Scientifically-Based, Decision-Support Groundwater Modelling:

A Conceptual Framework

Draft of Part 1

Preface

The first part of this book describes the way in which decision-support groundwater modelling is commonly conceptualised. It then proposes a better way. Chapters which comprise the second part of the book discuss how this better way can be implemented in a variety of real-world contexts. The first part is written by John Doherty. Authorship of chapters comprising the second part is shared.

The term “site conceptual model” is normally used to describe a qualitative understanding of processes that are operative, and hydraulic properties that prevail, in a groundwater system that is the subject of current management interest. Salient processes can include recharge, interaction of groundwater with surface water bodies, sources of contamination, geological layering and structure, and many other aspects of the subsurface that affect movement of groundwater and contaminants. The site conceptual model is therefore a kind of mental image of the subsurface, sometimes with some of its process and geometrical elements at least approximately quantified.

In this book we attempt to show that the ability of numerical simulation to support groundwater management requires more than simply respect for this mental image. Often, it requires that this mental image be challenged. It certainly requires that development of this image not take place in isolation from the numerical modelling process that follows it, for the two should not be considered as totally separate activities.

Between development of a site conceptual model and development of a numerical model there exists a workflow element whose importance often goes unrecognised. In Part 1 of this book we identify this element, and give it a name. Because it precedes construction of a numerical model, it is included in a suite of activities that can jointly be referred to as “modelling conceptualisation”. Included in modelling conceptualisation is the development of a site conceptual model. To distinguish the latter activity from the superset of conceptualisation activities which are required before numerical modelling can begin, we will always prefix the term “conceptual model” with “site” when referring to the mental image of processes and properties that are the outcomes of site characterisation.

This book is organised as follows.

Part 1 describes in detail how numerical simulation can provide scientific support for groundwater management. It identifies some sharp differences between the scientific method and ways in which decision-support groundwater modelling is generally undertaken. It points out that there is no “right way” to engage in groundwater modelling, as many subjective decisions must be made. However, it also points out that if groundwater modelling is scientifically-based, it is more likely to achieve its decision-support objectives than if it is based on an alternative set of, often unstated, principles.

Part 2 presents real world examples. For some of these examples, models have already been built. For others they have not. The purpose of these chapters is to demonstrate how concepts that are developed in Part1 1 can illuminate design of a decision-support modelling workflow that is appropriate for each case.

Recognition of the importance of scientifically-informed subjectivity has some immediate consequences. The first of these consequences is that a modeller must take responsibility for his/her own actions, for there is no universal roadmap. This, in turn, requires that he/she has the necessary expertise to take these actions. A second consequence is that a modeller should be free to exercise his/her creativity in designing a context-specific modelling workflow that places important information at the disposal of environmental decision-makers. A third consequence is that a modeller must be able to clearly articulate his/her modelling strategy, and demonstrate how it is derived from the scientific method.

This book attempts to show that the metrics for success of these endeavours do not lie in resemblance of model-generated graphics to “subsurface reality” as it is presently conceptualised. In fact, the harvesting of information from site data is often better served if the pictures that emerge from model-based processing of these data are somewhat abstract.

A term that is used repeatedly throughout this text is “structurally simple and parametrically complex”. The benefits of this approach to decision-support groundwater modelling are explained in Part 1. It is enabled by use of high-end, model-independent, parameter manipulation software such as PEST/ PEST++ in conjunction with standard numerical simulators. Use of this software marks a departure from traditional groundwater modelling workflows which were developed at a time that preceded its availability.

Newer workflows that are facilitated by PEST/PEST++ enable simulator-based data assimilation in ways that were previously impossible. However, decision-support possibilities that are afforded by coordinated use of PEST/PEST++ with numerical simulators can only be realised if deployment of these packages is preceded by astute conceptualisation of a scientifically-based modelling workflow. It is the purpose of this book to explain the principles behind this conceptualisation, and to provide some examples of its practical implementation.

Part 1: The Decision-Support Modelling Workflow

John Doherty

April, 2022

Contents

1. A Famous Experiment

Ernest Rutherford was born in New Zealand in 1871. In 1911, while employed as the Langworthy Professor of Physics at Manchester University, he conducted a famous experiment. The outcomes of that experiment revolutionised the conceptual model of the atom.

At the time, it was believed that mass and charge are uniformly distributed throughout the volume of an atom. Not quite knowing what they would find, Rutherford and his co-workers Hans Geiger and Ernest Marsden devised a simple but elegant experiment – one of the earliest experiments in particle physics. Using a natural radioactive source, they aimed a beam of alpha particles at an extremely thin sheet of gold foil. (Alpha particles are positively charged; they are identical to the nuclei of helium atoms). According to the prevailing view of the atom, these particles should have passed straight through the foil with little, or minor, deflection of their paths. Particles were detected using a screen coated with zinc sulphide. Zinc sulphide fluoresces if struck by an alpha particle. Bursts of light were detected using a viewing microscope attached to the back of the screen.

Unsurprisingly, when the screen was placed behind the gold foil, alpha particles were detected in abundance. What took the researchers by surprise was that some alpha particles were detected at large angles to the beam. A very small number of particles (about 1 in 8000) bounced back towards the source.

The repercussions are profound. An atom is mostly empty space. Negatively charged electrons occupy this space. Positive charges are concentrated in a tiny nucleus. Any alpha particle that bounced back towards the alpha source had struck the nucleus of a gold atom and was repelled by its positive charge.

This experiment illustrates science at its best – and at its most exciting.

2. Science and Philosophy

2.1 Inference and Logic

Over the years, philosophers have written much about the thought processes that lead to the design of ingenious experiments, and the sometimes even more ingenious interpretation of experimental outcomes. Baker (2017) and Caers (2018) provide easy-to-read summaries of these discussions; they also discuss how the scientific method is best framed in the hydrological sciences.

At the root of any journey of scientific discovery, and of the drawing of scientific conclusions, is the exercise of logic. In discussions about the philosophy of science, logic is often classified as belonging to one of three types.

Inductive logic draws conclusions from regularities that are apparent in the behaviour of a system. Repeated observations, possibly under controlled conditions, may allow a scientist to attribute causes to various facets of system behaviour. A conceptual model of a system thereby arises. However, information gaps may oblige the scientist to recognise that some or all aspects of the conceptual model are accompanied by uncertainty. Specification of this uncertainty then becomes part of conceptual model characterisation. It may be possible to reduce this uncertainty through acquisition of further data. The conceptual model may suggest the nature of these data, and the types of experiments that are most likely to yield it.

Bayesianism (discussed below) is said to be based on inductive logic. It accepts that much of what we know is accompanied by uncertainty, and must therefore be expressed probabilistically. It describes how uncertainty may be reduced through acquisition of pertinent data. Uncertainties that prevail before the acquisition of supplementary data are referred to as "prior uncertainties"; revised uncertainties are referred to as "posterior uncertainties".

In contrast to inductive logic, deductive logic establishes connections between premises. Mathematics is deductive. The veracity of a mathematical assertion rests entirely on the veracity of previously established assertions, and on the chain of logic that joins them. There is no room for uncertainty.

Finally, there is abductive inference, or simply, abduction. This term is used to characterise "lightbulb moments" such as those experienced by a master detective who finds an important clue and realises something that he/she had not realised before. It characterises the reaction of a scientist who, when confronted with a surprise (like Rutherford in the previous story), proposes a whole new hypothesis. It is the "lateral thinking" that occurs when an experienced and creative mind is presented with new facts.

2.2 The Scientific Method

Philosophers of science have also had much to say about the scientific method. Perhaps the best-known exposition of this method is that chronicled by Karl Popper (1959, 1963).

Suppose that a hypothesis has been proposed that provides a suitable explanation, or conceptual model, of a natural phenomenon. A scientist may design an experiment to test this hypothesis. The experiment may require the establishment of carefully controlled conditions so that there can be no confusion about causes of system behaviour that are observed during the experiment. Perhaps it is found that observations of system behaviour are in accordance with those predicted using the current conceptual model. This does not mean that the hypothesis that is embedded in this conceptual model is thereby accepted as the sole explanation for system behaviour. Other hypotheses may yield the same predictions. However, a hypothesis, and the conceptual model that underlies it, can be rejected if its predictions are not in accordance with experimental observations.

This view of the scientific method has the notion of "falsification" at its heart. It implies that knowledge is asymmetric. We cannot know exactly how a system will behave in the future; we can only know how it will not behave. By refusing to unconditionally accept one particular explanation of the behaviour of a system, the scientific method allows the existence of competing hypotheses. The door to new knowledge remains rigorously open, as the scientist's mindset is one of perpetual doubt. The demeanour of the scientist is that of humility, born of an awareness that he/she could be proved wrong at any time.

Despite its rigour, the Popperian "falsification" view of the scientific method is not universally accepted as a complete description of scientific thought and processes – especially in application of the scientific method to natural systems. Natural systems can never be controlled environments. Where measurements are made of system behaviour, the relationship between cause and effect may be obscure. Hence a measurement of system behaviour may not allow a scientist to unequivocally reject a particular hypothesis or conceptual model, even though it may contribute to mounting evidence against it.

Despite the fact that natural systems are not directly amenable to deductive logic applied to carefully controlled experiments, the need for scientific rigour in understanding them, and managing them, is not diminished. However a description of the scientific method as it must operate in these environments must accommodate its application to everyday problems that are somewhat perplexing, but are still amenable to discovery.

Quine (1951) and Kuhn (1962) argue that the neat separation between concepts and data, and even between objects of the scientist's attention and the scientist him/herself, are somewhat simplistic. An experienced scientist that is immersed in a system may notice something intriguing. This may not be the outcome of a controlled experiment that tests a hypothesis; it may simply be the outcome of a search for unanticipated clues that may elucidate one or a number of unresolved issues pertaining to a system's behaviour. A scientist will know the clue when he/she sees it; it is his/her experience and expertise that turns a nondescript observation into an important piece of evidence. It is his/her prior knowledge of the system (and a sense of frustration born of insufficient knowledge of it), that can turn this evidence into a new hypothesis that can then be tested by rigorous deductive analysis.

Central to this view of the scientific method is the role of abductive inference - not as a replacement for deductive and inductive inference, but as a complement to it. Baker (2016) points out the importance of abduction in the study of hydrologic systems. He argues that the scientific method applied to such systems often comprises a search for clues. Central to the uncovering of clues is the processing of environmental data in innovative ways that may allow a scientist to notice something that has not been noticed before. Meanwhile, an open mind, a thorough knowledge of the system, and a willingness to learn from the unexpected, allows a scientist to recognise the importance of the clue and to make the cognitive leaps which it inspires.

2.3 Groundwater Modelling and the Scientific Method

We argue below that abduction should also comprise a central part of scientifically-based processing of data in support of groundwater management. This does not negate the utility of other elements of the scientific method, for one is impossible without the others. However, it does call into question the utility of basing a numerical model on a multi-facetted suite of hypotheses of system behaviour that are embedded in a single site conceptual model in a way that suppresses questioning of these hypotheses, or that hinders the emergence of unexpected clues that may precipitate extensive revision of these hypotheses. Nor does invocation of a set of "mutually exclusive and collectively exhaustive" conceptual models as recommended by (amongst others) Refsgaard (2012) and Enemark et al (2018) constitute a suitable expression of the scientific method in the groundwater decision-support context. As will be discussed below, natural systems are far too complex for an approach that is so limited in its conceptual dimensions.

Application of the scientific method to natural systems can be as complex and diverse as natural systems themselves. We demonstrate below that, implied in the decision-support imperative, is a system management hypothesis that simulator-based data analysis can attempt to falsify using classical Popperian logic. Deployment of a relatively simple model that is dedicated to this task can be optimised for performance of it. While directing in this task, a scientifically-based decision-support modelling workflow may expose aspects of system behaviour and/or properties that were previously unanticipated, and that may have far-reaching implications for management of that system.

3. Groundwater Modelling: The Conventional Workflow

Figure 3.1 illustrates the conventional decision-support groundwater modelling workflow as suggested by many textbooks and guideline documents. It can be argued that this workflow has stood the test of time. However, it can also be argued that the time has come to revise this workflow in light of advances in data assimilation technology, and in thinking that these advances have inspired.

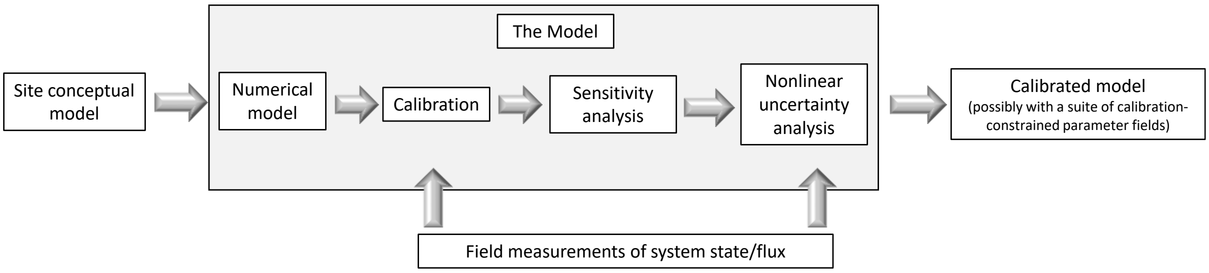

Figure 3.1. The traditional decision-support groundwater modelling workflow.

The workflow begins with development of a site conceptual model. This draws together data from many sources. It generally includes a geological model. It also includes a description of those facets of the hydrologic cycle that are important for management of a study area. These include sources and sinks of groundwater, as well as specification of locations where groundwater interacts with surface water. It identifies anthropomorphic system stresses such as pumping, mining and artificial recharge. Where groundwater contamination is an issue, it identifies possible contaminant source locations.

Next, a numerical model is built. This simulates hydraulic processes that are described by the site conceptual model, while respecting the locations and geometries of features that control the disposition of subsurface hydraulic properties.

The next stage of the workflow admits field measurements of system states and fluxes into the modelling process. Hydraulic property values are ascribed to hydrogeological units that are identified in the site conceptual model through calibration of the numerical model. Back-calculation of hydraulic properties from measured system behaviour in this way ensures that hydraulic property values that are assigned to different hydrogeological units pertain to the scale at which system management must take place.

The calibrated model is next subjected to sensitivity analysis. Model parameters are varied while model outputs are monitored. Parameters may represent system hydraulic properties ascribed to conceptual-model-identified hydrogeological units; alternatively, they may represent system stresses. Some of these parameters may have been adjusted during the preceding calibration process. If one or more history-matched model outputs are sensitive to a particular adjustable parameter, this suggests that the calibration-assigned value of that parameter is credible. If not, the parameter is likely to have a high degree of post-calibration uncertainty. Sensitivity analysis may also explore relationships between model parameters and model predictions. If a decision-pertinent model output is sensitive to a particular parameter, it will inherit uncertainty from that parameter to the extent that uncertainty remains after calibration.

Finally, nonlinear uncertainty analysis may be performed. In many cases this is achieved by generating a suite of “calibration-constrained” random parameter fields; decision-critical model predictions are then made using all of these parameter fields. Generation of these calibration-constrained random parameter fields generally requires use of specialist software; see, for example, Chen and Oliver (2013) and White (2018).

4. Some Explanations

4.1 General

Before providing a critique of the workflow that is depicted in Figure 3.1, a number of terms and concepts that enable an understanding of it, and of an alternative workflow that is proposed later in this document, are briefly discussed. Refer to Doherty (2015) for further details.

4.2 Model Structure and Model Parameters

In this document the word “structure” is used to describe those aspects of a numerical model that are an immutable part of its construction. These include the processes that it simulates, many of which are encapsulated in partial differential equations for which it provides a numerical solution. They also include the model’s gridding and layering, as well as the locations and types of its boundary conditions. They may include zonation that is used as a basis for hydraulic property adjustment during history-matching, as well as those of its properties or inputs that are assigned rather than estimated.

In contrast, parameters are aspects of a numerical model that are altered when undertaking history-matching, sensitivity analysis and uncertainty analysis. Often they include hydraulic properties that are assigned to different parts of a model domain. They can also include system stresses and boundary-condition specifications of which a modeller is unsure.

Doherty and Moore (2021) describe how the structure of a model can be conceptualised as providing the “shell” or “framework” which plays host to model parameters. A parameter has no meaning without this framework. However the framework can exist independently of parameters.

4.3 History-Matching: Calibration

Adjustment of model parameters in order to allow certain of its outputs to replicate field measurements of system states and fluxes comprises an inherently ill-posed inverse problem. Calibration, by definition, seeks a unique solution to this problem. Hence uniqueness must be manufactured. The mathematical term for seeking uniqueness where none exists is “regularisation”.

Regularisation methodologies can be classified into three broad categories.

The first category is manual. Guided by the principle of parsimony, a modeller may subdivide a model domain into broad zones in each of which hydraulic properties are assumed to be uniform. Hydraulic property values are then estimated for a limited number of these zones. Meanwhile, other hydraulic properties, as well as uncertain model inputs, are fixed at assumed values. The fact that these values are assumed does not signify that they are known; it signifies only that the dataset that is available for model calibration does not allow their unique inference.

The other two regularisation categories are numerical. They can be broadly labelled as subspace methods and Tikhonov methods. Generally they are used together.

Subspace methods work by reducing the number of entities that must be estimated in solving the inverse problem. They accomplish this through automatic formulation of a limited number of uniquely estimable “super parameters”. These super parameters are composed of orthogonal combinations of native model parameters; they are often weighted averages of these parameters. The remainder of parameter space is subdivided into an orthogonal-complementary set of super parameters that cannot be estimated on the basis of the current calibration dataset. These combinations of native model parameters retain their original values during subspace-based regularised inversion.

Tikhonov regularisation asks a modeller to specify a preferred parameter condition. This can be one of parameter smoothness and/or parameter homogeneity. Alternatively, it may be a condition in which each parameter is assigned a value that is expected on the basis of expert knowledge arising from site characterisation. Solution of the inverse problem then seeks a parameter field that departs minimally from this pre-defined condition. It can be shown that this minimum departure constraint creates the necessary conditions for parameter uniqueness.

Numerical regularisation has many advantages over manual regularisation. These include the following.

1. A modeller can assign as many parameters to a model as he/she desires. Parameterisation density can be high in areas where hydraulic property heterogeneity may be high and/or measurement spatial density is high. Numerical regularisation will find a unique solution to the inverse problem of model calibration regardless of the number of parameters that are featured in that problem.
2. If regularisation is properly implemented, solution to the inverse problem of model calibration can be shown to be of "minimised error variance". This means that predictions made by the calibrated model will lie somewhere near the centre of their posterior probability distributions. Their potential for parameter and predictive wrongness is minimised by ensuring that this potential is symmetrical with respect to the predictions themselves.
3. The necessity for subjective and ad-hoc definition of a zone-based parameterisation scheme, accompanied by the assumption of hydraulic property uniformity within each zone, is eliminated; these assumptions are often made when implementing manual regularisation.

In this text, we refer to calibration based on numerical regularisation as "highly parameterised inversion". Additional benefits accrued through use of a multiplicity of parameters are outlined below. However, at this point we would like to dispel the myth that use of a large number of parameters inevitably leads to "over-fitting" of model outputs to field measurements. To be sure, a modeller can over-fit field data if he/she wants to. However modern-day regularisation methods that are implemented in software packages such as PEST and PEST++ provide functionality through which this is easily prevented.

4.4 History-Matching: Bayesian Methods

Bayesian methods are most easily discussed by describing the terms that appear in Bayes equation. Bayes equation is conceptual in nature and easy to understand. However its numerical implementation may be challenging in some history-matching contexts. Nevertheless, once the programming is done, software which implements Bayesian methods is not difficult to use.

Suppose that **k** is a vector of model parameters. (A vector is a list of numbers.) Suppose that **h** is a vector containing field observations of system state. That is, **h** comprises a history-matching dataset. Bayes equation can be written as follows:

$$P(\mathbf{k}|\mathbf{h}) \propto P(\mathbf{h}|\mathbf{k})P(\mathbf{k}) \tag{4.1}$$

The notation *P*() signifies "probability". *P*(**k**) specifies the prior probability distribution of parameters **k**. This is the probability distribution of parameters that emerges from site characterisation. (Note that while this probability distribution exists conceptually, it may be difficult or impossible to formulate mathematically because of the natural complexity of geological media.) *P*(**h**|**k**) is the so-called "likelihood function". This increases to the extent that model outputs fit corresponding field measurements. *P*(**k**|**h**) is the posterior probability distribution of **k**; that is, it is the probability of **k** conditional on measurements embodied in the vector **h**.

Equation 4.1 comprises a probabilistic formulation of the history-matching process. In contrast to calibration, Bayesian history-matching does not seek parameter uniqueness. In fact, it seeks the opposite. It seeks a diverse ensemble of parameter vectors **k** that are all "realistic" according to *P*(**k**), and that all allow model outputs to replicate field measurements to within measurement noise. (The likelihood function *P*(**h**|**k**) includes a characterisation of measurement noise.) If a prediction is made using all of these history-match-constrained parameter fields, the posterior probability distribution of that prediction can be explored.

4.5 The Benefits of Using Many Parameters

Whether history-matching is conducted for the purpose of model calibration, or whether it is conducted in a Bayesian framework, many benefits are accrued if a model is endowed with a multiplicity of parameters instead of just a few. These benefits include the following.

1. If a quantity that characterises some aspect of the subsurface is incompletely known, then a modeller must either assume a value for it, or declare its value to be adjustable; in the latter case, the quantity becomes a parameter. The latter course of action will probably not result in its unique estimation (particularly if there are many such quantities). However it will prevent possible parameter and predictive bias that may be incurred by fixing that quantity at a possibly erroneous value. Meanwhile, neither numerically-regularised inversion nor Bayesian history-matching is troubled by declaration of a large number of adjustable parameters that express gaps in a modeller's knowledge of the subsurface.
2. The subsurface is replete with hydraulic property "surprises" (Bredehoeft, 2005). These include areas of anomalously high or low hydraulic property values that are not featured in the site conceptual model. Some of these anomalies may impact important model predictions. A superfluity of parameters gives the history-matching process the ability to respond to information within a history-matching dataset that may expose the presence of unanticipated heterogeneity. Highly parameterised inversion therefore gives data a voice. It is far better for surprises to be encountered during history-matching than during management of a system.
3. Unexpected parameter values and patterns that emerge from history-matching may result from parameter compensation for model structural inadequacies. (This is further discussed below.) The use of many parameters often localises parameter compensatory behaviour to areas where these structural deficiencies exist. These deficiencies can therefore be identified. This gives the modeller the opportunity to improve model structure if he/she judges that model structural deficiencies and/or parameter compensatory behaviour may induce bias in important model predictions. Alternatively, if parameter compensatory behaviour is restricted to a non-critical part of the model domain, a modeller may decide that his/her model performs satisfactorily when making management-salient predictions in other parts of its domain.
4. On many, if not most, occasions of model deployment, the uncertainties of decision-critical model predictions are high. This is usually an outcome of data insufficiency, and hence of parameter nonuniqueness. Or, to put it another way, it is the outcome of a high-dimensional parameter null space (Moore and Doherty, 2005). It is important that a model be endowed with enough parameters to adequately represent this null space; this may avoid underestimation of the uncertainties of decision-critical model predictions. In a decision-support groundwater model, the parameters that cannot be estimated are just as important as those that can.

4.6 The Cost of Using Many Parameters

The use of many parameters does not come without a cost. Calculation of sensitivities required for highly parameterised inversion can become expensive if these sensitivities are calculated using finite parameter differences. However, ensemble-based Bayesian techniques do not require that sensitivities of model outcomes with respect to individual parameters be calculated; hence use of an abundance of parameters can be numerically cheap.

For both calibration and Bayesian analysis, appropriate prior parameter probability distributions must be defined before history-matching is undertaken in order to guarantee parameter reasonableness. However, should history-matching surprises occur, both of these methodologies allow violation of these priors to the extent required for fitting of a measurement dataset.

In the groundwater modelling literature, it is sometimes stated or implied that use of many parameters can induce more uncertainty in some model predictions than there needs to be. (We refrain from providing references out of politeness.) This does not occur if prior parameter probability distributions are properly defined. Excessive over-estimation of parameter uncertainty is an outcome of modeller carelessness. Ultimately, the magnitude of predictive uncertainty is determined by lack of information, and not by the parameters that inform a modeller of information paucity. Reducing the number of parameters in order to preclude high predictive uncertainties is the numerical equivalent of shooting the messenger.

4.7 The History-Matching Dataset

Model-to-measurement misfit is often quantified using a so-called "objective function". This forms an important component of the Bayesian likelihood function. The lower is the objective function, the better is model-to-measurement fit, and the higher is the Bayesian likelihood function.

Formulation of an objective function is something of an art. Each observation must be given a weight that reflects its credibility; the higher is its credibility, the greater is the weight that should be assigned to it.

It is often advantageous to subject observations and their model-generated counterparts to mathematical processing before matching them. Appropriately-processed observations can then comprise different components of the overall objective function that is reduced through history-matching. Naturally, the objective function can also include unprocessed observations; alternatively, processed observations can comprise the entirety of the objective function. Processing may include temporal, spatial and vertical differencing; it can also include more complex mathematical operations.

Doherty and Welter (2010), White et al (2014) and Doherty (2015) show that strategic processing of data and corresponding model outputs before minimizing differences between them through parameter adjustment may protect the history-matching process from parameter and predictive bias. While this can be demonstrated mathematically, the logic behind this strategy is simple. There is no need for a model to replicate all aspects of a system's behaviour; it only needs to reproduce those aspects of its behaviour that are important to a prediction.

As a general rule, models are better at predicting differences than absolutes. Hence, if at least part of a history-matching dataset is composed of observation differences, parameters can be protected from history-matching-induced bias that is incurred when a model is asked to replicate aspects of environmental behaviour that are beyond its capacity to simulate. At the same time, posterior uncertainty analysis often demonstrates that model predictive differences have much less uncertainty than the model-predicted values from which these differences are calculated. This suggests that model-based management of a groundwater system should sometimes be based on differences rather than absolutes.

In summary, model history-matching and deployment should recognise that a model cannot be expected to simulate all aspects of a system's behaviour equally well. This does not necessarily compromise its role in decision-support. However, a modeller should "play to a model's strengths" when comparing its outputs to observed system behaviour, and/or when predicting the future behaviour of a managed system.

5. Revisiting the Conventional Modelling Workflow

5.1 Model Assessment

5.1.1 Fit for Purpose

Implied in the conventional decision-support modelling workflow of Figure 3.1, and in the narrative that is generally associated with present-day decision-support modelling, is the belief that with sufficiently large amounts of time and money devoted to construction of a numerical model, simulation "accuracy" can be attained. With less investment than this, some accuracy may be lost, but a numerical model may nevertheless possess sufficient simulation integrity to be declared as "fit for purpose". This licenses it to make a variety of predictions on which management of a groundwater system can be based.

A problem with "fit for purpose" is that both "fit" and "purpose" require definition. At the present time, there is no industry-accepted definition of "fit". Hence disputes about the fitness of a numerical model are as common as numerical models themselves. The "low hanging fruit" for someone who is motivated by conviction, profit or ambition to criticise someone else's model is that it omits, or poorly approximates, certain processes or features that operate in the subsurface; these omissions, it may be stated, erode the model's claims to simulation authenticity. Relationships between the omitted/approximated features and predictions that the model is required to make are often considered to be of secondary importance. Looking like the "real thing" is what matters most; predictive accuracy will surely follow.

The traditional modelling workflow of Figure 3.1 delivers a standalone numerical entity that represents someone's attempt to digitally capture reality. Predictive uncertainty is acknowledged, but the narrative that underpins this workflow may not acknowledge how large these uncertainties may be for some predictions; after all, this is at odds with the colloquial meaning of the word "model". If a model is structurally complex (presumably in order to honour the complexity of the subsurface), then many of its elements are fixed. If something is fixed, it is presumed to be fixed at the right value notwithstanding the fact that it is probably fixed at the wrong value. This may inculcate predictive bias at the same time as its fixed status inhibits exploration of predictive uncertainty. The chances of encountering unpleasant surprises when management decisions are based on this model are thereby raised.

An additional problem with the workflow that is depicted in Figure 3.1 is that the optics of this approach to modelling are tarnished if a group other than that which built the model challenges it with a model of their own that is built according to a similar philosophy. The unseemly public spectacle that follows inspires little confidence in science-based groundwater management.

5.1.2 Modelling Metrics

Rectification of these problems should begin by recognizing that similarity of a numerical model to an uncertain perception of reality is a fallible metric for decision-support. Structurally complex models should not be awarded greater scientific integrity than structurally simpler models just because they are structurally complex, as is intimated by at least one widely-accepted model classification scheme; see Barnett et al (2012). Instead, "model" should be viewed as a verb rather than as a noun. It should be viewed as a process through which information is extracted from environmental data in order to depict that which can be known about the future behaviour of a groundwater system, while quantifying that which cannot be known without acquisition of further data.

With decision-support modelling viewed in this way, scientifically-justifiable metrics can be associated with it. These can provide a basis for review of the decision-support modelling process as it pertains to a particular study site by peers and stakeholders. The review process can therefore focus on something less ambiguous than whether a model is "fit for purpose".

Doherty and Simmons (2013) and Doherty and Moore (2019) address the issue of decision-support modelling metrics. Following Freeze et al (1990), they note that the making of a decision requires that risks be assessed, and that the purpose of groundwater modelling is to provide this assessment. Assessment of risk requires at least approximate quantification of predictive uncertainty, for risk can be roughly perceived as the probability of something going wrong multiplied by the cost of its going wrong.

A decision-support modelling workflow can be deemed to "fail" if the uncertainty margin of a decision-critical model prediction precludes the possibility of an unwanted event (often referred to as a "bad thing" herein), when that event may actually occur.

Of course, decision-support modelling failure can be avoided if uncertainty margins are set so large that the possibility of occurrence of a bad thing cannot be eliminated, even though information contained in site data may support its elimination. This may occur if a modeller employs a structurally and parametrically simple model, or even a back-of-the-envelope calculation, and then encases its prediction in an overly-conservative engineering safety margin. It may also occur if a modelling workflow eschews history-matching of information-rich data, and then associates prior probability distributions with decision-critical model predictions instead of posterior probability distributions. Because the outcomes of workflows such as these may lend little support to the decision-making process, they can be described as "useless". (Of course, if the potential for a bad thing to happen can be excluded with an overtly conservative predictive uncertainty margin, then the modelling workflow is not actually useless; uselessness occurs when reduction of the uncertainty associated with a particular model prediction may enable the making of a different management decision, but the opportunity to reduce its uncertainty is not taken.)

It follows from the above considerations that scientifically-based metrics of "failure" and "useless" can be assigned to any groundwater modelling workflow. Although their assignment will mostly be subjective, they are metrics nevertheless. Furthermore they are directly relevant to a model's task rather than to its appearance.

A complaint may be made that these metrics are negative, rather than positive. This can be defended by noting that their negativity aligns with the notion of falsification that is integral to the Popperian view of the scientific method. Perhaps a less pessimistic view of modelling would follow from the use of antonyms to "failure" and "useless" as descriptions of a modelling workflow. However, a claim that modelling is "successful" appears to be a little audacious, and not aligned with the demeanour of humility that should accompany implementation of the scientific method. Perhaps less so is the claim that modelling is "useful". In fact, such a claim echoes the oft-quoted phrase (originally coined by the British statistician George Box) that "all models are wrong but some are useful". However as "useless" has a stricter definition, we will retain it.

We now revisit the traditional decision-support modelling workflow that is depicted in Figure 3.1.

5.2 The Site Conceptual Model

5.2.1 Cause and Effect

A site conceptual model can be the product of a great deal of expensive hydrogeological investigative work. Depending on the site, and depending on the issue, many holes may have been drilled and geologically/geophysically logged. Groundwater head and/or chemistry data may be available from present and past investigations. Streamflow measurements may record details of the interaction of groundwater with surface water bodies.

At other sites, issue-specific investigations may be minimal, so that the site conceptual model may have to rely on information gleaned from regional studies.

It is understandable that if much work has been devoted to conceptualisation of local hydrogeological properties and processes, then those who commissioned that work, and those who undertook that work, would like to see the outcomes of this work reflected in a numerical model. In fact, it may be construed as either an insult, or proof of a modeller's incompetence, if details that are dear to a hydrogeologist's heart are not on display in a numerical model. This is especially the case if site studies suggest that some of these details may impact flow of groundwater, and movement of contaminants, at locations that are important to groundwater management. These details may include the presence of aquitards, geological structures such as faults, and the nature and distribution of groundwater recharge.

Obviously, management-salient hydrogeological details revealed by site characterisation studies should not be ignored in a numerical model which attempts to support management of a site. However it must be remembered that it is the effects of these details that are important, with particular emphasis being placed on their undesirable effects. Simulation of the range of their possible effects does not necessarily require that their representation in a numerical model resemble pictures from a geological textbook; greater flexibility than this may be required, particular if their effects are uncertain. Furthermore, the representation of known structural details in a numerical model must not be such as to preclude the possibility that measurements of system behaviour, if replicated by a model, may reveal the existence of other important features, or may require a re-thinking of the way that known features affect groundwater and contaminant movement.

5.2.2 The Importance of Stochasticity

There is a paradox at work here, a paradox that will be discussed in greater detail later in this text. It cannot be denied that at many sites hydrogeological details are important. However even the most expensive site investigation can provide only a few details about the hydrogeological details that matter.

Of particular importance to groundwater and contaminant movement is connected permeability (Renard and Allard, 2013). This can arise from many sources. It can be the result of primary or secondary geological processes that prevailed during recent or ancient sedimentary deposition; it can be incurred by faulting or fracturing following deep burial, or by chemical processes that accompany shallow differential weathering. Regardless of its origin, the patterns and properties of features that exhibit connected permeability are likely to be complex and highly heterogeneous. Invariably, more is unknown about them, than is known. This has an important repercussion. If they are to be represented explicitly in a numerical model, then they must be represented stochastically. It follows that their geometries, properties and continuity relationships must be specified probabilistically. Hence many realisations of these features require representation in a stochastic modelling workflow rather than just a single representation in a deterministic modelling workflow. This is because representation of only one (probably incorrect) realisation may induce predictive bias.

History-matching of a numerical groundwater model often reveals that there are some aspects of groundwater behaviour that cannot be explained by the site conceptual model on which the numerical model is based. Where a groundwater model's structure is complex in order that it can give numerical expression to site conceptual model details, failure of the model to replicate system behaviour may be an outcome of the fact that its structure expresses an incorrect realisation of this detail.

In fact, subtleties in the measured behaviour of a groundwater system may reveal something that site characterisation studies have missed, or that they did not consider to be important. It follows that a decision-support modelling workflow should not preclude the possibility that history-matching can influence the design of a site conceptual model. The first arrow in Figure 3.1 should therefore be double-ended.

Conceptualisation of recharge processes is particularly problematic. Details may matter at every temporal and spatial scale. In shallow systems, recharge is episodic. Its spatial distribution is affected by flow of ephemeral streams and flooding of permanent streams. It is affected by soil type, topography and land use; the latter can vary with season. The greater is the detail with which recharge processes are represented, the greater is the importance of representing them stochastically in order to avoid representing them incorrectly. Alternatively, as is discussed below, it may be possible to represent them in a spatially/temporally upscaled manner in ways that afford simpler stochastic characterisation.

5.3 The Numerical Model

Later in this document, we argue that complexity of a site conceptual model does not necessarily imply that the decision-support imperative is best served by structural complexity of an attendant numerical model. However the narrative that often accompanies the workflow that is depicted in Figure 3.1 suggests the opposite, for everyday usage of the noun "model" implies this. In fact, if decision-support modelling is to take its lead from the colloquial meaning of words rather than from the scientific method, there is little other choice.

From this, it follows that if site geology is complex, then site management should be based on a complex numerical model. If this complexity represents one realisation of many possible dispositions of important structural and sedimentary features, then this matters less than that all of these features are explicitly represented in the numerical model. If groundwater levels vary seasonally, then the time-varying nature of recharge must be respected, even if its numerical representation bears little resemblance to the fast-moving, complex array of natural processes that accompany major recharge events, and even if the historical details of transient groundwater behaviour do little to inform its future behaviour when the system is subjected to very different stresses. And if groundwater extraction has the capacity to reduce streamflow, then flow within streams must be represented explicitly in a numerical model even if the complex upstream and instream processes that affect daily flow are imperfectly simulated, and even if the volume of water that is lost from a stream (or fails to flow into a stream) is governed more by local groundwater conditions than by conditions in the stream itself.

It must not be forgotten that even the most "faithful" attempts to simulate subsurface processes can only be approximate. The matrix equations that are solved by a groundwater model emerge from partial differential equations that are applicable on the scale of millimeters, but that are applied on the scale of model cells. The literature of upscaling shows that use of cell-averaged hydraulic properties in these matrix equations (setting aside the question of how averaging is done) is possible under only a small range of conditions. Where these conditions are violated, upscaled hydraulic conductivities are flow-direction-specific and require tensorial representation.

Upscaling of recharge processes, and of processes that govern interaction of groundwater with surface water, is even more problematic, for these processes are highly nonlinear. The same applies to simulation of contaminant movement. Here the requirement for upscaling is made explicit by the need to assign values to variables such as horizontal and transverse disperssivity as surrogates for contaminant spreading incurred by flow of water through media whose heterogeneity is unrepresented.

These simulation abstractions do not, of course, invalidate the assistance that numerical modelling can provide to decision-making. However they demonstrate that "purity" of simulation can never be attained; at the same time, they cast doubt on the extent to which the quest for purity is even necessary.

5.4 Calibration

5.4.1 What Calibration Can Achieve

As has already been explained, calibration seeks a single parameter field that allows a model to replicate past system behaviour. Uniqueness is attained by introducing constraints to the inverse problem that model calibration attempts to solve. Imposition of these constraints as history-matching is implemented can expose the flow of information that is harvested from field measurements of system behaviour.

For example, the calibration process may seek a parameter field that departs minimally from a condition of (rock-type-specific) homogeneity. Therefore, where heterogeneity is introduced to a model domain by software which supervises regularised inversion, it is because this heterogeneity “has to be there”. Alternatively, prior to model calibration, a modeller may assign hydraulic properties to a model domain by spatially interpolating those derived from pumping tests. The inversion process may then be asked to respect these prior hydraulic property values to the extent that it can, while nevertheless pursuing a good fit of model outputs to field measurements. Where alterations to a user-assigned parameter field are required in order to achieve this fit, the calibration process may be asked to do this preferentially at locations where structural features such as faults are known to exist. However, it may be given permission to violate these recommendations if necessary. These violations indicate locations within the model domain where important information has “landed”.

5.4.2 Calibration and the Traditional Modelling Workflow

The benefits of endowing a model with many parameters are discussed above, and will be further discussed below. Those who advocate structural model complexity in accordance with the narrative that is often associated with the workflow of Figure 3.1, often populate their models with somewhat parsimonious parameter fields. This expresses a model design philosophy that places great confidence in the outcomes of site characterisation while failing to recognise the inherently stochastic nature of these outcomes. The inferred dispositions of three-dimensional geological units are therefore hardwired into the structure of the numerical model despite the fact that these dispositions are uncertain except where intersected by drillholes. Hydraulic properties are then assigned to each of these hydrogeological units under the often dubious assumption of intra-unit hydraulic property uniformity. These properties are then varied on a unit-by-unit basis during model calibration.

Endowment of a structurally complex model with a simple parameterisation scheme may also be an outcome of “complexity fatigue” experienced by a modeller.

If a modeller decides to employ more complex parameterisation devices such as pilot points instead of, or together with, zones of assumed piecewise hydraulic property constancy, he/she will often impose upper and lower bounds on calibration-emergent parameter values that reflect prevailing expectations of hydraulic properties. While this strategy is pursued with good intentions, there are times when it may be useful to widen these bounds. For example, the upscaled vertical anisotropy of a single layer which represents a multitude of interfingering sand and clay layers, is often far greater than the vertical anisotropy of either sand or clay. Furthermore, if calibration is implemented using regularised inversion (as it should be), detection by calibration-supervision software that user-specified parameter ranges must be violated for a model to replicate system behaviour is a message that is worth listening to. This implies that the site conceptual model is in need of refinement.

More often than not, calibration of a structurally complex model yields a mediocre fit with a calibration dataset, particularly if structural complexity is not complemented by parametric complexity. It follows that information which is resident in a calibration dataset has been denied entry to the decision-support process. Unperturbed by this, a modeller may deem a mediocre model-to-measurement fit to be acceptable. He/she may justify this claim by reference to a spurious statistic such as normalised RMS. He/she may even claim that his/her failure to achieve a better model-to-measurement fit is an outcome of studious avoidance of over-fitting. A more likely explanation is that he/she has grown weary of attempting to calibrate the complex, slow-running model, and is seeking ways to declare it as “fit for purpose” so that this odious task can be over.

Most global misfit statistics are meaningless. They take no account of the benefits (discussed above) of processing field data and corresponding model outputs before fitting them. Nor do they account for the fact that an overall statistic may be deemed as “good”, while certain aspects of groundwater behaviour may be poorly replicated by a calibrated model. If these aspects of its behaviour are rich in prediction-pertinent information, then that information is lost to the modelling process.

5.4.3 Calibration and the Site Conceptual Model

If attempts to calibrate a numerical model make it obvious that important aspects of system behaviour cannot be replicated by that model, then either it, or the underlying site conceptual model, is in need of refinement. However the refinement process becomes troublesome when uncertain site complexity is embodied in a single realisation of numerical model structure. Furthermore, it may be difficult for a modeller to identify the structural element or elements that are causing the problem. For example, a modeller may not know whether to alter layering details, whether to revise parameter zonation, or whether a structural feature such as a fault should be moved. Revision of a numerical model's complex structure, and population of the revised structure with a new set of parameters, may be a laborious undertaking, especially if features such as faults are loci of model grid refinement. The temptation to declare a mediocre level of model-to-measurement fit as satisfactory becomes very strong indeed.

Manual adjustment of model structure has philosophical implications. Elements of a model's structure become de-facto parameters if they undergo adjustment. However they lack the versatility of real parameters. Calibration achieved through manual adjustment of model structure cannot be guaranteed to achieve a minimum error variance solution to an inverse problem, even if it enables a better fit with a calibration dataset. Furthermore, structure that is manually adjusted during model calibration may not get adjusted during post-calibration uncertainty analysis. This is in spite of the fact that the calibration process has revealed that at least some model structural elements cannot be uniquely specified through site characterisation.

5.4.4 Calibration and Structural Complexity

Conceptually, failure to complement structural complexity of a numerical model with parametric complexity makes little sense, for one generally implies the other. Adjustment of spatial hydraulic property details may be required in order to attain a good fit between model outputs and measurements of system state, and to thereby harvest the information that they contain. Inestimable, but real, hydraulic property heterogeneity may affect the values of decision-critical model predictions; it therefore requires expression in model predictive uncertainty analysis.

Unfortunately, however, the burden of parameterising a structurally complex numerical model may be heavy. If a model has many layers, then parameters are required for all of these layers. These must be endowed with appropriate prior probability distributions. The information required to do this is rarely provided by site characterisation studies. The cognitive and numerical burden of model construction, history-matching and uncertainty analysis becomes very high.

Because a model's structure is immutable, or can be adjusted only with difficulty, it embodies a hypothesis; actually, a structurally complex model embodies many hypotheses. The collective testing of all of these hypotheses during calibration hardly constitutes a carefully-designed scientific experiment. Failure to replicate measured system behaviour does not inform a modeller of the hypotheses that require rejection. In practice, a mediocre fit with a calibration dataset is tolerated. Strong claims are nevertheless made that groundwater management based on such a model is "scientific". These claims are not based on an attempt to implement the scientific method. Instead, they are based on the premise that the model's structure bears some resemblance to one realisation of a hypothesised subsurface. The stochastic nature of subsurface characterisation is ignored. Fondness for pictorial representations of the subsurface displayed on a computer screen take precedence over extraction of information from measurements of system behaviour. Predictions made by such a model are probably biased; in common with all models, they are almost certainly wrong. However, their potential for wrongness is unquantifiable, and probably underestimated. At the end of a laborious and expensive modelling process, risk assessment is difficult at best, and impossible at worst.

The above conceptual difficulties are compounded by severe practical difficulties. Structurally complex numerical models have long run times. They are often numerically delicate. This makes their use with software such as PEST/PEST++ difficult. Hand-calibration may therefore become the only history-matching option. Links with the scientific method then grow tenuous indeed.

5.5 Sensitivity Analysis

5.5.1 Global Sensitivity Analysis

"Sensitivity analysis" means different things to different people. Global sensitivity analysis is sometimes applied to large models that attempt to simulate complex economic, social or environmental systems, and that undergo little, if any, history-matching. In these contexts, the purpose of sensitivity analysis is often didactic. It allows stakeholders to recognise those subprocess which contribute most to aspects of system behaviour that are of particular interest to them. It also provides an understanding of how different components of a complex system interact with each other. This knowledge may be useful in itself. It may also provide the basis for design of simpler models. See Saltelli et al (2008).

5.5.2 Local Sensitivity Analysis

In the groundwater modelling context, "sensitivity analysis" often carries a different meaning. It is normally undertaken following model calibration. Furthermore, it is often "local". This refers to the fact that sensitivities are calculated by perturbing parameters by small amounts from their calibrated values. Basic statistics that are forthcoming from local sensitivity analysis can be used to guide manual regularisation. They may inform a modeller if too many parameters are being estimated, while identifying those that are inestimable. According to traditional advice, the latter should be held at their prior values; see Hill and Tiedeman (2007) for details.

With the availability of regularised, highly parameterised inversion, sensitivity analysis, undertaken for this reason, is no longer required.

Sensitivity analysis may also be undertaken in order to identify parameters which influence model predictions that stakeholders care about. However, in exploring this issue, sensitivity analysis does not constrain parameters by their posterior uncertainties, for this is the task of nonlinear posterior uncertainty analysis. The post-calibration prediction-salience of some parameters may therefore be overstated.

5.5.3 Alternatives to Traditional Sensitivity Analysis

Highly parameterised inversion provides opportunities for alternative types of sensitivity analysis. These can provide deep insights into flow of information during history matching. These analyses are based on the so-called "Jacobian matrix" of local model-output-to-parameter sensitivities. Utility programs supplied with the PEST, PEST++ and PyEMU suites perform tasks that include:

- approximate parameter and predictive uncertainty quantification based on first order second moment (i.e. FOSM) analysis;
- worth of information;
- pre- and post-calibration parameter contributions to predictive uncertainty;
- parameter identifiability;
- solution and null space contributions to parameter and predictive error;
- errors incurred through model simplification; and
- analysis of failure to fit.

Visual outcomes of these analyses, such as maps of post-calibration parameter uncertainty, can provide modelling stakeholders with easily-understood insights into what history-matching has achieved, and what it has failed to achieve because of information paucity. They can be used to guide further site characterisation studies by ensuring that future data acquisition is maximally effective in reducing the uncertainties of predictions that matter.

5.6 Nonlinear Uncertainty Analysis

As has already been discussed, nonlinear predictive uncertainty analysis can be implemented by generating a suite of calibration-constrained, stochastic parameter fields. This can be done efficiently using an ensemble smoother such as PESTPP-IES. Normally the process commences with a user-provided set of parameter fields that sample the prior parameter probability distribution. Through an iterative process, these parameter fields are adjusted until all of them allow model outputs to fit field measurements. Once this has been achieved, they comprise samples of the posterior parameter probability distribution.

Theoretically, use of an ensemble smoother for history-matching dispenses with the need for model calibration. However better fits with a history-matching dataset can often be attained if “prior” parameter ensembles are centred on the parameter field achieved through model calibration, or are even samples of a linearised approximation to the posterior parameter probability distribution.

The ensemble smoother process implements Bayes equation. In doing so, it recognises that model parameters, and predictions that are sensitive to these parameters, are uncertain. At the same time, the ensemble smoother process is numerically efficient. The number of model runs required per iteration of this process is of the same order as the number of parameter realisations that are simultaneously adjusted; it is independent of the number of parameters that comprise each of these realisations. Hence use of a large number of parameters incurs no numerical cost. Conceptually, this frees a modeller from having to make a litany of assumptions about aspects of a model of which he/she is unsure; if something is uncertain, then let it wiggle.

It must be remembered, however, that uncertainty analysis is only as good as the prior probability distributions that are assigned to model parameters. Use of a large number of parameters requires that care be taken in specifying these prior probability distributions, including spatial correlation between parameters of the same type, and correlations that exist between parameters of different types at the same location. If there is a penchant for connected permeability to exist in certain parts of a model domain, then this too must be respected in realisations of posterior parameter fields. If not, the uncertainties of predictions that are influenced by permeability connectedness may be seriously underestimated.

Because of their numerical efficiency, ensemble methods allow structurally complex models to be equipped with complex parameter fields. Hence they may make highly-parameterised history-matching possible for these models despite their long run times. However this process is not immune from practical problems. Structurally complex models sometimes behave badly when provided with random parameter fields; solution convergence failure may become a persistent problem as these fields are subjected to history-matching adjustment. Another problem is that model-to-measurement fits attained with an ensemble smoother are rarely as good as those attained through inversion based on local sensitivities; some information that is resident in a calibration dataset may therefore remain unharvested. Also, if a history-matching dataset is large, it is a mathematical requirement of the ensemble smoother process that the number of realisations that comprise an ensemble be large enough to hold the information that these data contain. This increases the number of model runs required for history matching. (It can be shown that an ensemble should possess at least as many realisations of parameter fields as the dimensionality of the calibration solution space, and then some more.)

5.7 Summary

Implementation of the traditional decision-support modelling workflow that is depicted in Figure 3.1 requires that groundwater simulation give numerical voice to a site conceptual model whose design precedes it, and is independent of it. If the numerical model that emerges from this process is unable to replicate measured system behaviour very well, this may be seen as unfortunate. However the decision-support modelling process is not deemed to suffer loss of credibility because of this. In contrast, in order to defend workflow credibility, the bar for model-to-measurement fit acceptability is normally set low. Failure of a model to replicate important aspects of measured system behaviour may even be seen as praiseworthy, as the demons of over-fitting have been kept at bay.

One has to ask oneself why the groundwater industry is satisfied with this workflow. To be sure, a decision-support modelling process must respect knowledge that emerges from site characterisation. However, the quality of information that site characterisation yields is often overstated by failing to recognise its inherently stochastic nature. Rarely is groundwater site characterisation accompanied by a detailed geostatistical characterisation of lithological distributions and properties. (This is in contrast to petroleum industry workflows.) Belief in the integrity of a single realisation of the subsurface often persists even when a numerical model whose structure respects that realisation cannot be calibrated to replicate important details of system behaviour. Cognitive and workflow inflexibility inhibit formulation of a suitable response to this discovery.

Despite the fact that a numerical model that emerges from this process is replete with hypotheses, the workflow of Figure 3.1 can neither isolate these hypotheses nor test them. The premise for construction of a numerical model that is intended to simulate future system behaviour is that respect for measurements of past system behaviour is less important than respect for a single, detailed image of the subsurface. The demeanour of doubt and the spirit of humility which characterise application of the scientific method are not aligned with this premise.

5.8 Nothing is Black and White

The above criticism of the traditional decision-support modelling workflow may be construed as unduly harsh. Perhaps it is. It is written as a critique; critiques tend to be harsh.

When it comes to groundwater modelling, compromises are required at every turn. Shortly, an alternative decision-support modelling workflow will be suggested. It is based on structural simplicity complemented by parametric complexity. There are those who will argue that numerical models that emerge from this approach are “unrealistic”, and that omission from their structure of certain complexities may severely bias decision-salient model predictions. They may argue that this can inflict heavier damage on the decision-making process than failure to extract information from measurements of system behaviour. They may further argue that a numerical model that is able to replicate the spatial and temporal details of historical system behaviour very well is either too structurally simple to be considered as other than a machine-learning algorithm, or too structurally complex to be subjected to history-matching and uncertainty analysis. Information contained in many nuances of system behaviour is therefore better left untapped.

Perhaps there are situations where these arguments have validity.

The workflow that is presented shortly comprises one end of a continuum of decision-support modelling workflow designs. We argue that this end has been ignored too often in favour of alternative workflows that offer visual appeal but are lacking in scientific integrity. However we also recognise that decision-support groundwater modelling requires that many subjective choices be made by individuals whose approach to modelling is moulded by their own experiences and personalities. Furthermore, there may be difficult situations where a multiplicity of approaches is warranted. Deployment of alternative approaches is facilitated by the fact that the workflow that we now outline can be implemented with relatively little expense in many groundwater management contexts.

6. Experimental Design

6.1 Hypothesis-Testing

Fundamental to the scientific method is the notion of an experiment. An experiment is undertaken in order to acquire data. Ideally, unambiguous conclusions can be drawn from processing of these data. Perhaps an experiment is designed to test a particular hypothesis. Alternatively, a formal hypothesis may not yet exist; the experimental setup nevertheless strives for minimisation of data interpretation ambiguity so that it can provide fertile ground for abductive reasoning to posit new hypotheses that relate cause and effect. See the discussion of Ernest Rutherford's experiment in Chapter 1.

Groundwater systems are not easy targets for experimental design. This is because it is difficult to isolate an individual component of a groundwater system and impose carefully-considered boundary conditions on it at a scale that is salient to its management. However, the decision-support context may provide the opportunity to isolate one or a small number of hypotheses that pertain to its management. As has already been discussed, management of a natural system seeks avoidance of a bad thing. This bad thing is often clearly defined; for example, the water level in a sentinel well should not fall below a certain level. This implies a hypothesis. The hypothesis is that the bad thing will happen if a contemplated course of management action proceeds. The task of decision-support modelling is to test whether this hypothesis can be rejected.

Often more than one bad thing is used to set constraints on system management. Therefore multiple hypotheses may require testing. This, in turn, may require the construction of multiple models. In the following discussion we assume only one hypothesis. The discussion is easily extended to multiple hypotheses.

Conceptually, the hypothesis that proposed management of a groundwater system will be accompanied by the occurrence of an unwanted event can be rejected if it is shown that occurrence of the event is incompatible with either site knowledge and/or with the way in which the system behaves. Site knowledge is embodied in a site conceptual model (including stochastic descriptors that should accompany it). The behaviour of a system is recorded in measurements of its state and flux, i.e. in a history-matching dataset.

Ideally, a numerical experiment can be designed to test a specific bad thing hypothesis. In doing this, it should also be capable of accommodating surprises that may trigger abductive insights that challenge the site's prevailing conceptual model. Partnered numerical simulation and high-end data assimilation can accommodate both of these experimental exigencies.

6.2 Some Repercussions

By testing one hypothesis at a time, the cognitive load of decision-support modelling is reduced while its scientific integrity is increased. Tractability of the workflow is enabled by the fact that a model does not need to simulate all aspects of system behaviour. The aim of the numerical experiment is clearly defined; aspects of simulation whose removal does not compromise this aim, can be eliminated from it.

An important repercussion of visualizing decision-support groundwater modelling as prediction-pertinent hypothesis-testing, is that the metrics by which decision-support modelling must be judged become aligned with the way in which it is undertaken. Decision-support failure occurs if a hypothesis is falsely rejected. It is therefore incumbent on a modeller to prove to regulators and stakeholders that the modelling process which he/she implements does not provide a false sense of confidence that a specified bad thing will not occur.

Because of the large uncertainties that are associated with many groundwater model predictions, it will rarely be possible to unequivocally reject a bad thing hypothesis. However it may be possible to show that its likelihood of occurrence is low. This assessment may be subjective. However, because the design of the decision-support modelling process is focussed on the testing of a specific hypothesis, the choices on which subjectivity must operate are clearly defined.

If a bad thing hypothesis cannot be rejected, this is presumably because a mechanism has been exposed for its occurrence. This can guide acquisition of further data. In the regulatory framework it can inspire the design of an adaptive management regimen that requires monitoring on the one hand, and management reaction to the crossing of monitoring thresholds on the other hand.

With the decision-support modelling process conceptualised as prediction-specific data processing, all aspects of model design have a reference point. Data interpretation requires that model history-matching be supervised by PEST/PEST++. This requires rapid simulation times and simulator numerical stability. These may place limits on the structural complexity of the simulator. For reasons that are discussed below, reduced structural complexity may instigate parameter and/or predictive bias at the same time as it enables uncertainty reduction and quantification. The reference point for resolving this issue is the prediction of management interest on which modelling is focussed. If predictive bias is small compared with predictive uncertainty reduction, then decision-support modelling objectives have been met.

Hypothesis-based decision-support modelling tends to focus on the pessimistic end of the posterior probability distribution of a decision-critical prediction. There may be no need to explore the entirety of this distribution. This may relieve a modeller of the necessity to define a comprehensive prior parameter probability distribution. Meanwhile, the emergence of unrealistic parameter values and patterns during history-matching can be used to constrain overly pessimistic predictions. It follows that the unrealistic status of emergent parameter fields must be recognisable.

6.3 Implications for Experimental Design

Where the focus of decision-support modelling is on the processing of data that informs one or a small number of decision-pertinent predictions, the decision-support modelling workflow can be optimised for this task. This reduces its chances of falling prey to either failure or uselessness – the two metrics discussed above according to which the decision-support modelling process should be judged.

Naturally, a simulator must be capable of making the prediction that is the focus of the decision-support modelling workflow. It must also be capable of representing the full prior uncertainty range of this prediction (or at least the pessimistic end of this range). Other aspects of system behaviour are of interest only insofar as their inclusion in a model allows harvesting of information that is pertinent to the prediction of interest, and transmission of that information with little loss or distortion to the prediction. Simulation of further aspects of system behaviour is either unnecessary, or can be accomplished simplistically. This provides many opportunities for model structural simplification.

Further gains in model simplification may be realised if a decision-critical prediction is data-driven. See Doherty and Moore (2021) for a discussion of these types of prediction. Under these circumstances, decision-support modelling accomplishes the same thing as prediction-specific machine learning. This can be a numerically rapid process that employs a relatively simple model.

7. Parameters

7.1 The Importance of Parameters

In the traditional decision-support modelling workflow of Figure 3.1, parameters are considered as something of an afterthought. The model structure is the important thing. Parameters are used to represent hydraulic properties of materials that occupy spaces that are defined by this structure. The model structure can demarcate units in which hydraulic properties are thought to be relatively homogeneous. In traditional workflows, an assumption of piecewise hydraulic property uniformity may be encapsulated in a parsimonious parameterisation scheme that relies on this demarcation. Claims to integrity of the decision-support modelling process are firmly rooted in resemblance of this structure to "the real thing" as it is presently conceived. This sets the stage for a modelling process which is fertile ground for confirmation bias. Doubt, the cornerstone of the scientific method, is an unwelcome guest.

We have argued above that implementation of the scientific method in decision-support modelling requires that environmental simulation be viewed as a component of a more comprehensive workflow whose primary purpose is environmental data analysis. This workflow requires that software such as PEST/PEST++ runs the simulator repeatedly according to the dictates of an algorithm that extracts information from data. Information extraction is effected through history-matching. Use of the history-matched simulator to make a decision-salient prediction assures transmission of this information to decision points.

It is a model's parameters that store information that is extracted from data, and that transmit this information to decision points. However their roles do not end there. Not only do parameters transmit information to decision points; they also transmit lack-of-information. The latter is required for quantification or predictive uncertainty.

Information must have context to be recognisable as information. When decision-specific processing of environmental data is undertaken using a numerical model, it is the structure of the model, together with the decision-salient prediction that the model is required to make, which creates this context. An inappropriate model structure may result in loss of information, or in distortion of information that is carried by parameters.

We argue below that design of a decision-support modelling workflow should focus on the roles that parameters play in this process. In doing so, it must also ensure that the model structure which plays host to parameters allows them to harvest prediction-pertinent information from data, and transmit that information with limited distortion to one or more predictions of management interest.

7.2 How Parameters Work

By definition, parameters are adjustable. They can be awarded values that allow a model to replicate measurements of system state and fluxes. Subject to the constraints of allowing a model to replicate these measurements, they can undergo stochastic sampling in order to allow a model to explore the future.

Parameters can only exist within a structure. This structure is provided by the model that hosts them. Parameters earn their names (such as "hydraulic conductivity") from this structure. It is with reference to this structure that a user attributes prior probability distributions to parameters. The latter are used in regularisation when a model undergoes calibration, and in stochastic sampling during Bayesian history-matching. In either case, these probability distributions constrain the movement of parameters as they are adjusted. However these constraints are not hard. This allows parameters to respond to history-matching surprises. Their ability to do this is essential to scientifically-based decision-support modelling.

Linear analytical methods such as singular value decomposition illuminate the way in which parameters hold and transmit both information and lack-of-information. A calibration dataset affords unique estimation of a finite number of combinations of parameters; these are often averages of native model parameters, with averaging weights implicitly defined by the measurement dataset. These combinations of parameters span a finite-dimensional subspace of parameter space wherein uncertainty is limited to that which is inherited from measurement noise. To the extent that a prediction is sensitive to combinations of parameters that occupy this subspace, it is informed by the calibration dataset; hence its uncertainty is reduced through history-matching. If a prediction is completely sensitive to occupants of this subspace, it is data-driven. We refer to this subspace of parameter space as the “solution space”.

Orthogonal to the solution space is the null space. This subspace is composed of combinations of parameters that are uninformed by the calibration dataset. The uncertainties of these combinations of parameters are inherited from the prior parameter probability distribution. To the extent that a prediction is sensitive to combinations of parameters which occupy this subspace, its uncertainty is undiminished through history-matching.

It is important to note that this subdivision of parameter space into solution and null spaces relies on combinations of parameters rather than on individual parameters. However individual parameters to which all calibration-pertinent model outputs are insensitive can reside in the null space.

The orthogonality of the null space to the solution space has important repercussions. Parameter combinations that occupy the null space are effectively “detached” from those that occupy the solution space. They can therefore wiggle independently of those that occupy the solution space, with no effect on model outputs that feature in the calibration process. This can be exploited when undertaking post-calibration uncertainty analysis (Tonkin and Doherty, 2009). It can also provide insights into how optimal model simplification should take place. White et al (2014) show that combinations of parameters that are implicitly defined when a structurally complex model is emulated by a structurally simple model should not straddle the boundary between the solution space and the null space, for this can precipitate calibration-induced predictive bias. Unfortunately, this is not always easy to avoid.

7.3 The Relationship Between Parameters and Structure

7.3.1 General Considerations

It is not possible to provide a structure for a numerical model domain without introducing assumptions. Many of these assumptions are inherited from the site conceptual model. While those who develop a site conceptual model may acknowledge that it is replete with uncertainties, the ability to express these uncertainties is lost once they are encapsulated in the structure of a numerical model, and thereby become part of its architecture. This architecture includes the numerical model’s layering and boundary conditions, as well as zonation that is designed to reflect the disposition of hydrogeological units.

Implied in a model’s structure, particularly in its layering, is the means by which measurable hydraulic properties must be upscaled for representation in a numerical model. Upscaling is not as big an issue in groundwater modelling as it is in petroleum reservoir modelling because the same attention is not given to measuring or inferring hydraulic properties from laboratory cores and geophysical logs in groundwater modelling as it is in petroleum reservoir modelling. However it does affect prior parameter probability distributions. In general, the greater is the degree of upscaling, the simpler do these prior probability distributions become because of the averaging that is implied in upscaling.

The structure of even the most complex numerical model comprises a simplified representation of reality. This, and the need for concomitant upscaling, may induce errors in model predictions. The propensity for predictive error is compounded if geological errors are hardwired into a model’s structure. This can happen if units that are represented in a geological model are not well constrained by drillhole intersections. This applies particularly to vertical features such as faults whose occurrence is somewhat random.

Boundary conditions are also represented simplistically even in complex numerical models.

7.3.2 Structure, Parameters and Predictive Bias

Model structural errors can bias model predictions in two ways. (Think of bias as unquantifiable predictive error.) We refer to the first of these was as "direct". This occurs when the structure of a model, and/or structure-required parameter upscaling, cause model-calculated system states and fluxes to differ from those which prevail in the real geological medium which the model represents.

In contrast, indirect predictive bias may be incurred when a model undergoes history-matching. The reason for this has already been discussed. Errors and simplifications that are hardwired into a model's structure may not necessarily compromise its ability to match a calibration dataset. However model parameters may need to adopt roles which compensate for model structural errors and simplifications in order to achieve this match; history-matching may therefore induce bias. Predictions that are sensitive to compensatory parameters may inherit their bias. Alternatively, they may not; see the remark on data-driven predictions above.

Because errors and simplifications in a model's structure can induce direct and indirect predictive bias, it can be concluded that the structure of a numerical model should be designed with as much care as possible. However this presents a paradox. Careful design of a model's structure may be interpreted to mean that this structure should reflect the details of "hydrogeological reality" as closely as possible. The paradox is that this is not possible. The more detailed is a model's structure, the more likely it is that unknown or uncertain aspects of hydrogeological reality are hardwired into that structure. Many structural details will almost certainly be wrong. Direct and/or indirect predictive bias follows.

Another problem that is incurred by increasing structural detail is the need to introduce more parameters to a model domain. For example, if the number of model layers is doubled, so too must be the number of model parameters. Model parameters are therefore less upscaled. This requires that they be endowed with more complicated prior probability distributions that reflect the closer relationships that these parameters bear to the complex disposition of rock types and properties that govern the subsurface movement of water. Failure to do this may compromise predictive uncertainty analysis while introducing its own predictive bias. At the same time, use of a greater number of parameters can slow the process of model calibration because of the necessity to compute a greater number of finite-difference derivatives.

It is apparent from this discussion that structural complexity may introduce bias to model predictions in a number of ways. Obviously, the same applies to excessive structural simplicity. Design of a model's structure therefore poses an optimisation problem. Solution of this problem can only be subjective, and is never entirely satisfactory. Inspiration on how to solve it must be sought from the scientific method.

8. Resolving the Paradox

8.1 Revisiting the Scientific Method

Groundwater management is an inexact science. It is a science nevertheless. Its inexactness requires that some aspects of experimental design, and of interpretation of information that is forthcoming from processing of environmental data, require subjective interpretation. Subjectivity must be informed by a mindset that is humbled by doubt, and that constantly seeks opportunities to question currently-held concepts that processing of environmental data have made vulnerable.

Implementation of the scientific method also requires that processing of environmental data render the outcomes of that processing amenable to abductive insights by modellers and their collaborators. It requires that these outcomes be projected onto a metaphorical canvass or screen (not unlike the screen of the Rutherford experiment) so that they can be inspected by those who understand the experiment, and who possess knowledge of the system.

Inherent in the traditional decision-support modelling workflow is the assumption that images which this workflow creates on the computer screen require little interpretation, for they can be construed as images of the subsurface. In contrast, the workflow that we now discuss does not make this assumption. The images that it creates on its viewing screen may not always resemble pictures from a hydrogeological textbook; sometimes they may be more akin to unexpected flashes of phosphorescence that announce an unexpected result. To use another metaphor (borrowed from Doherty, 2011), the pictures that it paints should not be viewed as representational art; they may convey deeper insights through abstraction than those which can be carried by "realistic" pictures of a misunderstood subsurface.

Another point of departure of the decision-support modelling workflow that we discuss below from the traditional decision-support modelling workflow, is its focus on parameters rather than structure. The importance of structure is not denied. However it is parameters that house information that is harvested from data, and that convey this information to decision-points. This places some important requirements on model design. The specifications of a model's structure must be such as to allow the model's parameters to play their important roles. Hence structure serves parameters, and not the other way around.

We present our alternative decision-support workflow in pictorial form in Figure 8.1. However, we warn against interpreting this figure as a flow chart, as modeller subjectivity and the exigencies of different decision-support contexts will require many variants of it. Furthermore, the narrative that supports this workflow is more important than its details.

We devote the remainder of Part 1 of this book to discussing the concepts which underpin the workflow that is pictured in Figure 8.1, and to explaining why and how it leads to superior decision-support than the workflow that is pictured in Figure 3.1.

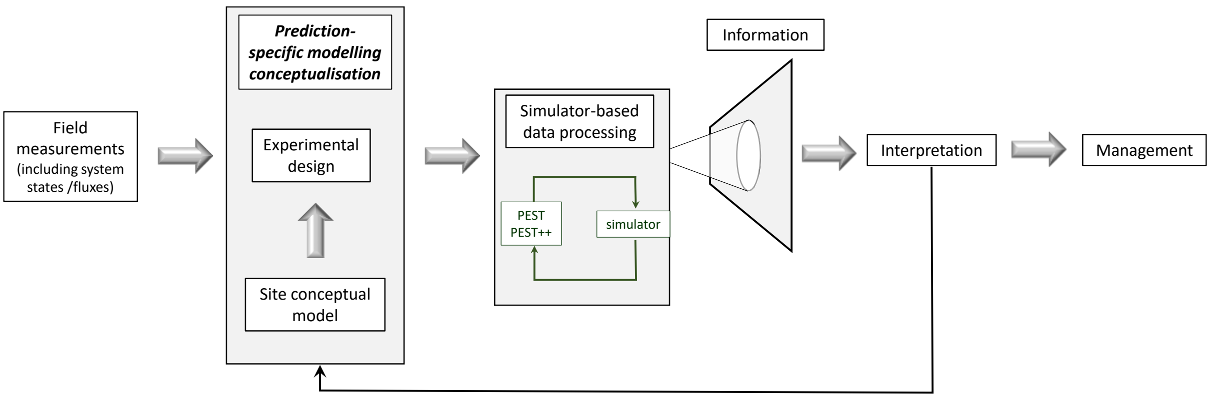

Figure 8.1 A scientifically based decision-support groundwater modelling workflow.

8.2 Structurally Simple and Parametrically Complex

The design philosophy behind the decision-support modelling workflow that is pictured in Figure 8.1 is that of structural simplicity and parametric complexity. For reasons that are now explained, this provides a practical means of implementing the scientific method in the groundwater management context, with the necessarily subjective nature of its implementation taken into account. Models with a simple structure generally run quickly, and are numerically stable. History-matching, either regularised or Bayesian, is easily undertaken.

Another benefit of structural simplicity is that a modeller does not become overwhelmed with the panoply of moving parts that beset a complex model. It is easy for a modeller to lose site of the goal of decision-support modelling when he/she is constantly distracted by the need to ensure that all of these parts maintain their movement in a coordinated way that conserves mass and preserves solution convergence under all parameterisation circumstances.

8.3 Parameterisation

Harvesting and processing of information, and quantifying the ramifications of lack of information, requires use of a large number of parameters. The reasons for this have already been explained. Ideally, any facet of a groundwater system to which a prediction is sensitive, and of which a modeller is unsure, should be represented as uncertain, and hence as a parameter. A joint prior probability distribution should be assigned to all parameters. However, as will be discussed below, a modeller should be ready to revisit this prior probability distribution if history-matching outcomes render this necessary.

8.4 Model Structure

The structure of a decision-support groundwater model must not be such as to artificially preclude predictions that may be possible according to the site conceptual model. This sets the minimum requirement for model structural complexity.

As discussed above, a simple model structure can induce direct and indirect predictive bias. In the former case, a prediction carries bias as a direct result of a model’s structure. In the latter case, bias is induced because of the compensatory roles that some parameters are forced to play during history-matching.

Ideally, design of a model’s structure should be such as to reduce both of these types of bias while recognizing that neither of them can be completely eradicated (even in a structurally complex model). The task of bias reduction is made easier where a model is focussed on only a single, or small number, of predictions; bias incurred in other predictions is irrelevant.

A model’s structure should also be complex enough for parameters to have a recognizable link with system hydraulic properties, taking the upscaled nature of those properties into account. This allows them to be assigned a prior probability distribution, including preferred values.

Representation of important geological features such as faults requires special consideration. If a modeller is sure of their locations, dimensions and throws, then he/she may represent them explicitly in a model’s structure. Alternatively, he/she may introduce special fault-aware parameters in strategic parts of a model domain whose structure otherwise ignores them. These fault-aware parameters may modify the values of other parameters so that water can move between and along model layers in a similar way to that enabled by faults. Meanwhile the prior probability distributions of these parameters may encourage the emergence of directed, connected (im)permeability. The flexibility of the latter option may assist the harvesting of information from fault-affected observations, and analysis of uncertainty resulting from a paucity of this information.

8.5 Calibration

8.5.1 Regularised Inversion

If a model's structure can be kept reasonably simple, parameters can be limited to a number (ten to fifteen thousand at most) that allows calculation of model output sensitivities using finite parameter differences. At the same time, model run times are likely to be low and model numerical stability is likely to be high. Regularised inversion then becomes easy. Meanwhile, parameter numbers should be high enough to facilitate the attainment of a good fit between model outputs and field measurements. This allows harvesting of information contained in nuances of system behaviour that cannot be reproduced by a more cumbersome, structurally complex model. The benefits of this are obvious if this information is salient to predictions of management interest. Meanwhile, variables which govern implementation of numerical regularisation can be set to preclude over-fitting.

Numerical regularisation normally seeks minimum parameter departure from a preferred set of values or conditions. These values/conditions may be illuminated by the prior parameter probability distribution, and hence by site characterisation. Meanwhile, the objective function that is minimised through regularised inversion should include components which compare appropriately-processed field measurements with their appropriately-processed model-generated counterparts.

8.5.2 Parameter Values and Patterns

If history-matching requires that parameters adopt values or patterns that are inconsistent with expectations, this can mean one of two things. Either they are adopting roles that compensate for model structural defects, or these patterns are indicative of anomalous (but real) subsurface hydraulic properties. As has already been discussed, the use of many parameters allows anomalous values and/or patterns of heterogeneity to arise in an unencumbered manner. Implementation of numerical regularisation dictates that they arise only to the minimum extent necessary for a model to replicate a calibration dataset.

There are occasions when it may be difficult for a modeller to distinguish between the above two sources of emergent heterogeneity. On other occasions the reason will be obvious. It is part of the skill of modelling to interpret anomalous parameter patterns and take appropriate action where necessary.

If the magnitudes, locations and shapes of anomalous parameter patterns suggests that parameters are responding to information contained in the calibration dataset that is reflective of unexpected subsurface conditions (which is often the case), then much has been learned from simulator-based data interpretation. These emergent parameter patterns may not depict the exact shapes of the anomalous subsurface bodies to which the inversion process is responding; where information is scarce, and where many details of subsurface hydrogeology are unknown, calibration-emergent parameter patterns that reflect these details are as much an outcome of numerical regularisation specifications as they are of subsurface hydraulic property dispositions. Nevertheless, the modeller (and modelling stakeholders) have learned something that they did not know before. Conclusions may be far-reaching. For example, if unexpected hydraulic property heterogeneity has been detected in one part of a study area, perhaps it exists in other parts of the same study area, but remains undetected because of data scarcity. Analysis of the uncertainties of decision-critical model predictions may then require that the prior probability distribution of model parameters be modified to accommodate this.

On the other hand, if unexpected parameter patterns are thought to reflect model structural inadequacies, a modeller has a number of choices.

One option is to modify the model's structure (for example by adding another layer, or by introducing/removing a boundary condition) in order to mitigate the need for calibration-induced parameter compensatory behaviour. Emergent patterns of parameter heterogeneity often suggest the locations and nature of alterations to a model's structure that are required in order to achieve this. Meanwhile, history-matching has demonstrated the salience of at least one aspect of a model's structure to the behaviour of the groundwater system. To the extent that details of this structure are imperfectly known, this may have a bearing on predictive uncertainty.

A second option is to acknowledge that the model's structure requires that parameters adopt compensatory roles in order to allow model outputs to fit field measurements. If the prediction required of a model is similar in some respects to measurements comprising the calibration dataset (if, for example, it is of the same type but is required at a different place and/or at a later time), then a modeller may accept the necessity for calibration-induced parameter compensation, and adjust the prior parameter probability distribution to accommodate this. This adjusted prior parameter probability distribution may embody more variability than that which is expected to prevail in hydraulic property types after which parameters are named. While abstraction of parameters in this manner separates a model to some extent from its physical basis, the alternative strategy of altering the model's structure may have worse predictive consequences if specifications for these alterations are unclear. Recall that structure is fixed and therefore cannot express uncertainty. Furthermore, the addition of extra structure requires the addition of extra parameters; assignment of a prior probability distribution to these extra parameters may need to respect their less upscaled status; see above. In the end, what matters most is that the posterior probability distributions of predictions of management interest are not understated while not being unnecessarily large. A modeller must decide whether this purpose is best served by retaining the current model structure and accommodating parameter abstraction, or by complexifying model structure in ways that may not be fully informed by the current site conceptual model.

A third option is to modify the calibration dataset. A modeller may judge that certain components of a calibration dataset contain little information that is pertinent to the decision-salient prediction that his/her model is required to make, but that they create the need for compensatory parameter behaviour. These components can be removed from the calibration dataset. If the modelling workflow is properly formulated, the loss of any prediction-pertinent information that they contain will be reflected in higher posterior predictive uncertainties. This may be a small price to pay for the removal of potential predictive bias. At the same time, information-rich signals contained in other components of a calibration dataset may be rendered more visible by appropriate processing of that dataset before matching them to similarly processed model outputs, and then weighting them for visibility in the overall objective function.

For example, in some modelling circumstances it is beneficial to include temporal changes in heads over time as a significant component of a calibration dataset, while omitting the heads themselves or giving them only a small weight. This may enhance a model's ability to predict future transient behaviour of the system. The withdrawal or de-weighting of absolute head measurements at certain locations within a model domain may increase uncertainties in local hydraulic conductivity parameters. This may be justified if predictions of interest are relatively insensitive to these parameters, or if the calibration process may induce bias in these parameters because of apparent errors in neighbouring boundary conditions for which the site conceptual model can provide little explanation.

A fourth option is to do nothing. There may be two reasons for this. If a prediction is very similar in nature and location to observations which comprise the calibration dataset, and if system stresses in the future will not be too different from those which prevailed in the past, then the prediction is probably data-driven. (That is, the prediction is entirely solution space dependent). Under these circumstances it can be shown that calibration-induced parameter bias does not inculcate predictive bias. The past is therefore the key to the present. All that is required for a relatively accurate prediction of the future is that the model is capable of accurately replicating the past. See Doherty and Moore (2021) for details.

Another reason for doing nothing is if parameter compensatory behaviour occurs in one part of a model domain but decision-salient model predictions are required in other parts of the model domain and have limited sensitivity to compensatory parameters. In this case, compensatory parameters may be seen as comprising a distributed boundary condition. They become "fitting parameters" which soak up information contained within a calibration dataset which would otherwise have nowhere to go.

8.5.3 The Scientific Method

Model calibration, conducted in ways that are described above, bears much similarity to a scientific experiment. While the natural environment cannot be controlled, considerable control can be exercised over simulator-based processing of environmental data in order to optimise the harvesting of decision-pertinent information contained therein, while filtering out signals that may contaminate that information. Model calibration paints the information which it harvests from a calibration dataset on the screen/canvass of parameter space. As a scientist, a modeller interprets the pictures that emerge in this space, taking into account the experimental setup that produced them. He/she can then draw conclusions, and take appropriate action.

8.6 Linear Analysis

Post-calibration linear analysis is briefly described in Section 5.5.3. It draws on the Jacobian matrix which is calculated as a by-product of regularised inversion. Linear analysis can be turned to the production of maps which characterise the distribution of post-calibration parameter uncertainty throughout a model domain. These maps may sometimes demonstrate that areas of low uncertainty coincide with areas of anomalous hydraulic properties. This is partly because numerical regularisation suppresses the emergence of anomalous heterogeneity unless a good fit with the calibration dataset cannot be attained without it. Similarly, these maps can identify those parts of a model domain where hydraulic properties are uninformed by measurements of system behaviour.

Nicol and Doherty (2020) show how linear analysis can be used to investigate whether model structural defects may have biased decision-salient model predictions. Such an analysis (which is easily performed) can assure modelling stakeholders that model structural simplifications which have enabled implementation of scientifically-based environmental data processing do not invalidate the decision-support integrity of a model.

8.7 Nonlinear Predictive Uncertainty Analysis

Calibration-constrained Monte Carlo analysis requires that a multiplicity of parameter fields be generated, all of which allow model outputs to replicate field measurements (to within a user-specified tolerance), and all of which comply with the prior parameter probability distribution. As has already been discussed, the PESTPP-IES ensemble smoother is capable of delivering these fields with a high level of model run efficiency.

It can be argued that Bayesian methodologies such as that implemented by the PESTPP-IES ensemble smoother make model calibration redundant. Their task is to sample the posterior parameter probability distribution. Model predictions can then be made with thus-sampled parameter fields. Prediction statistics can thereby be calculated. The prediction of minimised error variance provided by the calibrated parameter field is no longer needed. We suggest, however, that science-based decision-support often benefits if nonlinear uncertainty analysis is preceded by model calibration. This follows from the resemblance that calibration bears to a scientific experiment, and because calibration may influence analysis of prior and posterior uncertainty.

For example, as has already been discussed, lessons learned from calibration of a structurally simple, parametrically complex model may induce a modeller to modify the prior probability distribution of model parameters. This modified distribution can then be employed in subsequent Bayesian analysis. If posterior uncertainty is explored using an ensemble smoother, samples of this modified prior parameter probability distribution are history-match-adjusted until they become samples of the posterior parameter probability distribution.

As another example, calibration performed using regularised inversion reveals the level of fit that can be expected with a measurement dataset. Using pertinent regularisation levers, a modeller may temper this fit in order to prevent the emergence of distasteful parameter fields. In doing so, he/she ascertains the effective level of “noise” to associate with field measurements during subsequent Bayesian analysis.

It is therefore apparent that the two terms on the right of Bayes equation (equation 4.1) can benefit from lessons learned through a calibration process which precedes Bayesian uncertainty analysis.

This loss of strict “Bayesian purity” is not something that should be lamented, nor used to criticise a decision-support modelling workflow that implements it. This is because Bayesian purity is not possible in decision-support groundwater modelling anyway. Prior probability distributions of parameters which represent, in an upscaled manner, the complex heterogeneity that prevails in the shallow subsurface can only be characterised in a very approximate manner. Furthermore, geologically realistic patterns of connected permeability are difficult for conventional probability distributions to express, and even more difficult to maintain as they are adjusted during history-matching.

At the same time, model-to-measurement misfit is always dominated by structural noise rather than by measurement noise; structural noise cannot be characterised statistically (Doherty, 2015). This further compromises the “Bayesian integrity” of nonlinear posterior uncertainty analysis. This does not undermine its utility; however, Bayesian methods need all the help that they can get.

9. Direct Predictive Hypothesis Testing

9.1 General

A methodology for model-based decision-support is now described that attempts direct implementation of the scientific method. Literature-documented applications of this methodology are rare, but not absent; see Siade et al (2015) and Moore et al (2010). Rarity of application can be partly attributed to the groundwater industry's infatuation with the workflow of Figure 3.1. However, it can also be attributed to numerical difficulties that sometimes attend its implementation; see below. The methodology applies principles that are encapsulated in the workflow that is depicted in Figure 8.3.

9.2 Implementation Philosophy

9.2.1 The Hypothesis

As has been remarked previously, environmental management presents the decision-support modelling process with ready-made hypotheses. These are that one or a number of bad things will accompany a particular course of management action.

The hypothesis that a bad thing will occur can be rejected if its occurrence is demonstrably incompatible with one or more of the following:

- the principles of groundwater flow and contaminant transport;
- knowledge of the system that has emerged from site characterisation; and
- measurements of present/historical system behaviour.

9.2.2 Lessons Learned from Model Calibration

Suppose that a modeller has just calibrated a structurally simple, parametrically complex groundwater model against present/historical measurements of system behaviour. The cognitive outcomes of this process are multifaceted. The modeller has learned that it is possible to replicate (appropriately processed) field measurements with an acceptable level of model-to-measurement misfit. This fit is probably better than could have been achieved using a more complex model for reasons outlined above. However it will normally be greater than that which would be anticipated from measurement noise alone. The "structural noise" which normally dominates model-to-measurement misfit arises from an inability of numerical simulation to replicate all nuances of a system's behaviour.

Regularised inversion introduces patterns of parameter heterogeneity to a model domain. The values assigned to parameters, and the patterns that they exhibit, may not be entirely in accordance with expectations born of the prior parameter probability distribution. Nevertheless, their existence is integral to attainment of an acceptable fit with the calibration dataset. Regularisation ensures that contraventions of the prior parameter probability distribution are no larger than they need to be. The calibration-emergent parameter field thus expresses the heterogeneity that must exist to explain the behaviour of the system. For reasons that are described above, a modeller may decide to accept these values and patterns, in spite of the possibility that some of them may be an outcome of adoption by parameters of roles that compensate for model structural inadequacies.

It follows from the above considerations that regularised inversion has not only provided a modeller with a parameter field of minimised departure from a prior parameter reference field. It has also given the modeller the opportunity to decide on the level of model-to-measurement misfit that he/she can tolerate. It has done this while providing him/her with the opportunity to decide on the values and patterns of parameters that are tolerable outcomes of history-matching.

(It is worth noting that the exercise of subjectivity in acceptance of model-to-measurement misfit, and in acceptance of the calibrated parameter field, is unavoidable. It is a consequence of applying the scientific method to an inexact science. This does not diminish the importance of the scientific method. In fact, it makes its adoption as a philosophy for model-based data processing even more pressing.)

9.2.3 Testing the Hypothesis

Suppose now that the calibrated decision-support model is used to make a prediction. Decision-support integrity demands that the uncertainty of this prediction be analysed. Bayesian methods, applied in ways that have already been discussed, can be turned to this task.

An alternative way of learning about the pessimistic end of the posterior probability distribution of a prediction of management interest, is to treat an unwelcome value of this prediction as a hypothesis. This requires that the calibration dataset be supplemented with a single observation, this being the observation that a hypothesised bad thing actually occurs. The model is then re-calibrated under a regularisation regime that seeks minimum departure from the previously calibrated parameter field. If inclusion of the hypothesised prediction in an expanded calibration dataset incurs too much misfit with the historical component of the calibration dataset, and/or if heterogeneity that emerges in the updated calibrated parameter field is considered to be excessive or unrealistic, the hypothesis of occurrence of the bad thing can be rejected.

A difficulty in implementing this procedure is that of identifying "too much" and "excessive" when assessing calibration-emergent model-to-measurement misfit and parameter values and patterns. These are subjective choices. However reference misfits, and reference parameter patterns, are available from the previous calibration process. The modeller's decision to accept the outcomes of the previous calibration process implies a decision to accept model-to-measurement misfit and parameter patterns that emerged from that process. The modeller must now decide whether a "line has been crossed" when deciding whether he/she should accept model-to-measurement misfit and parameter patterns that emerge from direct predictive hypothesis testing. Meanwhile the use of regularised inversion as a basis for predictive hypothesis-testing guarantees that model-to-measurement misfit, and parameter departures from those attained during the previous calibration process, are only as great as they need to be.

9.2.4 Parameters Again

It is important to note that the direct predictive hypothesis-testing methodology suffers a loss of integrity if it is not undertaken in a highly-parameterised context. Parameters must be free to establish the patterns and values that are necessary for the occurrence of a bad thing. Introduction of connected permeability to a parameter field must not be precluded by a lack of parameters in critical parts of a model domain. If unrealistic connectivity, or unacceptably high or low parameter values emerge, a modeller can feel safe in rejecting the tested hypothesis. However integrity of the methodology rests on the premise that it is the modeller, and not parameter insufficiency, that forms the basis for hypothesis rejection.

9.3 Some Benefits

The hypothesis-testing methodology described above acknowledges the fact that numerical models can neither simulate the behaviour, nor represent the components, of a natural groundwater system very well. However it also acknowledges the fact that some measurements of system behaviour may be rich in management-salient information, and that replicating these aspects of its behaviour is therefore worthwhile. This may induce some parameters to adopt roles that compensate for model structural defects. However if history-matching is conducted in a controlled numerical environment that maximises the ability of a modeller to make necessarily subjective decisions about how best to implement it, this information can be harvested and directed to where it can do the most good.

Direct predictive hypothesis-testing implements the Popperian view of the scientific method. It is not Bayesian, because it allows re-definition of criteria for model-to-measurement misfit and parameter field acceptability in light of lessons learned from model-based experimentation. Nor is it formulaic, because modeller discretion is required. It seeks to facilitate the exercise of this discretion, while pursuing philosophical alignment with the scientific method.

An ancillary benefit of direct predictive hypothesis-testing is that, in exposing the possibility of an unwanted occurrence (i.e. a bad thing), it also exposes a mechanism for its occurrence – for example the existence of connected permeability that is undetectable by the current network of observation wells. This may suggest locations for additional monitoring wells. These may be installed either before proposed management plans are approved, or as an important component of adaptive management following acceptance of a proposed plan.

9.4 Some Drawbacks

The methodology outlined above allows the testing of only one hypothesis at a time; that is, it allows a modeller to define one particular value of a management-salient prediction, and to then test the hypothesis that this particular value eventuates. However, a site manager is generally more interested in a continuum of values, as he/she would like to know how quickly a prediction becomes unlikely as its value becomes more extreme.

At the time of writing, both PEST_HP and members of the PEST++ suite provide functionality for continuous appraisal of a prediction as it gets more and more undesirable. However further software development is required in order to increase numerical efficiency.

A more pressing problem besets use of the direct predictive hypothesis-testing methodology in an adaptive management framework. When deployed in this context, a model must accomplish more than simulation of groundwater flow. It must also simulate groundwater management. For example, it may simulate the fact that pumping from an extraction well is reduced if the head in a monitoring well falls below a certain level. The programming required to achieve this is not difficult. However the outputs of models which manage themselves are not continuously differentiable with respect to their parameters. This may compromise the ability of regularised inversion to fit a calibration dataset that includes a hypothesised prediction under an adaptive management scenario.

10. Multiple Conceptual Models

10.1 General

In the groundwater literature, the use of multiple conceptual models is a topic whose popularity seems to ebb and flow. The same applies to related topics such as Bayesian model averaging. Most applications are academic.

Before completing Part 1 of this book, we offer a short critique of this methodology, as it is salient to the book's subject matter.

10.2 Brief Description

The multiple conceptual model methodology recognises that there is much about the subsurface that remains unknown after site characterisation studies have been completed. Therefore, the outcomes of these studies must be probabilistic. As has already been mentioned, proponents of the multiple conceptual model approach recommend that a suite of numerical models be built that are "mutually exclusive and collectively exhaustive".

In scientific papers that discuss the multiple conceptual model methodology, models that are used to exemplify this methodology tend to be parametrically simple. The different structural concepts to which they give numerical incarnation sometimes comprise alternative dispositions of a limited number of zones of assumed parameter constancy. For reasons that have already been discussed, their parsimonious parameterisation schemes equip these models with limited capacity to respond to information that is resident in measurements of system behaviour. Difficulties in information harvesting are exacerbated by the fact that differences between their structures are discontinuous. Furthermore, the number of conceptual models that are tested is usually limited to just a few. Under these circumstances, claims to being mutual exclusive and collectively exhaustive are rather hard to justify.

In most implementations of the methodology, each model is individually calibrated. Some models may be rejected because parameter values cannot be found that allow model outputs to replicate historical conditions. Other models may also attain a mediocre fit with field measurements but are nevertheless retained. Ranking of retained models is based on model-to-measurement fit; sometimes parameter parsimony is also rewarded in statistics that are used for comparative model ranking.

Predictions of management interest are made using all retained models. If a single predictive value is required, this is calculated as a weighted combination of predictions made by individual models; greater weights are awarded to models which achieve a better fit with the calibration dataset with a reduced number of parameters. Variation of predictions between models is taken as a measure of predictive uncertainty. In some implementations of the method, individual models are subjected to calibration-constrained uncertainty analysis so that within-model predictive uncertainty can be added to between-model predictive uncertainty. However within-model uncertainty is likely to be underestimated because of parameter parsimony.

10.3 Brief Critique

10.3.1 Review of Decision-Support Modelling Requirements

Decision-support modelling is never satisfactory; compromises must always be sought. The best compromise for one situation may differ from that which is best for another situation. However, regardless of its context, the task of decision-support modelling is the extraction of information from data, and the transmission of that information to decision points. Modelling must also convey lack-of-information to decision-points so that this can be expressed as predictive uncertainty.

The present document stresses the importance of parameters to the decision-support modelling process, for these are the bearers of information. Parameters should also carry the stochastic load of lack-of-information where data is scarce. By definition, parameters are "born to wiggle".

As has been discussed in previous sections, the use of parameters as the principal conveyors of both information and stochasticity comes at a cost when this is enabled by structural simplicity (as it often must be). The cost is simulation abstraction. Parameters may need to represent upscaled hydraulic properties rather than those that can be measured in a laboratory or through aquifer testing. During history-matching, some parameters may adopt roles that compensate for simplifications in a model's structure. However, as long as the model's structure is not too simple, both of these can be accommodated. Meanwhile a modelling workflow can be developed that is tuned to the harvesting of information that is pertinent to a particular prediction. Good fits can then be attained with field measurements, or at least with those aspects of field measurements that a modeller deems worthy of fitting. If necessary, predictive uncertainty can be quantified with the compensatory roles of parameters taken into account.

10.3.2 Why the Multi-Model Approach Falls Short of these Requirements

The philosophy behind the alternative conceptual model methodology is very different from this. It requires that model structure bears much of the stochastic load. A benefit of this strategy (possibly its only benefit) is that pictures of the numerical model may resemble pictures of a geological model.

In documented use of the multiple conceptual model methodology, alternative concepts that are embodied in alternative numerical models are limited to just a few. These concepts are purposely very different. Stochasticity is therefore discontinuous and low-dimensional. The former characteristic impedes the search for a good model fit with a measurement dataset; important prediction-salient information may therefore go unharvested. The second characteristic may result in significant underestimation of predictive uncertainty. Together, they provide an inappropriate basis for processing of data pertaining to a complex system with many dimensions of uncertainty.

If the structure of a particular member of a multi-model suite is inappropriate, this is identified through failure to attain a good fit between model outputs and field measurements when that model is subjected to calibration. However where each member of the suite of models is equipped with a parsimonious parameter field (often based on zones of assumed piecewise constancy) it is possible that failure to parameterise within-zone heterogeneity may be the real reason for failure to attain a satisfactory level of model-to-measurement fit. This raises the possibility that a valid conceptual model may be falsely rejected. In practice, this possibility is circumvented by acceptance of a mediocre fit between model outputs and field measurements under the premise that a better fit would constitute over-fitting. This argument has little merit when the reason for model-to-measurement misfit is manufactured.

History-matching problems are compounded by the fact that parsimonious parameterisation limits the ability of any member of a model suite to quantify uncertainties that result from its inability to harvest information from data. Use of the multiple conceptual model methodology rests on the premise that predictive uncertainties can be calculated from inter-model predictive variability. However, there is no proof that this is the case. Furthermore, it is far from impossible that if any single member of the model suite were endowed with many parameters, then the uncertainty quantified by this model alone would be greater than that quantified by the suite of parsimoniously parameterised models collectively.

If, instead of a multi-model approach, a modeller was to adopt the approach to decision-support modelling that is recommended herein, inappropriateness of model structure would be expressed as either model-to-measurement misfit, or as the emergence of unlikely parameter values and patterns. The latter outcome normally permits ready identification of causative model structural inadequacies. As has been discussed, these can be rectified if a modeller deems this to be necessary. In contrast, the multi-model approach forces structural inadequacies to be expressed as model-to-measurement misfit alone. Identification of structural deficiencies from patterns of model-to-measurement misfit is often much harder than identification of structural deficiencies from calibration-emergent parameter patterns. In practice, however, the multi-model approach does not seek a reason for model-to-measurement misfit; the model is simply rejected (or given a low ranking) under the premise that another member of the model suite is more likely to possess the “correct” structure. This assumption is often baseless.

Finally, the multi-model approach removes from the history-matching process the ability of a modeller to gain abductive insights that may lead him/her to revise system concepts in previously unforeseen ways. It presents no screen or canvass onto which data can paint information. The information content of data is constrained to express itself in a small number of ways that are designed ahead of the history-matching process.

For all of the reasons presented above, it is difficult to view the multiple conceptual model approach, as commonly presented in the literature, as an implementation of the scientific method. Nor is it possible to view this approach as being capable of providing robust decision-support.

10.4 A Better Way to Deploy Multiple Models

It cannot be denied that there will be occasions when site uncertainties are such as to warrant consideration of entirely different numerical model structures. This will be required where parameters alone (even when defined in ways that allow them to emulate the effects of geological structures) cannot express the full potential for predictive variability that arises from hydrogeological conceptual uncertainty.

Where alternative numerical models are required, then, in accordance with advice provided herein, these models should (if possible) be structurally simple and parametrically complex. Each should be designed with the same specific prediction in mind. Each should be calibrated, and then subjected to nonlinear uncertainty analysis and/or direct predictive hypothesis testing. Limitations of one or more of these alternative models may become apparent during these processes; these model(s) can then be abandoned. Alternatively, it may be discovered that a predictive hypothesis that can be rejected using one of the retained models cannot be rejected using another. The hypothesis therefore remains unrejected. The possibility of false rejection of a bad thing hypothesis is therefore reduced, as is modelling failure according to the decision-support metric presented earlier in this document.

11. Conclusions

11.1 The Missing Ingredient

This book addresses the issue of decision-support modelling conceptualisation. This issue includes, but is not limited to, the relationship between a site conceptual model and the numerical model that it inspires.

A vital element is lacking in traditional decision-support modelling workflows. This element (together with a term to describe it) should be placed between the site conceptual model and the numerical model. It should specify how a numerical model, deployed in conjunction with a data assimilation package such as PEST/PEST++, can be used to formulate a scientific experiment that:

1. converts data into prediction-pertinent information through history-matching; and
2. delivers that information to one or a number of people who can interpret it in order to formulate appropriate groundwater management strategies.

Use of information that is harvested by scientifically-based decision-support modelling may vary from context to context. Information may be used to assess the risks associated with a contemplated course of groundwater management. Alternatively (or as well), it may precipitate a lightbulb moment in which current understandings of a groundwater system are changed forever.

The scientifically-based decision-support modelling workflow that is recommended herein is depicted figuratively in Figure 8.1. We use the term “figuratively” because this diagram describes a philosophy as much as it does a workflow; incarnation of the philosophy as a workflow will differ from site to site.

Figure 8.1 includes a component that is labelled “experimental design”; this is the ingredient that is missing from traditional groundwater modelling workflows.

The modelling process itself is depicted as a numerical activity in which simulation and data processing are inextricably linked; we denote this process as “simulator-based data processing”. The outcome of this numerical activity is depicted as information projected onto a screen or canvass. Interpretation of that information is undertaken by those who designed the numerical experiment; it is their job to decode its meaning so that it can be clearly comprehended by decision-makers and stakeholders.

11.2 Implications

Figure 8.1 makes it clear that the information outcomes of model-based environmental data processing may require expertise to interpret. This has some important implications. The first of these implications is that design and implementation of a model-based, environmental data processing workflow requires expertise; as in all other branches of science, this is acquired through education and training. A second implication is that there is no reason to expect that the products of a decision-support modelling workflow should comprise picture-perfect views of the subsurface at future times. The product is information; this can be used by decision-makers in ways that they see fit.

In contrast, the traditional decision-support modelling workflow that is depicted in Figure 3.1 delivers a stand-alone numerical model. Supposedly, stored in this numerical model, is all information that can be gleaned from site conceptualisation studies and measurements of system behaviour. This model can be used to generate pictures of the future state of a groundwater system under a variety of management conditions; these pictures, it is hoped, can be supplemented by uncertainty bounds that span the full range of predictive possibilities.

We believe that this endeavour is destined for failure. Science is more subtle than this; natural systems are more complex than this; harvesting and transmission of information requires more planning than this.

11.3 The Importance of Experimental Design

Figure 8.1 makes it clear that a heavy conceptualisation workload must precede construction and deployment of a numerical model. The importance of site conceptualisation has long been recognised. Unfortunately, the same does not apply to the importance of experimental design. In the traditional decision-support modelling workflow it is omitted under the premise that the sole purpose of simulation is to give numerical voice to the site conceptual model. In contrast, experimental design is key to implementation of the workflow that is depicted in Figure 8.1. It is this aspect of the workflow, more than any other, on which its claim to implementation of the scientific method rests.

Inclusion in Figure 8.1 of experimental design and site conceptual model development as components of an overarching conceptualisation activity acknowledges the close connection between the two. A high degree of familiarity with current site concepts is required for formulation of experiments to test these concepts, and to understand the extent to which processing of field measurements of system behaviour may require revision of these concepts. A high degree of familiarity with current site concepts is also a prerequisite for formulation of an experiment to test a sharply focussed hypothesis that is salient to management of a system. Experimental design requires that a modeller regard the study site, and what is known about the study site, not just from the perspective of the past, but also from the perspective of the future. Numerical modelling can then link the two.

11.4 Modelling, Modellers and Science

The scientifically-based decision-support modelling workflow of Figure 8.1 blurs boundaries that are implied in the traditional modelling workflow of Figure 3.1. Model development and model calibration are not seen as separate activities; nor are simulation and uncertainty analysis. Nor is it required that site conceptual model development entirely precede, and be completely independent of, numerical model development. Feedback loops are everywhere. Numerical modelling must be conducted in a way that welcomes, rather than eschews, conceptual surprises. Recognition that some outcomes of simulator-based groundwater data processing may, in fact, be surprising requires a thorough understanding not just of numerical modelling and simulator-based data processing, but of the current site conceptual model. Opportunities that the modelling process may offer for abductive leaps in understanding will not be realised if a modeller’s comprehension of the simulated system is poor.

Importantly, decision-support modelling must be seen as an activity rather than as a deliverable. This activity should be undertaken by personnel who are humbled by the complexity of the natural system whose management they are supporting, and who are committed to uncovering some if its secrets by innovative processing of its measured behaviour.

The complexity of natural systems exceeds the possibility of digital reproduction. This renders the traditional decision-support modelling workflow of Figure 3.1 unsatisfactory. However, scientifically-based numerical experiments may enable isolation and study of those aspects of a system’s behaviour that are salient to its future management. This process necessarily has a large subjective component at the same time as it has a significant numerical component. It should be designed in such a way that the two components complement each other. Uncertainty accompanies every step. However realistic assessment of uncertainty, and formulation of management plans that can accommodate uncertainty, depend on the scientific integrity of this endeavour.

12. References

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

Barnett et al, 2012. Australian Groundwater Modelling Guidelines, Waterlines Report, National Water Commission, Canberra.

Bredehoeft, J., 2005. The conceptual model problem – surprise. *Groundwater*, 13: 37-46.

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

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

Doherty, J., 2011. Modeling: picture perfect or abstract art? *Ground Water*, 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. and Moore, C., 2019. Decision Support Modeling: Data Assimilation, Uncertainty Quantification and Strategic Abstraction. *Groundwater*, 58(3) 327-337 doi: 10.1111/gwat.12969.

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. DOI: 10.25957/p25g-0f58.

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 Resour. Res*., 46, W05525, doi:10.1029/2009WR008377.

Enemark, T., Peeters, L.J.M., Batelaan, O. and Mallants, D., 2018. Hydrogeological conceptual model building and testing: a review. *Journal of Hydrology*, 569, 310-329. https://doi.org/10.1016/j.jhydrol.2018.12.007.s

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

Hill, M.C. and Tiedeman, C.R., 2007. Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions and Uncertainty. John Wiley and Sons.

Kuhn, T.S., 1962. The Structure of Scientific Revolutions, Univ. of Chicago Press, Chicago.

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

Moore, C., Wöhling, T., and Doherty, J., 2010. Efficient regularization and uncertainty analysis using a global optimization methodology. *Water Resources Research.* Vol 46, W08527, doi:10.1029/2009WR008627.

Nicol, C. and Doherty J., 2020. Exploring Model Defects Using Linear Analysis: A GMDSI Worked Example Report. National Centre for Groundwater Research and Training, Flinders University, South Australia.

Popper, K. R., 1959. The Logic of Scientific Discovery, Basic Books, New York.

Popper, K. R., 1963. Conjectures and Refutations: The Growth of Scientific Knowledge, Routledge, London.

Quine, W. V. O., 1951. Two dogmas of empiricism. *Philos. Rev.*, 60, 20–43.

Refsgaard J.C., Christensen, S., Sonnenborg, T.O., Dorte S., Højberg, A.L., and Troldborg, L., 2012. Review of strategies for handling geological uncertainty in groundwater flow and transport modelling. *Water Resour. Res*., 36, 36-50.

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

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S., 2008. Global Sensitivity Analysis : the Primer. John Wiley.

Siade, A., Nishikawa, T. and Martin, P., 2015. Natural recharge estimation and uncertainty analysis of an adjudicated groundwater basin using a regional-scale flow and subsidence model (Antelope Valley, California, USA). *Hydrogeology Journal* 23:1267-1291.

Tonkin M., J. and Doherty, J., 2009. Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques. *Water Resour. Res*., 45, W00B10, doi:10.1029/2007WR006678.

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