outcome can be rejected if its probability of occurrence, after processing all pertinent data, is evaluated to be low. Alternatively, scientific analysis may demonstrate that currently-available data are insufficient to allow rejection of the hypothesis.

1.3 Model Design and Flow of Information

It is information that reduces uncertainty. It is therefore the task of decision support groundwater modelling to capture information, and to channel it to where it is needed. Throughout this document we emphasise that a groundwater model is unlikely to simulate a natural system very well. However, if designed and deployed in the right way, it may nevertheless provide a coherent channel through which information can flow. This is what matters, for simulation is not the primary purpose of decision-support modelling. The primary purpose of decision-support modelling is enablement and facilitation of flow of decision-relevant information.

This has far-reaching consequences for how decision-support groundwater modelling should be undertaken, and for how it should be judged. As a book should not be judged by its cover, a decision-support groundwater model should not be judged by whether screen-captured visualisations of its input and output datasets look like “the real thing”. In fact, for reasons that we outline in this document, a claimed resemblance to the “real thing” may constitute logical grounds for treating a groundwater model with deep suspicion.

Decision support modelling is data processing. Data are processed because they contain information. This information may reduce the uncertainties associated with some decision-critical predictions. Numerical simulation comprises an important component of a scientifically-based data processing strategy. However, it is not the only component of this strategy, and should not be viewed as a deliverable in itself. Instead, it should be viewed as an essential element of a larger process, orchestrated by a modeller, that requires concomitant use of other software packages that undertake tasks such as history-matching and uncertainty quantification. The outcomes of this collective processing are quantification and reduction of the uncertainties associated with decision-critical predictions of future system behaviour.

1.4 Bias and Uncertainty

This view of decision-support modelling poses an important question. How can a modeller convince sceptical stakeholders that the uncertainty bounds that he/she associates with a decision-critical prediction span what will actually happen if a proposed management strategy is implemented? This may not be an easy task; but it has some vital ingredients.

Firstly, a modeller should welcome scepticism, because scepticism is the fertile ground in which science can flourish; in fact it is the only ground in which science can flourish. Secondly, a modeller should remind his/her listeners that processing of environmental data can never reveal what will happen in the future, for that task belongs to soothsayers. Scientific inquiry can only reveal what will NOT happen in the future (and even then, with an associated level of confidence).

Thirdly a modeller should explain to a sceptical audience that data processing of any kind runs the risk of inducing error (i.e. bias). The propensity for bias is especially high when processing is enacted on environmental data for the purpose of predicting the environmental future. Unfortunately, evaluation of bias is a difficult matter. As is well documented in the scientific literature, in any data processing context the proclivity for processing-induced bias can be traded off against acceptance of increased, but quantifiable, uncertainty. When testing the hypothesis that a contemplated course of management action will induce unwanted system behaviour, quantifiable uncertainty should always be chosen over unquantifiable bias.

While bias is difficult to quantify, its potential sources are sometimes easy to identify. For example, most people would agree that some predictions made by a simple model are susceptible to bias. However, while this may seem intuitive, it does not, on its own, constitute grounds for eschewing model simplicity. If simplicity-induced predictive bias is significantly less than uncertainty born of information scarcity, and if the latter can be more reliably quantified with a simple model than with a complex model, then the simple model provides superior decision support.

What is less commonly realised is that predictions made by a complex model are also subject to bias. Furthermore, its long run times and numerical delicacy render the litany of assumptions on which construction of a complex model is based difficult to test against field measurements. Nor is it easy to give rapid numerical voice to alternative, but equally complex, assumptions. Bias mitigation and uncertainty quantification are therefore difficult to implement. Such a model cannot easily implement tasks that can be characterised as “data processing”. It has little capacity to trade bias against uncertainty. A modeller has no ability to assure a sceptical stakeholder group that uncertainty intervals associated with its predictions span the right answer.

1.5 This Manuscript

In this manuscript we attempt to provide a set of principles that can enable a modeller to select a level of groundwater model complexity that is appropriate for his/her particular decision-support context. These concepts link model design to the data analysis tasks that modelling must accomplish. We try to explain these concepts in ways that are understandable by modellers and non-modellers alike. We see this as important, because an essential component of model-based environmental data-processing is the conversation that goes with it. Modellers contribute to this dialogue by attempting to bracket the unknown “right answer” in an uncertainty interval that is wide enough to be truthful, and narrow enough to be useful. They then bear the onus of proof that this interval is not understated. In doing so, they may have to convince sceptical stakeholders of non-scientific background that production of animated, “picture perfect” cartoons of vaguely-known subsurface processes on the one hand, and facilitation of information flow on the other hand, are two very different things.

In Section 2 of this document we provide a decision-support modeller’s job statement. In doing so, we examine just what modelling can, and cannot, contribute to the decision-making process.

Section 3 considers information. We look at where it comes from, how modelling can harvest it, and where it is held inside a model. Though simple (and somewhat obvious), the concepts that we present in this section are often forgotten. A modeller imperils model-based environmental management by disregarding them, for these concepts should comprise the foundations of decision-support model design.

Section 4 focusses on history-matching. Through this mechanism, information that is resident in field-acquired data can be assimilated into a model. We examine the important role that parameters play in a decision-support model. We show that the term “model complexity” must be dissected to be properly understood; in particular, structural complexity must be distinguished from parameterisation complexity. We show why parameters should be flexible enough to absorb information while being numerous enough to express lack of information. We also point out that things can go very wrong if model structural defects force parameters to distort the information that they are meant to host. We discuss ways in which this problem can be identified, and ameliorated.

In Section 5 we briefly examine decision-support modelling as it has been traditionally undertaken. A “calibrated model”, often boasting a complex structure fashioned as a “digital replica” of the subsurface, plays host to a limited number of parameters. This model is used to make a variety of predictions. Its credibility in making these predictions rests on the premise that its status as “a model” earns it this entitlement. We show how decision-support services rendered by such a model are mediocre at best and illusory at worst. We contrast this with an alternative approach to decision-support modelling that is encapsulated in the phrase “starting from the problem and working backwards”. When the modelling process, including the simulator itself, are tailored to processing of data that can most effectively reduce the uncertainty of a particular decision-critical prediction, the model design process gains a valuable point of reference. Without a reference point, no criteria can be developed for optimising model design, nor for justifying the design that is finally adopted.

In Section 6 we follow our own advice. We work backwards along the information pathway that the decision-support modelling process establishes - from the prediction on which this information should be focussed to the sources of that information in expert knowledge and field data. In doing so, we define a spectrum of prediction types.

While this spectrum is continuous, its end members are encountered often enough to warrant special attention. These are discussed in Sections 7 and 8 of this manuscript. Each of these end members embodies a relatively forgiving modelling context in which model design strategies are relatively easy to delineate, while their costs and benefits are relatively easy to analyse.

Section 9 considers the middle part of the prediction spectrum. Here, it is incumbent on the modelling process to access and mix information of very different types that emerges from very different sources, and for a simulator to host that information, sometimes in ways that have the potential to threaten its decision-support integrity. Decision-support modelling becomes a process of strategic compromise, where a modeller’s insights and experience are as important as the technology on which he/she relies. We discuss the costs and benefits of these compromises, why they are necessary, and how they can be justified.

Section 10 concludes this manuscript with a summary of its contents, and a re-iteration of its most important points.

2. A MODELLER'S JOB STATEMENT

2.1 Risk and Decision-Making

Natural systems are infinitely complex. The hydraulic properties of the geological media through which groundwater flows are heterogeneous at every scale. Recharge varies with time and space in ways which cannot be characterised using equations, nor replicated faithfully on a computer. Interaction of groundwater with rivers and streams is intricate; its interaction with wetlands requires that flow of energy, as well as that of water, be characterised.

When a change is made to a groundwater system, impacts may eventually be felt in distant parts of that system. Groundwater management seeks to ensure that these impacts are restricted to levels that society can accept, or that those who are impacted are suitably compensated. Sometimes impacts are designed to be beneficial; for example, they may be designed to reverse or contain existing groundwater contamination.

It is never possible to calculate the outcomes of a proposed groundwater intervention strategy exactly, for these depend on too many unknown details of the groundwater system. However, it is often possible to calculate them approximately. In recognition of the approximate nature of these calculations, prognoses of future system behaviour should be accompanied by estimates of their uncertainties. Stakeholders are therefore made aware of the full range of possibilities that system intervention may portend.

At the time of writing, the importance of associating uncertainties with decision-critical model predictions is rapidly gaining groundwater industry acceptance. However, the pivotal role that uncertainty analysis should play in decision-support modelling was described many years ago in a series of landmark papers published by Allan Freeze and his co-workers; the first of this series is Freeze et al (1990). They noted that decision-making, whether for environmental management or other purposes, requires enumeration of risk. When managing a groundwater system, it is modelling that enables evaluation of the risk associated with a contemplated course of action. In its most simple terms, risk can be characterised as the cost of something going wrong times the probability of it's going wrong. It is obvious that characterisation of risk requires analysis of predictive uncertainty. See Figure 2.1.

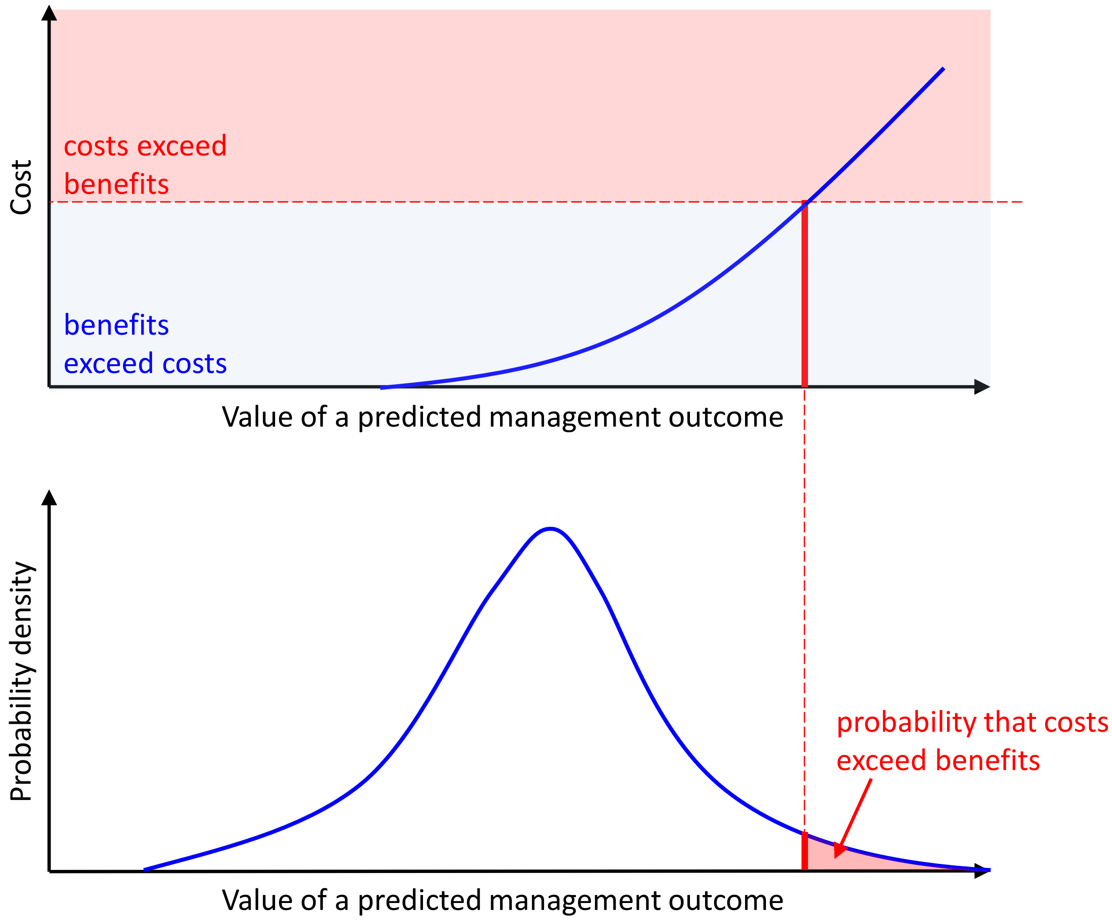


Figure 2.1. The benefits of a proposed groundwater management strategy (for example dewatering a pit or cleaning up a contaminant) are generally known. The task of decision-support modelling is to establish whether financial/environmental costs exceed these benefits.

While this depiction of the role of modelling in groundwater management is somewhat obvious, it has often been ignored since publication of the Freeze et al papers. Part of the reason for this is that uncertainty quantification is not an easy undertaking – especially when predicting the future behaviour of natural systems that undergo unnatural stresses.

The fundamental concepts that were elaborated by Freeze et al have repercussions that extend beyond the requirement that uncertainty intervals be associated with model predictions. They require that, before any modelling is actually undertaken, system stakeholders and managers identify the management impacts that they fear, for these define the predictions that the modelling process must address. These predictions, in turn, provide a basis for design of a model, and for definition of the data processing tasks that modelling must accomplish.

For reasons that we explain later, groundwater modelling requires many compromises. In modelling, as in everyday life, the making of a compromise requires evaluation of what is gained and what is lost by its making. Without a clear definition of what it is that a model must predict, there is no way to evaluate these gains and losses, and hence no way to make a compromise. It follows that a model, and the processing of data that modelling enables, must be designed from the ground up to make a particular prediction, and to quantify the uncertainty associated with that prediction. If stakeholders fear that a proposed intervention may incur more than a single unwanted outcome, then this may require the construction of more than one model.

2.2 What Decision-Support Groundwater Modelling is Not

Prediction-specificity of decision-support groundwater model design, construction and deployment stands in stark contrast to popular perceptions of computer simulation as it should be applied to natural systems. Much confusion surrounds use of the term “model”. Those with no modelling experience cannot be blamed for judging the worth of a numerical simulator by its perceived capacity to create pictures that look like “the real thing”.

In this section we present metrics by which the process of decision-support modelling should be judged. These metrics are based on the premise that it is the decision-support *process* that is enabled by modelling which must be judged, and not a single component of this process, namely a model. This single component has no role or meaning outside of the overall process which it enables.

The temptation to build a complex, “realistic”, numerical model is easy to understand. A modeller who builds such a model, or someone who commissions the construction of such a model, may perceive that it will more readily gain acceptance from traditionalist reviewers, and from the public at large, than a more abstract simulator whose focus is on processing data that is most salient to a single prediction. This may be an important consideration if decisions are made in the court of public opinion. However, for reasons that we discuss below, construction of a model which attempts to digitally replicate subsurface processes that are only vaguely known is not necessarily a scientific enterprise.

Nevertheless, the temptation is real. It is motivated by the idea that if a model gives numerical voice to partial differential equations that describe movement of groundwater and contaminants, and if it approximately represents the disposition of the major geological units that determine the three-dimensional distribution of hydraulic properties within the subsurface, then its predictions cannot be too wrong. Furthermore, once it has been built, such a model can be used to make a variety of predictions. The specifics of all of these predictions need not be known prior to construction of the model. The model can therefore be built and sold as an all-purpose “predictor”.

As alluring as this idea sounds, it is the authors’ experience that this approach to decision-support groundwater modelling does not work in practice.

Regardless of how much hydraulic process and property detail is represented in a groundwater model, even more detail is omitted. The litany of assumptions and approximations that underpin design of a complex model are better concealed than those on which a simple model is based. However they are present in abundance. A model with even a million cells requires boundaries where none exist in nature. It must simulate diffuse recharge, and interaction with surface water systems, in ways which are far less nuanced than those which prevail in reality. It must employ a discretisation scheme whose dimensions are coarser than features which affect the movement of groundwater. A discrete number of layers replaces continuous and complex vertical interfingering of lithologies.

Representation of system detail presents groundwater modellers with a paradox. A hydrogeologist may affirm that the nature and geometric relationships of different geological units can exert a considerable influence on movement of groundwater and dissolved contaminants. However while details may be important, their specifics are often unknown; hence a numerical model cannot represent them with fidelity. If it attempts to represent hydrogeological detail, it can only represent one possible manifestation of that detail. In doing so, it probably represents the wrong manifestation.

The challenge of decision-support groundwater modelling is to represent the *effects*, rather than the specifics, of hydrogeological detail that may enable or inhibit the occurrence of unwanted outcomes of a proposed management strategy. In fact, representation of just one manifestation of hydrogeological detail is likely to impede its ability to represent the effects of all possible manifestations of that detail, and hence to quantify the uncertainty of a decision-critical model prediction. Furthermore, it may inhibit the ability of the decision-support modelling process to harvest information from data that may constrain the inferred range of predictive possibilities.

2.3 The Scientific Method

2.3.1 Hypothesis-Testing

Implementation of the scientific method requires that hypotheses be tested. A hypothesis cannot be accepted; it can only be rejected. Grounds for its rejection are incompatibility with system properties and observed system behaviour. In classical experimental science, a component of a complex system is isolated, and then subjected to a carefully constructed stress. Its response to that stress is monitored. Any theory of its behaviour that fails to account for that response is rejected, regardless of its ability to explain other aspects of its behaviour.

Application of the scientific method requires that measurements be made of system states and/or fluxes, because these comprise its response to system stresses. In a laboratory, the stresses to which a system is subjected, and monitoring of its response to these stresses can be carefully controlled. This is less the case for groundwater systems. Nevertheless, the principles of experimental design can still be applied when acquiring and processing groundwater data.

In a typical decision-support context, the aspect of a groundwater system in which we are most interested is not so much a particular nuance of its properties, or even some details of its present behaviour. It is its future response to a stress that has not yet been imposed. Generally, of most interest is whether this response will transgress an acceptable threshold such as that illustrated in Figure 2.1. This constitutes the hypothesis that processing of existent groundwater data is required to test.

The hypothesis that a proposed system intervention will have a specific unwanted outcome (i.e. will result in the occurrence of a “bad thing”) can be rejected if this hypothesis is demonstrably incompatible with known properties of the system, and/or with measurements of its current or historical behaviour. Fundamental to the ability of the decision-support modelling process to test this hypothesis is its ability to simulate the future response of the system to the proposed stress as it depends on the range and spatial disposition of system hydraulic properties that are compatible with our current knowledge of them. It must also be capable of simulating the response of the system to stresses that prevailed when historical measurements of its state and/or fluxes were made.

Rejection of a bad thing hypothesis may not be possible. This is the case if historical measurements of system state and our current knowledge of system properties leave prognoses of future system behaviour wide open. In this case, it is up to system managers and stakeholders to decide whether they can accept the risk. Hopefully, decision-support modelling can assist their decision-making by associating some level of likelihood with the hypothesised occurrence of unwanted system behaviour. Alternatively, decision-support modelling may demonstrate that it is not possible for a simulator to predict an unwanted future while replicating the known past, or that it can do so only if it is endowed with hydrogeologically unreasonable hydraulic properties. This constitutes grounds for rejection of the bad thing hypothesis.

2.3.2 Scientific Instrument

Environmental hypothesis-testing requires more than simply an ability to simulate the behaviour of an environmental system. It requires concomitant use of simulator-partner software. This software must be capable of running a simulator many times. On each occasion that it runs the simulator, it must provide it with a set of hydraulic properties. It may need to adjust these properties so that the simulator can replicate historical system behaviour, while exploring the boundaries of predictive credibility that are compatible with that behaviour. Taken together, the simulator and its partner software comprise the scientific instrument through which decision support modelling implements the scientific method. We note that if simulation is to provide a useful component of this instrument, the numerical model which implements simulation of groundwater flow must be capable of conjunctive use with partnered software. This generally requires that it run reasonably fast and be numerically robust.

At the time of writing, two software suites provide model partner software that is suitable for use in groundwater decision-support; these are the PEST and PEST++ suites. To varying degrees, and using a variety of methodologies, members of these suites implement history-matching, uncertainty quantification and outcome optimisation. Fundamental specifications for model-partner software packages such as these are:

1. A non-intrusive interface with a model; that is, the model partner package must communicate with a model through the model's own input and output files; and
2. The ability to conduct model runs in parallel.

The first of these specifications (illustrated in Figure 2.2) makes it easy for a modeller to complement a simulator with parameter preprocessors and model outcome postprocessors that meet the demands of the decision support context in which he/she is working. Together with the simulator, these pre/postprocessors comprise part of the "scientific instrument" that a modeller must build in order to deliver scientifically-based decision-support. Running of the overall model then requires the calling of a batch or script file in which preprocessors, the simulator, and postprocessors are run sequentially. This call is made by the pertinent member of the PEST or PEST++ suite on each occasion that it runs the model.

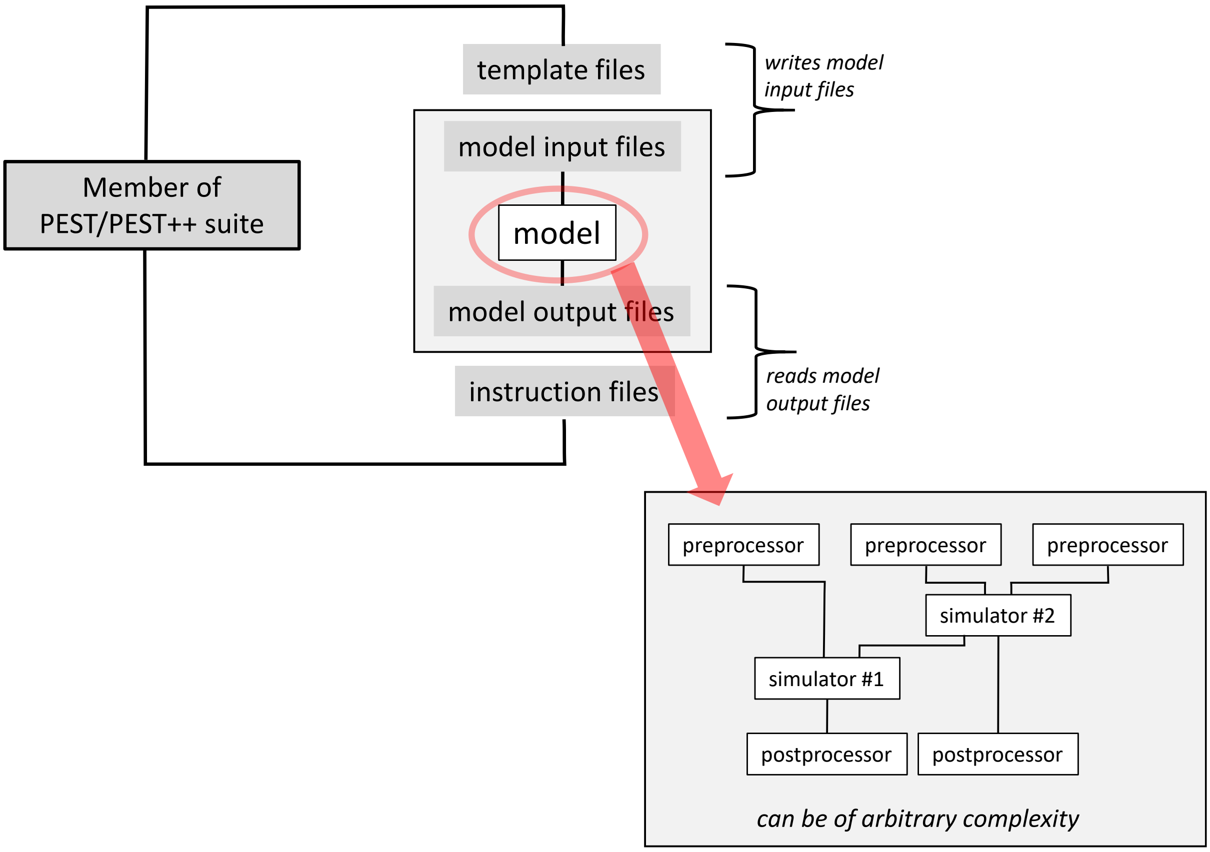


Figure 2.2. Members of the PEST and PEST++ suites have a non-intrusive interface with a model.

The second specification allows the model partner package to run a model many times without incurring the time penalties that accompany sequential model runs. Fortunately, the numerical algorithms through which highly-parameterised history-matching and uncertainty quantification are implemented by these packages are amenable to model run parallelisation. This facilitates their use in modern computing (including cloud) platforms.

2.4 Uncertainty Quantification and Reduction

The provision of scientifically-based decision-support requires recognition that predictions of future groundwater behaviour are uncertain. It must also acknowledge that it may be possible to reduce these uncertainties by accessing information that resides in measurements of present and historical system behaviour. As will be discussed later in this document, this is done through history-matching.

While decision-support modelling must accommodate uncertainty, it does not follow that its implementation requires characterisation of the full probability distributions of all predictions of management interest. Often it is only the pessimistic end of a predictive probability density function that must be explored; this is the end at which a hypothesised “bad thing” resides. This is the right end of the probability distribution depicted in Figure 2.1.

Nor do the imperatives of science-based decision support necessarily require that the uncertainty associated with a decision-critical prediction be reduced as much as available information may allow. If a bad-thing hypotheses can be rejected using a demonstrably conservative uncertainty interval, then the hypothesis is indeed rejected. Under circumstances which we discuss later in this document, implementation of a decision-support modelling strategy that is designed to compute conservative uncertainty intervals can, in some instances, reduce the numerical burden of decision-support modelling considerably.

Characterisation of predictive uncertainty is never straightforward, for there is uncertainty in uncertainty. Furthermore, as stated above, the decision-support imperatives of quantifying uncertainty (to the extent that this is necessary) and of reducing uncertainty (to the extent that this is warranted) generally require a certain level of simulation abstraction. This enables adjustable representation of system properties – a necessary prerequisite for history-matching and uncertainty quantification. As has been stated, the latter activities also require numerical stability of the simulator, as well as rapid simulator execution.

Unfortunately, abstract representation of groundwater system processes and properties can limit the ability of the modelling process to reduce the uncertainties of some decision-critical predictions. It may even introduce its own contribution to predictive error – a contribution that must be accounted for in stated predictive uncertainty intervals. This is a price that must be paid for the ability to analyse uncertainties associated with the environmental future. Design of the decision support modelling process is therefore one of compromise; it is a process that requires considerable expertise.

2.5 Metrics for Decision-Support Groundwater Modelling

In recognition of the uncertainty that accompanies any prediction of future environmental behaviour, and in acknowledging that its quantification (to the extent that this is necessary) may induce errors that must be included in its evaluation, Doherty and Simmons (2013) expanded the discussion of Freeze et al (1990) to include a definition of decision-support modelling failure. This is false rejection of the hypothesis that a bad thing can accompany a proposed groundwater management strategy. (This constitutes what is commonly referred to as a “type 2 statistical error”.) The consequences of decision-support modelling failure are obvious. They are illustrated by the second row of Figure 2.3.

Avoidance of decision-support modelling failure requires that a modeller overstate, rather than understate, the widths of decision-critical predictive uncertainty intervals. It also requires that the potential for predictive error that is introduced by the numerical methodologies that he/she implements to facilitate uncertainty quantification and uncertainty reduction be accounted for, and included in evaluated predictive uncertainty intervals. Furthermore, it is incumbent on the decision-support modeller to convince stakeholders in environmental management that this has been done properly.

While failure is easy to define, definition of decision-support modelling “success” is not so easy. It cannot be simply defined as “absence of failure”.

A sure way to ensure that predictive uncertainty intervals are conservative is to inflate them. Use of an over-simplified model whose uncertainties cannot be reduced through history-matching, and whose predictions must be accompanied by an “engineering safety margin” that is wide enough to include its proclivity for error, may indeed avoid decision-support modelling failure. However this decision-support modelling strategy may not allow rejection of a decision-pertinent bad thing hypothesis, even where enough information is available to allow this. A decision-support modelling strategy that relies on evaluation of grossly conservative predictive uncertainty intervals does not comprise insightful deployment of the scientific method. Nor is it useful. Decision-support modelling uselessness is illustrated by the third row of Figure 2.3.

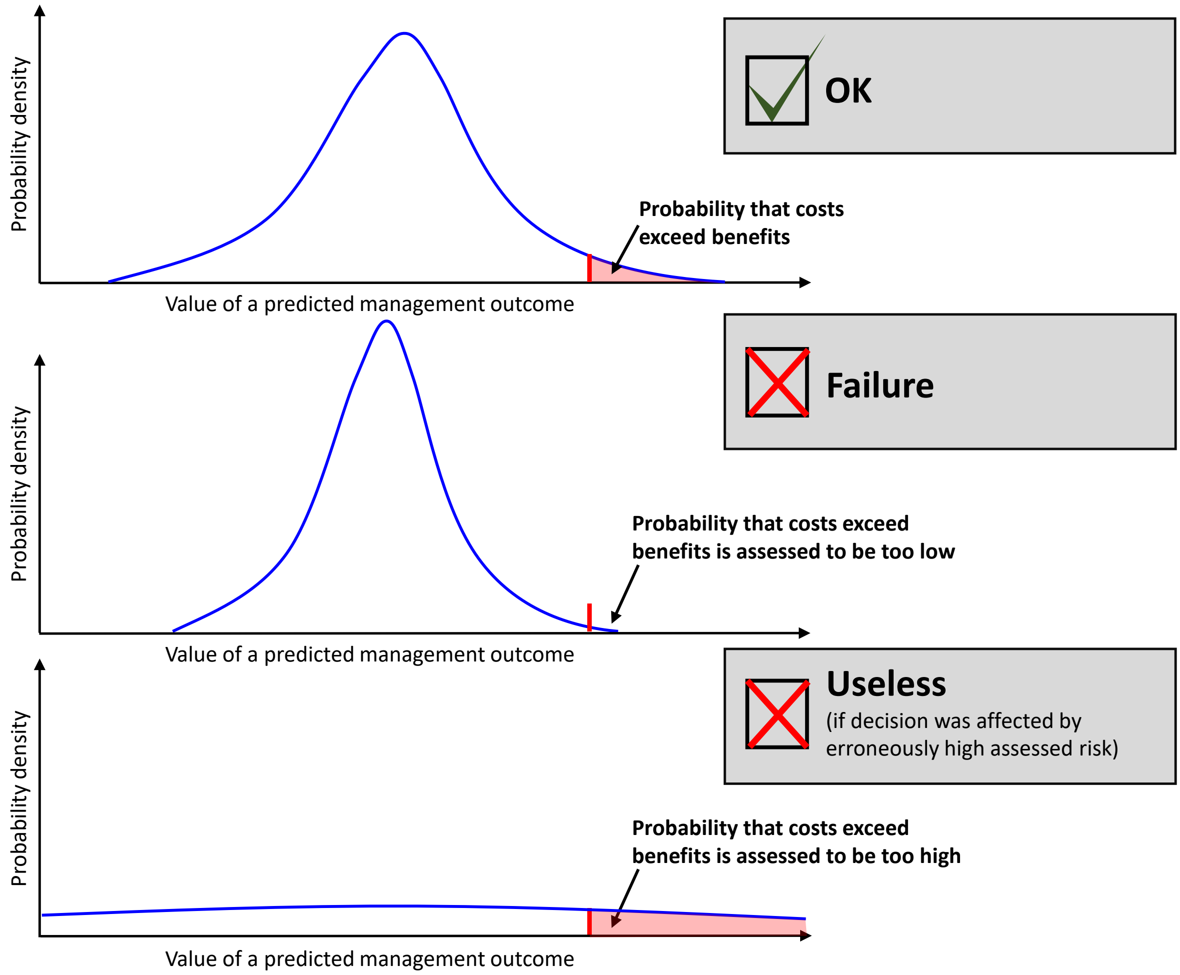


Figure 2.3. Decision support modelling fails when risks are underestimated. It is useless when (a) uncertainties are not reduced as much as available information allows, and (b) the consequential inflated assessment of risk affects a decision.

It is therefore apparent that any particular instance of decision-support modelling occupies a place on a rather unsatisfying continuum that extends between failure and uselessness. Failure must be avoided at all costs; however the price that must be paid for its avoidance may be a deterioration in decision-support modelling usefulness. Finding the best place on this continuum is a context-specific and prediction-specific challenge. Responses to this challenge will be subjective. It requires modeller expertise. It is part of a modeller's job statement to acquire this expertise.

3. FLOW OF INFORMATION

3.1 Introduction

This section briefly considers information. In Section 1 of this document we identified information as "that which reduces uncertainty". In Section 2 we characterised the goal of decision-support groundwater modelling as that of quantifying and reducing the uncertainties of decision-critical predictions. It follows that the decision-support modelling process must provide a means for information to flow from where it resides to where it is needed. Simulation is performed in support of this task. This is illustrated metaphorically in Figure 3.1.

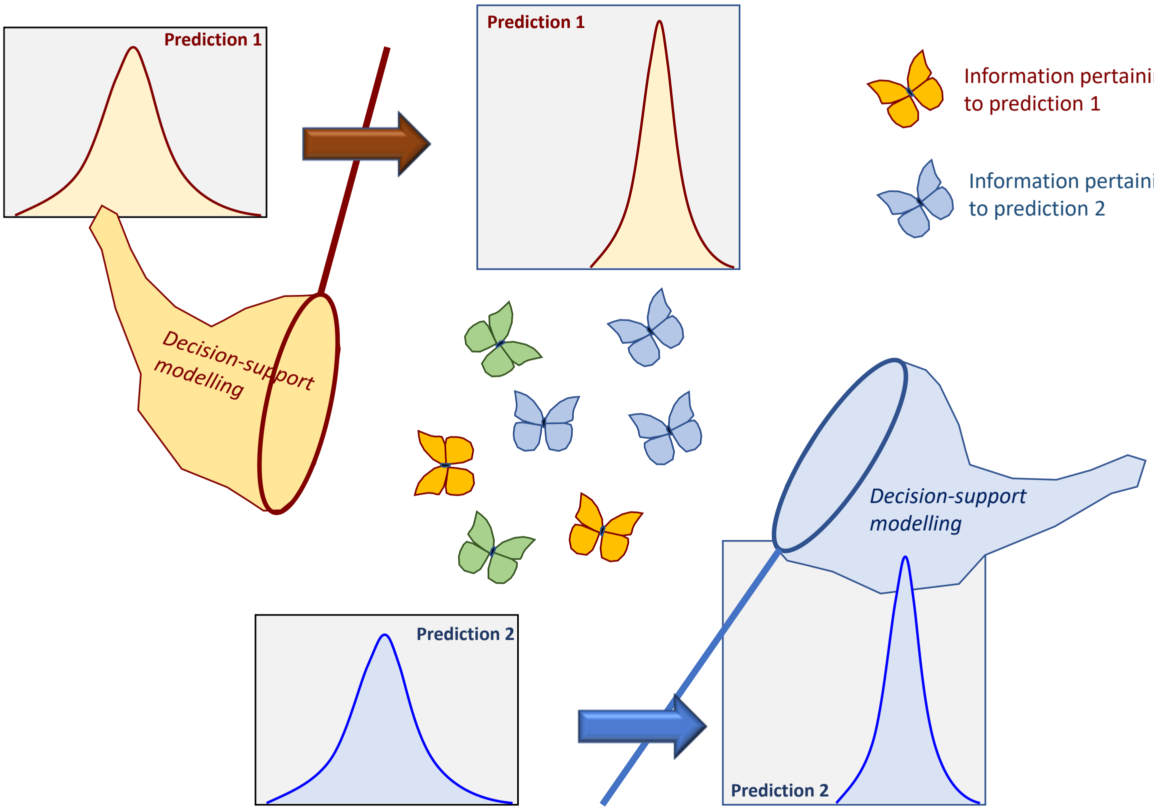


Figure 3.1. The aim of decision-support modelling is to gather information. This process is implemented most efficiently when it is tuned to reducing the uncertainty of one or a small number of predictions.

3.2 The Conceptual Model

3.2.1 General Considerations

Conventional wisdom dictates that construction of a conceptual model precede that of a numerical model. A conceptual model exists only "on paper". It identifies system components that are believed to impact system management. These are generally things that produce or capture water, or that control its flow. These items must therefore be represented in a numerical model of the system.

A conceptual model may also describe how items of hydraulic interest should be represented in a numerical model. For example, it may prescribe the type of boundary condition that is used to represent interaction of groundwater with a surface water body, or the disposition of model layers that is used to represent stratigraphic layering.

Conceptual model development is often seen as the work of one or a number of hydrogeologists. According to this view, the work of a numerical modeller begins once the conceptual model is complete. The job of the numerical modeller is to give numerical voice to the conceptual model.

In the opinion of the authors, this separation between conceptual and numerical model development often launches numerical modelling on a trajectory whereby it fails to live up to its decision-support potential. This is because the conceptual model which forms the blueprint for numerical model design should rest on concepts that extend beyond the flow of water; the conceptual model should also address the flow of information.

As has already been explained, simulation is only one component of the decision-support modelling process. The way in which it is implemented, and the ways in which a simulator must interact with software such as PEST/PEST++ to facilitate flow of information, must be addressed at every stage of development of a modelling strategy that best serves a particular management objective. The conceptual model on which numerical simulation is based must therefore identify where relevant information resides, and how it should be transferred to the decision-critical predictions whose uncertainties it will attempt to reduce. Furthermore, the possibility that different modelling strategies may be required to facilitate information transfer to different predictions should always be considered. (See Figure 2.1.)

A conceptual model should recognise that aspects of the subsurface that influence flow of groundwater, but whose details are uncertain, be represented in a numerical model in a way that reflects their uncertainties. It should specify that their numerical representation be responsive to information that is resident in measurements of system state. It should then identify how this information is best directed to them. Most importantly, it should endeavour to ensure that the numerical model can accommodate conceptual surprises once information starts to flow. All too often, a conceptual model dictates construction of a numerical model that is immune to new information, rather than open to it. We discuss this matter further in Section 9 of this document.

3.2.2 The Point of Reference

The imperatives of decision-support modelling require that a numerical model not only be responsive to information, but that it be capable of expressing the consequences of information insufficiency. If model design is prediction-specific (as it should be), then, naturally, a numerical model should represent, in an abstract or explicit way, those aspects of the subsurface to which a prediction of management interest is sensitive. If information scarcity does not allow unique representation of these features, then the conceptual model on which the numerical model is based should suggest how to represent them stochastically (i.e. probabilistically).

Traditionally, modellers have often been advised to adopt a parsimonious model construction and parameterisation strategy in decision-support contexts where information is scarce. We note here that such an approach does not necessarily respect the imperatives of decision-support modelling. For reasons that are elaborated in the following sections of this manuscript, the authors recommend that complexity be avoided unless it is necessary; however, a lack of ability to express subsurface details uniquely is not an argument for failure to represent them in one way or another. Decision-support modelling failure (see the previous section) is the inevitable consequence of an over-parsimonious modelling strategy. It must not be forgotten that those aspects of the subsurface that cannot be known are just as important as those that can be known if they influence a prediction of management interest. It is the latter that determine its uncertainty.

Recognition of this aspect of decision support model design reinforces the notion that a model should be built to make one, or only a small number, of predictions. Given the importance of representing in a model all factors which may impact an important prediction regardless of their estimability, it follows that a model which is required to make many different predictions must represent many different aspects of the subsurface, often in a stochastic sense. The numerical and cognitive burden of developing, history-matching and deploying such a model may erode the contribution that it can make to management of a groundwater system.

3.2.3 Where Information Resides in a Model

This is a subject which we address in greater detail later in this section. However we initiate the discussion now in order to provide context for the next subsection.

Numerical models solve partial differential equations that are defined over irregular two- and three-dimensional domains. Their use embodies information on groundwater flow and transport processes that prevail in a study area. Featured in these partial differential equations are the hydraulic properties of the medium through which groundwater flows; these include hydraulic conductivity and specific storage. These, and other properties of the subsurface, can vary in space and sometimes in time. If these were known everywhere, then predictions of future groundwater behaviour could be made with relative certainty (provided that system stresses such as pumping and recharge rates are also known). Unfortunately they are not. Furthermore, hydraulic properties such as hydraulic conductivity can exhibit large variations over short distances.

In most (but not all) decision-support contexts, it is lack of knowledge of the complex, three-dimensional distribution of hydraulic properties that has most influence on the uncertainties of decision-critical model predictions. It is in giving values to these properties that information finds a place in a groundwater model. Unfortunately, information is never sufficiently abundant to enable values to be assigned to all hydraulic properties at all points within the domain of a numerical model. Hence many (but not all) predictions of future system behaviour are likely to be accompanied by uncertainty.

3.3 Sources of Information

3.3.1 Information Source Categories

In common with the rest of this document, the discussion presented in this subsection is brief, somewhat abstract, and necessarily incomplete. However the categorisation of information sources that it provides is useful for understanding the flow of information in decision support modelling, and for implementing aspects of model design that enable its harvesting.

The two principal sources of information to which the decision-support modelling process must seek access are:

1. site characterisation data (including expert knowledge as it pertains to the site); and
2. measurements of system behaviour.

3.3.2 Site Characterisation

Insights emerging from site characterisation are often featured in a conceptual model. They include direct and indirect measurements of system properties, as well as three-dimensional mapping of rock types, and of hydraulically-important features such as faults. Site characterisation investigations also identify historical groundwater extraction rates, river flows, land uses and soil types. These enable identification of historical and future system stresses, or enable assignment of hydraulic properties to pertinent system submodels (such as ancillary recharge models).

An important feature of knowledge that emerges from site characterisation is its stochastic nature. Direct measurements may allow hydraulic properties to be ascribed to a few points within a study area. However, spatial interpolation between those points is uncertain – as is inference of lithologies between them. Hence a modeller cannot assign hydraulic properties to every cell of a numerical model's grid with certainty. Instead, he/she can only assign a joint probability distribution to cell-based properties. Generally, this probability distribution is complex – far too complex to represent using simple equations. Nevertheless, conceptually at least, as an outcome of site characterisation studies, a modeller possesses the ability to generate realisations of hydraulic properties that respect field measurements, as well as known and inferred geology. (The science of generating random realisations of subsurface lithologies and hydraulic properties based on information acquired from field measurements is called "geostatistics".)

3.3.3 Measurements of System Behaviour

"System behaviour" includes system states (for example heads and concentrations measured in boreholes) and fluxes (for example groundwater effluxes that comprise creek baseflow). Measurements of system behaviour may have been made one or many times at a few or many locations within a study area.

The behaviour of a natural system is, of course, determined by the stresses to which it is subjected, as well as by its hydraulic properties. It thus caries information pertaining to these stresses and properties – information that can be assimilated into a numerical model. Conceptually, assimilation of this information is achieved by adjusting a model's representation of hydraulic properties until it is able to replicate historical field measurements when simulating the system's response to historical stresses.

3.3.4 Other Types of Measurements

In many areas of interest, ancillary data are available that may not, at first sight, appear to fit neatly into either of the above categories. These include geophysical and remote sensing data. Nevertheless, for the purpose of the present discussion, the types of information that are hosted by these data can generally be assigned to either a "site characterisation" or "measurements of system behaviour" category.

Consider airborne electromagnetic (AEM) measurements. In coastal areas, electrical conductivity may be directly related to salt concentration. AEM measurements therefore comprise indirect measurements of system state. If the relationship between salt concentration and electrical conductivity (including the uncertainty in this relationship) is included in the history-matching process, the information that is hosted by these data can be used in hydraulic property inference.

In other contexts, measured responses to external geophysical excitations may be more directly related to system properties and/or system boundary conditions than to system states. Where this is the case, they may be used to develop stochastic descriptors of numerical model components which represent these properties and boundary conditions.

3.4 Model Receptacles for Information

3.4.1 Model Design Categories

We now continue our discussion of where information resides in a numerical model. In order to do this, we distinguish between two fundamentally different aspects of a model's design. We characterise these aspects using the terms “structure” and “parameters”.

By “structure” we mean those aspects of a model's design that pertain to its size, geometry and discretisation scheme. These include the number of its layers, the dimensions of its cells, its topology of cell connectivity, designation of its cells as active or inactive, disposition of its boundary conditions, as well as other aspects of its geometrical design. Collectively, these can be visualised as providing a kind of “wireframe image”, or “skeleton” of a numerical model.

In contrast, “parameters” specify a model's representation of system hydraulic properties. Numerical models often require that hydraulic properties be provided to it on a cell-by-cell basis. However in many modelling contexts, parameters are represented using higher order spatial devices. For example, a single value of a particular hydraulic property may be ascribed to an entire two- or three-dimensional zone. Alternatively, where pilot points are employed as a parameterisation device, hydraulic properties are ascribed to points that are strategically distributed in two- or three-dimensional space. These properties then undergo spatial interpolation to the model grid. The use of these and other parameterisation devices allows a modeller to characterise the distribution of hydraulic properties within a model domain with fewer parameters than the number of active cells in the model grid. A software package that calculates cell-by-cell hydraulic property values from values assigned to model parameters, and then records them on simulator input files is referred to as a “parameter preprocessor” herein. Obviously, one or a number of such preprocessors must be run prior to running the simulator if parameter values change. (Note that groundwater simulator preprocessors can also include ancillary submodels such as recharge models.)

The term “parameter” can be broadened to include any aspect of a model's input dataset that is not structural, and that may require alteration from model run to model run. These may include stresses such as historical and future pumping rates; they may also include coefficients that determine the rate of movement of water into and out of model boundaries.

In summary, the defining features of parameters are as follows.

- They are devices that allow easy adjustment of certain non-structural model inputs – often (but not always) inputs that pertain to hydraulic properties of a groundwater system or of its boundaries.
- While a spatial model may (and often should) have thousands of parameters, they are often far fewer in number than the number of model cells. Hence an individual parameter often possesses an area of influence that exceeds the spatial discretisation interval of a numerical model.
- Files which record parameters are rarely native simulator input files. The latter must be populated by parameter preprocessors that are part of an overall “model”, but that are executed ahead of the simulator.

3.4.2 Parameters as Receptacles for Information

Numbers that are assigned to parameters comprise information. Parameters therefore comprise receptacles for information. Their ease of adjustment is key to their ability to host information, and to transfer this information to decision-critical model predictions.

Because the values assigned to model parameters are easily varied, they can, in fact, be strategically random. That is, they can be samples of a probability distribution. Hence they can express information forthcoming from site characterisation studies. At the same time, their capacity for easy variation gives them agency in history-matching, whereby information that is resident in historical system behaviour is introduced to a model.

As has already been discussed, the task of decision-support groundwater modelling is to quantify and reduce the uncertainties of one or a number of decision-critical model predictions. This is achieved by transferring information from where it resides in data to where it is needed in a prediction. This information is extracted and delivered by the decision-support modelling process; while resident in a model, information is hosted by the model's parameters. The importance of model parameters cannot therefore be understated. In fact, they should be considered as the reason for construction of a numerical model in the first place.

Common perceptions of numerical modelling relegate parameters to a level of importance that is inferior to that of model structure. Where models are used to support decisions, this hierarchy should be reversed. Parameters are the reason for a model's existence. A model's structure is important to the extent that it plays host to parameters that house the information that the decision-support modelling process must assimilate, and then transfer it to decision-critical model predictions.

3.5 Transfer of Information

3.5.1 Bayes Equation

It is not possible to discuss transfer of information without mentioning Bayes equation.

Though the concepts on which Bayes equation rests are easy to understand, its implementation can sometimes be awkward. Fortunately, in many decision-support contexts (including those in which groundwater modelling operates), its deployment is supported by software packages which lighten the intellectual and numerical burden of its use.

Bayes equation 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 **k** is a vector of model parameters, while **h** is a vector of field measurements of system behaviour. (A vector is simply a collection of numbers.) *P*() signifies probability. A reading of Bayes equation from right to left mimics the flow of information which it describes. The components and operation of Bayes equation are illustrated in Figure 3.2.

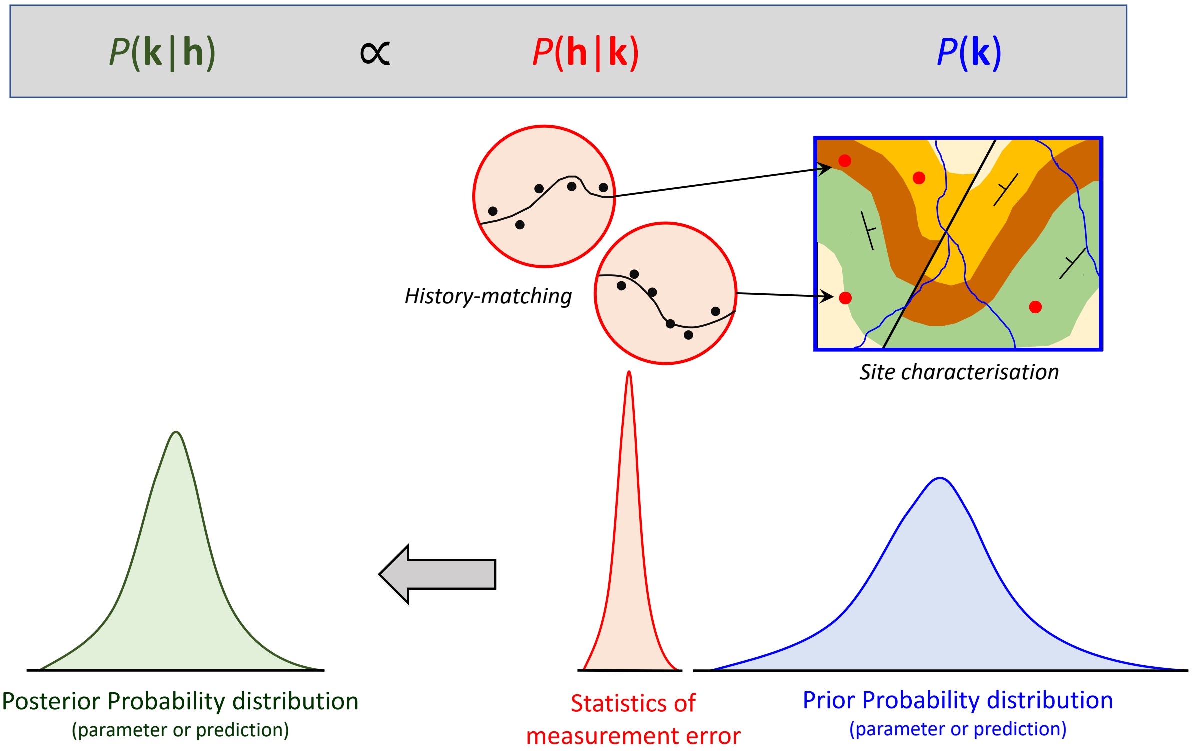


Figure 3.2. Conceptual representation of Bayes equation.

P(**k**) on the right side of Bayes equation refers to the so-called prior probability distribution of model parameters. This is what emerges from site characterisation studies. Its identification as a probability distribution expresses the fact that site characterisation cannot imbue the subsurface with unique hydraulic properties whose values are known everywhere. However, knowledge emerging from site characterisation can identify statistical relationships between properties of different types, and between properties of the same type at different locations. These relationships are born of the geological media to which these properties pertain. This knowledge can also bracket hydraulic property values between geologically-based lower and upper bounds.

The notation *P*(**k**|**h**) on the left side of Bayes equation is read as "the probability of **k**, as constrained by **h**". This denotes the so-called posterior probability distribution of **k**. It is the probability distribution of **k** after elements of **k** have received information contained in the history-matching dataset **h**. Information is transferred from that dataset to these parameters by constraining them such that pertinent model outputs are able to reproduce **h** to within limits set by measurement errors associated with **h**. Measurement error is itself a stochastic term. It is expressed by the first term on the right of equation 3.1, i.e. *P*(**h**|**k**). Where measurement errors are small, a tighter fit is warranted between pertinent model outputs and field measurements embodied in **h**. Where they are large, a looser fit should be sought so that the constraints on **k** are correspondingly looser; the posterior probability distribution of **k** is therefore broader.

Bayes equation makes it clear that information transfer from measurements of system behaviour to the parameters of a model takes place through history-matching. After history-matching constraints have been enforced, parameters are not as free as they previously were, for their values must now render model outputs compatible with field observations. However, depending on the information content of these observations, the constraints imposed by history-matching may still result in some parameters having a high degree of variability.

Regardless of their information content, Bayes equation makes it plain that the outcomes of history-matching are probabilistic. Parameter values are stochastic; they are not unique – either before or after history-matching.

Equation 3.1 refers to model parameters and not to a model prediction. The extent to which history-matching reduces the uncertainty of a decision-critical model prediction depends on the sensitivity of that prediction to parameters whose uncertainties may, or may not, have been reduced through history-matching. The posterior uncertainties of some predictions may be considerably reduced from their prior uncertainties. This applies particularly to predictions which are similar in nature and location to observations which comprise a history-matching dataset. In contrast, history-matching may reduce the uncertainties of other model predictions only slightly. Moore and Doherty (2005) show how little impact history-matching based on heads can have on the uncertainties of some contaminant transport predictions.

3.5.2 Subspace Analysis

The history-matching process can be analysed from a different, but complementary, perspective. This perspective enables explicit tracking of the flow of information that decision-support modelling enables. It allows us to classify predictions according to the extent to which they are informed by the two sources of information that are discussed above, i.e. information forthcoming from site characterisation and information forthcoming from measurements of system state. As will be discussed later in this document, this has important repercussions for model design.

Suppose that the relationships between a model's outputs and its parameters are linear when the model is operating under history-matching conditions. They can therefore be expressed using a matrix. This matrix is referred to as a sensitivity matrix, or a Jacobian matrix. We denote it using the letter **J**. This matrix contains the sensitivities of model outputs to which there are complementary field measurements to the parameters **k** of a model. If we subject this matrix to a numerical procedure known as singular value decomposition (SVD), it can be decomposed into three matrices whose product equals the original matrix. That is:

$$\mathbf{J} = \mathbf{USV}^{t} \tag{3.2}$$

(The "t" superscript designates matrix transposition wherein a matrix is flipped on its side.) This is not the place to explore the beautiful properties of the **U** and **V** matrices (they are, in fact, orthonormal), nor the elegance of SVD. It is sufficient to say that SVD allows parameter space to be subdivided into two orthogonal subspaces. Each of these subspaces is spanned by *combinations* of the original set of parameters **k**. In general, an individual member of **k** (i.e. an individual parameter) has a non-zero projection onto both subspaces. As is further discussed below, we refer to these two subspaces as the "solution space" and the "null space" of the original matrix (**J** in this case). In most history-matching contexts the dimensions of the null space are far greater than those of the solution space.

Let us represent the true, but unknown, hydraulic properties of a system by the vector $\mathbf{k}_0$. If history-matching is implemented using SVD, a unique parameter set $\underline{\mathbf{k}}_0$ is obtained. Mathematically, $\underline{\mathbf{k}}_0$ is the projection of unknown $\mathbf{k}_0$ onto the solution subspace. If a model is endowed with the $\underline{\mathbf{k}}_0$ parameter set, its outputs match the history-matching dataset **h** (to within limits set by measurement noise). The $\underline{\mathbf{k}}_0$ parameter set is the simplest parameter set that can achieve this level of model-to-measurement fit. This, and the fact that it is unique, endows $\underline{\mathbf{k}}_0$ with special status. It is the parameter set that is said to "calibrate" the model.

Now consider any parameter set $\mathbf{k}_n$ that lies within the null space of the Jacobian matrix. Because it lies in the null space of this matrix, it has no effect on model outputs under history-matching conditions (by definition). So if it is added to $\underline{\mathbf{k}}_0$, then the ($\underline{\mathbf{k}}_0$+$\mathbf{k}_n$) parameter set also allows the model to replicate field measurements **h**. Because the null space usually has many dimensions, an infinite number of $\mathbf{k}_n$ parameter sets can be found. The outcomes of history-matching are therefore nonunique; so (as Bayes equation teaches us) they must be expressed probabilistically. However the "calibrated parameter set" $\underline{\mathbf{k}}_0$ is unique, not because it is correct, but because it is the simplest parameter set that allows the model to replicate a history-matching dataset.

Concepts outlined in the above paragraphs are summarised in Figure 3.3.

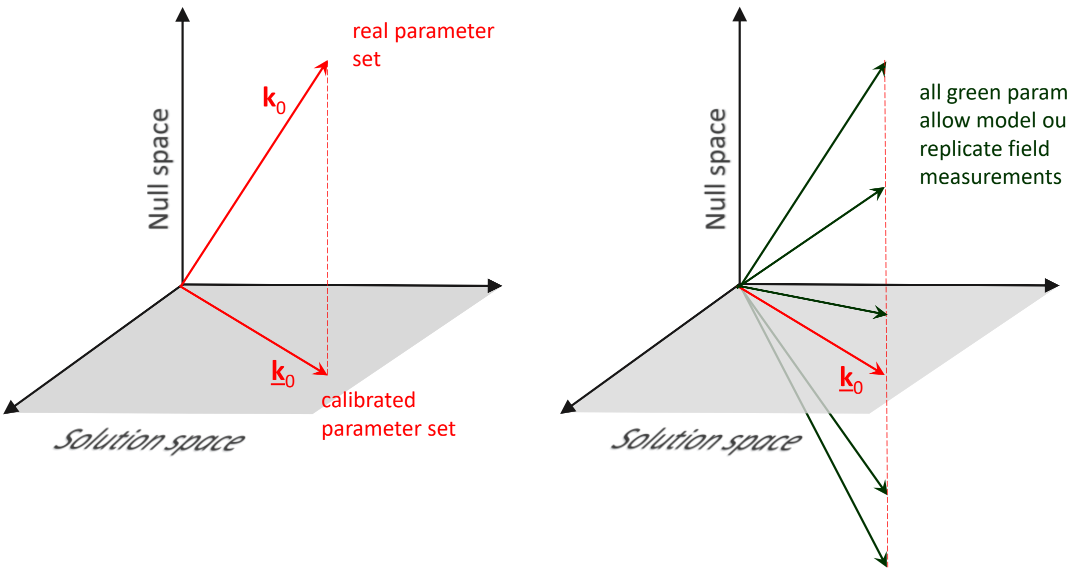


Figure 3.3. Subspaces of parameter space obtained through singular value decomposition.

Let us now consider a decision-critical prediction that a model is required to make. If this prediction is sensitive only to $\underline{\mathbf{k}}_0$, then it can be made uniquely. If it is sensitive only to parameter components which reside in the null space, then its uncertainty is not reduced at all through history-matching. If it is partly sensitive to $\underline{\mathbf{k}}_0$ and partly sensitive to null space parameter components, then its uncertainty is somewhat reduced through history-matching; however some uncertainty still remains.

3.5.3 Insights from SVD

Though at first sight esoteric, insights offered by SVD are profound.

Firstly, SVD puts the notion of the "calibrated model" into perspective. It shows that, notwithstanding traditional groundwater modelling practice, the calibrated model should not form the sole basis for model-assisted decision-making. A prediction made by a calibrated model is unlikely to be correct. However in being wrong, it can claim that its potential for incorrectness has been minimised. This is because history-matching achieved using SVD (or a similar numerical regularisation device such as Tikhonov regularisation) endows a model with an ability to make predictions that are somewhere near the centres of their posterior probability distributions; the potential for wrongness of any such prediction is thus rendered symmetrical with respect to the prediction itself.

Secondly, SVD (and its Tikhonov counterpart) provide insights into the relationship between the real (but unknown) parameter set $\mathbf{k}_0$ and the calibrated parameter set $\underline{\mathbf{k}}_0$. (By "real parameter set" we mean that which most closely represents the hydraulic properties of the real-world system that the model simulates.) It can be shown that calibrated parameter values are derived from real-world hydraulic property values through spatial averaging. The less information that is hosted by field measurements **h**, the greater is the amount of spatial averaging required for derivation of the unique, calibrated parameter set $\underline{\mathbf{k}}_0$, and the "blurrier" is the calibrated model's representation of hydraulic property reality; "blurriness" is a direct outcome of information insufficiency.

Thirdly, SVD identifies where information that is resident in a history-matching dataset flows as it enters a model. It flows to the solution subspace of parameter space; the null subspace is left untouched. Suppose that SVD reveals that the dimensionality of the solution space is D. Then the history-matching dataset **h** contains D separate pieces of information. These allow unique estimation of D combinations of parameters; these define, and span, the solution subspace. (Note that while these combinations of parameters can be estimated uniquely, they are not estimated without error, for they inherit error from field measurements.) If a model is endowed with a total of M parameters, then the null subspace of parameter space has $M – D$ dimensions. Combinations of parameters that span this subspace inherit stochasticity (i.e. a capacity for randomness) from the prior parameter probability distribution. Their capacity for randomness is undiminished through history-matching.

Finally, let us explain the term "combinations of parameters". The subdivision of parameter space that emerges from SVD is not normally such that a single parameter falls on either side of the boundary that separates the solution space from the null space.

Consider, for example, the situation shown in Figure 3.4. This represents a steady state model in which the history-matching dataset is comprised of a single head measurement. Parameter space is comprised of two parameters, These are hydraulic conductivities k_1 and k_2. As shown in Figure 3.4, these hydraulic conductivities are assigned to the two zones that exist between the observation well on the right of the model domain and a river on the left, the latter being represented by a fixed head boundary. Let us characterise the inverse of these hydraulic conductivities, as r_1 and r_2 ("r" stands for "resistance"). Obviously r_1 and r_2 cannot be estimated uniquely. However, it is possible to uniquely estimate the total resistance r between the river and the well using the single head measurement. The solution space therefore has a dimensionality of 1. It is spanned by r_1+r_2, the combination of parameters that is uniquely estimable on the basis of the calibration dataset. Meanwhile, the null space has a dimensionality of 1. It is spanned by r_1-r_2. The single head measurement that comprises the history-matching dataset has nothing to say about this combination of parameters.

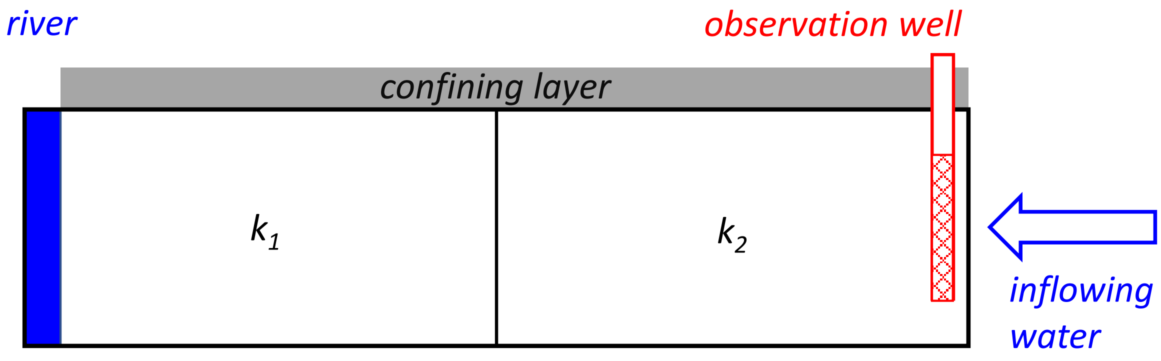


Figure 3.4 Water flows from right to left through a confined aquifer towards a river.

It is of interest to note that conclusions drawn from this simple example can be extended to far more complex examples. In general, the solution space is comprised of parameter averages, whereas the null space is comprised of parameterisation detail. This is hardly surprising; there is a limit to what can be inferred about a system from a limited number of observations of system state. These inferences tend to pertain to broadscale system properties, and not to system detail.

This example also makes an important point about parameterisation of a model – a point which we made previously, and to which we will return in later sections of this manuscript. It could be argued that the model of Figure 3.4 is over-parameterised because only one parameter is needed to fit the single head measurement. Suppose, however, that the purpose of the model is to predict the head half way between the observation well and the river. If a modeller suspects that the aquifer between the observation well and the river may be heterogeneous, then he/she should endow the model with at least two parameters so that the repercussions of information insufficiency on the posterior uncertainty of this prediction can be explored. This forestalls failure of the decision-support modelling process. As stated in the previous section, underestimation of predictive uncertainty incurs decision-support modelling failure.

4. HISTORY-MATCHING

4.1 Introduction

In this section we briefly discuss history-matching. In keeping with the theme of this manuscript, we address only those aspects of it that are salient to considerations of model complexity.

4.2 The Benefits of Many Parameters

In the previous section of this document we drew attention to the fact that a model may need more parameters than can be uniquely estimated. When it comes to quantification of predictive uncertainty, the (combinations of) parameters that cannot be estimated are as important as those that can. If they are excluded from a decision-support model, the uncertainties of decision-critical predictions may be understated.

Benefits that accrue from the use of a large number of parameters actually go deeper than this. Traditional wisdom once advised that adjustment of many parameters may promulgate over-fitting of model outputs to field data. This advice no longer applies. Modern-day inversion and uncertainty analysis software is well equipped to prevent this.

Where a model is undergoing calibration, numerical regularisation strategies such as Tikhonov and singular value decomposition ensure that heterogeneity introduced to a model's domain is the minimum required for model outputs to match field measurements. Their ability to do this in hydrogeologically meaningful ways increases with the number of parameters at their disposal. At the same time, model-to-measurement fit is restrained so that it does not exceed that which is expected. See Doherty (2015) for details.

Where highly parameterised methods are used to explore posterior uncertainty, the measurement dataset is augmented with random realisations of measurement noise. This prevents over-fitting to the unaugmented measurement dataset and ensures that uncertainties that arise from measurement error are taken into account. At the same time, the parameter adjustment process is numerically regularised to reduce departures from parameter fields that can be construed as samples of the prior parameter probability distribution. See Chen and Oliver (2013) and White (2018) for details.

A model cannot attain a good fit with a measurement dataset if it is endowed with too few adjustable parameters. The history-matching process therefore fails to extract as much information from that dataset as it actually contains. However, even if an adequate fit with a calibration dataset can be attained, information within that dataset may be directed to the wrong locations if parameterisation is too parsimonious. This can occur if spatial parameters are too broad in areal extent. If a decision-critical model prediction is sensitive to hydraulic properties that are assigned to parts of a model domain whose properties should not have been altered, then the history-matching process may have actually biased that prediction. History-match-induced parameter and predictive bias is a subject to which we shall return shortly.

A distinct benefit that is accrued through the use of many parameters is that it endows the history-matching process with the flexibility that it needs to accommodate surprises. It is the authors' experience that history-matching often introduces heterogeneity to a model in unexpected places. Unless parameters are spread liberally throughout a model domain, they may not be capable of expressing history-matching surprises. Sometimes unexpected heterogeneity is indicative of geological structures of which neither a modeller nor his/her hydrogeological colleagues were previously aware. On other occasions, the emergence of unexpected heterogeneity may reflect the need for parameters to adopt values that compensate for model structural defects as a good fit with a measurement dataset is pursued. This may indicate that some structural elements of the numerical model (and perhaps some elements of the conceptual model on which the numerical model's structure is based) are in need of refinement.

4.3 The Cost of Many Parameters

Use of many parameters can incur a high numerical cost where an inversion process requires that parameter-to-model-output sensitivities be computed using finite parameter differences. The numerical cost is that of running the model at least once for each parameter during every iteration of a parameter adjustment process. This cost can be reduced using methodologies such as "SVD-assist" (in which a limited number of SVD-based "super parameters" are adjusted after the first iteration of the history-matching process), or by using random parameter increments to compute an approximate Jacobian matrix. Members of the PEST and PEST++ suites provide these options.

The latter option in particular is exploited by the PESTPP-IES ensemble smoother. Exploration of the posterior parameter probability distribution requires that samples (i.e. realisations) of the prior parameter probability distribution be first generated. Parameters comprising these realisations are then adjusted until model outputs fit field measurements; in doing so, they thereby sample the posterior parameter probability distribution.

The computational efficiency of PESTPP-IES is achieved through use of an approximate sensitivity matrix on which parameter adjustment is based. This is calculated by running the model using the realisations themselves. The number of realisations that are required to ensure a good fit with a measurement dataset on the one hand, and to provide adequate characterisation of parameter and predictive uncertainty on the other hand, need only be slightly larger than the dimensionality of the solution space (see the previous section). This is despite the fact that each realisation can be comprised of hundreds of thousands – or even millions – of parameters. The number of model runs required per iteration of the realisation adjustment process is only slightly greater than the number of realisations.

It is apparent, then, that the numerical cost of using a large number of parameters is not prohibitive. However, it is incumbent on a modeller to endow parameters with a geologically plausible statistical characterisation. The greater the number of parameters, the more important does this become. When undertaking model calibration, numerical regularisation relies on this characterisation to maintain parameter sensibility in a minimalist way. When undertaking posterior uncertainty analysis through ensemble-based history matching, it is used to generate prior parameter realisations.

However, conceptual difficulties can sometimes arise when attempting to ascribe realistic stochastic descriptors to large numbers of parameters. Current technology does not allow statistical representation of geologically meaningful expressions of heterogeneity at the same spatial scale as that at which it is possible to introduce adjustable parameters to the domain of a numerical model. For example, it is difficult to give statistical voice to the notion that one or a number of hydraulically-significant faults of unknown length and width may (or may not) exist here (or there) within a model domain, so that history-matching can introduce fault-shaped heterogeneity to that domain if it needs to. Hence, while use of a large number of parameters poses few numerical problems, maintaining "stochastic control" of parameters as they are estimated or adjusted may become problematic.

4.4 Parameter Value Distortion

4.4.1 An Example

Consider the model depicted in section view on the left of Figure 4.1, and in plan view on the right of Figure 4.1. Two aquifers are separated by an aquitard. Water is extracted from well A, which is open to the lower aquifer. Loss of water incurred by this extraction is monitored in the stream. Suppose that this loss is small (a consequence of the aquitard that separates the stream from the extraction well), but is nevertheless detectable enough to be used in history-matching.

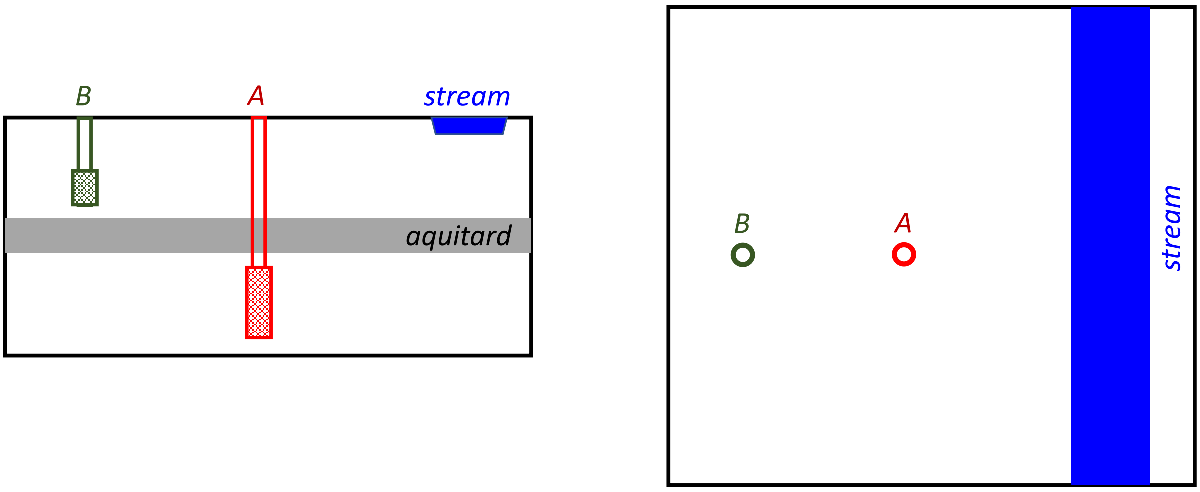


Figure 4.1. Two aquifers separated by an aquitard viewed in section (left) and in plan (right).

Let us suppose that site characterisation studies fail to reveal the significance of the aquitard. Conceptual/numerical modelling may therefore be based on the premise that flow of water is predominantly horizontal, and can be simulated using a single model layer; see Figure 4.2. History-matching then introduces a vertical band of low hydraulic conductivity between the extraction well and the river. Hydraulic conductivities assigned to this band allow the model to replicate loss of water from the stream while reducing this loss to its low observed value. (While hydrogeologists may raise their eyebrows at the outcomes of the model calibration process, they may judge that linear structural features are not unknown in the area, and that streams tend to follow them.)

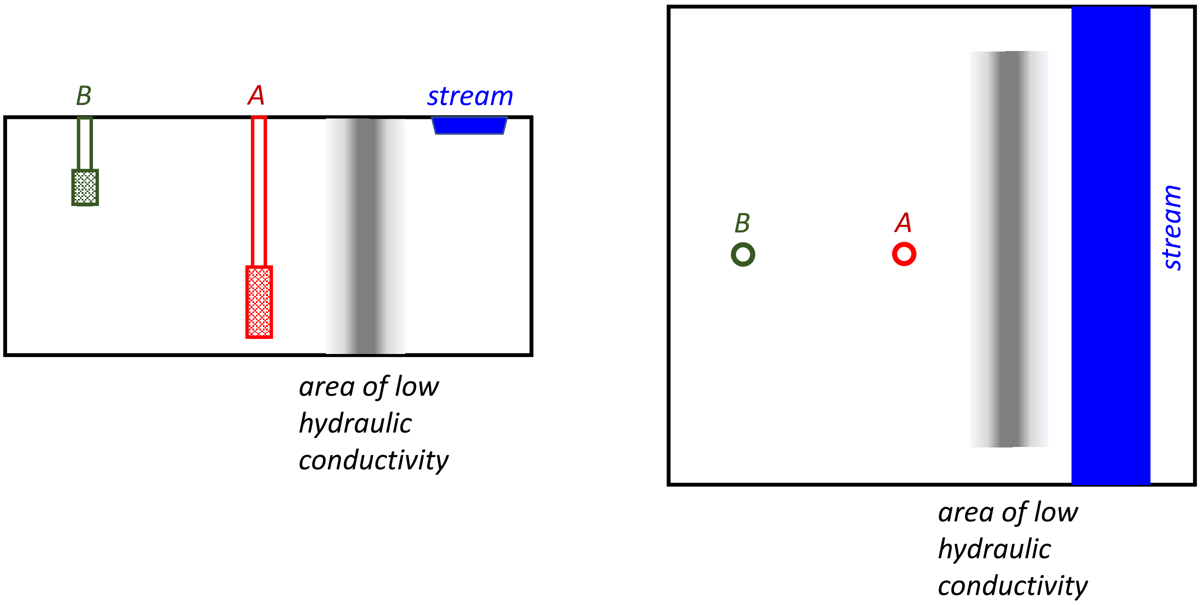


Figure 4.2. One-layer model calibrated against streamflow depletion incurred by extraction of water from well A.

Suppose that this obviously defective model is now used to make two predictions. Suppose firstly that it is asked to calculate loss of water from the stream if pumping from extraction well A is doubled. This is a prediction that the model is likely to make with a relatively high degree of accuracy. This is because the calibrated model represents the existence and strength of the hydraulic barrier that separates the extraction well from the river. The fact that this barrier is oriented in the wrong direction, and situated in the wrong place, makes little difference to its efficiency as a barrier that separates the source of stress from its point of impact. The model will successfully predict that if extraction from well A is doubled, then loss of water from the stream is eventually doubled.

Suppose now that the model is asked to predict the effect of extraction from a new well (well B situated in the upper aquifer) on the stream. According to the model, the effect of this extraction is minimal. Obviously, such a prediction is considerably in error.

4.4.2 Parameter Surrogacy

The above example illustrates the interplay between model structure and model parameters. The importance of parameters was discussed in the previous section of this manuscript. There, the role of model structure was characterised as that of playing host to model parameters. The above example illustrates that there is no “right” and no “wrong” structure. Use of a single layer does not impair the model’s ability to replicate historical measurements of system state, nor its ability to make useable predictions of a certain type. However for these things to happen, some of the model’s parameters must adopt values that compensate for its structural inadequacies. While this sustains the model’s ability to make some predictions, the requirement that some of its parameters adopt surrogate roles impairs its ability to make other predictions.

Doherty and Welter (2010), Doherty and Christensen (2011), White et al (2014) and Doherty (2015) explore the ramifications of model structural inadequacy in some detail. They demonstrate methodologies for analysing its potential to bias certain predictions, and draw the following conclusions.

- Model structural deficiencies do not necessarily impair a model's ability to replicate field measurements. Parameters may still be capable of receiving information that is resident in these measurements. However in doing so, some parameters may play roles that compensate for these structural deficiencies.
- For some predictions, this is of little consequence. These are predictions that are sensitive only to solution space parameter components of the "correct" but unknown model. (This is the two layer model of the preceding example). These predictions tend to resemble, in location and type, measurements that comprise the history-matching dataset. These predictions retain their integrity regardless of compensatory roles played by parameters to which they are sensitive. For these predictions, model structural defects can effectively be "calibrated out".
- Other predictions may incur bias through history-matching. These predictions are at least partly sensitive to some null space parameter components of the correct, but unknown, model. They are sensitive to solution space parameter components of the structurally-deficient model whose values have suffered history-matching-induced bias.
- In many circumstances, the propensity for history-matching to induce predictive bias can be ameliorated through "strategic" history-matching. A modeller may decide to omit those aspects of a history-matching dataset that are likely to induce parameter surrogacy. Alternatively, he/she may process measurements and corresponding model outputs before matching one to the other so that damaging components of measurement signals are removed, and so that those parts of measurement signals that are rich in prediction-pertinent information are retained and amplified.

We now discuss some of these points in greater detail.

4.4.3 Prediction Specificity

The benefits of tuning decision-support modelling to the making of one, or a small number, of predictions have already been outlined. Matters that are discussed in the present section of this manuscript reveal further benefits. These are central to recommendations that are made later in this document regarding appropriate model complexity.

To emphasise its importance, we rephrase one of the points that is listed above.

If a prediction is similar in location and type to one or a number of measurements that comprise a history-matching dataset, then that prediction is unlikely to suffer bias through history-matching, regardless of any bias that may be suffered by parameters to which it is sensitive.

It follows that the principal requirement of a decision-support model that is built to make predictions of this type is that it is able to replicate similar aspects of past system behaviour. At the same time, model parameters must be of sufficient number to support replication of this behaviour by the model. Predictive uncertainties are probably small, as these types of predictions are unlikely to be sensitive to null space parameter components. The need for adequate representation of the null space is therefore diminished.

4.4.4 The Art of History-Matching

The need for decision-support modelling to extract information from wherever it resides, and to direct that information to where it is needed, has been discussed at length. Hence, when subjecting a model to history-matching, a modeller must ask him/herself what parts of a measurement dataset are information-rich with respect to the prediction that he/she is required to make. He/she must also establish what parts of a measurement dataset may compromise the model's ability to make that prediction by forcing parameters to which it is sensitive to adopt compensatory roles. The former parts should be rendered highly visible in the history-matching process while the latter components should be removed from it, or their effects diluted. Both of these can be achieved through adoption of either or both of the following strategies.

- Processing of field data and complementary model outputs before fitting one to the other; and
- Employing an observation weighting scheme that penalises misfit of information-rich components of a measurement dataset, and that reduces or eliminates penalties incurred by failure to fit data that may compromise a model's ability to make a prediction of management interest.

A common example of the former strategy is the matching of temporal and vertical head differences in addition to, or instead of, heads themselves. Temporal head differences are highly informative of model storage and/or recharge parameters, while vertical head differences are highly informative of vertical hydraulic conductivity. These "difference measurements" should be weighted for visibility in the overall objective function that depicts model-to-measurement misfit.

An example of a strategy that eliminates measurements from a history-matching dataset where matching of these measurements may reduce the integrity of a decision-critical prediction, is provided in a GMDSI worked example report; see Doherty et al (2021). The authors of this worked example report show how it may be prudent to disregard near-stream head measurements when history-matching a model that is tasked with defining the capture zone of a near-stream production well. The complexities of alluvial sedimentation are unlikely to be accurately reflected in the layering of a numerical model that is charged with this task. Inferred vertical hydraulic conductivities and streambed conductances may therefore incur bias through history-matching. This can have a severe impact on model-calculated capture zones. By excluding near-stream head measurements from the history-matching dataset, the information that they hold is denied access to the decision-support modelling process. While the chances of predictive bias may thereby be reduced, the uncertainties of important model predictions are simultaneously increased.

This trade-off of bias against uncertainty is common to data processing of all types. For reasons that have already been outlined, predictive uncertainty is to be preferred over predictive bias when modelling to support environmental management. This is because uncertainty can be quantified while bias cannot.

4.4.5 The Art of Predicting

The point was made above that processed model outputs (for example temporal or vertical differences) may be less affected by model structural defects than raw model outputs. In order to reduce the propensity for history-matching-induced parameter bias, it was suggested that these be matched with similarly-processed field measurements in addition to, or instead of, raw model outputs.

Similar considerations apply to decision-critical model predictions. White et al (2014) show that predictive differences are often less uncertain than predictive absolutes; at the same time, they may have greater immunity from structural and/or history-matching-induced bias. A model's superior ability to predict differences over absolutes is something that appeals to the intuition of most modellers. Hence, where possible, it may be better to base system management on model-predicted alterations to system states and fluxes, than on model-predicted system states and fluxes themselves. The enhanced bias-immunity of these predictions may reduce the requirement for high levels of model complexity. The benefits of prediction-specific model design are again demonstrated.

4.5 The Dimensions of Complexity

In the previous section of this document, we dichotomised model design into two separate categories – namely structure and parameters. The two are related, for model structure plays host to model parameters. The relationship between the two was further investigated in the present section where we discussed how imperfections in model structure may induce bias in values assigned to model parameters during history-matching. For some predictions, history-matching-induced parametric bias may exacerbate bias that arises from structural deficiencies alone; for other predictions, one may cancel the other.

The purpose of the present document is to examine the issue of model complexity, and to provide concepts that can assist a modeller in choosing a level of complexity that is appropriate for his/her decision-support needs. Discussions that have been presented so far suggest that “complexity” is too general a term to be useful. Design of the decision-support modelling process must address the twin issues of structural complexity and parameterisation complexity, as well as the relationship between the two. We do this shortly.

5. TRADITIONAL DECISION-SUPPORT MODELLING

5.1 General

In the following section of this manuscript we formulate a "roadmap" that can guide modellers in selecting levels of model structural and parameterisation complexity that are appropriate for their decision-support needs. Before doing this, however, we devote the present section of this document to a brief discussion of problems that beset decision-support modelling as it is currently undertaken. Many of these problems are a direct consequence of unnecessary structural complexity.

5.2 Some Practical Considerations

As we have outlined in previous sections of this manuscript, the decision-support modelling process should be one of information transfer. This requires that model parameters be plentiful, adjustable, and capable of stochastic representation. It also places history-matching at the heart of the decision-support modelling workflow. At the same time, it mandates that numerical simulators be used in conjunction with model-value-adding software packages such as those provided by the PEST and PEST++ suites.

In order to perform their data assimilation and uncertainty quantification tasks, programs such as those belonging to the PEST and PEST++ suites must run a simulator, together with parameter preprocessors and model output postprocessors, many times. On each occasion that a program from this suite runs a numerical model, it provides the model with a set of parameters that are appropriate for that run. Sometimes the values of some parameters are unusually high or unusually low as uncertainty limits are explored, or as parameters assimilate information from field measurement datasets.

For this process to work, groundwater simulation must be numerically stable and relatively fast. The faster that a model runs, the more options are available for the exercise of creativity in decision-support modelling. While modern numerical methodologies that rely on random parameter generation to fill approximate Jacobian matrices can indeed tolerate moderate to high model run times (especially if model runs are undertaken in parallel), these Jacobian matrices are, unfortunately, too approximate to provide a basis for linear analysis. Methodologies based on linear analyses can be used to evaluate data worth, track the flow of information, determine contributions made by different model components to the uncertainties of decision-critical model predictions, infer the costs and benefits of various model simplification strategies, and detect potential predictive bias; see Moore and Doherty (2021) for a recent example of their deployment. Analyses such as these require that Jacobian matrices be filled using finite parameter differences. At least one model run is therefore required for each parameter.

The ability to fill a Jacobian matrix using finite parameter differences leaves open the possibility of model calibration as a precursor to posterior uncertainty analysis. As is explained in Section 3 of this document, calibration provides a solution to the history-matching problem that suppresses the emergence of any heterogeneity that is not directly supported by a measurement dataset. As such, it provides a mechanism for assessing a model's design and structure. If calibration-emergent parameter patterns suggest that parameters must adopt surrogate roles that compensate for model structural defects, a modeller has the opportunity to revise this structure if he/she feels that this will improve the decision-support utility of his/her model.

5.3 Physically-Based Models

5.3.1 Model Structural Complexity

A common justification for model structural complexity is that a structurally complex model is more “physically-based” than a structurally simple model. Implied in this argument is that greater model structural complexity somehow promulgates greater predictive accuracy. It must not be forgotten, however, that it is information, and information alone, that can reduce predictive uncertainty. It follows that if a complex model structure can express information that emerges from site characterisation studies, then it does indeed have the potential to reduce the uncertainties of at least some decision-critical predictions. However, it is salient to recall that information that is typically forthcoming from site characterisation studies should be expressed stochastically. Adoption of only one realisation of a range of possible structures can bias decision-critical model predictions – either on its own, or because it forces parameters to adopt values which compensate for its flaws as it undergoes history-matching.

Model structure is a problematic receptacle for site information as it is neither stochastic nor adjustable. In contrast, parameters are both of these. Unfortunately, however, the pervasive use of parameters to replace structure also has its problems, for there is some information that parameters simply cannot hold.

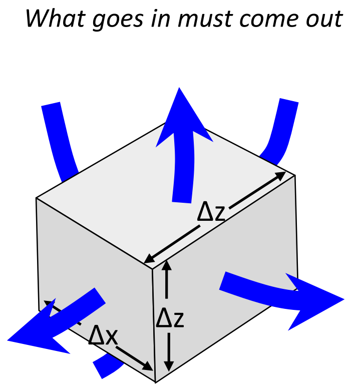
Use of parameters instead of structure as the primary receptacles for site information, inasmuch as this is possible, may require somewhat abstract representation of subsurface processes and properties. We argue below, and in Doherty (2011) and Doherty and Moore (2019), that this is not necessarily a bad thing. Furthermore, in many management contexts it provides vastly superior decision-support utility than use of a complex, slow-running, numerically unstable, “physically-based” model.

5.3.2 Hydraulic Property Upscaling

Even the most structurally complex numerical model is replete with abstractions – abstractions that can erode its claim to being physically-based.

Hydraulic properties that are ascribed to the grid cells of a structurally complex model are necessarily upscaled representations of hydraulic properties that prevail in the real world. Upscaling is a matter that has received considerable attention in the petroleum and groundwater modelling literature; see, for example, Farmer (2002) and more recently Canon (2018).

The partial differential equations on which numerical simulation is based describe flow of fluid in and out of “representative elementary volumes”. These tiny volumes are occupied by porous media whose properties are homogeneous; see Figure 5.1a. However, in a numerical model, discretised forms of these equations are applied to comparatively large cells of a model grid; see Figure 5.1b. The material that occupies these considerably larger volumes is heterogeneous. Upscaling refers to the averaging process through which model-cell-scale hydraulic properties are derived from point-scale hydraulic properties in a way that preserves the validity of the discretised partial differential equations of groundwater flow. Theory shows that this averaging process can only be approximate. Furthermore, in most practical cases, the averaging process is flow-regime-dependent; see, for example, Whitacker (1969), Dagan (1979) and Zhang et al (2008). Cell-based hydraulic properties, and the “physically-based” models which host them, are therefore tainted with abstraction.



$$\frac{\partial}{\partial x}\left(K_x \frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_y \frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_z \frac{\partial h}{\partial z}\right)+q_s=S_s \frac{\partial h}{\partial t}$$

Figure 5.1a. A representative elementary volume.

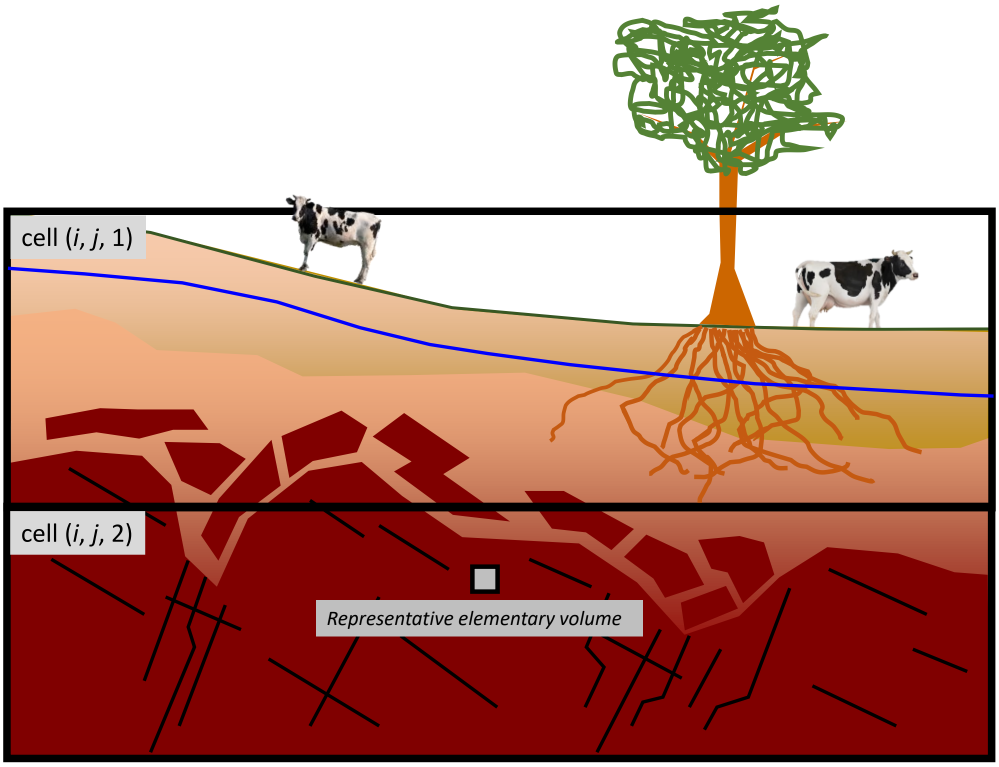


Figure 5.1b. Two cells of a model grid.

Properties such as conductance that are assigned to model boundary conditions bear even looser relationships with the properties of the real-world system whose behaviour a physically-based model seeks to emulate. Other model properties, such as hydrodynamic dispersion, dual domain transfer rate, and even specific yield, are upscaled by definition; it is often a difficult matter to relate their values to measurable material properties.

5.3.3 Geological Structure

A major problem in many groundwater modelling contexts is representation of discrete, continuous features such as faults, shear zones and buried river channels that can propagate drawdown and/or contaminants over considerable distances. Faults are particularly troublesome, as they can impede flow of groundwater in one direction and enhance it in another (Bense and Person, 2006; Bense et al, 2013). At the same time, displacement along a fault plane can juxtapose aquifers that are otherwise separated by an aquitard. However, the locations, dispositions and geometric/hydraulic properties of faults are often only poorly known. Sometimes it is possible to learn something of their existence and properties through history-matching; more often it is not.

An attraction of a “picture perfect” rather than abstract approach to decision-support groundwater modelling, is that important structural features such as faults can be represented explicitly. Use of an unstructured grid facilitates representation of cross-layer connections incurred by fault-induced stratigraphic offsets.

A problem with explicit representation of faults in a groundwater model is that any such representation is unlikely to be correct. Drilling and/or seismic data may reveal a fault’s location and throw at a small number of places. This may support integrity of its representation at those places. However, its representation at other places is likely to be approximate. Meanwhile, suspected faults, and smaller faults associated with large faults (of which there may be many) are not represented at all. Furthermore, while it may be possible to represent faulting-induced stratigraphic offsets (to the extent that these are known) in the structure of a model grid, properties of the fault that affect flow of water parallel to its dip, along its strike and across its displacement plane can only be guessed.

Any explicit representation in a numerical model of important structural features such as faults is as likely to be incorrect as it is to be correct. Given that model structural errors can induce bias of their own accord, or when model parameters are adjusted to compensate for them, a question that naturally arises is whether explicit representation of uncertain but important structural features has the potential to do more harm than good. It is incumbent on decision-support groundwater modelling that it represents the full range of *effects* of known and possible structural features on predictions of management interest. This may not be too difficult if these features are represented parametrically rather than structurally (more on this below). Errors incurred by adjustment-enabling abstraction may be small compared with the large predictive uncertainties that their (possible) existence incurs.

In contrast, explicit representation of all known and suspected features of hydraulic significance in a model’s structure is likely to be prohibitively expensive, especially if this is done stochastically (as it should be). Difficulties of explicit stochastic representation are exacerbated if each realisation of complex subsurface structures requires re-calibration of the model.

5.3.4 A Common (and Flawed) Approach to Decision-Support Modelling

Despite its many problems, the allure of model structural complexity is strong. A groundwater model that resembles a figment of geological imagination is less likely to be criticised by hydrogeologists, and indeed by the public at large, than a more abstract model. This is partly because there is a tendency for these groups to take the meaning of “model” too literally. It also suggests that people are more likely to trust an inanimate object, with some hydrogeologic features which they recognise, than the expertise of a modeller.

A recurrent theme of this document is that decision-support modelling is an activity rather than a deliverable, and that this activity is one of information acquisition and information transfer. This activity has metrics for failure and usefulness; see Section 2.5. These metrics do not include pictorial replication of an imagined subsurface.

Inevitable consequences of succumbing to the allure of model complexity are long model run times and numerical delicacy. For reasons that have already been stated, these make data assimilation and uncertainty quantification difficult, if not impossible. Furthermore, design errors that are inevitably hardwired into a complex model structure may obstruct achievement of a good fit with field measurements, and therefore assimilation of the information that they bear.

In many circumstances where modellers opt for structural complexity, their response to the near-impossibility of performing effective history-matching and uncertainty analysis is to reduce the number of model parameters to that which they decree to be a "manageable number". This only exacerbates the deficiencies of the decision-support modelling process that they have created, as it removes receptacles for information from the simulator. A combination of long model run times, numerical instability, parsimonious parameterisation and model structural defects may then render achievement of a good fit between model outputs and field measurements impossible. Failure to assimilate important information is often then proclaimed to be a virtue rather than a failing. This self-serving argument is based on the premise that over-fitting of model outputs to field data has been studiously avoided. Meanwhile, the complexity and expense of the modelling exercise, together with a passing resemblance that the model's structure bears to prevailing concepts of subsurface conditions, constitutes its claim to decision enablement.

A model such as this does not satisfy the metrics that we have outlined in Section 2.5 for avoiding decision support failure on the one hand and decision-support uselessness on the other. Instead of providing a voice for information, the modelling process effectively denies field measurements of system behaviour any opportunity to challenge the many conceptual details on which construction of the structurally complex numerical model rests.

5.4 Benefits of Abstraction

In this and previous sections of this manuscript we explain that:

- The primary task of decision-support groundwater modelling is acquisition and transfer of information.
- Where possible, this requires adjustable/stochastic representation within a model of components of its design that have the potential to receive and express information.
- Where possible, it also requires adjustable/stochastic representation of components of its design for which information is scarce, but that may nevertheless influence decision-critical predictions.

As has already been discussed, a model's parameters are, by design, capable of adjustment by software such a PEST/PEST++ in order to receive incoming information. At the same time, they are capable of stochastic representation of information insufficiency. It follows that, to the extent to which this is possible, decision-support model design should lean towards structural simplicity and parametric complexity.

Naturally, this approach to model design requires abstraction. It may require, for example, that parts of a model where structural features such as faults are known or suspected to exist be endowed with a superfluity of parameters. It may require that these parameters be endowed with a stochastic characterisation that encourages history-matching-emergence of enhanced vertical hydraulic conductivity in a horizontally continuous manner. This can enable along-fault flow of groundwater at the same time as it simulates displacement-induced cross-aquitard connectivity. This mode of fault representation provides maximum receptivity to expert knowledge and information forthcoming from site characterisation, at the same time as it provides a basis for stochastic exploration of the repercussions of information shortfalls.

Use of parameters to do the stochastic and data assimilation work required of decision-support modelling is not restricted to parametric representation of discrete geological features. As has already been stated, it is the task of decision-support modelling to represent the *effects* of hydrogeological complexity, rather than its details. This provides ample opportunities for parameter-based abstraction in representation of the many nuances of geology that are important to flow of water and transport of contaminants.

When assessing the impacts of anthropogenic stresses on a groundwater system, hydrogeological details often matter. However, the specifics of hydrogeological details are often only vaguely known. Stochastic representation of subsurface hydraulic properties, and an openness to information that is resident in measurements of system behaviour, therefore comprise essential elements of the decision-support modelling process. A relatively simple (but appropriate) model structure that is able to host a sophisticated, flexible parameterisation scheme offers a far better means of exploring the decision-pertinent repercussions of hydrogeological detail than a complex (and possibly expensive) model structure that embodies a single realisation of this detail. It is also superior to a limited number of structurally-complex models that provide only sparse stochastic expression of this detail. The expense of the latter approach is amplified when it is remembered that each of these structurally-distinct models must be individually history-matched.

5.5 Problems with Abstraction

There are, of course, limitations to the extent to which a complex parameterisation scheme embedded in a relatively simple model structure can support the imperatives of decision-support groundwater modelling. The predictions that are required of a decision-support model determine its minimum level of structural complexity. For example, a steady-state model cannot predict groundwater levels pursuant to an extreme recharge event. Nor can a single layer model predict propagation of drawdown across an aquitard.

Adoption by parameters of surrogate roles that compensate for model structural defects has already been discussed. Generally, this is easily recognised as history-match-emergent parameter values and patterns that are not in accordance with hydrogeological expectations. The extent to which this is problematic depends on the prediction; as we have explained, for some predictions it matters, while for others it does not. Where it matters, the locations and character of aberrant parameter patterns will normally indicate where model structural refinements are required; they may also suggest the nature of these refinements.

There will be occasions where, despite evidence of parameter surrogacy in a model that must make a prediction that is not entirely informed by a measurement dataset, a modeller may decide that he/she will not alter its structure. Perhaps time/budget precludes this, or perhaps ways in which a model's structure can be improved are unclear. In this case the surrogate roles that some parameters must play during history-matching must be taken into account when defining their prior probability distributions. As Bayes equation shows, the posterior parameter probability distribution (and those of predictions that the model is required to make) depends on the prior probability distribution of model parameters. Parameter values and patterns that emerge from history-matching may guide a modeller in assigning appropriate prior probability distributions to parameters that acquire compensatory roles during the history-matching process. Ideally, their revised stochastic characterisation allows their compensatory roles to be taken into account when evaluating posterior predictive uncertainty.

It is acknowledged that model abstraction can diminish the didactic value of numerical simulation, particularly for non-modelling stakeholders. This is a matter that requires a sociological rather than a numerical remedy. It is not addressed in this manuscript.

5.6 Starting from the Problem and Working Backwards

The authors of this manuscript see strategic abstraction as the key to successful decision-support groundwater modelling in most modelling contexts. However, any decision to simplify or abstractify a model must have a point of reference. This reference point can only be the prediction that a model is required to make.

The metrics for design of a model are the same as those that are outlined in Section 2.5 for decision-support modelling in general. Hence if structural simplification induces bias in a prediction, then that bias must somehow be included in model-quantified predictive uncertainty intervals (notwithstanding the difficulty of calculating bias). If simplification-induced uncertainties are greater than uncertainty reductions that are accrued through use of a stable, fast-running model to process and assimilate site data, then a more complex model structure is warranted.

Everything depends on the prediction. Once it is clear what prediction a model is required to make, it can be designed accordingly. If groundwater system management requires the making of a number of different predictions, a number of different models may be required, the construction and deployment of each of which is tuned to the prediction that it must make.

In the next section of this manuscript we propose a “model complexity roadmap”. This roadmap respects the dichotomy between model structural and parameterisation complexity. It also respects the need to start from a prediction and work backwards when formulating a modelling process that is capable of quantifying and reducing the uncertainty of that prediction.

6. THE PREDICTION SPECTRUM

6.1 SVD Revisited

Section 3 of this manuscript discussed singular value decomposition (SVD). It described how, by performing SVD on the Jacobian matrix that relates model outputs used in history-matching to parameters employed by a model, parameter space can be subdivided into two orthogonal subspaces. One of these subspaces is the null space. This is composed of combinations of parameters that are uninformed by currently available measurements of system state and flux. These parameter combinations often denote system hydraulic property detail. The other subspace is the solution space. This subspace is spanned by combinations of parameters that are uniquely informed by the measurement dataset. These combinations of parameters often pertain to broadscale, or spatially-averaged, system properties.

As a numerical tool, SVD is fundamental to the operation of programs of the PEST and PEST++ suites that estimate or alter the values of parameters in order to allow model outputs to replicate field measurements. However in this manuscript, our interest in SVD focusses on insights that it offers into flow of information from measurements that comprise a history-matching dataset to parameters that are hosted by a numerical model, and ultimately to predictions that are required of the decision-support modelling process.

Recall that the purpose of decision-support modelling is to quantify and reduce the uncertainties of decision-critical predictions. It is information that reduces uncertainty; it is ultimately to model predictions where information that is harvested by the decision-support modelling process must flow.

6.2 Predictions

The value of a model prediction depends on the values of its parameters. Because model parameters are uncertain, so too are model predictions. The posterior uncertainty of a prediction depends on the posterior uncertainties of parameters to which it is sensitive.

Suppose that a prediction is sensitive mainly to combinations of parameters that lie within the history-matching solution space. Then its uncertainty is likely to be relatively small for, by definition, it is a recipient of information. Furthermore, to the extent that it is uncertain, most of its uncertainty is inherited from errors (sometimes referred to as “noise”) that accompany field measurements. That is to say, its uncertainty is not a result of lack of information, but of contamination of information. For ease of reference, we refer to this type of prediction as a “data-driven prediction”.

Consider now a prediction that is sensitive solely to combinations of parameters that occupy the history-matching null space. This prediction is uninformed by the measurement dataset. Its posterior uncertainty is therefore undiminished from its prior uncertainty. We refer to this type of prediction as an “expert-knowledge-driven prediction”. This is because expert knowledge, and information that emerges from site characterisation, are the only sources of information for this prediction.

Many predictions of management interest are sensitive to combinations of parameters that occupy both the solution and null spaces. Depending on their relative sensitivities to parameter combinations that span both of these spaces, their posterior uncertainties may be greatly reduced, or only mildly reduced, by history-matching. We refer to these types of predictions as “mixed predictions”.

6.3 Two Models

In order to reap the benefits of insights provided by SVD, and in order to embody these insights in strategies for decision-support model design, it is necessary to draw a distinction between two models. The first is the “reality model”. This is a notional model – a model that has no defects because it is as complex as reality. This model possesses as many parameters as are required to express the heterogeneity of real-world hydraulic properties. The second model is the decision-support model that exists on a modeller’s computer; we refer to the latter as the “modeller’s model”. The classification of predictions presented in this section actually pertains to SVD as notionally undertaken on the reality model.

This is not as esoteric as it may at first appear, for understanding the reality model is a prerequisite for designing the modeller’s model. A modeller is normally intuitively aware of which predictions are data-driven, which predictions are expert-knowledge-driven, and which predictions are mixed. He/she can use this awareness to develop strategies for information flow that is enabled by a “modeller’s model” that is appropriate for a particular prediction’s type.

It is in the nature of data-driven predictions that they are often required at the same locations as those at which measurements comprising a history-matching dataset were made. Furthermore, it is anticipated that system stresses that will prevail under predictive conditions will be similar to those that prevailed when these measurements were taken. Hence when a model is required to make a data-driven prediction, its task is to draw inferences about the future behaviour of a system directly from measurements of its past behaviour. As we will discuss, under these circumstances the decision-support modelling process is actually a type of machine learning process.

Intuition and experience can also guide a modeller in recognizing expert-knowledge-driven predictions. Predictions that depend on system process details and on hydraulic property details are often of this type. Predictions of contaminant fate and transport also often fall into this prediction category, particularly if a history-matching dataset is comprised only of water level measurements.

In theory, a modeller can use linear analysis tools, such as are provided with the PEST and PyEMU suites, to determine the solution and null space dependencies of a prediction. Use of these tools assumes the existence of a Jacobian matrix. However, this matrix can only be calculated using the modeller’s model, and not the reality model. Hence caution must be exercised when using them to characterise a prediction. Nevertheless, it may be possible to recognise expert-knowledge-driven predictions and mixed predictions based on calculations which employ this matrix. The same does not apply to data-driven predictions. This is because, as we will discuss in the following sections, a decision-support model that is used to make a data-driven-prediction does not need to even represent the history-matching null space.

In general, characterisation of a prediction is something that should be done prior to devising a decision-support modelling strategy. The design of this strategy can then exploit the benefits of this characterisation.

6.4 A Roadmap

The “roadmap” depicted in Table 6.1 suggests decision-support modelling strategies that are appropriate for different prediction types. Its details are the subject of the next three sections of this manuscript. Recommendations that comprise the lanes of this roadmap implement the principle of starting from the problem (i.e. a prediction required of the decision-support modelling process) and working backwards. The approaches to decision-support modelling that are described by these lanes distinguish between structural complexity and parameterisation complexity.

Prediction driven by	Complexity		Strategy
	Structural	Parameter	
Data	moderate	moderate	Pursue good fit with field measurements
Expert knowledge	high	high	Express expert knowledge using geostatistics
	low	low	Worst case scenario analysis
Mixed	high	high	Model structure expresses hydrogeological detail
	moderate	moderate to high	Fit first and ask questions later

Table 6.1. Decision-support modelling roadmap.

The roadmap is schematic only. It makes no claims to being an algorithm that can automate selection of an appropriate modelling strategy for any and every environmental management context, for each context has its own nuances. Its purpose is to show that, in any decision-support circumstance that a modeller may encounter, a chain of logic can be formulated that links the system-management hypothesis that he/she must test to the sources of information on which the testing process must rely. This chain of logic is schematised in Figure 6.1.

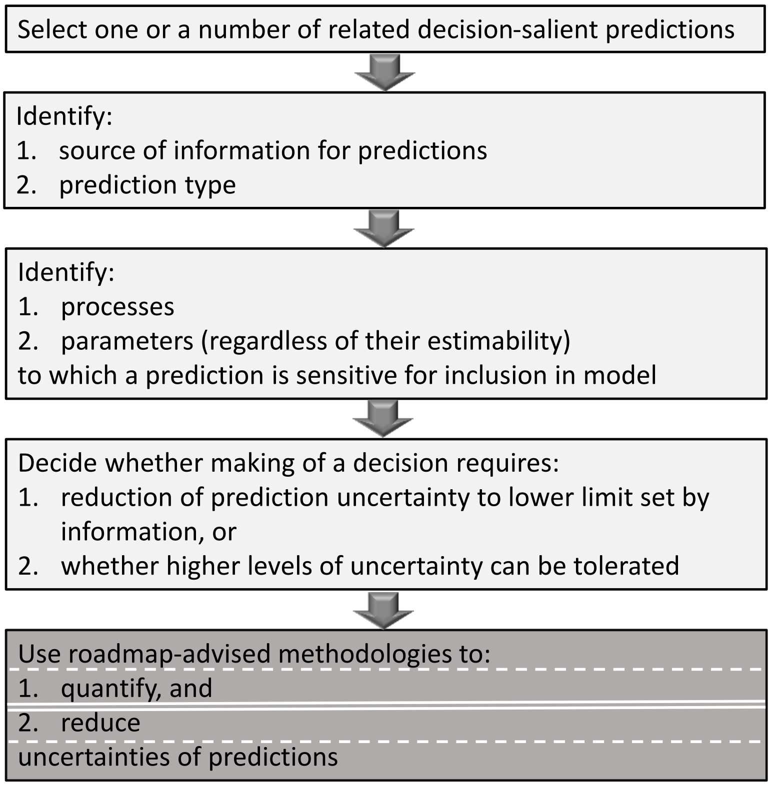


Figure 6.1. Schematic decision-support modelling workflow.

6.5 Model Assessment

The roadmap of Table 6.1 enshrines the notion that decision-support modelling is best understood and implemented if it is viewed through the lens of the prediction that it must make. It is also based on the premise that a prediction is best understood in terms of the data that inform it. These concepts apply not only to the way in which decision-support modelling should be implemented. They apply also to the way in which decision-support modelling should be judged by its stakeholders. These include peer-reviewers, regulators, the public, and possibly the legal profession. In the following sections of this document we reflect on criteria that should be used to review the decision-support utility of modelling that is focussed on each of the prediction types that are identified above.

Metrics for decision support failure and decision support utility were presented in Section 2.5 of this manuscript. These are universal; they are independent of prediction type. The roadmap of Table 6.1 is formulated with respect for these metrics in mind. Traditional model-appraisal criteria such as goodness of fit with a calibration dataset, and reasonableness of estimated parameter values, should serve these more universal metrics. Because of this, they have different relevance, and should be applied differently, when assessing decision-support modelling strategies that are designed for the making of predictions of different types.

7 DATA-DRIVEN PREDICTIONS

7.1 Examples

Modelling in support of data-driven predictions is relatively commonplace in other branches of environmental science. Rainfall-runoff modelling is an example. This is normally accomplished using so-called "lumped parameter" models for which only 10 to 15 parameters are open for calibration adjustment. These models are loosely physically-based. A key assumption in their use is that future weather patterns will not be too different from those which prevailed in the past. The historical behaviour of a watershed is thus the key to its future behaviour. Most of the uncertainty that is associated with predictions made by these kinds of models is inherited from their inability to exactly replicate measurements of historical streamflow, these being an outcome of errors in these streamflow measurements and approximations in model design. It does not arise from failure of the past to inform the future.

Groundwater models that are used to set yearly water allocations for groundwater users are also required to make predictions which are, in general, data-driven. Time intervals over which predictions are required are generally short – perhaps a season or two into the future. Data available for model calibration are often plentiful, being comprised of many water levels measured in many observation wells at many times.

Time series analysis of groundwater levels can also be considered to fall into this predictive category. The "prediction" that is yielded by these analyses is often identification of the principal sources of a complex groundwater signal, and/or the behaviour of a groundwater system in the short-term future. See, for example, Obergfell et al (2019) and Peterson and Fulton (2019).

A similar decision-support situation exists where groundwater drawdowns in a number of observation wells incurred by pumping from a number of extraction wells must be predicted, and where observations of past drawdowns incurred by historical pumping from the same extraction wells are plentiful. Regardless of prevailing hydraulic property heterogeneity, the calibrated "model" that is used to predict future drawdowns can be comprised of a set of analytical radial or linear flow models (depending on the local geology) – one for each extraction/observation well pair. However, unless the subsurface is perfectly homogeneous and boundary conditions are infinitely distant, this model must not be used for calculating drawdowns at locations other than those at which measurements of historical drawdown are available; a numerical model must be used for that.

Data-driven predictions are schematised in Figure 7.1.

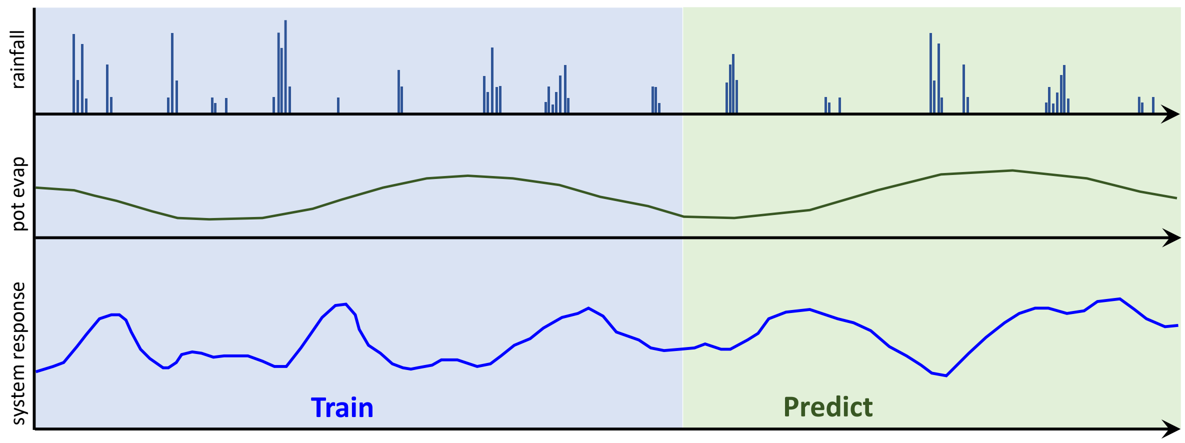


Figure 7.1 Conceptualisation of data-driven decision-support modelling.

7.2 Strategies

7.2.1 Structural Complexity

Modelling to make data-driven predictions is relatively forgiving. The principal model design criterion is that it be capable of being calibrated against historical field measurements, for these measurements are the primary source of information on future system behaviour. The model must be endowed with enough parameters to allow replication of not just the gross features of a measurement dataset, but also its temporal and spatial nuances as well if these details are required in predictions of future system behaviour.

In some environmental management contexts, data-driven simulation may be only loosely physically-based. In fact, it may be little more than a machine learning algorithm. Alternatively, it may be comprised of one or a number of analytical and/or lumped-parameter models that are inspired by reality, but do not attempt to simulate any more of reality than a single time series. Meanwhile, its loose relationship with reality ensures a design that includes enough processes, and enough accompanying parameters, to hold the information to which a measurement dataset plays host. Because this information is transferred to predictions of future system behaviour at the same or similar locations under similar circumstances, this information is sufficient to make these predictions with little uncertainty.

In other cases, the simulator may be more sophisticated than this. For example, a transient groundwater model that is used for evaluation of seasonal water allocations may be accompanied by one or a number of ancillary lumped-parameter recharge models. Parameters belonging to the main model and to all ancillary models may be history-match-adjustable. In these management contexts, the history-matching dataset is generally large. It may be comprised of lengthy time series of groundwater levels measured in widely-distributed observation wells. These may be accompanied by time series of groundwater contributions to streamflow, as well as independent estimates of long-term recharge, historical water usage and even satellite estimates of evapotranspiration. Collectively, these may provide a relatively complete record of the system's managed behaviour under a variety of climatic conditions. Replication of this behaviour supports credible predictions of short-term future behaviour under a variety of climatic scenarios that managers may wish to explore.

Once again, the past is the key to the future. The model must possess sufficient process and structural complexity to simulate important nuances of system behaviour, and sufficient parameterisation complexity to allow it to reproduce nuances of system behaviour that were measured in the past. A physically-based design can ensure both of these. At the same time, the model must run relatively quickly, and possess unwavering numerical stability. The latter qualities may require some compromises in its design that challenge its "physical basis" (for example use of a reduced number of model layers). However, these must not harm its ability to reproduce the measured past.

7.2.2 Parameterisation Complexity

A model that is used to make data-driven predictions may host more parameters than can be estimated uniquely. Hence its parameter space may include a null space. A superfluity of parameters may ensure that a good fit is attained with a measurement dataset. This is especially the case where parameters have a spatial connotation (for example pilot point parameters that host aquifer hydraulic properties), as it is not known in advance of the history-matching process where heterogeneity must be introduced to a model domain.

Note that just because a model has a superfluity of parameters, it does not follow that a data-driven prediction of management interest will be sensitive to null space parameter combinations. The prediction may be sensitive to individual parameters that lie partly in the solution space and partly in the null space. It can do this while being sensitive only to parameter *combinations* that span the solution space. We remind the reader of the role of parameter combinations in defining solution and null spaces; see Section 3.5 of this manuscript.

Alternatively, the history-matching process may reveal that a prediction of management interest is indeed partly sensitive to combinations of parameters that span the null space. If so, its reliance on measurement-uninformed parameter combinations will contribute (perhaps significantly) to its uncertainty. Naturally, this erodes the status of the prediction as "data-driven". A modeller must then decide how to deal with this situation. Perhaps it is possible to adequately characterise the uncertainty of a management-salient prediction using a low-dimensional null space; model structural and parameterisation complexity can thereby remain unaltered.

If a model is performing signal separation using methods that are based on time series analysis, exact characterisation of predictive uncertainty may not even be necessary; the history-matching process may simply need to reveal whether a signal can be ascribed uniquely to one signal source or another. However, in other cases where null space dependency is discovered, it may be decided that a higher level of structural complexity is warranted, and that the model should become more physically-based. This can allow expert knowledge to better inform the prior probability distributions of null space parameter combinations to which the prediction is sensitive.

7.2.3 Other Considerations

In decision-support contexts where data-driven predictions are required, use of software packages such as PEST and PEST++ which automate and optimise history-matching becomes essential.

When deploying packages such as these, a modeller should ensure that the objective function whose value is minimised through history-matching is carefully defined. It may need to include components that accentuate certain spatial and temporal subtleties of historical system behaviour so that predictions of future system behaviour can also include these subtleties. Inclusion of spatial and temporal head differences in the overall objective function has already been discussed. Other objective function formulation strategies that improve a model's ability to make a data-driven prediction include design of a weighting scheme that ensures visibility of prediction-pertinent measurements in the overall objective function, and recognition of temporal correlation in stochastic characterisation of model-to-measurement misfit.

As has already been discussed, the level of structural complexity with which a model of this type is endowed must be sufficient to host enough parameters of enough types to promulgate a good fit between model outputs and field measurements. However, the level of structural complexity does not need to be so high as to ensure that these parameters do not adopt surrogate roles that compensate for model structural simplifications as they are adjusted. Hence, when using software such as PEST and PEST++ to fit model outputs to field measurements, a modeller may endow parameters with generous lower and upper bounds. These can exceed the range of values that are expected for the hydraulic properties after which they are named.

7.3 The Roadmap Lane

Table 7.1 highlights the lane of the decision-support modelling roadmap to which data-driven predictions belong. Parameterisation complexity is depicted as “moderate”. Parameters must be plentiful enough to support a good fit between model outputs and the measurement dataset. In some management contexts, a superfluity of parameters may be warranted in order to allow the history-matching process to be selective in its introduction of heterogeneity to different parts of a model domain. The number of parameters may also need to be high enough to allow discovery of prediction sensitivity to null space parameter combinations if this possibility exists. However, the premise on which the data-driven prediction classification rests is that the null space dependency of a prediction is small. Hence expert-knowledge-based representation of prior parameter uncertainty is not required for evaluation of predictive uncertainty, as the latter is determined solely by the integrity (or otherwise) of the measurement dataset.

Prediction driven by	Complexity		Strategy
	Structural	Parameter	
Data	moderate	moderate	Pursue good fit with field measurements
Expert knowledge	high	high	Express expert knowledge using geostatistics
	low	low	Worst case scenario analysis
Mixed	high	high	Model structure expresses hydrogeological detail
	moderate	moderate to high	Fit first and ask questions later

Table 7.1. Decision support modelling roadmap. The lane for data-driven predictions is highlighted.

The level of model structural complexity need only therefore be moderate, as the link between the model and “reality” need only be weak. However, it needs to be strong enough to host parameters in sufficient numbers to carry out the tasks outlined above. Because structural complexity is only moderate, parameters may incur bias as they are adjusted during history-matching. However, it is the nature of data-driven predictions that they do not inherit this bias.

7.4 Model Assessment

7.4.1 Goodness of Fit

By definition, data-driven predictions are informed by field measurements of system states and fluxes. Model design must be such as to extract this information through history-matching. Ideally, model-to-measurement fit that is achieved through this process should be commensurate with measurement error.

It follows that in his/her reporting of the modelling process, a modeller should provide reviewers and stakeholders with an ability to assess the fit that he/she was able to attain between model outputs and field measurements. However, summary statistics such as root mean square misfit and Nash-Sutcliffe coefficient tend to hide more than they reveal. Success or otherwise in extracting prediction-pertinent information from historical system behaviour is best demonstrated by graphs which directly compare model outputs with field measurements in a temporal/spatial setting.

7.4.2 Parameter Credibility

For a model that is tasked with making a data-driven prediction, attainment of a good fit with a measurement dataset overrides all other considerations – including those of parameter credibility. As is stated above, the integrity of a prediction of this type does not rely on the integrity of parameters to which it is sensitive; parameters are strictly “information bearers”.

It follows that a model of this type should not be judged too harshly if values that are estimated for its parameters are “unrealistic” when compared with the range of values that are normally associated with hydraulic properties after which they are named.

8. EXPERT-KNOWLEDGE-DRIVEN PREDICTIONS

8.1 Examples

Groundwater management often requires that predictions be made of its future behaviour under conditions that will be very different from those which prevailed in the past. Under these circumstances, it may not be possible to estimate the values of parameters to which decision-critical predictions are sensitive by matching model outputs to field measurements of past system behaviour, as the past may provide little clue as to what will happen in the future. The values of prediction-sensitive parameters can only therefore be informed by expert knowledge and by site characterisation. Data emerging from the latter may include direct measurements of system properties at one or a number of locations, at scales that may, or may not, be applicable to future system management. Regardless of its composition, the defining quality of information on which decision-critical predictions must rely under these management circumstances is its stochastic (i.e. probabilistic) nature.

The impact of coal seam gas extraction on hydraulic pressures in other aquifers, on spring discharges, and on stream baseflows, exemplifies predictions of this type. Calculation of impact must rely on geological, geophysical and geochemical appraisal of possible impact pathways, and on an assessment of the hydraulic properties of materials which occupy these pathways. The latter cannot be known exactly. They can only be surmised.

Another context where measurements of historical system behaviour may have little bearing on management-salient predictions of future system behaviour is that of contaminant transport. Moore and Doherty (2005) demonstrate the limited capacity of head measurements to reduce the uncertainties of contaminant travel time predictions. This is because the parameter combinations to which these predictions are sensitive lie almost exclusively in the history-matching null space; by definition, information from field measurements does not infiltrate this space.

In the example provided by Moore and Doherty (2005), heterogeneity of hydraulic conductivity was characterised by a multiGaussian distribution. In real geological media, patterns of heterogeneity are more complicated than this. They are often narrow, extensive in one or two dimensions, and of limited extent in the remaining direction. They may be associated with tectonic features such as faults, or with alluvial features such as paleo river channels. Unless observation wells are plentiful, these features may be invisible to the history-matching process. However, they may exert a considerable influence on predictions of management interest.

8.2 Discussion

8.2.1 Costs and Benefits of History-Matching

When making expert-knowledge-driven predictions, history-matching can sometimes be dispensed with. This is because attempts to assimilate information that is resident in a history-matching dataset may bias decision-critical model predictions at the same time as it does little to reduce their uncertainties. As has been discussed, introduction of predictive bias is a risk that always accompanies history-matching, as a model’s structure is always simpler than that of reality. The need for parameters to be continuously-adjustable when undergoing history-matching, and therefore to be somewhat abstract in nature, may also contribute to history-matching-induced predictive bias.

8.2.2 A Note on History-Matching Technology

Algorithms on which modern-day data assimilation software are based assume a continuous relationship between model parameters and model outputs. They therefore assume that a parameter field that fails to fit a calibration dataset can be “morphed” into one that does by continuous adjustment of those components of it that compromise model-to-measurement fit. These assumptions fail where subsurface features of hydrogeological significance are discrete and discontinuous, and where part of their significance lies in their ability to connect or disconnect other features of hydrogeological significance.

With present technology, it is not possible to efficiently manipulate the geometries and connectivities of complex, elongate, structural and stratigraphic features in order for model outputs to match field measurements. This applies particularly to hydrogeological features that are categorical in nature; that is, they are discrete geobodies of unknown shape that may exist at one place but not at another. It is therefore difficult to explore the posterior uncertainty of a prediction that is sensitive to these features, for it is not possible to generate many stochastic realisations of them that all allow a model to replicate field measurements of system state and fluxes.

History-matching and posterior uncertainty analysis are best served by parameterisation devices such as pilot points. Hydraulic property values that are ascribed to these points are adjustable and continuously variable. Prior probability distributions that are assigned to them tend to be multiGaussian or piecewise multiGaussian (possibly after appropriate transformation). All of this enables their effortless use in conjunction with high-end inversion and uncertainty analysis software. Unfortunately, however, their ability to represent complex patterns of connectivity between categorical geobodies with convoluted shapes is limited.

History-matching therefore requires compromises. The cost of these compromises is small in many management contexts. In these contexts the benefits of predictive uncertainty reduction accrued through history-matching may be worth the cost of possible predictive bias incurred by use of “blunt” parameterisation instruments such as pilot points because the information content of a measurement dataset with respect to a prediction of management interest may be high. In contrast, it may not be worth the cost where the information content of a measurement dataset with respect to a prediction of management interest is low.

8.2.3 Options

Lifting of the burden of history-matching from a modeller’s shoulders opens up possibilities for decision-support modelling that are otherwise unavailable. One of these is the deployment of geostatistical methods that allow stochastic representation of complex, categorical geology. The other is direct exploration of the pessimistic end of a predictive probability distribution through worst case analysis. As we now discuss, each of these options has advantages and disadvantages.

8.3 Strategy 1: Geostatistical Methods

8.3.1 Description

Modern-day geostatistical software offers many options for stochastic generation of “realistic” representations of subsurface geology. Some packages specialise in representing particular geological features such as fracture networks, or layered sequences of lithologies. Others are more general. The most general methodology of all is multiple point geostatistics. This supports generation of two- and three-dimensional stochastic fields that can include both categorical and continuous variables. The nature and style of connectedness between different components of these fields are learned from one or more training images.

Conceptually, geostatistically-based decision-support modelling is not a difficult procedure. A realisation of geology is generated, and then transferred to a model. Individual geobodies or lithofacies within the model are then populated with hydraulic properties. These can be uniform, or spatially variable; in the latter case they may be based on an appropriate multiGaussian distribution. Once the model domain has been equipped with hydraulic properties, the model is run and a prediction is made. If this process is repeated a sufficiently large number of times, the probability distribution of a prediction can be established. This is schematised in Figure 8.1.

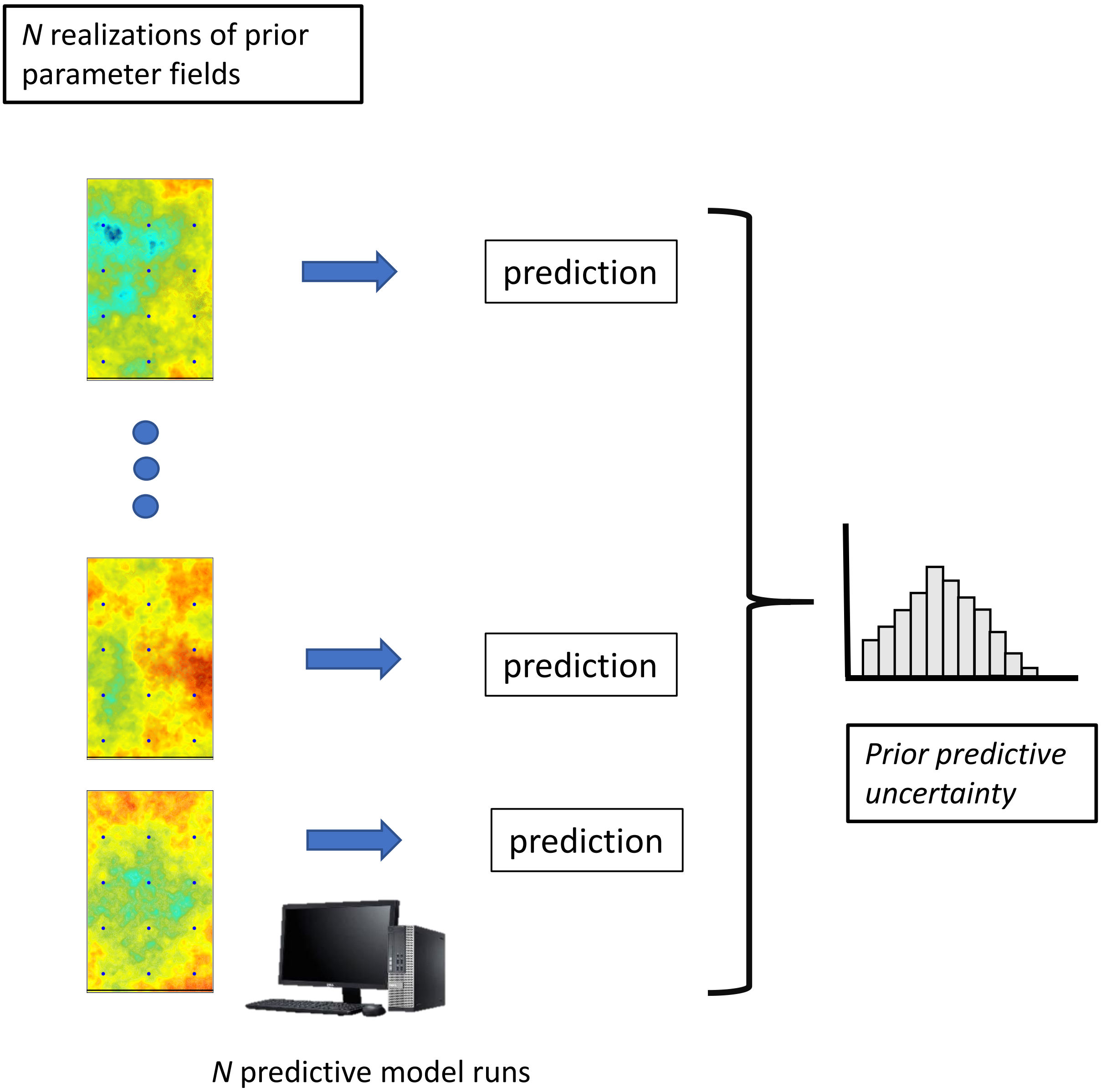


Figure 8.1. Using stochastic realisations of parameters to evaluate the prior uncertainty of a prediction.

The above process involves no history-matching. Hence the prior probability distribution of a prediction becomes a surrogate for its posterior probability distribution. The expectation that the history-matching dataset is almost devoid of prediction-pertinent information justifies this approach. Conservatism in predictive uncertainty evaluation accrued through failure to attempt history-matching respects metrics for decision-support modelling failure that are discussed in Section 2.5 of this document.

8.3.2 Prior-Data Conflict

Despite its limited utility in reducing predictive uncertainty, failure to compare model outputs with field measurements of system behaviour can have a significant negative consequence. It denies these measurements the opportunity to challenge the concepts on which a model's structure, and geostatistical population of it, are based.

It is therefore wise to run the model over the past, as well as over the future, on each occasion that it is populated by a geostatistical realisation. Pertinent model outputs can then be compared with complementary field measurements. It is not expected that these outputs will reproduce field measurements particularly well for any realisation. What is expected, however, is that the space spanned by all realisation-specific model outputs at a particular location should include field measurements that were made at that same location. If they do not, then the model's structure, and/or the conceptual basis for geostatistical field generation, is called into question. So too is the calculated predictive probability distribution of any prediction that is analysed using that model. This state of affairs is often referred to as "prior-data conflict". It indicates that the model's structure, and/or the geostatistical basis for populating it, must be reconsidered.

8.3.3 Primitive Rejection Sampling

Suppose, however, that a probabilistic comparison between model outputs and field measurements does not reveal prior-data conflict. Having verified the absence of conflict, a modeller may take the opportunity to discard geostatistical realisations of model hydraulic properties for which the departure between model outputs and field measurements is very high. This process comprises a primitive form of "rejection sampling". This is a Bayesian method that allows refinement of a prior probability distribution so that it is more aligned with a posterior probability distribution.

8.3.4 Data Space Inversion

If a model is populated with a suite of geostatistical realisations, and then run over the past and the future using each of these realisations, it may be possible to construct a joint probability distribution that links past system behaviour directly to future system behaviour. This joint probability distribution can then be directly conditioned by actual field measurements of past system behaviour. An approximation to the posterior probability distribution of a prediction of management interest can then be evaluated without the need to adjust model parameters, or run the model any more. See Sun and Durlofsky (2017). The DSI program provided with the PEST suite supports data space inversion.

By definition, the posterior probability distribution of an expert-knowledge-driven prediction resembles its prior probability distribution. Data space inversion may be used to verify that this is actually the case or, if not, which aspects of a measurement dataset inform it.

8.3.5 Some Problems

The principal problem facing geostatistically-based decision-support modelling is that of realisation diversity. Three-dimensional geostatistical realisations of subsurface properties and lithologies may not be intricate enough, nor diverse enough, to expose realistic possibilities of lithological connectedness. Hence, they may preclude representation of key impact pathways. Given the complexities of modern-day geostatistical algorithms, this problem may be difficult to detect, and even more difficult to rectify. In contrast, the range of predictive possibilities that may emerge from a more abstract representation of subsurface hydraulic properties based on continuous stochastic variables can often be increased simply by raising the variance or correlation length ascribed to much simpler probability distributions.

Another problem with geostatistical methods is their relevance to groundwater modelling. Many of these methods were developed for use in the geotechnical and petroleum industries where they work well in simulating fracture networks or sedimentary sequences that characterise a variety of depositional environments. However, they are not designed to simulate complex patterns of hydraulic property heterogeneity, such as those that are associated with differential weathering of igneous and metamorphic rocks.

8.3.6 Roadmap Lane

Table 8.1 highlights the roadmap lane occupied by geostatistically-based decision-support modelling. We characterise it as requiring high structural complexity and high parameterisation complexity. Nevertheless, in spite of ticking both complexity boxes, it comprises a relatively forgiving form of modelling as parameters do not require history-matching adjustment. The need for short model run times is therefore reduced; model numerical delicacy can be tolerated.

Prediction driven by	**Complexity**		**Strategy**
	Structural	**Parameter**	
Data	moderate	moderate	Pursue good fit with field measurements
Expert knowledge	high	high	Express expert knowledge using geostatistics
	low	low	Worst case scenario analysis
Mixed	high	high	Model structure expresses hydrogeological detail
	moderate	moderate to high	Fit first and ask questions later

Table 8.1. Decision support modelling roadmap. The lane for geostatistically-based modelling is highlighted.

8.3.7 Assessment: Goodness of Fit

As history-matching does not play a significant role in geostatistically-based decision-support modelling (except for primitive rejection sampling and analysis of prior-data conflict), this type of modelling should not be assessed by how well model outputs replicate past system behaviour.

A stakeholder may object that a model which cannot replicate the past cannot be expected to predict the future. At the same time, he/she may find it difficult to comprehend that failure to expose prior-data conflict constitutes at least partial validation of a geostatistically-based decision support modelling strategy. The response to this criticism is that it is the predictive probability distribution whose veracity requires assessment, and not the value of any particular realisation that is drawn from this distribution. “Validation” in this case comprises an assurance that, even though no realisation of subsurface properties allows the model to match field measurements at all places and at all times, there is a high likelihood that if the model were to be run an infinite number of times based on an infinite number of geostatistical realisations, then some of these realisations would, in fact, allow the model to replicate field measurements of system behaviour.

8.3.8 Assessment: Parameter Credibility

Geostatistically-based predictive uncertainty analysis relies almost exclusively on expert knowledge as it is expressed in probability distributions that dictate the dispositions and connectedness of hydraulically significant geobodies. An assessment of its credibility therefore requires acknowledgement by reviewers and stakeholders that patterns of heterogeneity that are exhibited by stochastic realisations of the subsurface are geologically credible, and are collectively diverse enough to cover all predictive possibilities.

8.4 Strategy 2: Worst Case Analysis

8.4.1 Description

In many environmental management contexts there is no need for decision-support modelling to define the entirety of a predictive probability distribution. Often it is only the pessimistic end of this distribution that requires characterisation. This becomes an easy task where a measurement dataset carries little prediction-relevant information, for history-matching can be dispensed with. A modeller's task then becomes that of finding a geologically credible impact pathway that has a small, but non-zero, chance of existing. In doing so, he/she may choose to generate geostatistical realisations of subsurface geology in order to establish worst case lithological connectedness between a source of groundwater stress and a point where impact must be calculated; see, for example, Renard and Allard (2013). A high impact, but low probability, pathway that is discovered in this way can then be encapsulated in a relatively simple model that calculates the effect of the stress on the receptor. Alternatively (and more commonly), the existence and properties of low probability impact pathways may be based on hydrogeological intuition. See Figure 8.2.

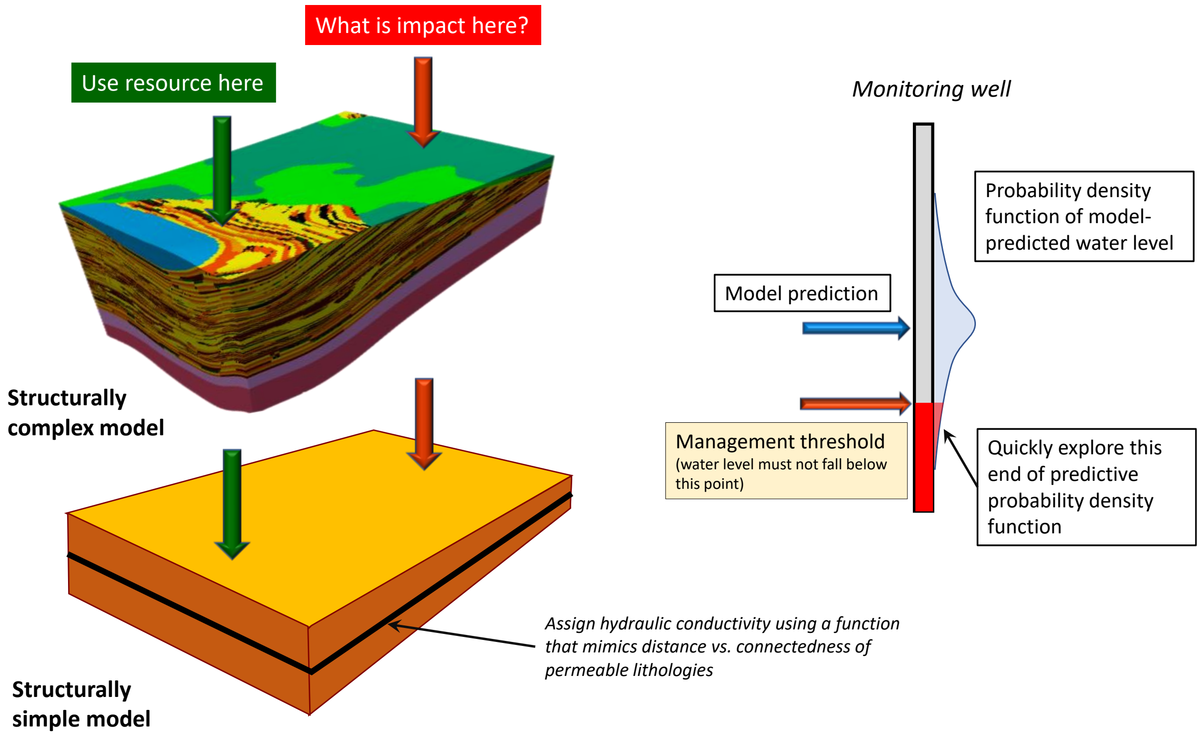


Figure 8.2 Conceptualisation of worst case scenario analysis using a structurally simple model.

In the authors' opinion, this approach to making expert-knowledge-driven predictions has much to recommend it. It is cheaper to implement than geostatistically-based decision-support modelling as it does not require use of a geostatistical software package nor a structurally complex groundwater model. While there is a risk that this type of modelling may fail to acknowledge some important hydrogeological control on groundwater movement so that "worst case" is not "worst case enough", this is something that is easily examined by reviewers and stakeholders, and can be challenged if necessary. Alterations to a model's design in order to accommodate such a challenge are easily made.

A high degree of subjectivity may accompany the design of a worst case model. This is both a strength and a weakness of this approach. A full geostatistical approach to decision-support modelling places the modeller at the mercy of complex geostatistical algorithms. As discussed above, these have their limitations; some of them may artificially preclude pessimistic predictive possibilities. In contrast, it is generally a simple matter for a modeller to make his/her worst case model perform as pessimistically as he/she desires. However, while this is a strength, it is also a weakness because the conceptual grounds for limiting pessimistic predictive behaviour may not be clear. Avoidance of decision-support modelling failure as defined in Section 2.5 of this manuscript may therefore be difficult to justify.

The model that is used in worst-case decision-support modelling is likely to be structurally simple. It may possess only a single layer, populated by one or a small number of parameters attributed to zones of assumed hydraulic property uniformity. Values ascribed to these parameters are likely to reflect upscaled hydraulic property extremes, with maximum connectivity of these properties along possible impact paths taken into account. The model may run under steady state conditions to allow infinite time for impact to propagate.

Despite its simple construction, its likely numerical stability, and its rapid run time, it would normally be unwise to subject a simple model that is built for worst-case scenario analysis to history-matching. Its abstract nature may preclude the establishment of direct relationships between measurements sites and model components. Its structural and parameterisation simplicity are likely to provide fertile ground for history-matching-induced parameter and predictive bias. Nevertheless, there may be occasions where a subjective comparison of pertinent model outputs with historical measurements of system behaviour may allow a modeller to test the conceptual basis of his/her model.

8.4.2 Roadmap Lane

See Table 8.2. This lane of the decision support modelling roadmap is characterised by both structural and parameterisation simplicity.

Prediction driven by	Complexity		Strategy
	Structural	Parameter	
Data	moderate	moderate	Pursue good fit with field measurements
Expert knowledge	high	high	Express expert knowledge using geostatistics
	low	low	Worst case scenario analysis
Mixed	high	high	Model structure expresses hydrogeological detail
	moderate	moderate to high	Fit first and ask questions later

Table 8.2. Decision-support roadmap. The lane for worst case scenario analysis is highlighted.

8.4.3 Assessment: Goodness of Fit

As no history-matching is required, model replication of field measurements does not comprise an assessment criterion for this type of decision-support modelling. As stated above, estimation of parameters through history-matching should probably be avoided, as it is likely to be meaningless at best, and bias-inducing at worst.

8.4.4 Assessment: Parameter Credibility

Parameter credibility is likely to be the focus of most arguments pertaining to the integrity of model-based worst-case scenario analysis. The abstract nature and subjective design of a model that is built for this purpose will probably provide fertile ground for argument. However, as stated above, this may actually be a strength rather than a weakness of this approach to decision-support modelling, for model assumptions are in plain view and can be readily queried. If a stakeholder argues that a particular impact pathway has not been taken into account, it may not be a difficult task to revise the model such that it provides abstract representation of this pathway, and then re-run it.

Contrast this with arguments that swirl around complex numerical models, where large investments of money, time and modeller emotion turn legitimate scientific inquiry into matters of bitter dispute. Under these circumstances, the hypothesis-testing process that should characterise decision-support modelling often ends before it begins. Hostilities cease when battle-weariness pervades the discussion. At this stage, the notion that modelling can support the ability of stakeholders to embark on collective scientific inquiry has long faded from view.

9. MIXED PREDICTIONS

9.1 General

Most predictions that are required of decision-support groundwater modelling lie between the two end members of the prediction spectrum which are the subjects of the previous two sections of this document.

Information that reduces the uncertainty of a mixed prediction resides partly in measurements of system behaviour, and partly in data acquired through site characterisation. As is discussed in Section 3 of this document, these sources of information are represented by the first and second terms on the right of Bayes equation. The making of a mixed prediction therefore requires that Bayes equation be evoked, in spirit or in fact.

The sources and types of information that are mixed through Bayes equation are very different. Combining them can be a challenge – a challenge that must be met through integration of simulation and history-matching. Use of tools such as PEST/PEST++ is essential.

9.2 Examples

Decision-support modelling requires the making of mixed predictions whenever plans are made to subject a groundwater system to stresses which it has not previously experienced. However, the new stress regime is not so different from the old one that information contained in the past behaviour of the system is irrelevant to its future behaviour. Planned system interventions may have potentially unwanted consequences. Alternatively, their purpose may be to repair damage caused by previous interventions.

Environmental impact assessment provides many examples of mixed predictions. The source of impact may be a mine; or it may be new or increased groundwater extraction from a confined or unconfined aquifer.

Examples of modelling that is undertaken to support groundwater repair are also plentiful. They include appraisal of improved agricultural practices in areas where groundwater and rivers are degraded by agricultural chemicals. They also include the design of groundwater extraction, treatment and reinjection systems whose purpose is to effect contaminant cleanup. In both of these contexts, measurements of groundwater heads and contaminant concentrations are likely to be plentiful.

Other contexts in which mixed predictions are likely to provide a basis for groundwater management include the following:

- prevention or reduction of seawater intrusion;
- long-term sustainability of a groundwater supply;
- definition of well head protection zones; and
- prediction of extraction rates for mine dewatering.

GMDSI worked examples illustrate modelling that has been undertaken in all of these contexts. See https://gmdsi.org/worked-examples/.

Because the range of groundwater management contexts in which mixed predictions are required is so broad, no advice is good for all occasions. However, discussion of this subject can be advanced by subdividing decision-support modelling strategies that are appropriate for this prediction type into two broad categories.

The first category is that in which a modeller feels that expert knowledge and the outcomes of site characterisation should be strongly reflected in the structure of the numerical model. Its structure should therefore resemble that of the site geological model. (Presumably, the latter can be trusted.) We use the epithet "hydrogeologically-based" to describe this modelling strategy.

We invoke the phrase "fit first and ask questions later" to describe the second modelling strategy. Notwithstanding the colourful way in which we describe it, this approach to the making of mixed predictions does not necessarily eschew information that is forthcoming from site characterisation studies. However, it embraces a higher level of model abstraction than the first category in the hope that parameter patterns that emerge from untrammelled history-matching can expose "conceptual surprises" as well as hydrogeologically important, but as yet only partially known, aspects of the subsurface.

We illustrate the first of the above strategies using a figurative example, and the second using a real-world example.

9.3 Hydrogeologically-Based Modelling

9.3.1 Model Structure

Model structure is discussed in Section 4 of this manuscript. There it is pointed out that the role of model structure is to host model parameters. If a model's structure is defective, this may not actually compromise the ability of model outputs to replicate measurements of system behaviour. However, in attaining a good fit with a measurement dataset, at least some parameters are likely to be assigned values that compensate for model structural defects. Some model predictions may thereby incur bias.

It was also pointed out that predictions that are sensitive only to solution space components of the "reality model", do not incur bias, regardless of their sensitivities to parameters of the "modeller's model" that do, in fact, incur bias. These data-driven predictions are the subject of Section 7 of this manuscript. Predictions that are the subject of the present section may, or may not, be susceptible to this problem. In general, it is difficult to know. Hence there is a strong argument for presuming the worst, and to therefore adopt a model structure that respects the hydrogeological realities of a particular site, insofar as these are known.

Unfortunately, insistence that the structure of a numerical model respect the details of hydrogeological reality can introduce problems. The dispositions and properties of hydrogeologically significant units are only ever approximately known. Nevertheless, site characterisation studies may reveal that their details matter when it comes to predicting impact. However, hardwiring one realisation of these details into a model's structure may be less beneficial than representing these details in a more abstract (and adjustable) manner. While the former strategy may render a numerical model pleasing to a hydrogeologist's eye, it will almost certainly require that parameters compensate for its incorrectness as they are adjusted in order to promulgate a good fit between model outputs and field measurements of system behaviour.

The problem of whether or not to explicitly represent hydrogeological detail in the structure of a numerical model is exacerbated where features such as faults are of hydrogeological significance. As is discussed in Section 5 of this document, potential misrepresentation of their properties and hydrostratigraphic offsets as they vary along its strike questions the wisdom of representing them in ways that are not subject to easy adjustment.

Previous sections of this document recommend that the burden of representing subsurface properties that are not fully known should fall on parameters rather than on structure. However, it is freely acknowledged that this strategy too has its limitations. For example, where drawdown or contaminants undergo vertical subsurface migration, simulation requires representation of multiple layers. Where a model supports seasonal groundwater management, then it must be transient rather than steady state; furthermore, one or a number of ancillary recharge models may be required in addition to the groundwater flow model itself. However, it is a paradox of numerical simulation of natural systems, that the more processes that a model attempts to simulate, and the more spatial and temporal dimensions that it attempts to represent, the more likely it is that these representations are compromised by the necessity to represent detail that is only vaguely known. The argument for ignoring these dangers is that omission of these details would compromise the modelling process even more than representing them in possibly erroneous ways. This is, of course, a valid argument – an argument which has some important consequences.

9.3.2 Model Parameters

Adoption of high levels of structural/process complexity in a model requires concomitant adoption of high levels of parameterisation complexity. Recall that the role of structure is to host parameters, and that parameters host information. A complex model structure can host many parameters – probably many more than can be estimated uniquely. However this does not justify their simplification or omission from a model. In fact it may require the opposite. Implied in a decision-support modelling strategy that relies on a complex model structure is recognition that at least some subsurface details are important. For reasons that have already been outlined, prediction-critical detail should be represented stochastically in a model so that the repercussions of its partially known status are properly reflected in predictive uncertainty.

Suppose, for example, that a new layer is introduced to a groundwater model. This requires the complementary introduction of new sets of spatial parameters. Suppose that the model employs pilot points as its spatial parameterisation device. Then introduction of the new model layer requires introduction of new families of pilot points to host values of horizontal hydraulic conductivity, vertical hydraulic conductivity and specific storage of the new layer. These new parameters must be accompanied by stochastic descriptors of their uncertainties, spatial correlation, and possibly of correlations that exist between hydraulic properties of different types. These stochastic descriptors are used in regularised inversion (i.e. model calibration), and in predictive uncertainty analysis.

The same applies where recharge is explicitly simulated in a transient groundwater model. Presumably, a separate ancillary recharge model is required for each land use and soil type that prevails within a model's domain. Each of these recharge submodels must possess at least four or five adjustable parameters in order to allow it to replicate the temporal nuances of processes that operate in the plant root zone and in the underlying vadose zone. These parameters cannot be estimated uniquely; hence their introduction to the overall model must be accompanied by stochastic characterisation of their variability and interrelationships.

Paradoxically, the need to complement structural complexity with parameterisation complexity is often disregarded in everyday modelling practice. As is discussed in Section 5 of this document, it is not uncommon for a structurally complex model to be endowed with a relatively parsimonious parameterisation scheme. This is naively justified on the basis that long run times and numerical instability that inevitably accompany model structural complexity present numerical obstacles to parameter estimation and posterior uncertainty analysis. Reduction of a model's parameter load is seen as one way to accommodate these problems. Unfortunately, this strategy is counterproductive for a number of reasons. It may prevent attainment of a good fit between model outputs and field measurements. At the same time, it will almost certainly instigate history-matching-induced parameter and predictive bias as estimated hydraulic property values are introduced to inappropriately large parts of the model domain. Analysis of predictive uncertainty will also be compromised through failure to represent local scale parameters to which predictions are sensitive but that are characterised by post-history-matching nonuniqueness.

9.3.3 Resolving the Paradox

The above discussion suggests that adoption of a high level of model structural complexity is a decision that should not be taken lightly. Adoption of this strategy should not be based on the premise that greater structural complexity allows a numerical model to look more like “the real thing”. It should be based on the proposition that removal of any its complexity would compromise a model’s ability to support environmental management. That is, it would degrade its ability to quantify and reduce the uncertainties of decision-critical predictions. The focus of this justification should be on the potential for parametric and predictive bias that would accompany complexity reduction. In presenting this argument, however, a modeller should demonstrate that the benefits of retaining structural complexity outweigh the numerical obstacles to uncertainty reduction and quantification that model complexity poses, and the potential for predictive bias that hardwired structural detail may induce.

In the authors’ opinion, the onus of proof should be on those who argue for structural complexity in a model rather than on those who argue for structural simplicity. We explain our stance using a simple conceptual example.

9.3.4 A Conceptual Example

Consider the conceptual hydrogeological cross section depicted in Figure 9.1a. A production well (shown in red) extracts water from a thick, steeply-dipping aquifer. A model is built to investigate whether this extraction may reduce baseflow in a river; the river is depicted in blue. A small amount of pumping-induced drawdown has been detected in the green observation well. Drawdown impact is transmitted to this well across a fault that offsets an aquitard. The fault is hidden beneath the weathering layer.

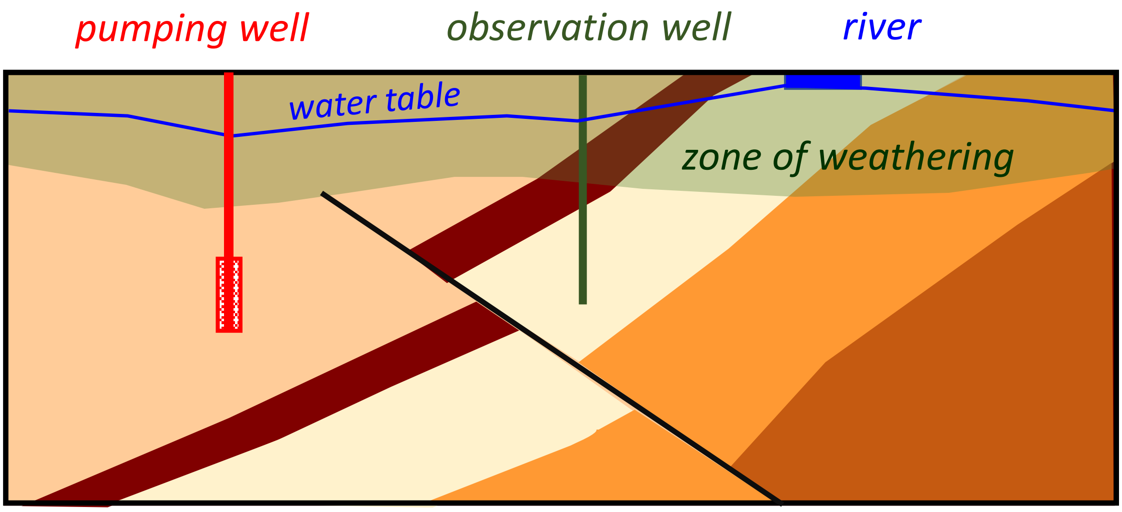


Figure 9.1a. A hydrogeological section.

Suppose that hydrogeologists are unaware of the existence of the fault. Current hydrogeological concepts are depicted in Figure 9.1b. The aquitard is continuous and uninterrupted according to these concepts.

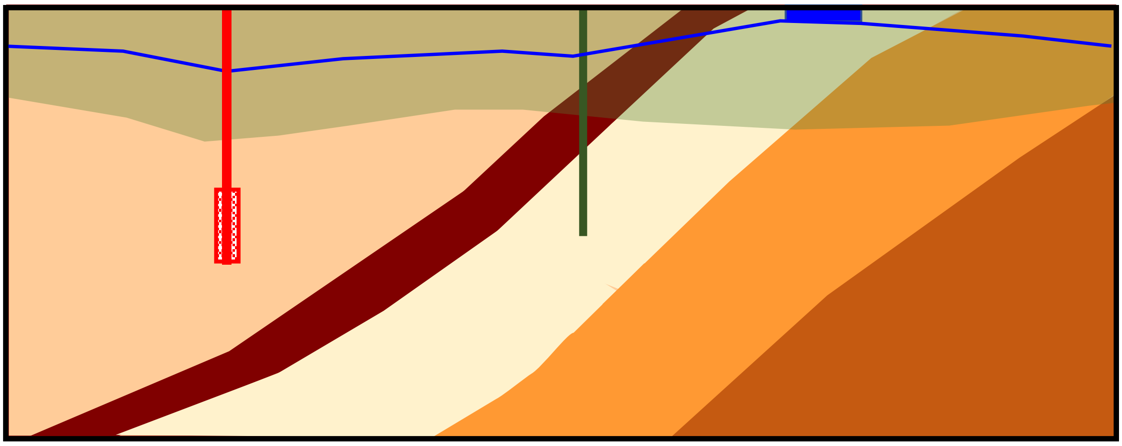


Figure 9.1b. Presumed hydrogeological section.

Model layering is depicted in Figure 9.1c. The modeller attempts to respect prevailing hydrogeological concepts by explicit inclusion in his/her model's structure of aquifers and the aquitard. This has some annoying numerical consequences. Model layers and accompanying saturated thicknesses become thin before they disappear up dip. This may increase model runs times while creating fertile ground for numerical instability. This, in turn, may inhibit the model's ability to be used with software such as PEST and PEST++. Replication by the model of small extraction-induced drawdowns in the observation well may therefore be threatened.

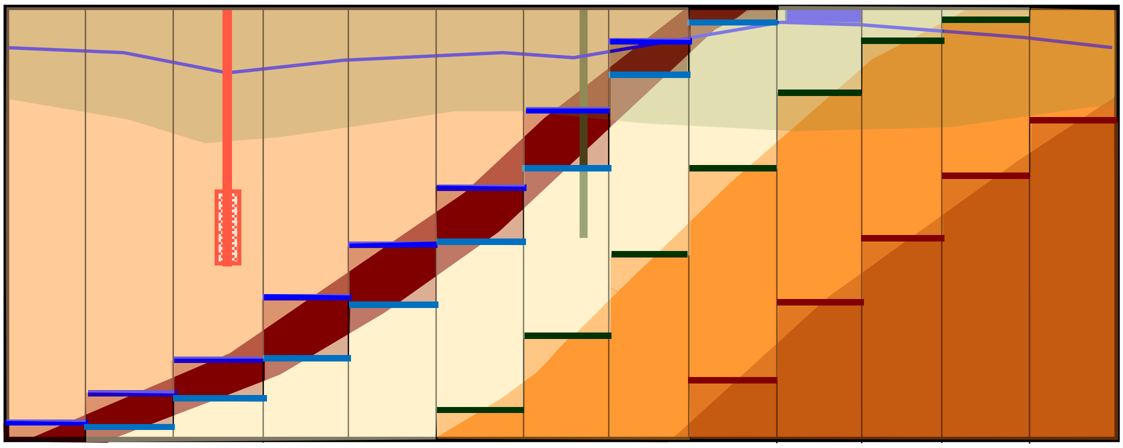


Figure 9.1c. Layering of a numerical model.

The modeller may exacerbate this problem by placing an upper bound on the vertical hydraulic conductivity that is estimated for the aquitard. This may be done under the premise that the model is physically-based, and that aquitards are aquitards because their vertical hydraulic conductivities are low. To make matters worse, the modeller may even ascribe a small number of parameters to the aquitard in the hope that reduced parameterisation complexity can relieve parameter estimation headaches incurred by numerical instability of the model.

All of these steps erode the ability of the history-matching process to challenge prevailing hydrogeological concepts. When modelling is over and the model report is written, failure to fit the small drawdowns in the green observation well may be glossed over. In fact, it may even be touted as worthy of a modelling badge of honour as the modeller has steadfastly resisted the temptation to "overfit" field data while calibrating a model whose decision-support credentials rest on "faithful representation" of prevailing hydrogeological concepts.

Obviously, this approach to modelling does not serve the imperatives of decision-support well at all (at this conceptual site at least). In attempting to respect prevailing hydrogeological concepts, it fails to represent the inherently probabilistic nature of those concepts. It excludes, rather than welcomes, information that is resident in measurements of system behaviour.

If history-matching exposes a conflict between measured system behaviour and prevailing hydrogeological concepts, it is obvious that the former must hold sway. This suggests that it may be better for the decision-support modelling process to begin with a strategy that is as open as possible to information that is resident in field measurements of system behaviour; this strategy can then be modified if assimilation of this information suggests the need for greater structural complexity. This “fit first and ask questions later” approach is discussed next.

9.3.5 Roadmap Lane

The roadmap lane which this hydrogeologically-based decision-support modelling strategy occupies is highlighted in Table 9.1. It requires both structural and parametric complexity. For reasons that are explained above, we do not see parametric simplicity as an option, for parametric complexity is a necessary adjunct to structural complexity. Numerical and conceptual problems that beset structural complexity are therefore exacerbated by the need to define many parameters, characterise them stochastically, and adjust them during calibration and uncertainty analysis.

Prediction driven by	Complexity		Strategy
	Structural	Parameter	
Data	moderate	moderate	Pursue good fit with field measurements
Expert knowledge	high	high	Express expert knowledge using geostatistics
	low	low	Worst case scenario analysis
Mixed	high	high	Model structure expresses hydrogeological detail
	moderate	moderate to high	Fit first and ask questions later

Table 9.1. Decision-support modelling roadmap. The lane for hydrogeologically-based modelling is highlighted.

It must be acknowledged, however, that every management context requires a minimum level of model structural complexity if the support that modelling provides to the decision-making process is to have integrity. High levels of model structural complexity are sometimes unavoidable.

9.3.6 Assessment: Goodness of Fit

It is a fact of modelling life that structurally complex models rarely fit history-matching datasets as well as simpler, smaller models that are dedicated to the history-matching task. Information that is resident in these datasets is thereby denied entry into the model. While this is not always a bad thing, there will be some occasions where failure of a complex model to replicate measured system behaviour should precipitate its replacement by one or a number of structurally simpler models.

A model may be multi-layered and structurally complex because the groundwater system that it simulates is large, and because the management issues that it addresses are regional. The history-matching dataset may include measurements that host information pertaining to the regional groundwater system; this information should be assimilated through history-matching. These measurements may also host information that pertains to local processes that the regional model does not represent. Naturally, the model should not attempt to replicate aspects of system behaviour that pertain to processes which it is incapable of simulating. History-matching can ensure this by adopting weighting and data processing strategies that have already been outlined.

There may be other cases, however, where failure to replicate measured system behaviour erodes the capacity of the modelling process to support decisions. A conceptual example is provided in the previous subsection. Because model structure is nonadjustable, and cannot therefore be expressed stochastically, it can just as easily host disinformation as it can host information. When history-matching a structurally complex model, a modeller should always be sensitive to indications that its structure may not be reconcilable with system behaviour. If evidence of this is found, then action must be taken.

The nature of this action depends on the modelling circumstances. The simplest action that a modeller can take is to raise or lower parameter bounds when using PEST/PEST++ for history-matching. Parameters are then free to play surrogate roles if this accomplishes a better fit with field measurements. Where this occurs, the nature of these roles can indicate the type of structural refinement that the model should undergo. In other cases, a modeller may decide that the aberrant values adopted by adjusted parameters do not diminish the integrity of decision-critical model predictions that the model must make. If some parameters can soak up information that would otherwise have nowhere to go because of conflicts between local nuances of model structure and real-world system architecture, this may free other parameters to play their information-hosting roles with integrity. To the extent that a prediction of management interest is sensitive to the latter parameters, its integrity is unimpaired.

In contrast, failure to achieve a good fit with aspects of system behaviour that are thought to be rich in information with respect to a prediction of management interest constitutes an indictment of the modelling process. A modeller then has no choice but to redesign this process. In doing so, he/she may adopt a model that is structurally simpler but parametrically more complex. Unfortunately, for reasons outlined above, structural simplicity may induce bias in a decision-critical model prediction. However, failure of the complex model to replicate prediction-relevant system behaviour evinces its own proclivity to induce predictive bias. Meanwhile opportunities for data assimilation afforded by decreased structural complexity and increased parametric complexity may accrue a reduction in predictive uncertainty, at the same time as stochasticity ascribed to model parameters may support superior quantification of predictive uncertainty. Obviously, the potential for predictive bias incurred by the simpler model structure must be included in the quantified predictive uncertainty interval, to the extent that this is possible.

For mathematical convenience, model-to-measurement misfit is often attributed to measurement error. As a further convenience, the latter is often decreed to be random and uncorrelated. This assists in its processing. In particular, it supports evaluation of the contribution that model-to-measurement misfit makes to predictive uncertainty. Unfortunately, model-to-measurement misfit incurred by model structural inadequacies is not statistically random; see Doherty and Welter (2010) for more details. Its contribution to predictive uncertainty and/or predictive bias is therefore difficult to quantify. It follows that there are benefits to be had in pursuing a modelling strategy such as that described in the preceding paragraph that seeks to avoid structurally-induced model-to-measurement misfit – as long as possible structurally-induced parametric biases that are induced by such a strategy can be included in quantified predictive uncertainty intervals.

9.3.7 Assessment: Parameter Credibility

Ideally, credibility of estimated parameter values should complement (and confirm) a hydrogeologically-based approach to decision-support modelling. Where values ascribed to model parameters through history-matching are not within expected credibility limits, then the hydrogeological basis of the model is eroded. Their occurrence can suggest strategies for model redesign (presumably to enhance its hydrogeological reasonableness). Alternatively, for reasons discussed above, a modeller may decide that parameter surrogacy in some parts of a model domain does not compromise the integrity of decision-critical model predictions that pertain to other parts of the model domain.

9.4 Fit First and Ask Questions Later

9.4.1 Discussion

The decision-support modelling strategy that we characterise as "fit first and ask questions later" requires construction of a model which is designed with history-matching in mind. Its run time must be low and its numerically stability must be high so that it can be used effortlessly with programs of the PEST and PEST++ suites, or other history-matching software. It should be endowed with many parameters. At the same time, its structural complexity should be sufficient for the parameters which it hosts to be capable of fitting field measurements well - either directly, or after processing of these measurements to extract signal components that can be replicated by the model in spite of its structural deficiencies. (Fitting to drawdown rather than to head is an example of appropriate processing.) However, the model need not be so structurally complex that some of its parameters do not adopt roles that compensate for its structural imperfections. These compensatory roles may be evinced in estimated parameter values and patterns that are not in accordance with prevailing geological expectations.

The surrogate roles that some parameters may adopt in achieving a high level of model-to-measurement fit can be accommodated in various ways. For example, a modeller may use parameter patterns and values that emerge from history-matching to construct an approximate "prior" parameter probability distribution that takes account of these roles. This prior parameter probability distribution can then be used in posterior predictive uncertainty analysis.

Alternatively, estimated parameter values and patterns may be seen as predictive outcomes in themselves, as a modeller may not be otherwise aware of the subsurface disposition of hydraulically significant geological units. He/she may therefore use the modelling process to seek hydrogeological illumination through history-matching against an information-rich measurement dataset; see the following example.

A significant advantage of the "fit first and ask questions later" approach to decision-support modelling is that model construction, deployment, history-matching and uncertainty analysis are all relatively inexpensive. Another advantage is that a good fit can generally be attained with a field-measured dataset, as the whole modelling process is designed to achieve this outcome. It is thereby designed to extract all information that resides in these data.

A disadvantage of the "fit first and ask questions later" approach is that this information is directed to parameter receptacles whose roles may be distorted by a simplistic model structure. A modeller must be prepared to accommodate the somewhat abstract nature of his/her model and the expressions of information that are embodied in values and patterns with which its parameters are endowed. So too must modelling stakeholders. Nevertheless, there will be many occasions where this type of modelling can serve decision-support imperatives well, either when undertaken on its own, or as a precursor to development of a more structurally complex model whose design has been inspired by history-matching of the structurally simpler model.

9.4.2 An Example

This example is fully documented in a GMDSI Worked Example report that can be downloaded from the following site:

https://gmdsi.org/blog/worked-example-simultaneous-interpretation-of-six-pumping-tests/

We describe it only briefly here. Refer to the report for full details.

The model which is the focus of this example simulates drawdown incurred in 21 observation wells by extraction from 6 production wells. Over a six month period in late 2014, each production well was pumped sequentially for between 5 and 11 days. Recovery was monitored before pumping the next well. These tests were conducted in order to gain insights into subsurface hydraulic properties that could assist assessment of dewatering requirements for an open cut mine (named OB31) that was planned for the area. This had repercussions for design of a pipeline that would convey extracted water elsewhere.

The local geology is complex. See Figure 9.2. Ore is extracted from the Precambrian Brockman Iron Formation – as is water for pit dewatering. This, and neighbouring formations, dip steeply. Some of these neighbouring formations are relatively impermeable. In contrast, the Paraburdoo Member is composed of permeable dolomite. At the time at which these pumping tests were carried out, it was not known whether structural features (which were known to exist in the area) create hydraulic connections between the Brockman Iron Formation and the Paraburdoo Dolomite. Such connections have serious consequences for pit dewatering requirements.

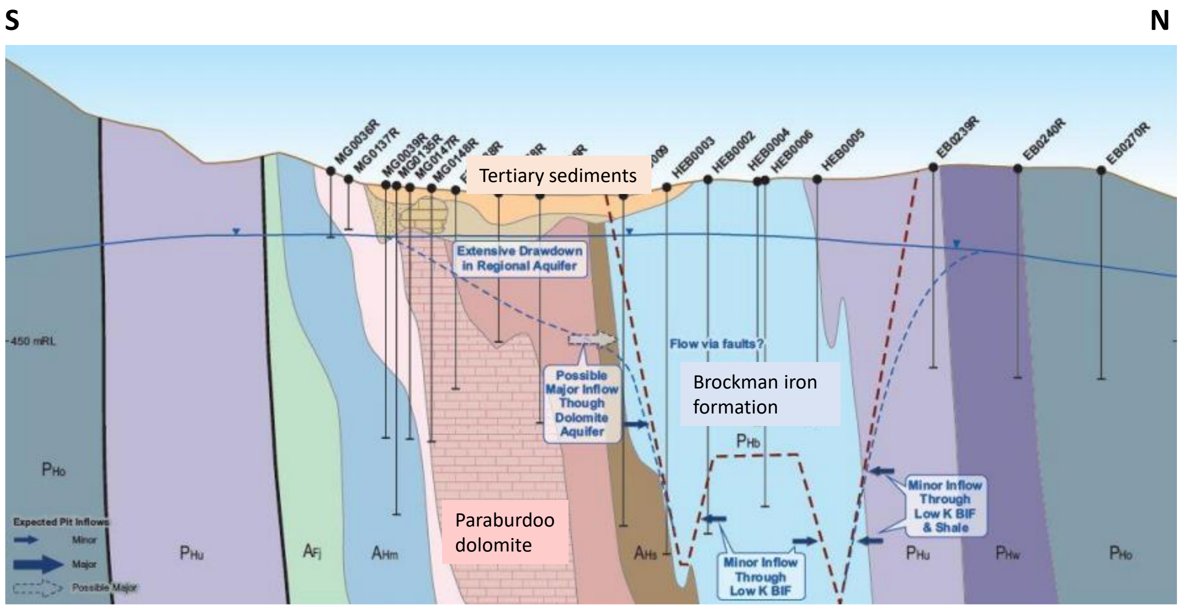


Figure 9.2. Geological cross section through proposed pit.

Figure 9.3 depicts a model that had previously been used to analyse drawdown data. A problem with use of this model for drawdown analysis is that it was charged with other decision-support responsibilities as well, these including impact assessment. The model had 7 layers. As is apparent from Figure 9.3, its structure attempted to encapsulate what was known about local hydrogeology at the time, in particular its steeply dipping layers. In spite of attempts at history-matching, the model was not able to replicate pumping-induced drawdowns well. Reasons for this included its long run time, its parsimonious parameterisation, and its delicate numerical health. Its predictions of pit dewatering requirements have been shown to be too low.

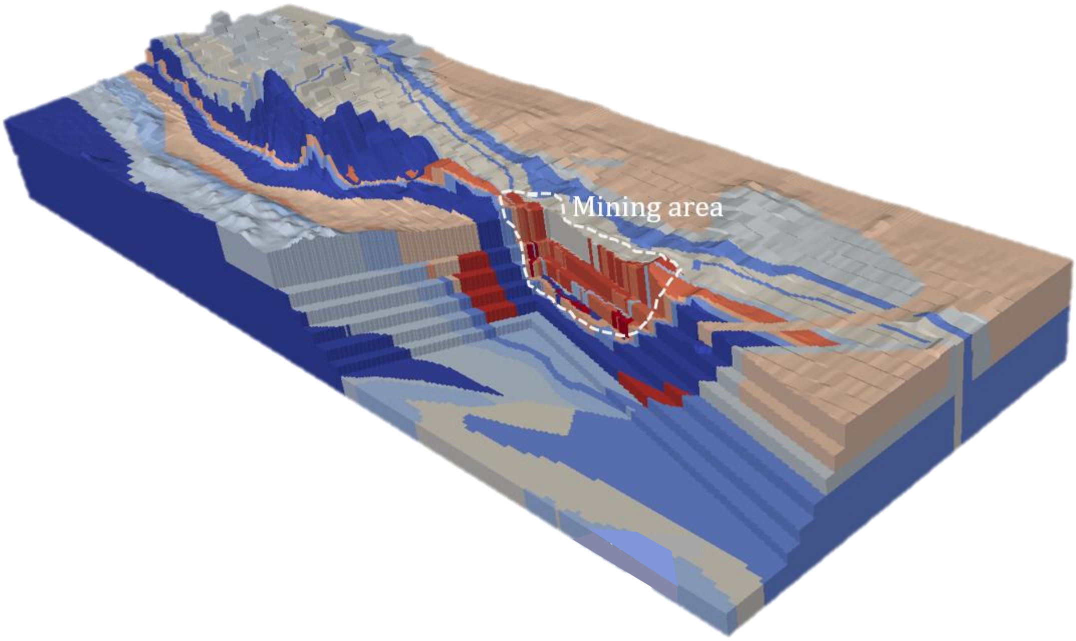


Figure 9.3. Previous model. Dimensions are 27 km × 10 km.

In 2020 GMDSI personnel built a new model to reinterpret the old pumping data. This was done in order to establish whether a dedicated model, designed according to the “fit first and ask questions later” principle, would have enabled better inferences to be drawn of local geology, and superior assessments to be made of pit dewatering requirements. Figure 9.4a shows the model grid, together with the locations of pumping and observation wells. Based on the premise that pumping-induced groundwater flow is predominantly horizontal, the model possesses only a single layer. However, it is endowed with 3526 pilot point parameters. Half of these pertain to hydraulic conductivity while the other half pertain to specific yield. Both of these parameter types share the same set of pilot points. The disposition of these points is depicted in Figure 9.4b.

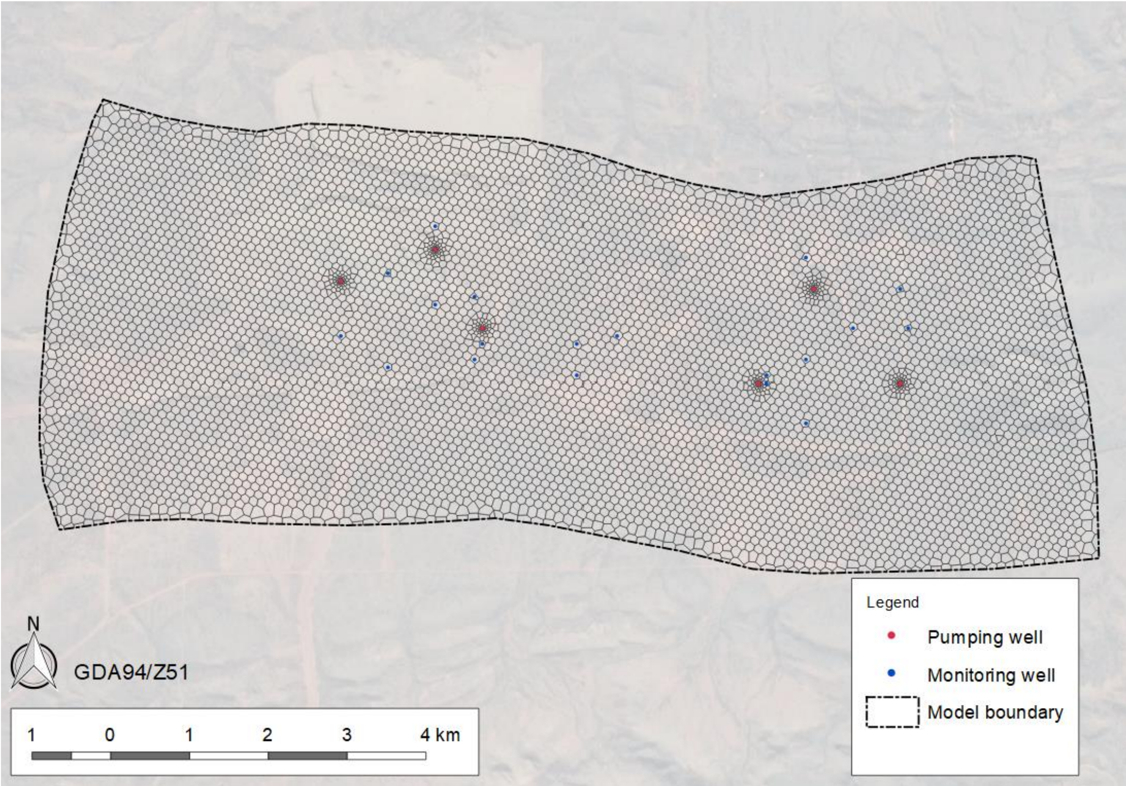


Figure 9.4a. Domain and grid of new model.

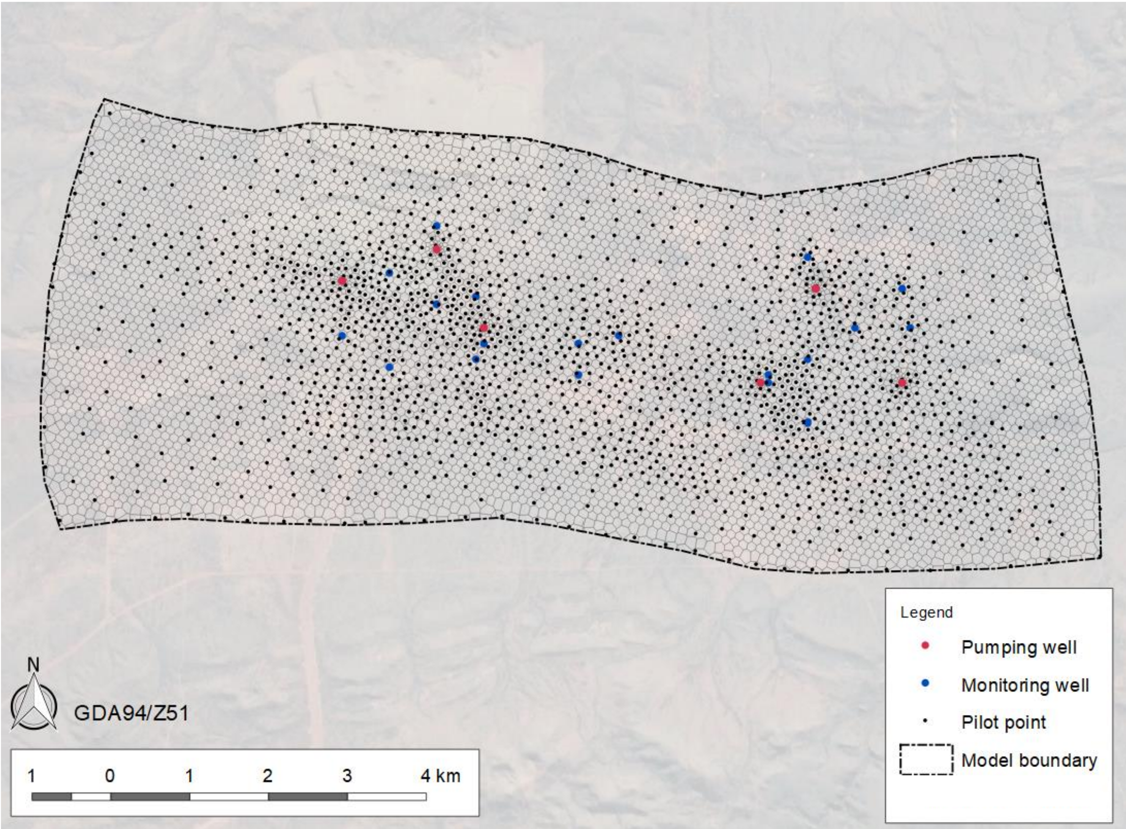


Figure 9.4b. Model domain and pilot points.

History-matching (conducted using PEST_HP) attained a good fit between modelled and field-measured drawdowns; Figure 9.5 displays some of these fits.

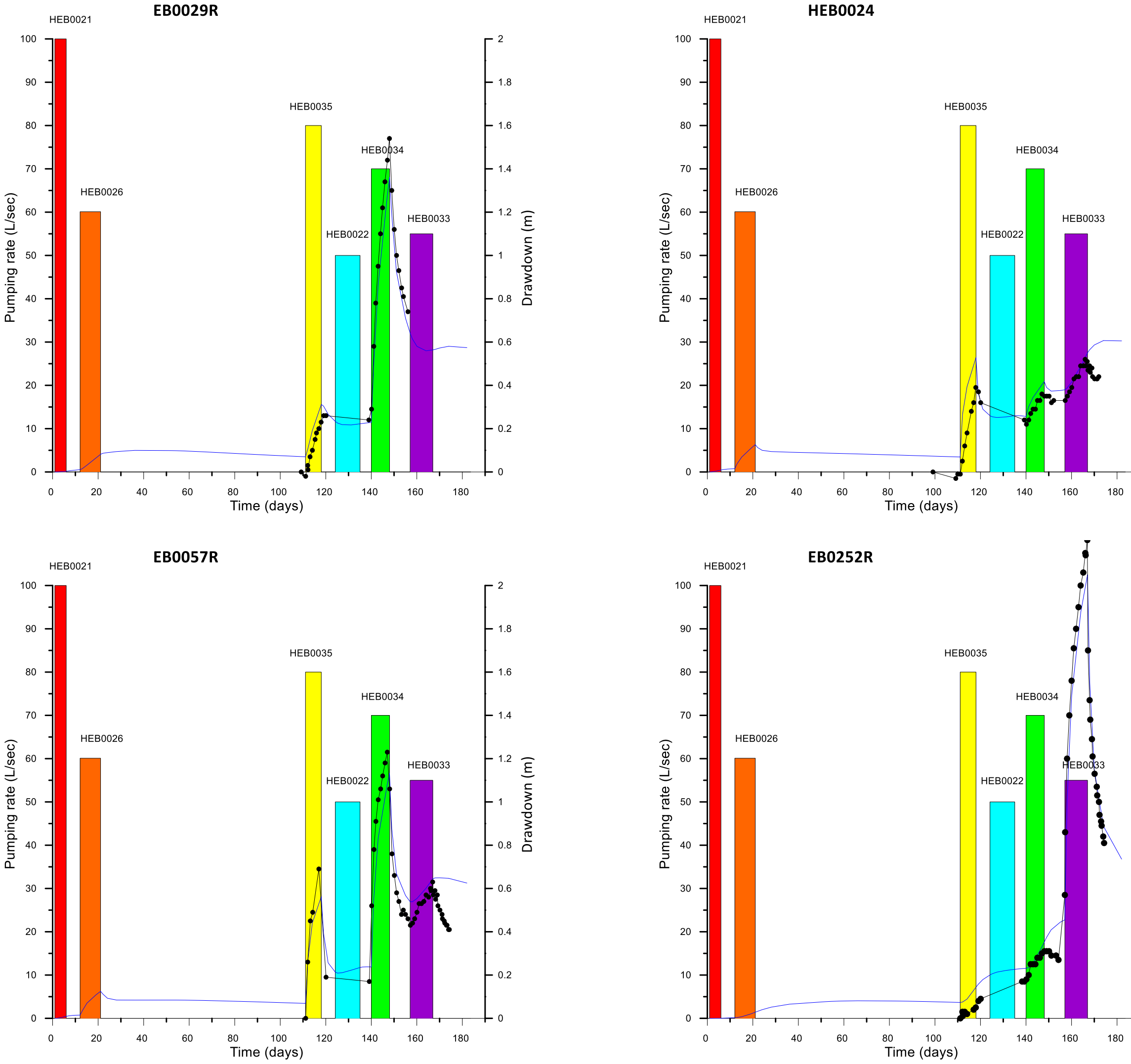


Figure 9.5. Measured (dots) and model-calculated (blue lines) drawdowns in four observation wells. Coloured bars depict extraction from pumping wells.

Figure 9.6 shows, in order:

- estimated hydraulic conductivity;
- estimated hydraulic conductivity superimposed on a satellite image (the OB31 pit is in the north of the model domain); overlain on this image are possible fault locations;
- estimated hydraulic conductivity superimposed on a map of hydrostratigraphic units as represented in the original model;
- relative reduction in hydraulic conductivity uncertainty incurred through history-matching; this map was composed using linear analysis tools provided with the PEST suite.

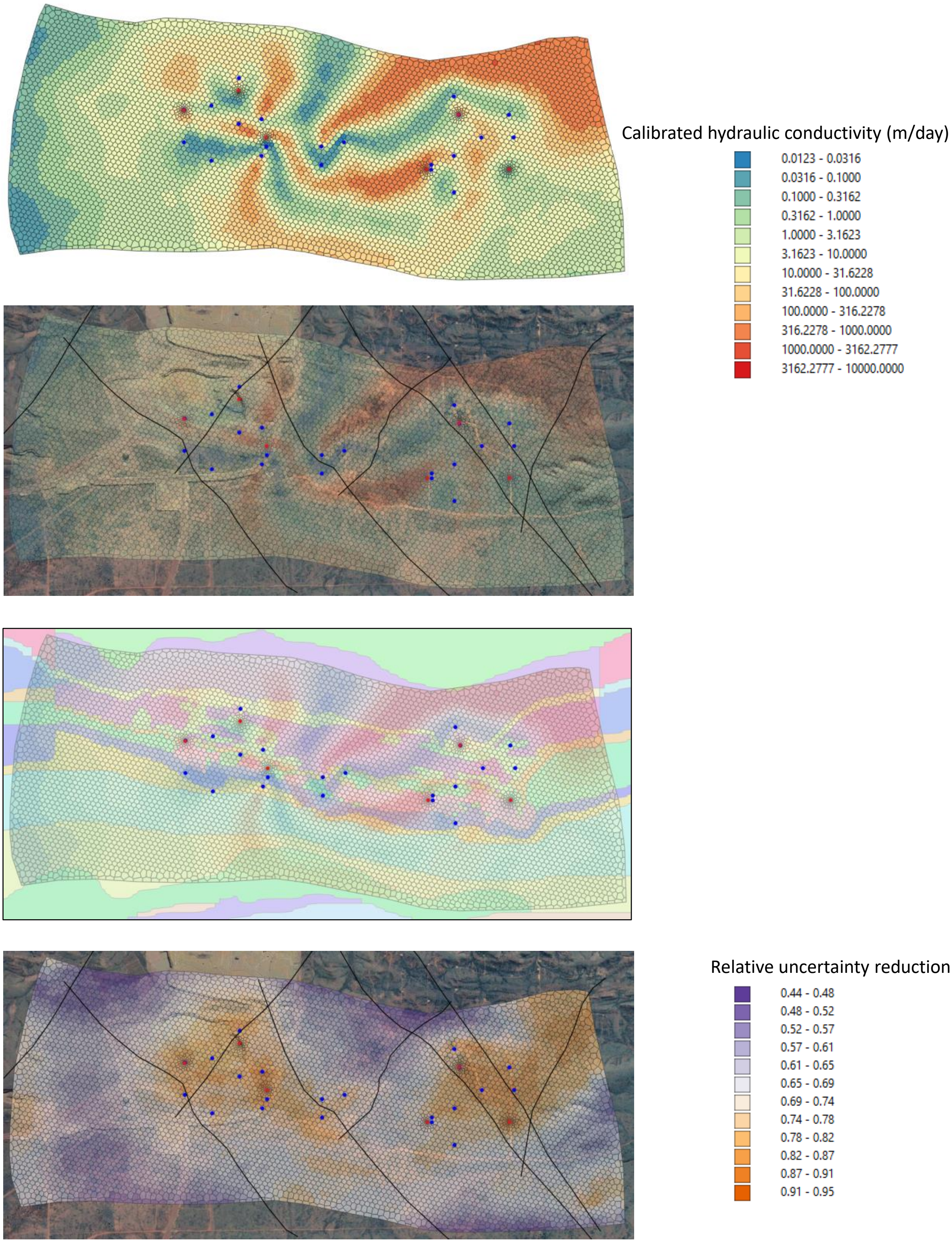


Figure 9.6. Estimated hydraulic conductivity and outcomes of linear uncertainty analysis.

Figure 9.7 does the same for history-matching-inferred specific yield.

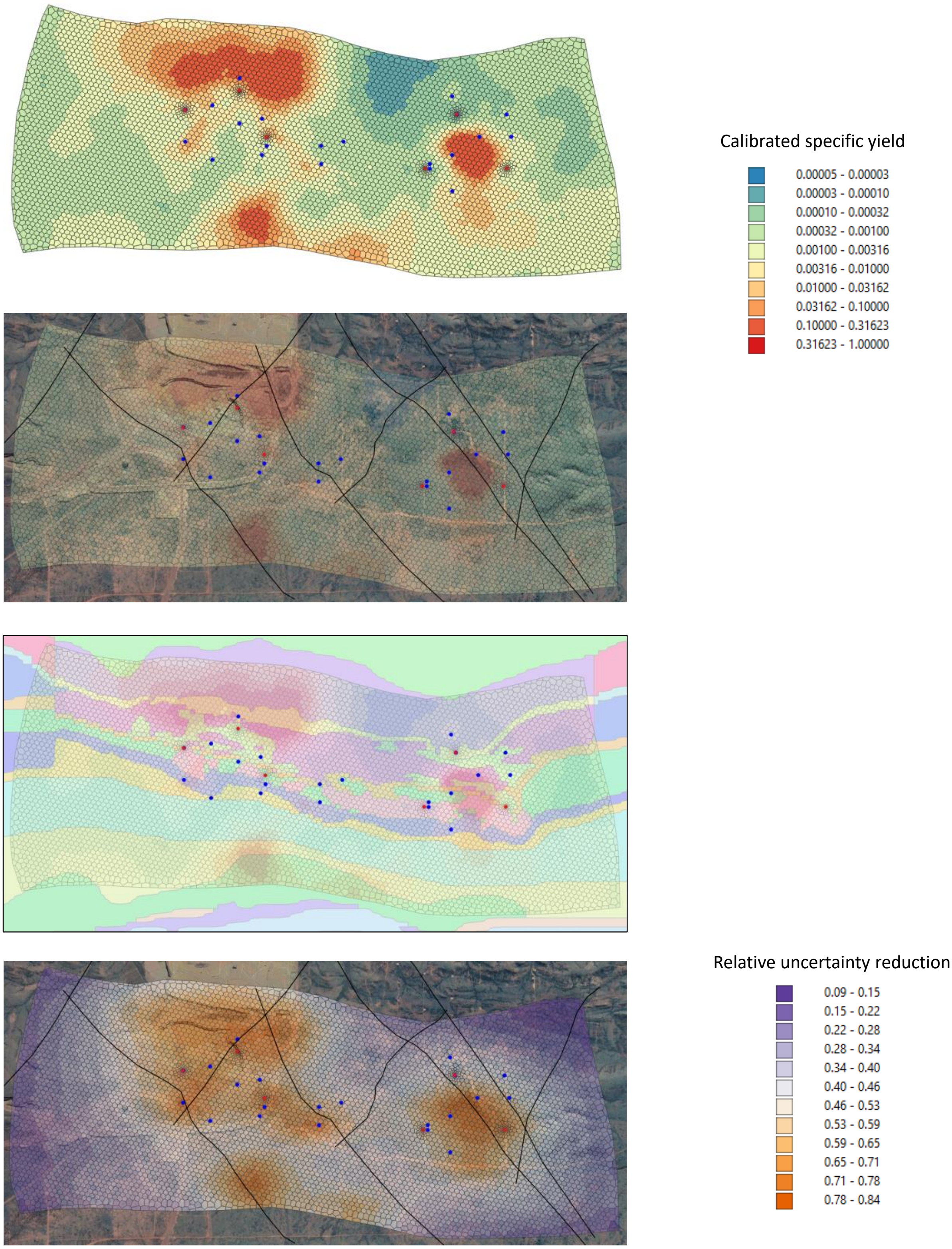


Figure 9.7. Estimated specific yield and outcomes of linear uncertainty analysis.

Given the somewhat abstract nature of the model, the parameter patterns that are depicted in Figures 9.6 and 9.7 should be interpreted with caution. Nevertheless, it is apparent that some regions of high inferred hydraulic conductivity are aligned with stratigraphic units that are known to be permeable; the same applies to some regions of low inferred hydraulic conductivity. More importantly, narrow zones of high hydraulic conductivity are inferred to transect units of low hydraulic conductivity; furthermore, they are aligned with suspected structural features. At the same time, uncertainties associated with hydraulically significant parts of these parameter patterns are relatively low; this provides strong evidence for their existence, and of their hydraulic importance.

Patterns of inferred specific yield cannot be interpreted as directly as those of hydraulic conductivity. They suggest high porosity of ironstone and dolomite formations. However, an area of high specific yield in the east-central part of the model domain is difficult to reconcile with known Precambrian geology; perhaps it reflects Tertiary alluvium to which hydraulic access is gained through permeable vertical structures.

The somewhat abstract outcomes of the history-matching process do not diminish their importance. They provide strong evidence of permeable pathways between the Brockman Iron Formation and the Paraburdoo Dolomite, and perhaps to other formations as well. The existence of these pathways, together with evidence of a considerable amount of stored water in these and other units, suggests that pumping rates required for OB31 dewatering are likely to be high. The model was used to calculate order-of-magnitude dewatering rate requirements (and uncertainties associated therewith). There is good agreement between these predictions and present day dewatering rates.

9.4.3 Roadmap Lane

Table 9.2 depicts the roadmap lane to which the “fit first and ask questions later” approach to decision-support modelling belongs. The level of structural complexity is denoted as “moderate” rather than “low”. For the example discussed in the previous subsection, one model layer is sufficient. Other contexts may require multiple model layers; this is the case where predictions pertain to inter-layer flow, and/or where vertical head differences are rich in decision-critical information.

Prediction driven by	Complexity		Strategy
	Structural	Parameter	
Data	moderate	moderate	Pursue good fit with field measurements
Expert knowledge	high	high	Express expert knowledge using geostatistics
	low	low	Worst case scenario analysis
Mixed	high	high	Model structure expresses hydrogeological detail
	moderate	moderate to high	Fit first and ask questions later

Table 9.2. Decision-support modelling roadmap. The “fit first and ask questions later” lane is highlighted.

The philosophy behind this approach to decision-support modelling requires that parameters be plentiful enough to match a measurement dataset in order to harvest the information that it carries, and to convey that information to wherever it should reside within a model domain. They should also be plentiful enough to express uncertainties in those parts of the model domain which receive no information, and to transfer these uncertainties to predictions of management interest.

9.4.4 Assessment: Goodness of Fit

Attainment of a good fit with a field-measured dataset is the defining feature of this approach to decision-support groundwater modelling. Failure to attain a good fit may require that greater structural complexity (and complementary parameter complexity) be introduced to a model. This may, in fact, be a beneficial outcome of the modelling process. If a modeller had previously thought that the structure of his/her model was adequate to host parameters that could absorb information contained in measurements of system state, demonstration through a dedicated history-matching process that this is not the case teaches a modeller, as well as modelling stakeholders, much about a local groundwater system.

9.4.5 Assessment: Parameter Credibility

It is not unlikely that the “fit first and ask questions later” approach to decision-support groundwater modelling will yield parameter values and patterns that compensate for model structural simplifications. This does not necessarily detract from the usefulness of the process, nor from the instructional value of the parameter patterns that emerge from it. As the above example demonstrates, it may be possible to draw some important conclusions from these patterns, even with their abstract nature taken into account. In other cases, the nature and locations of compensatory parameter behaviour may suggest strategic improvements in model design.

Where the surrogate roles that parameters must adopt during history-matching do not detract from the decision-support utility of a structurally simple model, “prior probability distributions” can be assigned to model parameters that take their compensatory roles into account. These may be based on patterns that emerge from the history-matching process itself. Use of these specialised prior parameter probability distributions in evaluation of posterior predictive probability distributions enables predictive bias incurred by model structural simplicity to be taken into account. Notwithstanding the possibility of history-matching-induced bias, posterior predictive uncertainty may be considerably smaller than prior predictive uncertainty because of the history-matching alacrity of this type of modelling.

10. CONCLUSIONS

10.1 Decision-Support Modelling Metrics

Decision-support modelling always requires compromise. There is no such thing as perfection. Nor is there even a clear definition of "success". As we explain in Section 2 of this manuscript, metrics can be developed for decision-support modelling failure on the one hand, and decision-support modelling uselessness on the other. "Success" comprises avoidance of these two conditions.

This leaves a lot of room for subjectivity. It also leaves a lot of room for discussion and debate. The concepts that we present in this manuscript do not provide a prescription for the way in which decision-support modelling should be implemented at this particular location or that, for no such prescription exists. However we hope that they provide common ground for discussion between stakeholders who may have different values and diverging opinions, but who are united in their pursuit of an approach to model-based processing of groundwater data that can be justifiably described as "scientific".

10.2 New Perspectives

In this manuscript we have tried to expose some misconceptions that routinely erode the utility of decision-support groundwater modelling.

We make the point that the primary task of decision-support groundwater modelling is not that of simulation. It is that of assimilation of information. It is not that of giving numerical voice to partial differential equations that characterise groundwater flow. It is that of giving numerical voice to the imperatives of Bayes equation. Bayes equation states, in a succinct mathematical way, that information from two different sources must be blended in order to reduce uncertainty when predicting the future behaviour of an environmental system. Numerical simulation facilitates this blending process. However, it can only serve this purpose if it is designed to do so.

We also make the point that the means through which simulation can best serve this purpose is not through adoption of a complex structure that attempts to replicate what we imagine reality to be. The imperatives of decision support require greater subtlety than this.

Conceptually, a numerical simulator possesses two fundamental attributes. These are its structure and its parameters. Either can be complex or simple; however a complex parameterisation scheme normally requires at least moderate structural complexity to host it. While neither of these attributes can exist without the other, it is parameters that are of primary importance, for these comprise the receptacles that a model provides for information. Information is, by definition, that which reduces uncertainty. A model's structure must be capable of hosting the parameters that are required to convey information from where it resides in site characterisation data and in measurements of system state to the decision-critical predictions that the model is required to make.

A further reversal of the way that both modellers and modelling stakeholders understand numerical simulation is warranted if decision-support modelling is to achieve its full potential. A model should not be viewed as a "deliverable". What a modeller should actually "deliver" is a process - a process that is described by Bayes equation. Simulation is therefore a numerical activity that is undertaken jointly with other numerical activities in support of a greater cause. These other activities have history-matching and uncertainty quantification as their foci. They are implemented by software packages such as those supplied with the PEST and PEST++ suites which are specifically designed to work in partnership with simulators. These activities require that a modeller associate prior uncertainties with model parameters. When undergoing calibration, prior parameter uncertainties are used for regularisation; when evaluating posterior uncertainty, prior parameter uncertainties provide the starting points for numerical analyses that are targeted at their reduction.

10.3 Modelling as a Compromise

We point out in this manuscript that while implementation of the precepts of Bayes equation is the decision-support modeller's primary objective, many obstacles separate him/her from attaining this objective. These cannot be ignored. The decision-support modelling process must be designed to accommodate them.

A simulator that is used in decision-support should run reasonably fast; it should be numerically stable. Absence of these traits makes its use with partner software such as programs of the PEST and PEST++ suites difficult. Design of a model for use in a particular decision-support context must take this into account. "Integrity" of simulation must sometimes yield to the numerical exigencies of speed and stability.

However, the notion that there is much "integrity" in numerical simulation to begin with is largely fictitious. Numerical simulation of natural systems is approximate at best. Hence in many decision-support circumstances, compromises that must be made in order to facilitate use of a model with members of the PEST and PEST++ suites are not great compared with approximations that are incurred by the simulation process itself.

The potential for simulator imperfections to compromise the outcomes of the decision support modelling process must be recognised when designing that process. Information that emerges from site characterisation, and that is resident in measurements of system behaviour, can be distorted on its passage through an imperfect model. This is because defects in a model's structure, and in its simulation capabilities, can distort the information-hosting calibre of its parameters. Information distortion can be mitigated by appropriate model design, and by appropriate processing of field measurements and corresponding model outputs before they are matched. Optimisation of both of these strategies is likely to be prediction-specific. Therefore, where the uncertainties of more than one prediction must be quantified and reduced in order to support management of groundwater at a particular site, more than one model, and more than one incarnation of the decision-support modelling process, may be warranted.

10.4 Parameter Subspaces

Insights into the passage of information from field data to model parameters can be obtained by considering the action of singular value decomposition (SVD) on a model sensitivity matrix. SVD is described only briefly in the present manuscript. References are provided to papers where its use in model calibration, and in understanding and evaluating model predictive bias, are described.

SVD subdivides parameter space into two orthogonal subspaces. One of these subspaces receives information from measurements of system state. The other is invisible to this source of information. The latter subspace can be informed only by expert knowledge and site characterisation; parameter combinations which occupy this so-called "null space" retain their prior stochasticity.

The subdivision of parameter space that is exposed by SVD can be used to define a spectrum of model prediction types. The position of a decision-critical prediction on this spectrum depends on the extent to which it is informed by expert knowledge and site characterisation studies on the one hand, and by measurements of system behaviour on the other hand. In this manuscript we classify the predictive end-members of this spectrum as “expert-knowledge-driven” in the first case, and “data-driven” in the second case. We refer to prediction types which lie between these endpoints as “mixed”.

Insights born of SVD that are explored elsewhere but are documented in this manuscript, illustrate the complex relationship that exists between model structural defects and history-matching. If a model’s structure is too simple, this may erode the ability of its outputs to replicate field measurements. Alternatively, it may leave its ability to replicate field measurements unimpaired. However, parameters may need to adopt roles that compensate for model structural imperfections as they are adjusted to enable model replication of observed system behaviour. This can bias estimates of their values. Under certain circumstances, it can bias the values of predictions that are made by a history-matched model.

10.5 Decision-Support Modelling Strategies

The position that a prediction of management interest occupies on the prediction spectrum is of fundamental importance. It can normally be assessed prior to designing the process by which modelling extracts information that is pertinent to that prediction from data wherein it resides, and transfers it to the prediction itself. It can thus be used to design and optimise that process.

Where a prediction is data-driven, the primary concern of the decision-support modelling process must be to attain a good fit between model outputs and field measurements (or appropriately processed field measurements) of system behaviour. The uncertainties of data-driven predictions are unlikely to be high. Furthermore, despite the fact that model parameters may incur bias through history-matching as they compensate for model structural imperfections, these biases are not transferred to data-driven predictions.

At the other end of the spectrum are predictions that are almost entirely uninformed by field measurements of system behaviour. Predictions of this type are informed only by expert knowledge and site characterisation. Information from these sources is stochastic in nature. In these circumstances, the task of decision-support modelling is to imbue a decision-critical prediction with stochasticity that it inherits from this source. There may be little need for history-matching. A decision-support modeller may turn to geostatistics for prediction uncertainty quantification. Alternatively, he/she may undertake worst case analysis, perhaps after undertaking some representative geostatistical simulations of local hydrogeological detail in order to establish worst case connectivities of permeable and/or impermeable lithologies.

Most predictions that are required of decision-support modelling are of the mixed type. Here the Bayesian imperative reigns supreme as the decision-support modelling process must attempt to express stochastic information emerging from site characterisation, while reducing predictive uncertainty through history-matching. Model run times must be kept reasonable to make this possible; maintenance of model numerical stability becomes a priority. Because the decision-support modelling process must include history-matching, and because decision-critical predictions may be sensitive to parameters which adopt surrogate roles to compensate for model structural defects as they are history-match-adjusted, there is a possibility that some predictions will incur bias. This establishes a tension between model design alternatives that are characterised by high and low structural complexity. The latter facilitates attainment of a good fit between model outputs and field measurements, possibly at the cost of predictive bias. Meanwhile, the former reduces the chances of history-matching-induced predictive bias at the same time as it facilitates expression of expert knowledge; it does this by imbuing a model with a numerical structure that is easily recognisable by site experts, and that can host parameters whose values bear a close (though upscaled) relationship to site hydraulic properties.

In this manuscript we discuss two possible strategies for resolution of this tension. One is demonstrated using a conceptual example, while the other is exemplified using a real-world case study. The first strategy attempts to minimise the possibility of predictive bias by maximizing the ability of model structure to represent hydrogeological information born of expert knowledge. Though sometimes necessary, this approach is shown to be fraught with difficulties - so much so that its benefits may sometimes be illusory. We describe the second alternative as “fit first and ask questions later”. While this strategy may incur some predictive bias, and while its history-matching outcomes may be somewhat abstract, it is simple to implement. In some decision-support circumstances, it may provide insights that have a direct bearing on system management; in others it may guide strategic construction of a more structurally-complex model.

10.6 The Roadmap

We summarise our understanding of decision-support modelling complexity using a simple roadmap. This roadmap differentiates between structural and parameterisation complexity, while recognizing interdependence between the two.

Use of the term “roadmap” does not imply that it can provide unerring guidance to the correct modelling destination in any particular decision support context, for there is no “correct” destination. The roadmap is intended to serve as a reminder of concepts that we present herein. Our hope is that these concepts will enable modellers to formulate decision-support modelling strategies that are appropriate for the management settings in which they work. We also hope that it assists modelling stakeholders to understand the metrics by which decision-support modelling should be judged, and the compromises that must be made in respecting these metrics.

11. REFERENCES

Bense, V.F., and Person, M.A., 2006. Faults as conduit-barrier systems to flow in siliciclastic sedimentary aquifers. *Water Resources Research*. 42, W05421, doi:10.1029/2005WR004480

Bense, V.F., Gleeson, T., Loveless, S.E., Bour, O. and Scibek, J., 2013. Fault zone hydrogeology. *Earth Science Reviews*. 127, p171-192.

Cannon, S., 2018. Reservoir modelling: a practical guide. John Wiley & Sons, Ltd. Print ISBN:9781119313465; Online ISBN:9781119313458; DOI:10.1002/9781119313458.

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.

Dagan, G., 1979. Models of groundwater flow in spatially homogeneous porous formations. *Water Resources Research*. 15(1): 47-63.

Doherty, J., 2011. Modeling: picture perfect or abstract art? *Groundwater*. 49(4):455-456.

Doherty, J., 2015. Calibration and uncertainty analysis for complex environmental models. Published by Watermark Numerical Computing, Brisbane, Australia. 227pp. ISBN: 978-0-9943786-0-6. Downloadable from www.pesthomepage.org.

Doherty, J., 2019a. Manual for PEST. Downloadable from www.pesthomepage.org.

Doherty, J., 2019b. Manual for PEST_HP. Downloadable from www.pesthomepage.org.

Doherty, J. and Christensen, S., 2011. Use of paired simple and complex models in reducing predictive bias and quantifying uncertainty. *Water Resources Research.* 47, W12534, doi:10.1029/2011WR010763.

Doherty, J. and Moore, C., 2019. Decision support modelling: data assimilation, uncertainty quantification and strategic abstraction. *Groundwater.* 58(3), 327-337 doi: 10.1111/gwat.12969.

Doherty, J., Rumbaugh, J. and Muffels, C., 2021. Probabilistic Contributing Area Analysis. A GMDSI Worked Example Report. National Centre for Groundwater Research and Training, Flinders University, South Australia.

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 Welter, D., 2010. A short exploration of structural noise. *Water Resources Research*. 46, W05525, doi:10.1029/2009WR008377.

Farmer, C.L., 2002. Upscaling: a review. *Int. J. Numer. Meth. Fluids*. 40:63-78. doi: 10.1002/fld.267.

Freeze R.A., Massmann J., Smith L., Sperling T. and James B., 1990. Hydrogeological decision analysis: 1, a framework. *Groundwater*. 28(5):738–766.

Moore, C. and Doherty, J., 2005. The role of the calibration process in reducing model predictive error. *Water Resources Research*. 41(5). W05050.

Moore, C. and Doherty, J., 2021. Exploring the adequacy of steady-state-only calibration. *Accepted for publication: Frontiers in Earth Science.*

Obergfell, O., Bakker, M. and Maas, K., 2019. Identification and explanation of a change in the groundwater regime using time series analysis. *Groundwater*. 57(6):886–894.

PEST++ Development Team, 2021. Manual for PEST++. Downloadable from https://github.com/usgs/pestpp

Peterson, T.J. and Fulton, S., 2019. Joint estimation of gross recharge, groundwater usage, and hydraulic properties within HydroSight. *Groundwater*. 57(6):860–876.

PyEMY Development Team, 2021. PyEMU is downloadable from https://github.com/pypest/pyemu.

Renard, P. and Allard, D., 2013. Connectivity metrics for subsurface flow and transport. *Advances in Water Resources*. 51:168-196.

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.

White, J.T., 2018. A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions. *Environmental Modelling & Software*. 109. 10.1016/j.envsoft.2018.06.009. http://dx.doi.org/10.1016/j.envsoft.2018.06.009.

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

Whitacker, S., 1969. Advances in the theory of fluid flow in porous media. *Ind. Eng. Chem*. 61(12):14-28. Doi: 10.1021/ie50720a004.

Zhang, P., Pickup, G and Christie, M., 2008. A new practical method for upscaling in highly heterogeneous reservoir models. *SPE Journal* March 2008, 68 – 76.

gmdsi

Groundwater Modelling
Decision Support Initiative

gmdsi.com

BHP
Flinders University
National Centre for Groundwater Research and Training
RioTinto