GROUNDWATER MODELLING UNCERTAINTY

IMPLICATIONS FOR DECISION-MAKING

NATIONAL CENTRE FOR GROUNDWATER RESEARCH AND TRAINING

iah aih

IAH AUSTRALIA

Flinders UNIVERSITY

Summary of the Groundwater Modelling Uncertainty Workshop - Australasian Groundwater Conference

10th July 2017, University of NSW, Sydney, Australia

Hugh Middlemis, Glen Walker, Luk Peeters, Stuart Richardson, Phil Hayes and Catherine Moore

Published by:	Flinders University - The National Centre for Groundwater Research and Training (NCGRT) GPO Box 2100, Adelaide SA 5000 +61 8 8201 2193 groundwater.com.au In collaboration with the International Association of Hydrogeologists PO Box 7634, Cloisters Square WA 6850
Authors:	Hugh Middlemis (HydroGeoLogic), hugh@hydrogeologic.com.au Glen Walker (Grounded in Water), Glen.Walker@internode.on.net Luk Peeters (CSIRO), Luk.Peeters@CSIRO.au Stuart Richardson (CDM Smith), RichardsonSB@cdmsmith.com Phil Hayes (Jacobs, now University of Queensland / Hayes GeoScience), Phil.Hayes@uq.edu.au Catherine Moore (GNS Science, now CSIRO), Catherine.Moore@CSIRO.au
Copyright:	
Citation:	For bibliographic purposes this report may be cited as: Middlemis H, Walker G, Peeters L, Richardson S, Hayes P, Moore C. 2019. *Groundwater modelling uncertainty – implications for decision making*. Summary report of the national groundwater modelling uncertainty workshop, 10 July 2017, Sydney, Australia. Flinders University, National Centre for Groundwater Research and Training, Australia. ISBN: 978-0-646-98608-1 DOI: https://doi.org/10.25957/5ca5641defe56

Acknowledgements

The work detailed in this report would not have been possible without substantial contributions from:

- Dr John Doherty (Watermark Numerical Computing), Dr Miriam McMillan (Office of Water Science), Dr Dougal Currie (CDM Smith) and Mr Chris Nicol (Groundwater Logic) for their co-authoring of the discussion papers prepared for the workshop process;
- Dr Peter Baker and Mr John Higgins (Office of Water Science) for their contributions to the workshop and the subsequent plenary at the Australasian Groundwater Conference, Sydney NSW, July 2017;
- Prof Craig Simmons (Director NCGRT), and the support team at NCGRT: Fiona Adamson and Andrea McCarthy;
- Dr Noel Merrick, Dr Ilka Wallis, Dr Juliette Woods, Prof Tony Jakeman, Dr Hashim Carey, Dr Kevin Hayley and Dr John Doherty who provided peer review and feedback.

FOREWORD

National Groundwater Modelling Uncertainty Workshop Notes, July 10, 2017, Sydney.

IAH and NCGRT are delighted to present these workshop notes from the National Groundwater Modelling Uncertainty Workshop, held prior to the 2017 IAH-NCGRT Australasian Groundwater Conference in Sydney.

Quantifying and communicating uncertainty in groundwater assessments, underpinned by groundwater modelling, is a major technical, management, policy and regulatory challenge. These notes are a summary of the workshop and the associated background papers that were presented at the workshop. They are not guidelines.

We believe that there is a need for improved treatment of hydrogeological uncertainty and that we need to explore methods for dealing with uncertainty in groundwater modelling. There is also a very strong appetite to discuss and debate this – the very reason for this foundational workshop and notes. The collective experience and expertise at the workshop represented a wide range of groups, including model users, model developers, government agencies, consultants, researchers and industry.

The workshop tackled key questions such as:

What are current practices and procedures for uncertainty analysis nationally and internationally?
Why is uncertainty analysis not ubiquitous already? What are the major impediments for uptake of improved practice?
How do we best communicate uncertainty, such that model results are interpreted appropriately with respect to implications for policy makers, approval authorities, proponents and the public?
How do we employ uncertainty analysis as part of a risk-based decision-making framework?
Are there gaps between uncertainty analysis research and practice, and if so, how do we close these gaps?

We are delighted that Hugh Middlemis (HydroGeoLogic) has led a team of experts to put this workshop and workshop notes together. We thank Hugh and the team for their hard work and leadership. The topics and leaders were:

- Problem Definition – Dr Glen Walker (Grounded in Water)
- Integrating Uncertainty with Model Workflows – Dr Luk Peeters (CSIRO)
- Effective Model Simplification – Dr Catherine Moore (GNS, New Zealand)
- Uncertainty Evaluation Methods – Dr Phil Hayes (UQ / Hayes GeoScience)
- Communicating Uncertainty – Stuart Richardson (CDM Smith)

The authors revised and reviewed these papers after the workshop. We have collated them here to produce a citable report. This workshop and workshop notes represent a part of an important discussion and a starting point for ongoing discussion and debate.

Further work includes practical worked examples and case studies, software development, making analysis tools accessible, textbooks, instruction manuals, education and training. The Menindee case study (Uncertainty Evaluation Methods) is a great example, and we are very grateful to Water NSW and Geoscience Australia for making it available. It shows that, even with the most comprehensive and complex data and knowledge base available, an overly complex model is not necessarily the best approach to investigate the effects of uncertainties on project objectives.

The field of uncertainty analysis is developing rapidly. We hope these workshop notes are a useful reference for the groundwater community and will inspire future work towards improved uncertainty analysis and communication in our industry.

Professor Craig T. Simmons
Director, NCGRT

Dr Ian Brandes de Roos
President, IAH Australia

EXECUTIVE SUMMARY

This report provides a summary of the outcomes from the 2017 national groundwater modelling uncertainty workshop convened by the National Centre for Groundwater Research and Training (NCGRT), and the subsequent modelling outlook panel session held at the Australasian Groundwater Conference (AGC) on 10 and 13 July 2017, respectively. The purpose of this report is to provide simple documentation of the workshop proceedings (inputs, outputs and discussions). It has been reviewed (refer to Acknowledgements) and has been subject to basic editorial procedures.

Modelling uncertainty issues are discussed in chapter 7 of the *Australian groundwater modelling guidelines* (AGMG; Barnett et al., 2012). However, there has been generally poor uptake of uncertainty methods and a perception of low value for money from the results of uncertainty analysis and/or poor appreciation of how to make decisions based on a range of predictive simulations and results expressed with probabilities rather than 'a single number'. This is despite the AGMG guiding principle 7.1: 'Because a single "true" model cannot be constructed, modelling results presented to decision makers should include estimates of uncertainty'.

To address these issues, the NCGRT is promoting the improved treatment of uncertainty in groundwater investigations generally and more particularly in modelling studies. This was initiated via workshop and panel session activities at AGC 2017 with training initiatives planned. The aim is to provide additional information and training opportunities that augment the AGMG on how groundwater modelling uncertainty methods can best support decision making.

Draft discussion papers were presented and discussed at the workshop and the subsequent panel session at AGC 2017. The final discussion papers, revised to account for the discussions, are attached to this report (the papers may also be published in an open access journal).

The reader is urged to consult the accompanying final discussion papers (Appendix A) for detailed information on uncertainty analysis issues, as this report provides a basic overview and some information on the consultation process.

After the workshop, a report on *Uncertainty analysis guidance for groundwater modelling within a risk management framework* for application to large coal mines and coal seam gas projects was issued by the Department of the Environment and Energy (Middlemis and Peeters, 2018). It was based largely on the discussion papers and the workshop process, but it also provides detailed information on a wide range of uncertainty issues, and it outlines some guiding principles and a workflow process for addressing uncertainty (i.e. it is not a step-by-step guide). It may be accessed at:
www.iesc.environment.gov.au/publications/information-guidelines-explanatory-note-uncertainty-analysis

The purpose of the workshop and panel session at AGC 2017 was to consider, discuss and expound upon practical methods for improving the treatment of uncertainty in groundwater modelling investigations that are applied to impact assessments. A secondary goal was to provide information to support water resources management and decision making.

In other words, this report on the workshop is not a step-by-step guide, but a set of discussion papers on key topics (Appendix A), along with this over-arching report. Additional NCGRT reports may be required in future to provide practical, step-by-step guidance on conducting uncertainty analyses, along with training initiatives to develop the requisite skills in the workforce.

From a management perspective, modelling is considered to have failed if there is sufficient bias for a poor decision to be made (e.g. by lack of uncertainty analysis or lack of transparency in documentation), especially if the consequence is large (Walker, 2017; Doherty and Moore, 2017). Put another way: if modelling is used to predict that an unwanted outcome won't happen (e.g. via a biased model that overlooks important causal pathways), but it can indeed eventuate (with non-trivial probability), then we should consider that the model study has failed.

A definition of the purpose of a modelling study is proposed: to provide information about uncertainties in the conceptualisations and model simulations in a way that allows decision makers to understand the effects of uncertainty on project objectives (echoing the ISO 31000:2009 risk definition) and the effects of potential bias. For low risk projects, it is generally acceptable to describe the effect of uncertainty on the project objectives in qualitative terms (e.g. Peeters, 2017). For high risk projects, a quantitative uncertainty assessment is required in addition to the qualitative assessment. Objective uncertainty analysis (qualitative or quantitative) gives end-users confidence that future potential impacts (e.g. threshold impacts exceeded) have been considered carefully in an unbiased way.

Explicit consideration of uncertainty is warranted for every groundwater project: it must be considered at the problem definition stage and it should be integrated with the workflow. A generic uncertainty analysis workflow is presented in Table 1 as an iterative work breakdown structure (after Middlemis and Peeters, 2018). It is based on the principles in the accompanying discussion papers (Appendix A) and provides detail on the key steps that are commensurate with a quantitative uncertainty analysis methodology embedded within a risk management framework (Walker, 2017).

Problem Definition	Modelling Investigation	Predictive Uncertainty
• **Define Objectives.** • **Collate/Analyse Data.** • **Define Environmental Values to protect.** • **Review Development & Climate Stressors.** • **Develop/Refine Conceptual Model(s):** o use all data, inc. geophysics/geochemistry; o include objective information (water balance, parameter distributions); o simple/complex balance; o design for purpose of quantifying uncertainties. • **Analyse Potential Impact Causal Pathways:** o depressurisation & dewatering, surface water & groundwater interactions, springs, subsidence, final pit voids/lakes, existing users; o direct, indirect and cumulative impacts; o consider attractive/effective risk treatments. • **Perform initial risk assessment and Qualitative Uncertainty Analysis.** • **Define model purpose:** o to reflect decision support purpose; o suitable to investigate uncertainties; o specific terms (e.g. thresholds/triggers). • **Design Uncertainty Assessment methods.** • **Prepare Problem Definition Report.** • **Engagement with Agencies;** o review & refine Problem Definition.	• **Refine Conceptual Model(s)** (see Problem Def'n). • **Design and Build Mathematical Model(s):** o Complexity-Simplicity balance commensurate with risk/consequence; o transparent parameterisation design; o transparent uncertainty design. • **For all simulations (history match calibration or deterministic or stochastic methods):** o constrain simulations to be consistent with **all** information/data: - heads and fluxes (both); - geophysics, geochemistry, tracers etc. o **determine acceptable level of model to measurement mismatch for each observation**, preferably based on measurement uncertainty and weight observations accordingly; o if doing history match calibration: - minimise error variance via pilot points and Regularisation; - minimise non-uniqueness with match to heads and fluxes across multiple distinct periods of hydrological stress (climate, pumping, etc). o **perform sensitivity analysis** to identify parameters that can be constrained by observations ("identifiability"). • **Refine Uncertainty Assessment methods.** • **Prepare Modelling Investigation Reports.** • **Engagement with Agencies;** o Review, Refine and Iterate Methods.	• **Identify parameters to include in Uncertainty Quantification.** • **Establish probability distributions** for identified parameters and covariance between parameters. • **Test and confirm model convergence** for the parameter ranges defined, based on justified choices/assumptions. • **Define acceptable model-to-measurement misfit** based on measurement uncertainty or project/model specific objectives. • **Select Uncertainty Quantification method:** o Scenario analysis / subjective probability. o Deterministic / linear probability. o Stochastic / Bayesian probability. • **Verify, justify and document assumptions** salient to uncertainty quantification method and include in qualitative uncertainty analysis. • **Present uncertainty in model outcomes** as an integral part of report, tables and figures. • **Prepare Comprehensive Reports.** • **Engagement with Agencies;** o Review, Refine and Iterate Methods & Results.

Table 1: Modelling uncertainty analysis iterative workflow summary

The following key **guiding principles** should be used to design a modelling workflow to objectively assess uncertainty. The principles are based on the discussion papers accompanying this report (Appendix A), and information in Middlemis and Peeters (2018). The guiding principles are all consistent with the AGMG (Barnett et al., 2012), including engagement with agencies at the outset and at key stages throughout the iterative workflow:

1. Uncertainty analysis is an integral part of a robust risk management framework (risk is defined as the effect of uncertainty on project objectives; AS/NZS 31000:2009), as it informs and complements other aspects such as risk assessment, mitigations/treatments, communicating outcomes and prioritising efforts to reduce uncertainty (e.g. data acquisition).
2. While all projects require a qualitative uncertainty analysis at a minimum (e.g. discussing how model assumptions can potentially affect simulations; Peeters, 2017), high risk projects also

require a quantitative uncertainty assessment to a level of detail commensurate with the potential risks and/or consequences of the project (i.e. a preliminary hydrogeological risk assessment is needed at an early stage in the project).

3. It is crucial to explicitly define project objectives and what the model needs to predict in specific and measurable terms that will support unbiased decision making (e.g. applying threshold or trigger impact terms which provide information on which decisions may be based objectively).
4. Modelling methods should be designed to investigate the causal pathways for potential impacts on water resources and water-dependent assets that may arise from a proposed development or management plan, and to quantify the related uncertainties, thus providing unbiased information to support decisions for groundwater management and policy.
5. The methodology should be designed in such as way as to provide information about the uncertainty in conceptualisations and model simulation outputs so as to allow decision makers to understand the effects of uncertainty on project objectives (i.e. consistent with the risk management methods of ISO 31000:2009) and the effects of potential bias. The model must be specifically fit for this purpose.
6. In developing the model(s), a balance must be struck between model simplicity and complexity for the purpose of uncertainty evaluation, commensurate with the risk/consequence profile of the project (i.e. more than one model may be required).
7. The model simulations should be constrained with available observations and information.
8. The range of model outcomes that are consistent with all observations and information should be presented (calibration-constrained model outcomes).
9. Reports should transparently and logically discuss modelling and methodology assumptions and choices, and how they affect simulations; uncertainties and potential bias; and present the results clearly such they are not prone to misinterpretation, to instil confidence in the model simulations with stakeholders and clients (elaborated in more detail in Richardson et al., 2017).
10. Uncertainty analysis results should be: (i) carefully tailored to decision makers' needs (i.e. based on consultation), (ii) focussed on the messages that are most likely to be relevant to their decisions; and (iii) presented in plain and clear (precise, non-jargon) language.
11. Project workflow should be iterative, revisiting objectives, assumptions, conceptualisations and simulations, as well as the risk assessment (with consideration of any risk treatments applied to mitigate impacts), in a process of engagement between proponents, water managers (agencies) and their technical experts and reviewers that begins at project inception.

Noting that the above principles are procedural rather than technical, another guiding principle on a key technical matter which was discussed at the workshop is also proposed:

- Faults may be included as specific model features only where explicit evidence exists, consistent with the AGMG principle of parsimony. Where some minor/inconclusive evidence exists, faults could be considered as part of a sensitivity/uncertainty analysis involving parameterisation of the fault features and consideration of probabilities.

1. Introduction and Context

This report provides a summary of the outcomes from the 2017 NCGRT national groundwater modelling uncertainty workshop and the subsequent modelling outlook panel session held at the Australasian Groundwater Conference (AGC) on 10 and 13 July 2017, respectively.

The discussion papers shared at the workshop have been refined and are attached to this report, which addresses talking points from the event. The reader is urged to consult the discussion papers (Appendix A) for details on methodologies to improve uncertainty analysis practice.

This report provides a basic overview of the workshop proceedings (inputs, outputs and discussions) and summarises the guiding principles, sets out some information on the workshop process and some arguments justifying the level of effort required to conduct uncertainty analyses.

This report is not a step-by-step guide.

It is a summary of a set of discussion papers on key topics that were workshopped. Further investment in this initiative may be required in future to provide practical, step-by-step guidance on conducting uncertainty analyses, along with training initiatives to develop the requisite skills in the workforce.

Decision making in the context of uncertainty is fundamental to natural resource management. Deterministic groundwater modelling with no uncertainty analysis is inconsistent with the *Australian groundwater modelling guidelines* (Barnett et al., 2012). A model should be able to quantify its own reliability by accompanying its simulations with an assessment of uncertainty so that model-users and decision makers have some assurance that uncertainty is not underestimated (paraphrasing Doherty and Moore, 2017).

It is well known in the groundwater community that predictive uncertainty can be large, principally due to poorly known sub-surface conditions (aquifer properties and flow systems). Continuing advances in computing power, software and expertise have improved our ability to assess groundwater uncertainty. Increased capability and demand led to an entire chapter (chapter 7) on uncertainty issues in the 2012 *Australian groundwater modelling guidelines* (Barnett et al. 2012).

However, there has been poor uptake of uncertainty methods by all but a few expert groundwater modellers in each state. There are perceptions by some of low value for money in the results of uncertainty analysis and/or poor appreciation of how to make decisions based on a range of predictive uncertainty simulations and results expressed with probabilities (noting that decision makers and/or policy-setters have traditionally been provided with a ‘single number’). Others have suggested that the low uptake is because the upskilling is too onerous for practical/commercial modelling projects (i.e., steep learning curve, complex methods and data processing tools, no easy to follow guides on conducting uncertainty analysis). Whatever the reasons, there is clearly a strong need for practical training in uncertainty analysis methods. This document does not provide that basic guide. Rather, it presents information on fundamental uncertainty issues (to assist with the learning curve), and it describes a workflow and a set of guiding principles for application to uncertainty analysis. The NCGRT and other organisations are planning training initiatives, and it is expected that a basic ‘how to’ guide will be prepared in due course.

Indeed, a ‘single number’ result is often demanded from modellers without the two-way engagement that is necessary between modellers, decision makers and/or policy-setters to make well-informed decisions and to develop, refine and implement evidence-based policy (Figure 1). This is despite the AGMG guiding principle 7.1: ‘Because a single “true” model cannot be constructed, modelling results presented to decision makers should include estimates of uncertainty’.

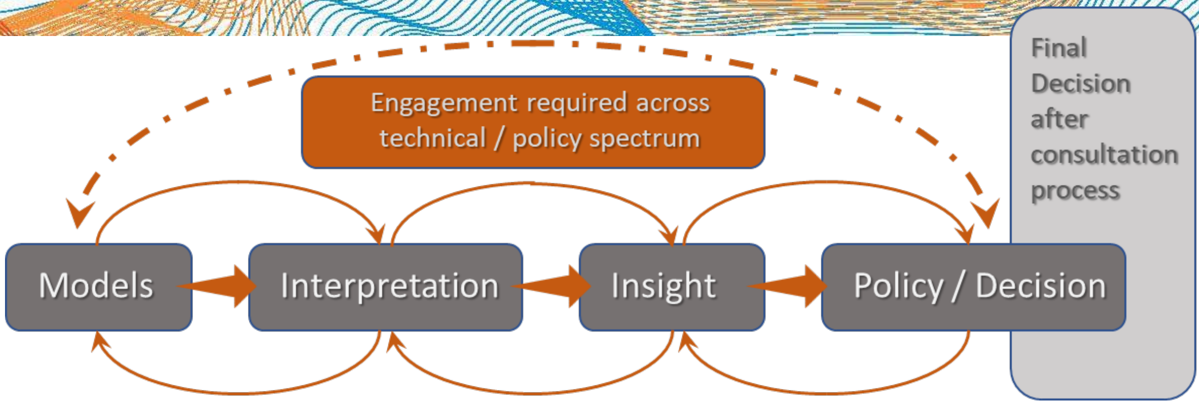


Figure 1: Iterative engagement between technical and policy

There are, however, examples of successful engagement, such as the revision of extraction limits for the Padthaway Prescribed Wells Area in South Australia, where community consultation was also an element of the multiple feedback loops across the technical–policy–community spectrum (Harrington et al. 2010). Although uncertainty analysis was not technically conducted for the Padthaway case, the methodology involved a wide range of modelling scenarios to investigate resource condition limit sensitivities under extraction and climatic stresses, and to identify a robust sustainable yield option.

In other cases, concerns are often raised that the modelling guidelines are being used inappropriately at times to justify ‘indiscriminate complexification’ of models, rather than the ‘effective simplification’ that is often needed to meet investigation objectives (Voss, 2011; Doherty and Moore, 2017). There are often claims that appropriately simple models are not ‘fit-for-purpose’ and that highly complex models are needed to ‘understand the dynamics of groundwater flow systems’. However, highly complex models are expensive to develop, and usually run slowly or are not numerically stable across feasible parameter ranges, which hinders the application of uncertainty analysis methods that can provide valuable information for decision making (the ‘simplicity/complexity dilemma’: Pappenberger and Beven, 2006; Beven and Young, 2013; Doherty and Simmons, 2013; Doherty and Moore, 2017).

Advances in groundwater modelling capability have coincided with expansion of the resource sector (e.g. CSG and unconventional gas); this has led to more intense public scrutiny of environmental impact assessments and the role of groundwater modelling of complex aquifer and aquitard systems and the effects of uncertainty.

Other than the Bioregional Assessments[1], (e.g. Janardhanaran et al., 2016) there are few examples of comprehensive uncertainty analysis applied to models of complex systems. Other notable examples include the GEN3 dual phase water and gas regional modelling of the Surat Basin (QGC, 2013) and the related study of Herckenrath et al. (2015); the Menindee Lakes groundwater modelling investigation (Jacobs, 2016; Hayes and Nicol, 2017); and the impact assessment for the Pilbara Mining Area C iron ore deposits (BHPB, 2017a, b).

These examples demonstrate that uncertainty analysis is no longer just in the academic realm. Uncertainty analysis can and indeed must be applied to decision making in relation to high value environmental assets and/or economically productive assets.

The burgeoning need for uncertainty analysis to guide decision making and the continued development of methods invokes challenges in selecting an appropriate uncertainty quantification method for the study objectives and in developing adequately skilled practitioners. There are also challenges in interpreting and communicating the results, all of which have implications for decision making, as explored in this report.

[1] http://www.bioregionalassessments.gov.au/methods/propagating-uncertainty-through-models

2. Aims, Process And Products

The AGC 2017 workshop and panel session process, including preparation of discussion papers beforehand and publishing this report, aims to promote the improved treatment of uncertainty in the groundwater modelling industry.

This report and the accompanying discussion papers (Appendix A) are philosophically integrated and consistent with the *Australian groundwater modelling guidelines* (Barnett et al., 2012). This report augments the guidelines with additional information on the practical implementation of uncertainty methods, highlighting the need for consultation at all stages during a model study, along with consideration of the risk context to identify the level of effort required in the qualitative and/or quantitative uncertainty analysis.

This report provides an overview of the guiding principles on groundwater modelling uncertainty analysis workflows and methodologies that have been drawn from the discussion papers; the reader is urged to consult the accompanying discussion papers for detailed guidance (Appendix A).

The process that led to this report involved:

- Preparing draft discussion papers for presentation and discussion at a workshop on 10 July 2017;
- Engaging with an influential cross-section of around 40 workshop delegates, including uncertainty experts, practising groundwater modellers, academics, industry and government users of modelling results;
- Workshopping alternative modelling workflows and uncertainty assessment methods, and discussing opportunities to reduce impediments to the uptake of improved practice, including training (being followed up by the NCGRT and ICEWaRM;
- Preparing a glossary of uncertainty analysis terms (augmenting the modelling guidelines glossary);
- Learning from the experience of Australian and international uncertainty assessments and providing advice on communicating uncertainty in ways that are useful for decision making, including:
    - The Menindee Lakes modelling study, where the workflow was designed as an uncertainty assessment with effective model simplification applied, even though comprehensive datasets were available from the extensive Broken Hill Managed Aquifer Recharge investigation (Lawrie et al., 2012; www.ga.gov.au/about/projects/water/broken-hill-managed-aquifer-recharge).
    - The Bioregional Assessments (www.bioregionalassessments.gov.au), where the uncertainty analyses applied are integrated with risk assessments and consideration of causal pathways for propagation of impacts, and innovative methods have been trialled for the communication of uncertainty analysis results.

In addition to the workshop, a plenary address at AGC 2017 by Dr Catherine Moore on 13 July 2017 expounded the need for and justification of optimal model complexity (or effective model simplicity) to ensure tractable uncertainty analysis.

The plenary reprised some of the issues set out in the accompanying discussion paper (Doherty and Moore 2017), and was followed by a modelling outlook panel session on uncertainty, which considered some key questions arising and/or outstanding from the workshop, as well as questions from the conference floor.

Notes from the workshop and panel sessions are presented in Appendix C.

Final discussion papers that account for the workshop and panel discussions are attached to this report (Appendix A). The papers may also be published in an open access journal (e.g. *Environmental Modelling and Software*).

The reader is urged to consult the final discussion papers (Appendix A) for detailed information on uncertainty analysis issues; this report provides a simple overview and some information on the process.

Subsequent to the workshop, a report on *Uncertainty analysis guidance for groundwater modelling within a risk management framework* (for application to large coal mines and coal seam gas projects) was issued by the Department of the Environment and Energy (Middlemis and Peeters, 2018): http://www.iesc.environment.gov.au/publications/information-guidelines-explanatory-note-uncertainty-analysis.

That report has drawn on information from these workshop discussion papers, and this report in turn has also drawn on some content from Middlemis and Peeters (2018), and has tried to avoid inconsistency.

3. Types Of Uncertainty And Other Definitions

There are different ways to categorise uncertainty, but it is often lumped into two main types (Barnett et al. 2012):

- Deficiency in our knowledge of the natural world (including the effects of error in measurements), and
- Failure to capture the complexity of the natural world (or what we know about it).

It is helpful, however, to consider four main types of uncertainty:

1. Structural/Conceptual
   Geological structure and hydrogeological conceptualisation assumptions applied to derive a simplified view of a complex hydrogeological reality (any system aspect that cannot be changed in an automated way in a model);
2. Parameterisation
   Hydrogeological property values and assumptions applied to represent complex reality in space and time (any system aspect that can be changed in an automated way in a model via parameterisation, including aquifer properties, boundary conditions, etc.);
3. Measurement error
   Combination of uncertainties associated with the measurement of complex system states (heads, discharges), parameters and variability (3D spatial and temporal) with those induced by upscaling or downscaling (site-specific data, climate data);
4. Scenario uncertainties
   Guessing future stresses, dynamics and boundary condition changes (e.g. mining, climate variability; land and water use change).

These four sources of scientific uncertainty result in predictive uncertainty – the bias and other error associated with model simulations (see **Figure 2**, after Richardson et al. 2017, and Doherty and Moore, 2017). Bias refers to systematic error, which displaces the model outputs away from the accepted 'true' value, and error refers to the difference (spread) between the average value of model simulations and the accepted true value. Bias and error affect the precision of model results, even when that model is consistent with the conceptual understanding of the system and the related observations and measurements.

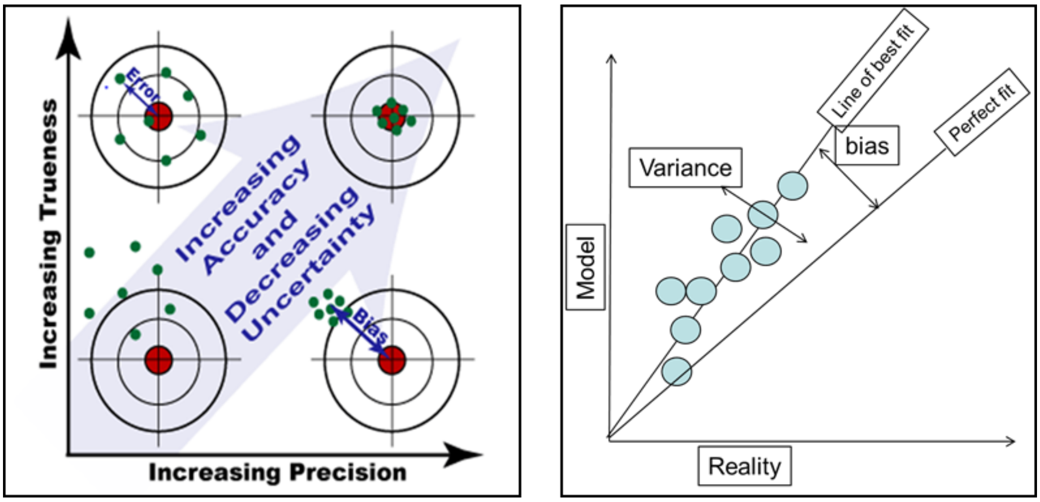


Figure 2: Errors, biases and influences on uncertainty (after Richardson et al. 2017; Doherty and Moore, 2017)

Being overcommitted to one conceptualisation over others (bias), perhaps one that is not adequately representative of key system features, could lead to simulations that overestimate or underestimate impacts. If uncertainty analysis focusses only on errors and neglects to account for or discuss biases, incomplete and distorted evidence of the modelling accuracy will be provided (Doherty and Moore, 2017). The effects of bias are also discussed in Richardson et al. (2017).

A list of uncertainty-related definitions is presented in Appendix B.

4. Decision-Making With Uncertainty And Risk

4.1 Uncertainty analysis and risk management

The sub-surface environment is heterogeneous and complex, difficult to directly observe, characterise or measure. Groundwater systems are open to influences from climate, topography, vegetation, hydrology and human activities. This means that uncertainty affects our ability to accurately describe existing or future states of hydrogeological systems.

Scenario modelling is used to investigate and support decisions for groundwater resource assessment, management and policy. This invokes further uncertainty due to the assumptions applied to conceptualise a simplified view of hydrogeological reality, due to parameterisation and scenario uncertainties (e.g. land/water use, climate variability).

The seminal paper by Freeze et al. (1990) characterises the role of models in decision support as quantifying the level of risk associated with management options. It follows that if a model is applied to support environmental decision making, its simulations of the consequences (impacts or outcomes) of management options must quantify the related uncertainties and probabilities (Doherty and Moore, 2017), noting that:

- Risk is defined as the effect of uncertainty on project objectives (AS/NZS 31000:2009);
- Risk is characterised/quantified as a function of the probability and consequences of an outcome.

As discussed in Walker (2017), uncertainty analysis is an integral part of a robust risk management framework, as it informs and complements other aspects such as risk assessment, mitigations/treatments, communicating outcomes and prioritising efforts to reduce uncertainty (e.g. by acquiring data on key processes). A simple example of high priority (and relatively low cost) data that reduces uncertainty in groundwater models is accurate LiDAR topographical data. Accurate definition of the interface between the surface and the sub-surface is critical for implementing boundary conditions in a model to represent surface-water features (creeks/rivers), evapotranspiration and spring features (Doble and Crosbie, 2017).

Risk assessment and management frameworks involve consideration of the balance between estimating uncertainty; reducing uncertainty through monitoring and scientific investigations; and managing or treating risk, at times in response to new information (Figure 3, after Walker, 2017).

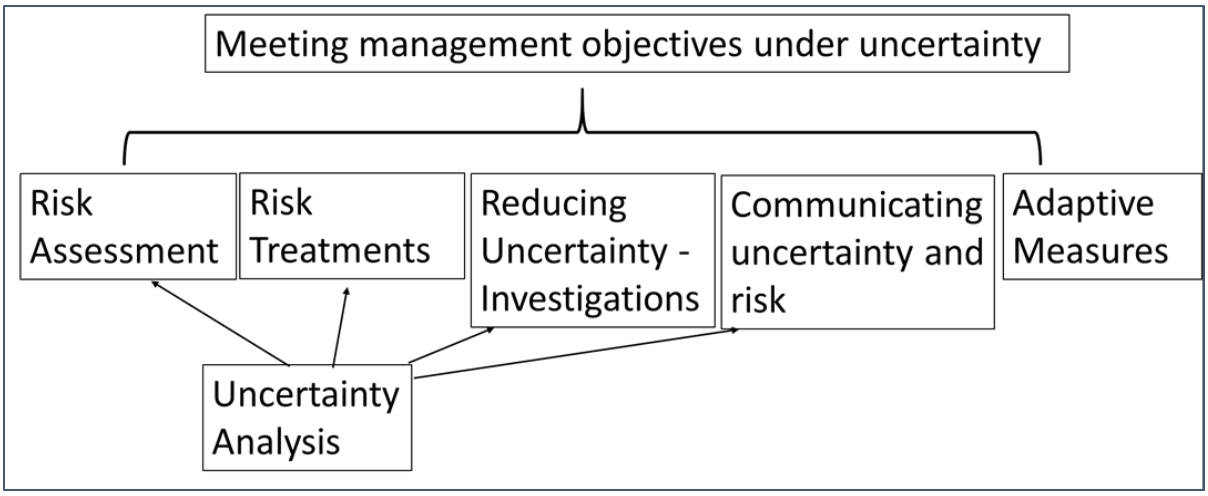


Figure 3: Schematic showing how uncertainty analysis may support different elements of a risk management framework (after Walker, 2017)

Each situation is different, and the balance of these components will vary. An uncertainty analysis is required to support each of these elements and conversely, the balance of these elements will determine the objective of the uncertainty analysis and the level of effort required. Further information is provided in the accompanying discussion paper (Walker, 2017; see Appendix A).

In environmental management, risk has negative connotations generally associated with the hazards or impacts of a development. In this sense, risk is one possible (negative) consequence of uncertainty. Other (positive) consequences of uncertainty can also be identified (Begg, 2013) by way of opportunities to achieve desired benefits (e.g. to justify expenditure on a mining project where sound environmental management can manage other project risks).

This highlights the point that value judgements are involved in all risk and uncertainty assessments and the value judgements depend largely on the economic, social and/or environmental values established in public policies, business cultures and community viewpoints, while scientific studies provide objective information on environmental risks, impacts, mitigations, benefits and management.

4.2 ESD and the precautionary principle

Ecologically sustainable development (ESD) and the precautionary principle are very important to decision making and environmental management, and have been tested in Australian law, notably in the Queensland Land Court case in 2015 regarding the proposed Adani Carmichael coal mine (QLC 48), where groundwater modelling issues and uncertainties were considered in detail. A subsequent case on the New Acland Coal project (QLC24) also considered modelling issues and judged elements of the modelling to be not consistent with best practice.

ESD principles establish that social/community considerations (including value judgements) are a key factor in decision-making processes, along with economic and environmental factors. The precautionary principle is incorporated in the principles of ESD, which are promoted by the objectives of the EPBC Act 1999, as explained in the 2009 independent review of the EPBC Act (Cth, 2009).

The precautionary principle may be summarised as follows: if a development raises the risk of harm to the environment (i.e. in non-trivial likelihood and consequence terms), then proportionate precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically.

There are two key pre-conditions for the application of the precautionary principle: the threat of serious or irreversible environmental damage, and scientific uncertainty as to the nature and scope of the threat of environmental damage. If both pre-conditions are established (noting that these conditions or thresholds are cumulative), the burden of proof as to the impacts/outcomes shifts to the proponents of the development (Cth, 2009; item 13.21).

This means that groundwater models should be designed to investigate the causal pathways for potential impacts on water resources and water-dependent assets that may arise from a proposed development or management plan, and to quantify the related uncertainties, thus providing unbiased information to support decisions for groundwater management and policy. In principle, causal pathways should be identified by conservatively considering potential connectivity between groundwater units and/or surface-water features and related ecological assets such as groundwater dependent ecosystems (GDEs).

4.3 Need for improved modelling practice to support decision-making

There has been a demand from government agencies for improved assessment of potential impacts that may be manifested via groundwater pathways. There is also increased awareness by stakeholders of the limitations of modelling and the potential effect of associated uncertainties. However, the users, purchasers and reviewers of model studies are seeing little uptake in the analysis and quantification of uncertainties of modelling studies, despite the encouragement over many years of guidelines (Middlemis et al. 2001; Barnett et al. 2012) and papers (Freeze et al. 1990; Moore and Doherty, 2005; Doherty, 2010). This is despite material changes in our capability to describe predictive uncertainty through greater computing power, better software and related documentation on the technical methodologies, and better-trained professionals.

Reasons for the generally poor uptake might include (i) rapid changes and developing maturity of the subject and time and effort required to achieve competency, (ii) lack of acceptance of the higher costs associated with uncertainty analyses, (iii) unclear workflow processes to successfully undertake uncertainty analyses for specified project objectives and to communicate them clearly, (iv) lack of wider communication amongst groundwater modellers, purchasers and regulators and issues of guideline adoption pathways, (v) the users of model study results expressing their preferences to model developers for ‘a singular number’ on which to base decisions for resource allocation or management, impact assessment and/or infrastructure and agriculture development. This is despite the AGMG guiding principle 7.1: ‘Because a single “true” model cannot be constructed, modelling results presented to decision makers should include estimates of uncertainty’.

As indicated above, modelling guidelines are sometimes being used to justify ‘indiscriminate complexification’ of models, which results in highly complex models that are expensive to develop, and usually run slowly or are not numerically stable, which hinders formal uncertainty analysis (the ‘simplicity/complexity dilemma’: Voss, 2011; Doherty, 2010; Doherty and Moore, 2017). Formal uncertainty and sensitivity analysis is required, based on effective or optimum model complexity, to help reduce or quantify predictive uncertainty so that model-users and decision makers have some assurance that uncertainty is not underestimated (see accompanying discussion paper by Doherty and Moore, 2017; Appendix A)).

5. Model Confidence Level And Modelling Workflows

The 2012 *Australian groundwater modelling guidelines* (Barnett et al., 2012) dedicate an entire chapter to uncertainty analysis. While this may lack the detailed methodologies presented in certain papers (e.g. Doherty, 2010), the AGMG is arguably more accessible in describing aspects of the philosophy of uncertainty assessments, including the simplicity/complexity dilemma and how to communicate the effects. This report is designed to augment the AGMG, and is philosophically integrated with its principles, despite one or two nominal divergences, as explored in this section. Additional reports will be required in future to provide practical, step-by-step guidance on conducting uncertainty analyses, along with training initiatives to develop the requisite skills in the workforce.

This section combines information from the accompanying discussion papers (Appendix A), the workshop and plenary sessions at AGC 2017. It incorporates some information from the *IESC Explanatory Note* (Middlemis and Peeters, 2018), which itself expands on issues outlined in the discussion papers.

5.1 Model confidence level classification

The AGMG (Barnett et al. 2012) state that objective consideration of uncertainty is warranted for every groundwater project. As described in Peeters (2017) and Richardson et al. (2017), a well-executed uncertainty analysis (qualitative and/or quantitative) means that a model should be able to quantify its own reliability, rather than relying on the AGMG 'confidence level' schema (Barnett et al. 2012, Section 2 and Table 2-1) which is prone to misinterpretation. For example, there is a tendency to 'cherry pick' a comment from the AGMG content around its Table 2-1 to undermine the model confidence classification, or to ignore the commentary and overreach the confidence level, rather than considering the balance of model performance against the entire table of attributes (Middlemis and Peeters, 2018). The table is itself not unreasonable, but the related commentary and guidance is poor and self-contradictory on some elements. For example, with reference to Table 2 (discussed below), the AGMG commentary indicates that a single Class 1 attribute is sufficient to classify the model as Class 1 overall, even though the weight of evidence in this example indicates a Class 2 model is reasonable. Other AGMG commentary also suggests that 'if a model has any of the characteristics or indicators of a Class 1 model it should not be ranked as a Class 3 model, irrespective of all other considerations', which implies that, while a model may not be Class 3, it could be labelled as Class 2, even though it has Class 1 characteristics, which is logically inconsistent with the point a single Class 1 attribute.

Table 2: AGMG model confidence level case study example

Class	Data		Calibration		Prediction		Quantitative Indicators	
1 (simple)		Not much.		Not possible.		Timeframe >> Calibration		Timeframe >10x
		Sparse coverage.	~	Large error statistic.		Long stress periods.		Stresses >5x
	√	No metered usage.		Inadequate data spread.		Poor/no validation.		Mass balance > 1% (or one-off 5%)
		Low resolution topo DEM.		Targets incompatible with model purpose.		Transient prediction but steady-state calibration.		Properties <> field values.
		Poor aquifer geometry.						No review by Hydro/Modeller.
2 (impact asessment)	√	Some.	√~	Partial performance.	√	Timeframe > Calibration	√	Timeframe = 3-10x
	√	OK coverage.	~	Some long term trends wrong.		Long stress periods.	√	Stresses = 2-5x
	~	Some usage data/low volumes.	~	Short time record.	√	OK validation.	~	Mass balance < 1%
	√	Baseflow estimates. Some K & S measurements.		Weak seasonal match.	√	Transient calibration and prediction.	~	Some properties <> field values. Review by Hydrogeologist.
	√	Some high res. topo DEM &/or some aquifer geometry.		No use of targets compatible with model purpose (heads & fluxes).	√	New stresses not in calibration.		Some coarse discretisation in key areas of grid or at key times.
3 (complex simulator)		Lots, with good coverage.		Good performance stats.		Timeframe ~ Calibration		Timeframe < 3x
		Good metered usage info.	√~	Most long term trends matched.	√	Similar stress periods.		Stresses < 2x
	√	Local climate data.	~	Most seasonal matches OK.		Good validation.	~	Mass balance < 0.5%
	~	Kh, Kv & Sy measurements from range of tests.		Present day data targets.		Calib. & prediction consistent (transient or steady-state).	√~	Properties ~ field measurements.
		High resolution DEM all areas.	√	Head & Flux targets used to constrain calibration.	√~	Similar stresses to those in calibration.	√	No coarse discretisation in key areas (grid or time).
	√	Good aquifer geometry.					√	Review by experienced Modeller.

(after Table 2-1 of Barnett et al (2012) Australian Groundwater Modelling Guideline)

Alternative methods of confidence level assessments have been tested in practice, based on indicating which attributes in the table are satisfied for a given model and explaining the reasoning in relation to the model objectives, outcomes and uncertainties (including why other attributes may not be satisfied). This requires modellers to justify assumptions and choices in technical reports in a manner that is open, transparent and amenable for scrutiny (a key guiding principle for effective uncertainty analysis). The overall confidence level may then be assessed via consideration of the weights in each class, ignoring the contradictory commentary in the guideline content referring to the table, but working directly with the table attributes themselves. An example is presented in Table 2, based on an original suggestion by Dr Noel Merrick (pers.comm.), as cited in Middlemis and Peeters (2018). This model is assessed as Class 2, despite the lack of metered usage data.

5.2 Generic modelling workflows

As explained in the accompanying papers (e.g. Walker, 2017; Peeters, 2017) uncertainty analysis must be considered at problem definition and each subsequent stage of the workflow. It must be integrated within a risk management framework, which will require an initial risk assessment and subsequent iterations to review and revise (e.g. as risk treatments are invoked). It must involve meaningful (‘without prejudice’) engagement/consultation between proponents and agencies on methodologies and assumptions.

This need not be a frequent/intensive consultation process, but it should occur at appropriate times during the three key stages of the workflow, as discussed at section 5.4 and **Table 3** later, and consistent with the AGMG guiding principles. For example, at the first stage, it may be prudent for a proponent to conduct an initial risk assessment and engage with a modelling consultant to investigate the hydrogeological conceptual model and potential causal pathways to identify whether risk treatments may be effective. That would lead to a modelling and uncertainty assessment methodology design report that would be appropriate to discuss with agencies before committing to the next stage (see **Table 3** later).

A conceptual example is illustrated in Figure 4 (after Walker, 2017, based on discussions in Guillaume et al. 2016, and Peeters, 2017). Initially, a preliminary risk assessment is done, possible risk mitigations are considered, and the model is conceptualised to meet the objectives. As the modelling and assessment workflow proceeds through its iterations, there is a winnowing of the objectives according to risk, and complexity may be added or refined as necessary. In the preliminary stages, there may not be any need for numerical modelling, and if risks are not high at any stage, nothing more may be required and resourcing the investigation may be curtailed.

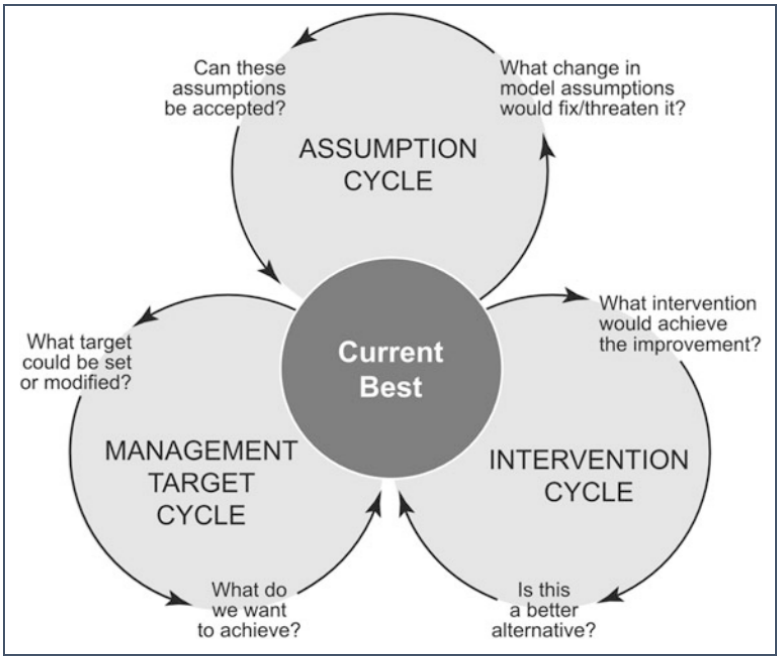


Figure 4: Schematic iterative approach for groundwater modelling involving setting objectives, risk mitigation options and modelling conceptualisation (Guillaume et al. 2016)

For low risk projects, it may be acceptable to qualitatively assess or describe the effect of uncertainties on the project objectives. As explored in Peeters (2017), this may involve the application of the model confidence level classification, as indicted in Section 5.1, along with a traditional workflow (conceptualise, design, build, calibrate, predict) that may include sensitivity analysis simulations (Figure 5).

For high risk projects, a quantitative uncertainty analysis is warranted (Walker, 2017; Peeters, 2017; Doherty and Moore, 2017), which will result in the model evaluating its own predictive reliability, without needing to invoke the model confidence level classification (Section 5.1). The modelling workflow required to support a quantitative uncertainty analysis differs conceptually from the traditional workflow (Figure 5), although it is still consistent with AGMG principles.

The conceptual flow charts for a traditional modelling workflow and an alternative workflow for improved uncertainty assessment, expounded by Ferré (2016), are illustrated in Figure 5. While the uncertainty analysis workflow includes traditional elements of model development (e.g. design, calibration and sensitivity), an uncertainty-driven approach differs conceptually in that it is designed and applied specifically to support decisions by exploring uncertainties within a risk and adaptive management framework.

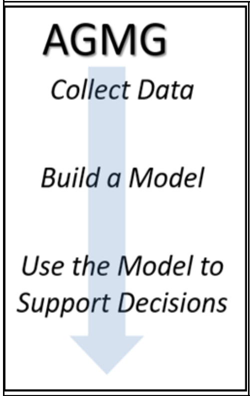
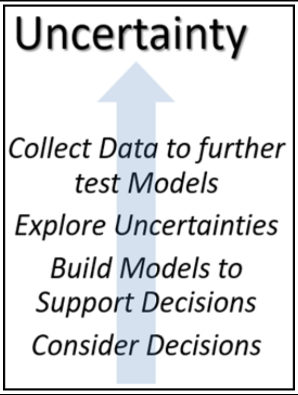
Traditional modelling workflow		**Uncertainty-driven modelling workflow**	
	Collect data, develop conceptual models, build and calibrate models and assess sensitivities, use predictions and/or uncertainty analysis to support decision making.		Consider objectives, decisions, risk context and conceptual model(s); build models to support decisions and explore uncertainties; collect/analyse data to test models; evaluate effects of uncertainties on objectives/decisions.

Figure 5: Conceptual/generic modelling workflows

The traditional workflow tends to favour the development of a complex model (i.e. typically one deterministic realisation), even though the AGMG encourages finding the right balance between complexity and simplicity for the project objectives, and it suggests exploring alternative conceptualisations and parameterisations and investigating uncertainty.

An uncertainty-driven modelling approach, however, requires carefully designed models with short run times for the large numbers of runs involved. Careful design can take many forms, such as ensuring stable model simulations and that complexity is included where it is relevant to the project objective ('effective simplicity'), while not using long run times as an excuse to avoid necessary complexity. Careful design can also mean that more than one conceptualisation or realisation is tested, as suggested by many leading hydrogeologists, including Darcy Distinguished Lecturers (Ferré, 2016; Poeter, 2006). While building multiple models is a challenge for any project in these budget-constrained times, there are examples of it in practice (e.g. Hayes 2017; Middlemis 2011, referring to examples including Prominent Hill, Tasmanian Sustainable Yields, McLaren Vale). There is also an example where a conceptual model has not been updated with new data and an environmental court has judged the work as poor practice (QLC17–24; see also http://envlaw.com.au/acland/).

5.3 Uncertainty analysis workflows

As discussed in Peeters (2017), the goal of an uncertainty analysis is to build confidence with clients and stakeholders by communicating what we do know, and what we can predict with an objective estimate of uncertainty. This means that we must communicate honestly and transparently that there can exist a wide range of model simulations that are all consistent with our current understanding of the system, even though gaps limit our system understanding. Part of an uncertainty analysis is therefore to assess how these knowledge gaps might affect the model simulations.
The common principles that emerge from examining these alternative modelling workflows, which also can be found in the AGMG (Barnett et al., 2012), are:

- Explicitly define project objectives and what the model needs to predict (i.e. not simply 'the development of calibrated model' for use in deterministic scenarios);

- To meet the objectives, define an appropriate study methodology, optimal model complexity and modelling workflow;
- Constrain the model simulations with all available observations and information;
- Discuss model assumptions and choices and how they affect simulations;
- Iterate and revisit model objectives, assumptions, workflow and simulations during the project in consultation with proponents and agencies.

These principles form the basis of the uncertainty analysis workflow depicted in Figure 6 (after Peeters, 2017). The left-hand side provides a high-level workflow, emphasising the need to explicitly define project objectives and the types of information to be provided by the simulations so that these can be taken into consideration when developing conceptual models. From that stage, there are two parallel processes, a qualitative and quantitative uncertainty analysis. A qualitative uncertainty analysis is a formal discussion of all model assumptions and choices and how they affect model simulations, which should be conducted and documented for all projects (low or high risk) as a minimum requirement. A quantitative uncertainty analysis seeks to find all model simulations that are consistent with (or constrained by) the observations; it is warranted for all high-risk projects. The workflow depicted on the right of Figure 6 provides more detail on quantitative uncertainty analysis aspects such as what change to make to an existing model, how to define initial parameter ranges and how to bring in observations. The thick blue arrows indicate the need to iterate between all the components of the workflow during the model project, including communication and reporting.

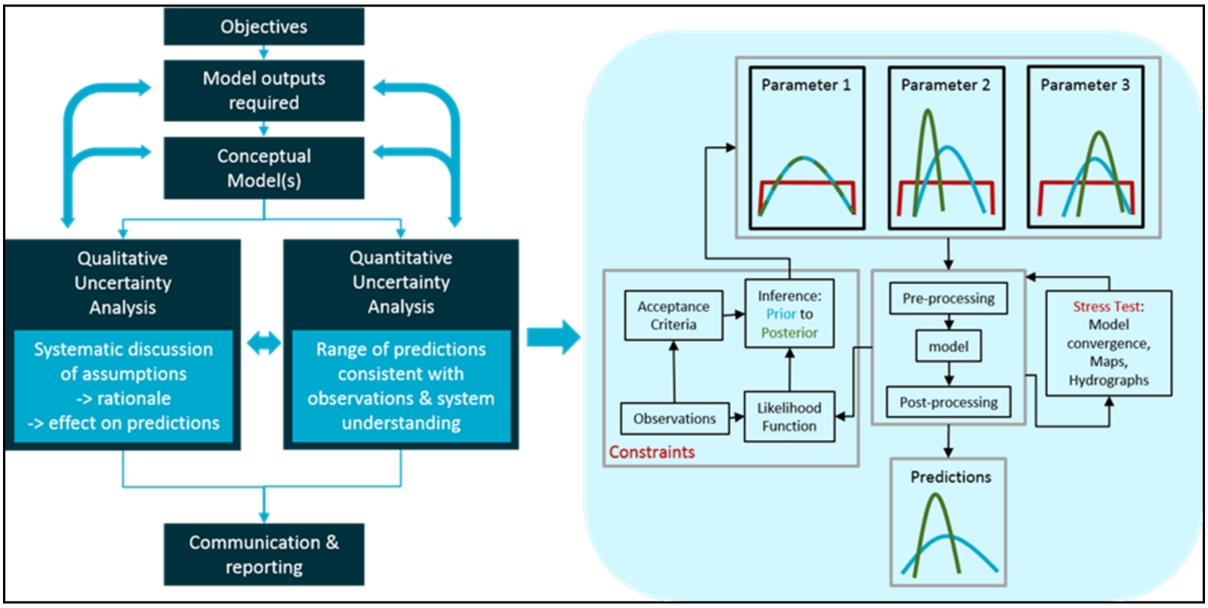


Figure 6: Qualitative and quantitative uncertainty analysis iterative workflow (after Peeters, 2017)

In the discussion paper by Peeters (2017), various aspects of the uncertainty workflow are discussed in an approachable way, based on questions pertinent to the modelling process, such as:

- Where to start?;
- Which assumptions matter?;
- What to change in the model?;
- How to bring in observations?; and
- Do initial parameter values matter?.

An emphasis is given to the need for the following (see also item 2.8 in Appendix C, Workshop notes):

- Specifying explicitly what simulations are required of the model and justifying all assumptions in the context of these simulations;
- Developing a stable model and well-defined acceptance criteria based on the available observations, especially for parameters that cannot be constrained by observations or are

prone to compensating for structural model issues (closely related to complexity issues discussed in Doherty and Moore, 2017);

- Transparently documenting and honestly reporting the results of an uncertainty analysis together with its limitations and assumptions, to instil confidence in the model simulations with stakeholders and clients (elaborated in more detail in Richardson et al., 2017).

5.4 Uncertainty analysis work breakdown structure

In Middlemis and Peeters (2018), the uncertainty analysis workflow is presented as an iterative work breakdown structure (**Table 3**). It is based on the principles in the accompanying discussion papers and provides more detail than Figure 6 (above) on the key steps that are commensurate with a quantitative uncertainty analysis methodology embedded within a risk management framework (Walker, 2017).

Table 3: Modelling uncertainty analysis iterative workflow summary

Problem Definition	Modelling Investigation	Predictive Uncertainty
• **Define Objectives.** • **Collate/Analyse Data.** • **Define Environmental Values to protect.** • **Review Development & Climate Stressors.** • **Develop/Refine Conceptual Model(s):** o use all data, inc. geophysics/geochemistry; o include objective information (water balance, parameter distributions); o simple/complex balance; o design for purpose of quantifying uncertainties. • **Analyse Potential Impact Causal Pathways:** o depressurisation & dewatering, surface water & groundwater interactions, springs, subsidence, final pit voids/lakes, existing users; o direct, indirect and cumulative impacts; o consider attractive/effective risk treatments. • **Perform initial risk assessment and Qualitative Uncertainty Analysis.** • **Define model purpose:** o to reflect decision support purpose; o suitable to investigate uncertainties; o specific terms (e.g. thresholds/triggers). • **Design Uncertainty Assessment methods.** • **Prepare Problem Definition Report.** • **Engagement with Agencies;** o review & refine Problem Definition.	• **Refine Conceptual Model(s)** (see Problem Def'n). • **Design and Build Mathematical Model(s):** o Complexity-Simplicity balance commensurate with risk/consequence; o transparent parameterisation design; o transparent uncertainty design. • **For all simulations (history match calibration or deterministic or stochastic methods):** o constrain simulations to be consistent with **all** information/data: - heads and fluxes (both); - geophysics, geochemistry, tracers etc. o **determine acceptable level of model to measurement mismatch for each observation**, preferably based on measurement uncertainty and weight observations accordingly; o if doing history match calibration: - minimise error variance via pilot points and Regularisation; - minimise non-uniqueness with match to heads and fluxes across multiple distinct periods of hydrological stress (climate, pumping, etc). o **perform sensitivity analysis** to identify parameters that can be constrained by observations ("identifiability"). • **Refine Uncertainty Assessment methods.** • **Prepare Modelling Investigation Reports.** • **Engagement with Agencies;** o Review, Refine and Iterate Methods.	• **Identify parameters to include in Uncertainty Quantification.** • **Establish probability distributions** for identified parameters and covariance between parameters. • **Test and confirm model convergence** for the parameter ranges defined, based on justified choices/assumptions. • **Define acceptable model-to-measurement misfit** based on measurement uncertainty or project/model specific objectives. • **Select Uncertainty Quantification method:** o Scenario analysis / subjective probability. o Deterministic / linear probability. o Stochastic / Bayesian probability. • **Verify, justify and document assumptions** salient to uncertainty quantification method and include in qualitative uncertainty analysis. • **Present uncertainty in model outcomes** as an integral part of report, tables and figures. • **Prepare Comprehensive Reports.** • **Engagement with Agencies;** o Review, Refine and Iterate Methods & Results.

6. Optimal Simplicity And Conditional Calibration

6.1 Simplicity–complexity dilemma

As explored in detail by Doherty and Moore (2017), highly complex models are usually expensive to develop, and usually run slowly or are not numerically stable. Many complex models fail to deliver the decision support promised because they are too complex and insufficiently agile to use in a decision-making process that is based on an uncertainty analysis requiring thousands of simulations. Whereas increased complexity does not necessarily translate directly into a stronger technical basis for regulatory decisions, the use of overly simplified models may result in erroneous decisions. A careful balance must be achieved between simplicity and complexity, requiring discussion and agreement between modellers, proponents and agencies.

Doherty and Moore (2017) suggest that models can provide better support for environmental decision making if the premise of their construction is altered from the common setting of a simulator of complex environmental processes to that of a tool for implementation of the scientific method. This method requires that hypotheses be tested, and maybe rejected, through evaluating their compatibility with information/data on the nature and properties of the groundwater system. In the decision-making context, hypotheses that require testing are the unwanted outcomes (‘bad things’) that we seek to avoid if a certain course of management action is adopted. For each such bad thing, it should be possible to construct a model, specific to that bad thing, which can provide receptacles for information against which the probability of its occurrence can be tested. Such a model will likely be ‘optimally simple’ (rather than indiscriminately complex), for its performance in carrying out the task for which it was built will be enhanced if it is unencumbered by unnecessary complexity and is not required to provide receptacles for unnecessary information.

Based on the principles outlined in the accompanying discussion papers (Appendix A), Middlemis and Peeters (2018) advocate an approach that goes beyond the platitudes of subjectively making a model ‘as simple or complex as required, but not too simple or complex’ as many guidelines recommend. Rather:

- The model must be designed to be specifically fit for the purpose of providing information about uncertainty in a way that allows decision makers to understand the effects of uncertainty on project objectives, and the effects of potential bias;
- Engagement with regulatory agencies is required from the outset and at all stages throughout the modelling study, to discuss and agree upon the uncertainty analysis methodologies and understand the implications of the results.

6.2 Conditional calibration

The traditional workflow has been characterised as a means of reducing parameter bias and uncertainty through calibrating a (deterministic) model against measured observations of historical hydrologic system behaviour. A model that is demonstrably consistent with monitoring data (especially if head and flux calibration targets are matched) is traditionally deemed to be a reliable deterministic simulator of future behaviour. However, neither the structure nor the parameter values of a deterministic model are unique. This ‘equifinality’ problem has long been recognised and is not simply one of identifying a system's ‘true’ model structure or parameter values (Beven, 1993). In fact, a ‘true’ model for a hydrologic system does not exist, due to the sources of uncertainty outlined previously. Even the most complex model can (by definition) only be approximate in its attempted simulation of environmental processes.

Doherty and Moore (2017) show that the calibration process does not reduce the uncertainty of a simulation where it is sensitive to parameters/combinations that lie within the ‘calibration null space’. The calibration null space here refers to those model parameters and combinations that are not informed by the available historical measurements. However, they also show that calibration is a valid first step in a two-step uncertainty analysis process using linear methods:

- The first step is finding a history match (inverse) solution of minimum error variance by fitting model outputs to the calibration dataset of heads and fluxes (preferably during a period of wide-ranging hydrological stress); this reduces non-uniqueness and can be achieved using the uncertainty analysis techniques of pilot point parameter estimation with Tikhonov regularisation (a means of ensuring that parameter estimates do not move far from initial estimates that are considered to be reasonable; Barnett et al. 2012).

- The second step is quantifying the error in simulations made by the history-matched model.

A model that is carefully calibrated (and/or subsequently validated) in this way should be qualified as a conditionally calibrated (validated) model in that it has not yet been falsified by tests against observational data (Beven and Young, 2013). Conditionally calibrated models are useful for running simulations within the range of the calibration and evaluation data (Barnett et al. 2012), while allowing for their updating in the light of future investigation/research and development or changes in catchment characteristics.

A conditionally calibrated model can be considered a ‘receptacle for expert knowledge’ (Doherty and Moore, 2017), or a ‘good representation of the system of interest’ (Barnett et al. 2012), in terms of:

- The conceptualisation and parameterisation used to represent real world hydraulic properties with effective simplicity (or appropriate complexity); and
- The historical behaviour of the system (as the history match [conditional calibration] constrains parameters to a narrow stochastic range).

Deterministic scenario analysis using a conditionally calibrated model and subjective probability assessment is discouraged as an uncertainty quantification approach due to its questionable subjectivity. However, if it can be established that the conceptualisation and parameterisation is conservative (i.e. over-estimates impact), then a deterministic scenario analysis can be used as a screening tool for further investigation and detailed modelling, or it may be used in qualitative uncertainty analysis for a low risk context (Middlemis and Peeters, 2018). However, it does not necessarily reduce sources of predictive bias that may be introduced via simplification assumptions or via a conditional calibration process that compensates for model defects via biased parameter values in the history-match (conditionally calibrated) model (Doherty and Moore, 2017).

7. Guiding Principles For Uncertainty Analysis

As outlined in Walker (2017) and Doherty and Moore (2017), from a management perspective, modelling is considered to have failed if there is sufficient bias for a poor decision to be made (e.g. by lack of uncertainty analysis or lack of transparency in documentation), especially if the consequence is large. Put another way, if modelling is used to predict that an unwanted outcome won't happen (e.g. via a biased model that overlooks important causal pathways), but it can indeed eventuate (with non-trivial probability), then we should consider that the model study has failed.

Objective uncertainty analysis (qualitative or quantitative) gives end-users confidence that future potential impacts (e.g. threshold impacts exceeded) have been considered carefully in an unbiased way. Explicit consideration of uncertainty is warranted for every groundwater project; it must be considered at the problem definition stage and it should be integrated with the workflow. For low risk projects, it may be acceptable to describe the effect of uncertainty on the project objectives in qualitative terms (e.g. Peeters, 2017). For high-risk projects, a quantitative uncertainty assessment is required in addition to the qualitative assessment.

The following key guiding principles should be used to design a modelling workflow to objectively assess uncertainty. The principles are set out in Middlemis and Peeters (2018) and are based on the discussion papers accompanying this report (Appendix A). The guiding principles are all consistent with the AGMG (Barnett et al. 2012), including engagement with agencies at the outset and at key stages throughout the iterative workflow:

1. Uncertainty analysis is an integral part of a robust risk management framework (risk is defined as the effect of uncertainty on project objectives; AS/NZS 31000:2009), as it informs and complements other aspects such as risk assessment, mitigations/treatments, communicating outcomes and prioritising efforts to reduce uncertainty (e.g. data acquisition).
2. While all projects require a qualitative uncertainty analysis at a minimum (e.g. discussing how model assumptions can potentially affect simulations; Peeters, 2017), high risk projects also require a quantitative uncertainty assessment to a level of detail commensurate with the potential risks and/or consequences of the project (i.e. a preliminary hydrogeological risk assessment is needed at an early stage in the project).
3. It is crucial to explicitly define project objectives and what the model needs to predict in specific and measurable terms that will support unbiased decision making (e.g. applying threshold or trigger impact terms which provide information on which decisions may be based objectively).
4. Modelling methods should be designed to investigate the causal pathways for potential impacts on water resources and water-dependent assets that may arise from a proposed development or management plan, and to quantify the related uncertainties, thus providing unbiased information to support decisions for groundwater management and policy.
5. The methodology should be designed in such as way as to provide information about the uncertainty in conceptualisations and model simulation outputs so as to allow decision makers to understand the effects of uncertainty on project objectives (i.e. consistent with the risk management methods of ISO 31000:2009) and the effects of potential bias. The model must be specifically fit for this purpose.
6. In developing the model(s), a balance must be struck between model simplicity and complexity for the purpose of uncertainty evaluation, commensurate with the risk/consequence profile of the project (i.e. more than one model may be required).
7. The model simulations should be constrained with available observations and information.
8. The range of model outcomes that are consistent with all observations and information should be presented (calibration-constrained model outcomes).
9. Reports should transparently and logically discuss modelling and methodology assumptions and choices, and how they affect simulations; uncertainties and potential bias; and present the results clearly such they are not prone to misinterpretation, to instil confidence in the model simulations with stakeholders and clients (elaborated in more detail in Richardson et al., 2017).

10. Uncertainty analysis results should be: (i) carefully tailored to decision-makers' needs (i.e. based on consultation), (ii) focussed on the messages that are most likely to be relevant to their decisions, and (iii) presented in plain and clear (precise, non-jargon) language.
11. Project workflow should be iterative, revisiting objectives, assumptions, conceptualisations and simulations, as well as the risk assessment (with consideration of any risk treatments applied to mitigate impacts), in a process of engagement between proponents, water managers (agencies) and their technical experts and reviewers that begins at project inception.

Noting that the above principles are procedural rather than technical, another guiding principle on a key technical matter, which was discussed at the workshop, is also proposed:

- Faults may be included as specific model features only where explicit evidence exists, consistent with the AGMG principle of parsimony. Where some minor/inconclusive evidence exists, faults could be considered as part of a sensitivity/uncertainty analysis involving parameterisation of the fault features and consideration of probabilities.

8. Engagement And Communication

A key guiding principle is that engagement with regulatory agencies is required at the workflow outset and at subsequent key stages, to discuss and agree the objectives and methodologies, and to understand the implications of the results. It should be no surprise that this key principle is established in the AGMG (Barnett et al. 2012) and it remains valid. It requires meaningful dialogue between modellers, decision makers (including the proponent), the regulatory agencies and their technical experts.

The subtle change in focus recommended in this report is that, given the complexities involved in uncertainty analysis, engagement can and should occur on a ‘without prejudice basis’, and it requires engagement throughout the investigation, not simply at the end to present the results (Richardson et al. 2017; Barnett et al., 2012).

Richardson et al. (2017) also explore the issue of bias (common forms include availability bias, confirmation bias, confidence bias and framing bias) and the need for transparent and honest communication of information to counter its pernicious effects.

As set out in Richardson et al. (2017), the key to successful engagement and communication is to design and undertake the investigation methodology and present the results and related information about uncertainty in a way that will allow decision makers to understand the effects of uncertainty on project objectives; that is:

- Based on agreed and transparent model objectives;
- Tailored to decision-makers’ needs;
- Focussed on the messages that are relevant to their decisions;
- Presented in plain and clear (precise, jargon-free) language, made fully transparent for independent scrutiny, and not prone to misinterpretation.

Effective engagement means that the modelling objectives need to be discussed early in the project workflow and may require agreement to develop more than one model in order to address what may be quite different objectives (e.g. mine dewatering options based on adequate site data and/or impact assessment at sensitive receptors where data may be sparse). Using one model to address all issues has often delivered sub-optimal results. However, recent advances in software (unstructured grids) and hardware (networked processors) mean that a well-designed one-model approach may be deemed adequate, provided it considers causal pathways and evaluates the effects of uncertainties.

Together, the model impact assessment results and uncertainty analysis should be used by decision-makers as a guide to the likelihood of consequences eventuating (be they beneficial or adverse) and to the assessment (by all parties) of attractive[2] and effective[3] management actions/options.

Positive or negative framing of the reporting syntax can be used in the decision-support context, meaning that expressions should take advantage of this priming (i.e. the direction and expression should be consistent) to reduce cognitive stress for all parties. For example, a 5% chance that drawdown will be greater than 0.2 m is the same as 95% chance that drawdown will be less than 0.2 m; both should be easy to understand by anyone, but the latter could be seen as a more positive framing (or possibly negative, depending on the reader’s viewpoint). Whether to use a positive or negative framing is an intentional decision in the context of a project as it can affect the interpretation of the results. Whichever framing is adopted, it should be applied consistently throughout the reporting.

2D and 3D visualisation of conceptual models and other data can be helpful in explaining complex scientific processes and communicating concepts and simulation results. The AGMG (Barnett et al. 2012) present comprehensive guidance on reporting and visualisation issues.

Richardson et al (2017) augment the AGMG with detailed discussions on effective communication principles (and an 11-point summary), and give examples of clear descriptions for the scale of

[2] Economically and socially acceptable

[3] Able to reduce risk and able to be implemented in a timely fashion

uncertainty using combinations of numeric and visual descriptors integrated with colour, along with narrative descriptors that apply 'calibrated language'. Middlemis and Peeters (2018) have taken the concept further, based on some real-world testing by Dr Merrick (a workshop participant), resulting in the schema outlined in Table 4. It is worth noting that the 10th and 90th percentiles are equally likely (each about 10% probability), or indeed unlikely for that matter. The colour coding from green to red is suggested in the context of likelihood of exceedance, and is designed to aid the narrative descriptions, indicating that the outcome is likely (80% probability) to lie between the green and the red indicators. Some projects may find it helpful reduce cognitive strain by using the alternative description, or perhaps adding it along with the default description. It is also worth pointing out that an 80% probability based on a set of 1000 simulations means that some 200 simulations will have predicted outcomes outside the criteria range selected.

Table 4: Example of a combined numeric, narrative and visual approach to describing likelihood

Probability class (Outcomes ranked from small to large)	**Colour code**	**Narrative descriptor** (In terms of likelihood of exceedance)	**Alternative description** (In terms of likelihood of non-exceedance)
<10%		It is very likely that the outcome is larger than this value.	It is very unlikely that the outcome is smaller than this value.
10-33%		It is likely that the outcome is larger than this value.	It is unlikely that the outcome is smaller than this value.
33–67%		It is as likely as not that the outcome is larger than this value.	It is as likely as not that the outcome is smaller than this value.
67–90%		It is unlikely that the outcome is larger than this value.	It is likely that the outcome is smaller than this value.
>90%		It is very unlikely that the outcome is larger than this value.	It is very likely that the outcome is smaller than this value.

9. Examples

As indicated above, the few examples of comprehensive uncertainty analysis applied to models of complex systems include the:

- Menindee Lakes groundwater modelling investigation (Hayes and Nicol, 2017; Appendix A);
- Bioregional assessments (e.g. Janardhanaran et al. 2016);
- Pilbara Mining Area C iron ore deposit impact assessments (BHPB, 2017a, b);
- GEN3 dual phase water and gas regional modelling of the Surat Basin (QGC, 2013) and the related study of Herckenrath et al. (2015).

The first two examples were discussed during the workshop to help demonstrate in practical ways how uncertainty should be integrated with modelling workflows (rather than as a separate task) and what benefits resulted for decision making. The Pilbara case study is summarised in Middlemis and Peeters (2018), while Herckenrath et al. (2015) investigate key uncertainty issues of the Surat study.

The Menindee Lakes case study (Jacobs, 2016; Hayes and Nicol 2017) used the comprehensive information and data from the well-resourced multi-disciplinary Broken Hill MAR study by Geoscience Australia (Lawrie et al., 2012). An uncertainty-driven workflow was applied by Jacobs (2016) to investigate groundwater supply options and river- and lake-aquifer interactions. Despite the unusually comprehensive information available, the Menindee modelling study implemented an effective simplification approach in model design and execution. Predictive uncertainty was successfully evaluated using PEST and Null-Space Monte-Carlo methods (1000s of realisations), with results presented in terms of probabilities of drawdown impacts and subsequent aquifer recovery. The Menindee study considered and addressed uncertainty issues to a much more extensive and detailed level than is typically undertaken. While not every study may be able to achieve this level of detail, it demonstrates the advantages of best practice methods to investigate uncertainty and guide decision making.

10. References

Barnett B, Townley L, Post V, Evans RE, Hunt RJ, Peeters L, Richardson S, Werner A, Knapton A and Boronkay A. 2012. *Australian groundwater modelling guidelines*. Waterlines report 82. National Water Commission. webarchive.nla.gov.au/gov/20151020044340/http://archive.nwc.gov.au/library/waterlines/82

Begg S. 2013. *Uncertainty, decisions, models and people*. 2nd AusIMM International Geometallurgy Conference (GeoMet 2013).

Beven K. 1993. 'Prophecy, reality and uncertainty in distributed hydrological modelling'. *Advances in Water Resources*, 16, 41–51.

Beven K and Young P. 2013. 'A guide to good practice in modeling semantics for authors and referees'. *Water Resources Research*, 49, 5092–5098.

BHP Billiton. 2017a. *Mining Area C Southern Flank proposal: hydrological impact assessment and water management principles summary* (Appendix 7). www.epa.wa.gov.au/sites/default/files/PER_documentation/Appendix%207%20-%20Hydrological%20Impact%20Assessment.pdf

BHP Billiton. 2017b. *Mining Area C: hydrological impact assessment and water management plan options summary* (Appendix 7; EMP rev6). www.epa.wa.gov.au/sites/default/files/PER_documentation/Appendix%207%20-%20Hydrological%20Impact%20Assessment.pdf

Commonwealth of Australia. 2009.. *The Australian environment act – Report of the independent review of the environment protection and biodiversity conservation act 1999*. October 2009. http://www.environment.gov.au/system/files/resources/5f3fdad6-30ba-48f7-ab17-c99e8bcc8d78/files/final-report.pdf

Doble R and Crosbie R. 2017. 'Review: Current and emerging methods for catchment-scale modelling of recharge and evapotranspiration from shallow groundwater'. *Journal of Hydrogeology*, 25, 3-23.

Doherty J. 2010. *Model-based environmental decision-making*. Watermark Numerical Computing, Australia. http://pesthomepage.org/

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

Doherty J and Moore C. 2017. 'Simple is beautiful'. Discussion paper no.3 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. National Centre for Groundwater Research and Training, Australia.

Ferré T. 2016. *Seeing things differently: Rethinking the relationship between data, models, and decision-making*. 2016 Darcy Lecture Series in Groundwater Science. https://www.youtube.com/watch?v=6YnYGwYSmQk

Freeze RA, Massmann J, Smith L, Sperling T, and James B. 1990. 'Hydrogeological decision analysis: 1 A framework'. *Ground Water*, 28.5,738–766. https://darcylecture2016.files.wordpress.com/2015/08/freeze-et-al-1991.pdf

Guillaume JHA, Hunt RJ, Comunian A, Blakers RS and Fu B. 2016. 'Chapter 28: Methods for exploring uncertainty in groundwater management predictions'. In Jakeman AJ, Barreteau O, Hunt RJ, Rinaudo J and Ross A (eds). *Integrated groundwater management*. Springer International Publishing, New York.

Harrington G, Richardson S, Evans R, Middlemis H and Conesa D. 2010. *Combining science with meaningful stakeholder consultation to achieve an Acceptable Extraction Limit for the Padthaway Prescribed Wells Area, SA*. National Conference Groundwater 2010, Canberra.

Hayes P and Nicol C. 2017. 'Uncertainty by Design: Sufficient model simplification to make predictive uncertainty analysis tractable'. Discussion paper no.4 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. National Centre for Groundwater Research and Training, Australia.

Herckenrath D, Doherty J and Panday S. 2015. 'Incorporating the effect of gas in modelling the impact of CBM extraction on regional groundwater systems'. *Journal of Hydrology*, 523, 587–601.

Jacobs. 2016. Menindee regional groundwater modelling. Prepared for WaterNSW. November 2016.

Janardhanaran S, Holland KL, Gallagher M, Aramini D, Davies P, Merrin LE and Turnadge C. 2016. *Groundwater numerical modelling for the Maranoa-Balonne-Condamine subregion*. Product 2.6.2 for the Maranoa-Balonne-Condamine subregion from the Northern Inland Catchments Bioregional Assessment. Department of the Environment and Energy, Bureau of Meteorology, CSIRO and Geoscience Australia, Australia. https://www.bioregionalassessments.gov.au/assessments/262-groundwater-numerical-modelling-maranoa-balonne-condamine-subregion

Lawrie KC, Brodie RS, Dillon P, Tan KP, Gibson D, Magee J, Clarke JDA, Somerville P, Gow L, Halas L, Apps HE, Page D, Vanderzalm J, Abraham J, Hostetler S, Christensen NB, Miotlinski K, Brodie RC, Smith M and Schoning G. 2012. *Broken Hill Managed Aquifer Recharge (BHMAR) Project: Assessment of conjunctive water supply options to enhance the drought security of Broken Hill, regional communities and industries*. Summary report. Record 2012/15. Geoscience Australia, Canberra. www.ga.gov.au/about/projects/water/broken-hill-managed-aquifer-recharge

Middlemis H, Merrick N, Ross J, and Rozlapa K. 2001. *Groundwater Flow Modelling Guideline*. Prepared for Murray–Darling Basin Commission by Aquaterra, January 2001. www.mdba.gov.au/sites/default/files/archived/mdbc-GW-reports/2175_GW_flow_modelling_guideline.pdf

Moore C, and Doherty J. 2005. 'The role of the calibration process in reducing model predictive error'. *Water Resources Research* 41.5.

Middlemis H and Peeters LJM. 2018. *Uncertainty analysis – Guidance for groundwater modelling within a risk management framework*. A report prepared for the Independent Expert Scientific Committee on Coal Seam Gas and Large Coal Mining Development through the Department of the Environment and Energy, Commonwealth of Australia. http://www.iesc.environment.gov.au/publications/information-guidelines-explanatory-note-uncertainty-analysis

Middlemis H. 2011. *Uncertainty*. IAH NSW Annual Conference.

New Zealand Government (Ministry for the Environment and Ministry for Primary Industries). 2014. *A draft guide to communicating and managing uncertainty when implementing the National Policy Statement for Freshwater Management*. www.mfe.govt.nz/publications/fresh-water/draft-guide-communicating-and-managing-uncertainty-when-implementing

Pappenberger F and Beven K. 2006. 'Ignorance is bliss: or seven reasons not to use uncertainty analysis'. *Water Resources Research*, Vol. 42, W05302.

Peeters L. 2017. 'Uncertainty analysis in groundwater modelling: where to start?' Discussion paper no.2 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. National Centre for Groundwater Research and Training, Australia.

Poeter E. 2006. *All models are wrong: how do we know which are useful?* Darcy Distinguished Lecture. http://inside.mines.edu/fs_home/epoeter/DarcyLecture2006/

QGC. 2013. *Stage 3 Water Monitoring and Management Plan*. 6.0 Numerical Groundwater Flow Modelling – GEN3. http://www.shell.com.au/about-us/projects-and-locations/qgc/environment/water-management/reports/_jcr_content/par/expandablelist_48b1/expandablesection_ea.stream/1498083754692/470349bbd120a02eb1e1eb20943e1fbc89fb72071d0ed88d2341dc0cad9fcdeb/qgc-stage-3-wmmp-dec-13-6.pdf

QLC24. 2017. *New Acland Coal Pty Ltd v Ashman & Ors and Chief Executive, Department of Environment and Heritage Protection*. http://archive.sclqld.org.au/qjudgment/2017/QLC17-024.pdf

QLC 48. 2015. *Adani Mining Pty Ltd v Land Services of Coast and Country inc. & Ors*. Land Court of Queensland. www.austlii.edu.au/cgi-bin/sinodisp/au/cases/qld/QLC/2015/48.html

Voss CI. 2011. 'Editor's message: Groundwater modeling fantasies — part 1, adrift in the details'. *Hydrogeology Journal*. 19, 1281–1284.

Voss CI. 2011. 'Editor's message: Groundwater modeling fantasies — part 2, down to earth'. *Hydrogeology Journal*. 19,1455–1458.

Walker G. 2017. 'Predictive uncertainty in groundwater modelling: how is it best used?' Discussion paper no.1 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. National Centre for Groundwater Research and Training, Australia

APPENDIX A

Discussion papers on groundwater modelling uncertainty:

1. Walker G. 2017. 'Predictive uncertainty in groundwater modelling: how is it best used?' Discussion paper no.1 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. National Centre for Groundwater Research and Training, Australia.
2. Peeters L. 2017. 'Uncertainty analysis in groundwater modelling: where to start?' Discussion paper no.2 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. National Centre for Groundwater Research and Training, Australia.
3. Doherty J and Moore C. 2017. 'Simple is beautiful'. Discussion paper no.3 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. National Centre for Groundwater Research and Training, Australia.
4. Hayes P and Nicol C. 2017. 'Uncertainty by Design: Sufficient model simplification to make predictive uncertainty analysis tractable'. Discussion paper no.4 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. National Centre for Groundwater Research and Training, Australia.
5. Richardson S, McMillan M and Currie D. 2017. 'Communicating uncertainty to decision makers and stakeholders'. Discussion paper no.5 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. National Centre for Groundwater Research and Training, Australia.

Predictive uncertainty in groundwater modelling: How is it best used?

Glen Walker[1], Member, Independent Expert Scientific Committee on Coal Seam Gas and Large Coal Mining Developments, [1] currently at Grounded in Water, Adelaide

Abstract

The external expectation is that statements on groundwater modelling should be made transparently, honestly and with precision. Given deficiencies in our knowledge of underground structures and properties, this would require some form of uncertainty analysis on the groundwater modelling. This uncertainty analysis would normally need to fit into a risk management framework and complement other aspects of this such as risk assessment, risk mitigation, reducing uncertainty, communicating outcomes and prioritisation of effort. While the output of uncertainty analysis is ideally a probability distribution, this may not be strictly feasible nor required. Embedding the risk assessment framework into an iterative or hierarchical process provides an approach for finding the appropriate level of resourcing required to meet a need, and for refining objectives, model conceptualisation and risk treatments. Over-complexity of groundwater models can impede the application of uncertainty analyses by requiring unnecessary overheads. Over-complexity can be perceived to be improving defensibility of models and modelling outputs yet possibly making projections worse. We may be better using a suite of simpler groundwater models rather than a single complex model to address a range of management objectives. The main impediment to appropriate uptake of uncertainty analyses may well be understanding how to incorporate uncertainty methods into the decision-making processes.

Making decisions under uncertainty

Risk assessment and mitigation and adaptive management frameworks (AS/NZS 2009) are now widely used and accepted for making decisions and managing impacts under uncertainty. They are designed to meet management objectives despite some deficiency of knowledge. While imperfect knowledge is often used as a reason for deferring decisions until more information is available – on the basis that it can prevent poor outcomes – it can also incur costs. Policies, such as the precautionary principle under the ecologically sustainable development (ESD) framework specifically address this by stating that scientific uncertainty should not be used as a reason to postpone actions to prevent irreversible environmental degradation. Risk frameworks often support flexible management arrangements that allow actions to be taken with imperfect knowledge, while collecting further information and enabling management to be adjusted in response to this information. Risk assessment and management frameworks for programs that affect the environment are required to address ESD principles more widely by balancing potentially conflicting economic, social and environment management objectives.

The high profile of global issues such as climate change and controversial local developments have raised awareness amongst environmentalists, industry, regulators and the community of uncertainty and risk and how this influences the balance between conflicting management objectives. This has raised expectations that scientific results are presented in an honest, precise and transparent fashion. It is difficult to see how this could be achieved without some form of analysis of scientific uncertainty and the resulting confidence of conclusions and recommendations. The need for transparency has been emphasised by the growing distrust of government, scientists and industry. There are drivers for and occurrences of both understatement and overstatement of uncertainties, in some cases deliberately aimed at undermining the science. If decisions on the appropriate balance of social, economic and environmental outcomes are to be made on the best evidence rather than distorted or ignored science, this issue needs to be addressed.

The climate science community, through the Intergovernmental Panel on Climate Change, has tackled this through clear and reviewed statements on the identification of uncertainties and the implications of these uncertainties. Recent court cases have shown that groundwater is not immune to these issues and are a timely reminder of the importance of reviewing current practices. Fortunately, community expectations reinforce scientific principles of analysing uncertainty and confidence of conclusions.

Risk management requires that the full range of impacts of decisions and their likelihood are considered. Modelling is usually required to support this by indicating the future likelihood of any impact and attributing responsibilities for this impact. Where groundwater forms an important pathway linking actions to future impacts, groundwater modelling is used, usually with numerical models on established platforms. The bias and spread associated with modelling projections is called predictive uncertainty and the process of determining this is uncertainty analysis. Predictive uncertainty includes the impact of our deficiency of knowledge of processes causing the impacts, the ability of the models to capture the main processes and the uncertainty due to random processes. From a management perspective, the modelling is considered to have failed if there is sufficient bias for a poor decision to be made on the basis of the bias. This failure becomes a greater concern if the consequence of this is large.

This paper is the first of a series of discussion papers that form the basis of a workshop to discuss current practices in uncertainty analysis for groundwater modelling and how they can be improved. It aims to provide a context for the other papers, which deal with the mathematics, algorithms and logic process to address predictive uncertainty and how to communicate the outputs of an uncertainty analysis.

This paper considers groundwater modelling to support risk assessment and management. While groundwater modelling can be used for purposes other than risk assessment (e.g. improvement of system understanding), major applications such as groundwater allocations, salt interception, mining and gas developments and seawater intrusion usually involve risk assessment and management, e.g. DEWNR, 2012).

The difficulty of assessing underground properties means that there can be quite different hydrogeological conceptualisations of the connectivity between different components of the hydrological system, the relative importance of different processes for most predictions and the estimation of parameters. This would suggest that predictive uncertainty should be an important part of the groundwater modelling process. However, until relatively recently, the normal approach has been to use a single ‘fit-for-purpose’ model without any uncertainty analysis. This means that it is difficult to provide definitive statements on the confidence of any predictions, and more than this, it connotes an overconfidence in the predictions. This is sometimes partially addressed through the use of a deliberately biased model is to provide the ‘best case’ or ‘worst case’ scenario.

Over the past ten years, there has been an increase in the use of uncertainty methods in Australia. There has also been increased availability of software, expertise and experience. Courses are being run in the use of such software. A chapter on uncertainty has been included in the *Australian groundwater modelling guidelines* (Barnett et al., 2012). There are also now some documented and practical case studies at reasonable scales including the Surat Basin (QWC, 2012) and Bioregional Assessments (e.g. Cui et al., 2017). A further example, the Menindee Lakes model is discussed in companion paper 4 (Hayes and Nicol 2017). Improvement in all these adoption processes is still required to raise the skill level of the groundwater modelling community. The inadequacy of uncertainty analyses in reports would suggest that the impediments to wider use of uncertainty analyses are not necessarily all technical in nature. This paper discusses some of the non-technical issues. To explore these, it is first necessary to understand in more detail how modelling and uncertainty analyses are integrated with the risk framework.

The application of risk and adaptive management frameworks to groundwater management

The management and mitigation of risk can occur in different ways, depending on the nature of the risks. These are discussed in more detail elsewhere (AS/NZS, 2009). This section gives a sense of how risk is managed in groundwater, using groundwater allocation to exemplify concepts.

One of the principles of risk frameworks is to gather information and adjust management on the basis of this information. This is often referred to as ‘adaptive management’. In one end member of risk management, the system under scrutiny is treated as a black box (a transfer function), in which the responses will lead directly to adjustments of the levers through rules, without direct reference to the details of the groundwater system. This is analogous to driving a car without knowing how the engine works. The rules may be reviewed periodically, and this may entail an investigation of the system. An example of this in groundwater management is water level response management (WLRM). (McIntyre and Wood 2011). Management is adjusted and even optimised in response to monitoring data to keep within the target range of objectives. There are commonly large time lags between actions and impacts, and this, together with high administrative costs, means that approaches such as WLRM can be difficult to implement. WLRM is an example of risk avoidance, using essentially one lever i.e. allocations.

The other end member of adaptive management is scenario planning. This is generally built explicitly upon hydrogeological conceptualisation and groundwater modelling. Scenarios refer to the means in which the future unknowns are handled. Scenarios could include assumptions on commodity prices, government policies, technological innovations, climate and social behaviour. Groundwater planning using plausible scenarios provides a transparent approach for dealing with uncertainties associated with the future. Scenario planning allows various contingencies within plausible futures to be considered. It also allows various management options to be compared under the same assumptions. Scenario planning is often used at the initial stage of development and then refined periodically as part of the planning cycle. Scenarios are not critical for conducting adaptive approaches such as WLRM although they are sometimes used to build robustness into the management plan.

An example of the use of scenario planning is a groundwater sharing plan that uses a sustainable yield determination. Groundwater sharing plans tend to have conflicting objectives: providing a secure operating environment to the industry, optimising operation costs, minimising government administration and intervention, and protecting environmental objectives. Adjustments are made as part of the planning cycle, often less frequently than for WLRM. Significant adjustments can be politically and socially difficult. Conceptualisation and some form of groundwater modelling is used to help determine an extraction regime that allows management objectives to be achieved.

Risk is managed in a range of different ways. A precautionary approach is often used, especially in early stages of development (Figure 1). This can occur by lower allocations and rules such as buffers. Monitoring triggers that indicate when the groundwater is responding outside of the predicted envelope become more important as development is increased. For example, New South Wales local management plan rules use water level and quality triggers based on reduction of saturated aquifer thickness, change in groundwater salinity, levels at sites of groundwater dependent ecosystems or in the vicinity of streams. When the monitored water level or quality passes some pre-determined threshold, it triggers a review of the plan. Often problems to do with groundwater allocation can occur at hotspots, rather than across the whole plan area. Local rules allow flexibility to deal with not only unpredictability, but also hotspots, while the extraction limit deals with the broader plan area. The use of sustainable yield allows simpler administration and by providing security to groundwater users, it supports investment. The political and economic difficulty in reducing allocations can cause a risk to environmental objectives, including protecting the groundwater management objectives. This is more likely to occur when the demand for groundwater places pressure on using what is perceived to be an overly precautionary approach.

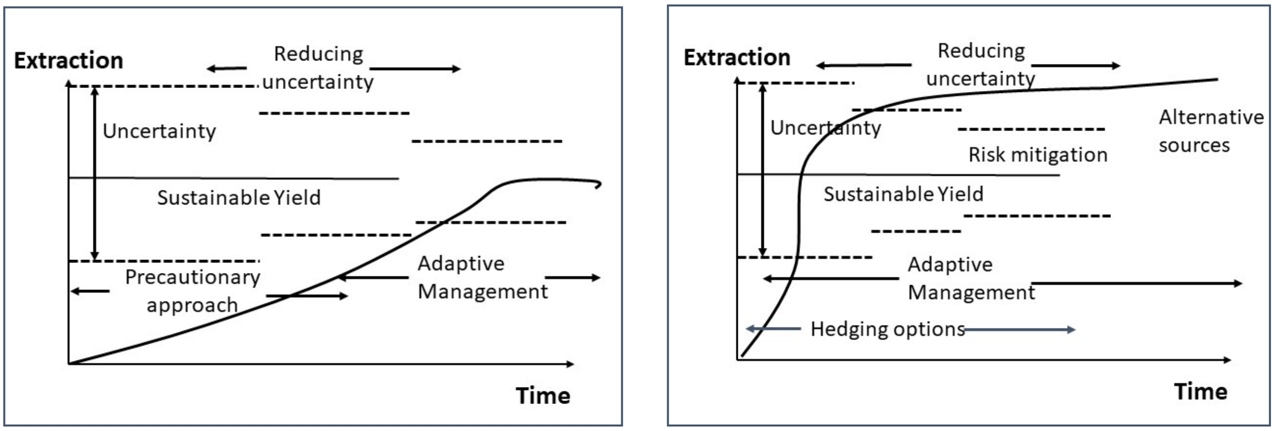


Figure 1: How risk management might vary along a timeline of increasing extraction in order to meet management objectives as represented by the sustainable yield. The graph on the left represents a situation in which there is a large number of decisions being made gradually, while the graph on the right shows a situation where a decision needs to be made on a major development.

Risk can also be managed through mitigation or acceptance, rather than by avoidance. All groundwater extraction will cause some impacts, but the benefits from groundwater extraction mean that the government and community may accept some adverse impacts. Even when a management objective is set through a resource condition, it may be agreed to later vary this for the greater good. The variation of resource conditions can be associated with compensation or offsets. In some cases, there may be agreement to maintain resource conditions but find alternative water supplies through desalination or transportation of water from elsewhere. In some cases, the risks may be mitigated. For example, salt interception schemes mitigate the salinity impacts on streams that might occur from surface-water irrigation, and managed aquifer recharge may mitigate the environmental impacts of high allocation.

The availability of attractive and feasible risk treatments allows more than one lever to be used and gives more opportunities for uncertainties to be addressed in an adaptive fashion. If monitoring and scientific investigations show that the groundwater is not responding as originally predicted, an alternative management option might be implemented. All these options will take some time to prepare, including obtaining government approval and social acceptance, purchase of any land required and design and financing of the mitigation. Fortunately, the large time lags associated with groundwater processes may enable sufficient time for some of this preparation if the monitoring was designed to provide early warning of any impacts. However, if the option was more contentious, some degree of approval will be required much sooner, possibly at the start of the development or program, to allow alternative options to be considered. Thus, alternative options are included in the initial stages of programs ('hedging') and as better information is obtained, options are narrowed. This may allow a less precautionary approach to be taken.

For some environmental assets, such as for groundwater-dependent ecosystems and especially for discharge springs, there may be no feasible or attractive mitigation scheme. Where ecosystem values are high and sensitive to groundwater pressure changes, the precautionary principle would suggest that the likelihood of a groundwater impact should be demonstrated to be very low before initiating any action. In such circumstances, large time lags can be a significant disadvantage, as turning off the pumps as impacts are observed may be too late to prevent environmental damage. The monitoring would need to be structured to provide sufficiently early warning to modify pumping strategies. If the likelihood of an impact at the start of the program is sufficiently high, there may be a need for closer scrutiny before program is approved and monitoring and management plans approved at the same time.

In contrast to sustainable yield approaches, WLRM emphasises environmental objectives at the cost of security of groundwater supply and administrative efficiency. Addressing different management objectives in different ways would generally reflect the risk associated with each objective. WLRM tends to be used where environmental values are high and the likelihood of impacts is greater. As demand increases, it may make sense, where high environmental values are important and time lags shorter, to abandon the precautionary approach by changing from a sustainable yield approach to WLRM. This would transfer the risk from the environment to groundwater users.

Thus, it can be seen that risk can be managed by avoidance, acceptance, transfer and mitigation. This allows uncertainty to be managed by the use of scenario planning, taking a precautionary approach, hedging of options and optimisation through adaptation. While adaptive management should allow adjustment of plans, this is reliant on active seeking of information and review of the conceptual understanding. Better knowledge will allow the reduction of available options, the availability of which can be expensive to maintain. Some upfront planning will always be required, including contingencies, but the level of required planning will depend on the residual risks, i.e. the risk assessed when management is considered. Where the residual risk is high, a greater upfront effort is required.

Defining the objectives of uncertainty analysis for risk management

To define the objectives of uncertainty analysis for an issue, we need to consider the risk assessment and management framework in its entirety, including the balance between determining uncertainty; reducing uncertainty through monitoring and scientific investigations; and managing or treating risk, at times in response to new information (Figure 2). Each situation is different, and the balance of these components will vary. An uncertainty analysis is required to support each of these elements and conversely, the balance of these elements will determine the objective of the uncertainty analysis and the level of effort required. This section describes this interaction in more detail.

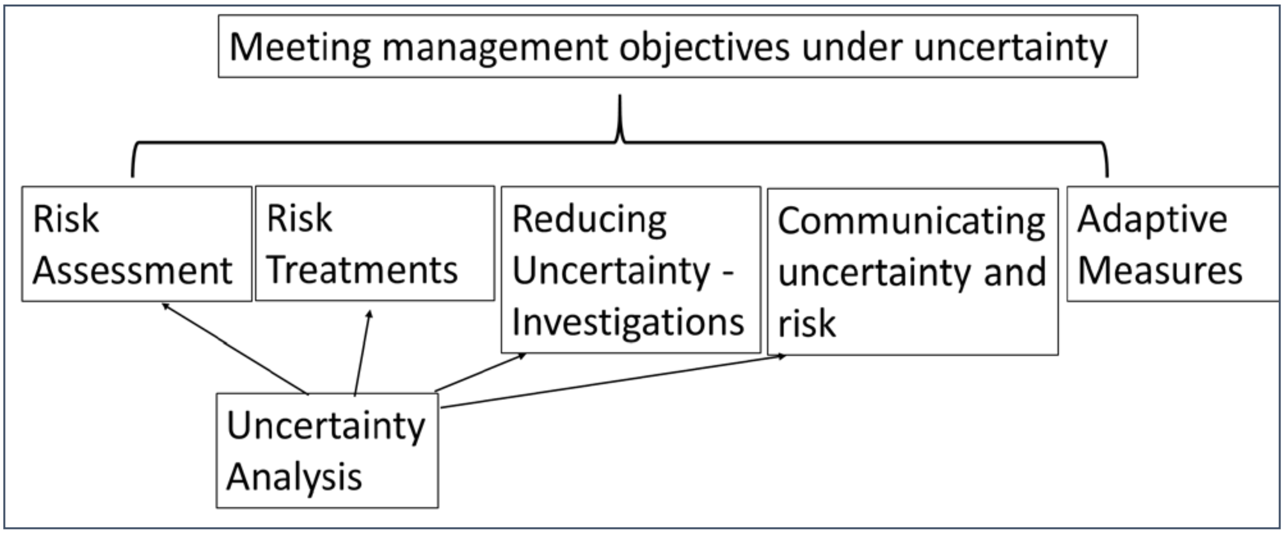


Figure 2: How uncertainty analysis may support different elements of a risk management framework

Groundwater modelling of some kind provides the only means for the future impact of an action to be assessed, including whether management objectives are met. Outputs from groundwater modelling that relate directly to a resource condition can be compared to thresholds that indicate a management objective. For example, if the management objective is not to have water table drawdowns of greater than 5 m, this can be tested using a model. Not all management objectives are in the form of groundwater pressures, fluxes and groundwater. In some cases, e.g. ecological resource condition limits, post-processing of groundwater model outputs using models or regressions will provide an indication of management objectives.

In testing management objectives, there needs to be recognition that outputs from a groundwater model will have a bias. This can happen as a result of deficiencies in knowledge or as a result of the modelling not adequately representing what we know. Figure 3 schematically shows how a groundwater models can wrongly indicate impacts on management objectives. The true solution represents what happens if the real world was subject to the management actions. Given that we do not know this, we do not know the bias. Thus, without any uncertainty analysis, we are unsure of whether the management objective is met or not. If we use only one model, as is often done in practice, there can be an undeserved level of confidence attached to the prediction. Often the one model output chosen is one that in some sense optimally fits historical data. However, as will be seen, that can still lead to a large bias. For risk assessment, we need some indication of the likelihood that the management target is met.

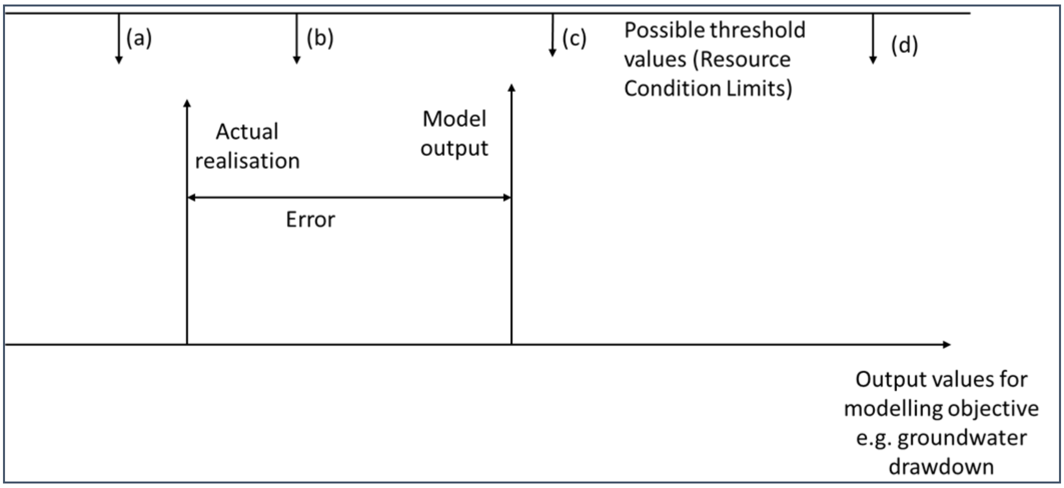


Figure 3: How an error in a modelling projection for a particular management objective might change a decision based on whether the projection is above a threshold value used for the decision. In (a), (c) and (d), the model projection would lead to the correct inference to base a decision upon. In the case of (b), the model would lead to an incorrect inference. Normally, we do not know the actual realisation and hence any estimate of the error.

One way of giving a sense of the uncertainty is to consider a range of models that generally fit our known knowledge of hydrogeology and other processes. This could be done by varying parameters within a possible range or varying the hydrogeological structure. The outputs of this are shown in Figure 4.

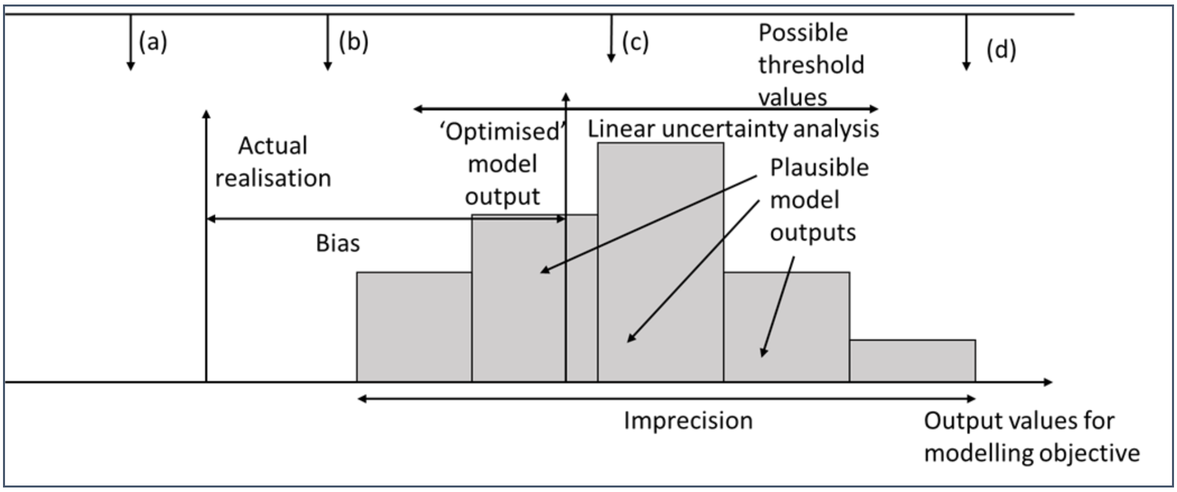


Figure 4: How an ensemble of model projections might affect the decision-making process. This gives both a bias and a spread. Normally we do not know the actual realisation and hence the bias. If the model is properly developed, the actual realisation should fit in the spread of projections. If not, as shown here, it may lead to a poor decision being made, because of a bias in predictions leading to the prediction being on the wrong side of a threshold value (case b).

A range of model outputs leads to both a bias and a spread. An uncertainty analysis should provide us with some indication of the spread and/or bias. As we do not know the true answer, we generally only know the spread. By consideration of appropriate structures and parameters, the true answer should lie within the spread of outputs. One can see that consideration of a spread produced by varying a small range of parameters in a poorly structured model may produce a spread that does not include the true answer. We would expect that the spread and bias would decrease as more information is obtained from monitoring.

Figure 5 shows different ways uncertainty may be displayed. Ideally, it would be good if the analysis led to a full distribution of possible outputs that relates to our current knowledge. While uncertainty methods may produce distributions, the discussion above shows that these may not represent the true uncertainty. A feature of interest is the extremes of our distributions that generally will be represented by the 5th and 95th percentiles. Scenarios help us represent plausible futures within the modelling. It is difficult to assign a probability to these scenarios, except as broad ranges of likelihood. This large uncertainty propagates to the prediction uncertainty. Similarly, we could provide some sense of a likelihood to faults or levels of connectivity between two points and hence represent it in the models, recognising we do not fully understand the hydrogeology or represent it correctly in the model. However, as we do not know what we do not know, it becomes difficult to provide a metric in regard to this uncertainty.

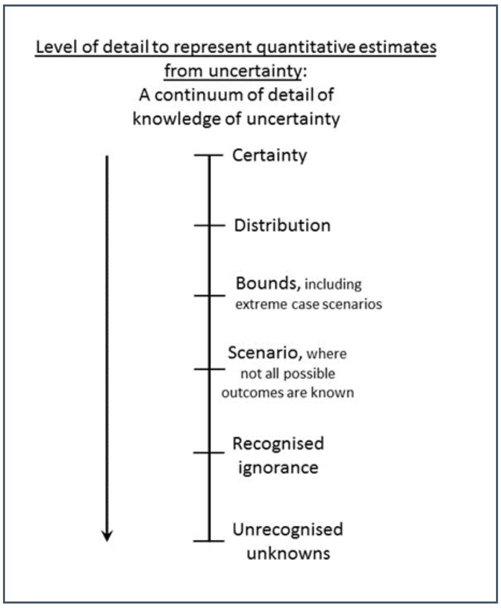


Figure 5: How different levels of uncertainty can be represented in different forms (from Guillaume et al., 2016)

The metric required for a risk framework will depend on the context. The broad aim of the risk framework is to make the program of actions robust against surprises. Sensitivity analyses can show whether an impact is sensitive to a specific factor and hence allows methodologies to be developed that facilitate robustness. Uncertainty analyses add to sensitivity analyses by using information on whether that factor is likely or not. For example, the modelling output may have some measured sensitivity with a given parameter. By adding information on the probability distribution of values for that that parameter, one can limit the range of outputs. Sensitivity analyses generally apply to scenarios and conceptualisations.

Scenario planning requires uncertainty analysis in order to understand the distribution of risk (risk assessment) and to develop contingencies to manage potential risks. Trigger levels to be applied and management responses need to be defined. Where there are no attractive or feasible management options, risk needs to be avoided. Precautionary approaches require information about the extremes. Under the precautionary principle, the objective is for a high unlikelihood of irreversible damage to an important asset. The higher the residual risk to the asset, the greater the confidence is required that management objective is met. To reduce the number of options, there often needs to be a high degree of confidence that an option is not required. Some modellers deliberately apply a bias ('conservatism') to provide an indication about the extreme (Ferré, 2017).

Adaptive management requires information to be continually gathered, often by monitoring. Uncertainty analysis can provide a means to assess where scientific uncertainty can be reduced in the most relevant areas. Choosing sites of monitoring to provide early warning systems relies on knowledge of structures that limit the rate at which pressure changes occur. More knowledge about faults or zones of high connectivity could significantly limit scientific uncertainty with a focussed piece of work. Determining the source of groundwater for springs and sensitivity of ecosystems to groundwater changes provide clearer information about the management objectives and the associated risk. The volume of water needed to be extracted to depressurise a coal seam affects not only the volume of co-produced water (and salt) that needs to be managed but also the impacts on other aquifers.

Approaches such as WLRM are also aided by groundwater modelling and uncertainty analysis. While such an approach protects environmental objectives, there is a risk to the security of other users and the efficiency of operating the plan. Groundwater modelling can help assess these other management objectives. Uncertainty is important for assessing these risks, which in turn, is required for the risks to water-dependent industries. Groundwater modelling is also required to define and review trigger levels, and together with sensitivity analysis prevents the water levels moving outside target range.

If other sources of water are being considered as part of regional planning, information will be required on the security to different water users. Confidence of model predictions needs to be communicated to the users more generally, particularly where there is contention. This requires appropriate language to communicate confidence, where there may be little trust. There is also a need for as much consistency as possible. Communication is considered more in the workshop discussion paper 5 (Richardson et al., 2017).

This discussion has shown that uncertainty analyses are not uniform in terms of objectives, outputs and level of effort. Risk and uncertainty need to be described clearly when defining the objectives of a groundwater modelling project or indeed the larger project for which the groundwater modelling is being done. This is likely to lead not only to better outcomes but also to better targeting of resources.

Lack of resources is a common reason for uncertainty not to be analysed. Low risks mean that greater resourcing would be difficult to justify. Systems where risks are high – and where there are a lack of attractive and effective risk treatments and time lags are long – demand a more detailed level of analysis. Without commensurate funding, there are reputational and political risks. In between these two limits, there is a range of possibilities.

While not comprehensive, this is a sensible level of analysis and associated burden on a development. It is important to also understand whether uncertainty analyses from the supply (modelling and hydrogeology) side can be scaled proportionately to the prioritisation of resources from the demand (purchaser and regulator) side. We address this issue later, after first discussing the sources of uncertainty.

Sources of uncertainty

Predictive uncertainty includes both our uncertainty about processes and the ability of the model to capture the processes that we do know about. We tend to divide the sources of uncertainty into the following categories (Figure 6):

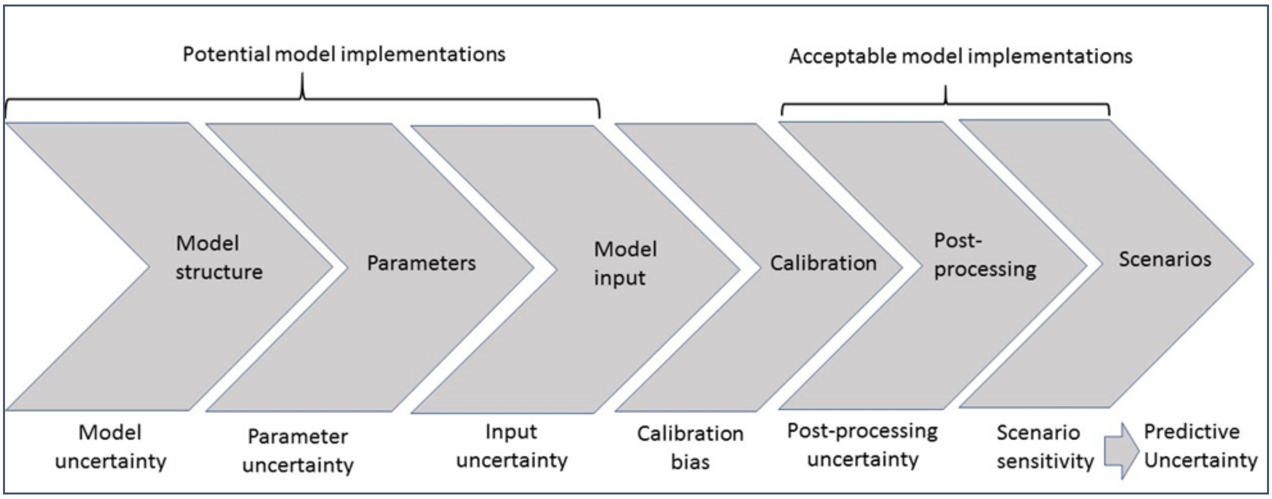


Figure 6: How different elements of the modelling process contribute to predictive uncertainty

- **Model structure:** We are generally interested in issues where future stresses are different from historical stresses. Hence, there is usually a need to use physically-based models. The groundwater model is a gross simplification of the real world, and structural error relates to the effect of the simplification on predictions. Structural error includes assumptions on which processes and features are included and how.
- **Parameters and inputs:** Hydraulic properties that affect groundwater movement are not well known. As the model is a simplified version of the real world, parameters tend to be ‘lumped’ rather than measurable quantities. Parameters can be derived or inferred in different ways: measurements at the site, transference from nearby sites, values from the literature and calibration. A prior distribution of parameters can be derived from the first three methods, while calibration can constrain these

distributions further. Groundwater models use time series of climate, recharge and groundwater extraction as inputs. Any bias in these inputs can lead to biases in predictions. Inputs can themselves be parameterised functions.

- **Post-processing:** The outputs from a groundwater model include pressures, fluxes and groundwater balances. However, management objectives may relate to impacts on groundwater users, ecological values, low flow, economic costs, sometimes at different scales to the modelling. Hence, there can be uncertainties due to downscaling, conversion to other values and aggregation. Fortunately, regional groundwater outcomes at a site are necessary for other impacts to occur. Hence, such uncertainties become irrelevant if regional groundwater impacts at the given site are very unlikely and groundwater policies tend to be written in terms of groundwater model outputs. However, care needs to be taken to ensure that the thresholds in these outputs reflect both risks and timeliness with respect to risk treatments.
- **Calibration**: This is the process in which the combinations of model structures, parameters and inputs are constrained to those whose outputs acceptably fit historical data. Optimisation can be used to find the combination that best fits this data. Such an optimisation can lead to problems of 'non-uniqueness' and bias. The optimisation process can be modified through regularisation to remove the worst features of non-uniqueness, but these modifications can then become a source of uncertainty. Parameters are varied based on the sensitivity done during calibration to provide a range of outputs as the basis for a metric for predictive uncertainty. Because future stresses often differ from historical ones, the suite of parameters that optimise the fit to historical data may not provide the best predictions. Monte Carlo methods select from the large number of possible combinations to identify a suite of plausible parameters and implementable models. The range of outputs from these various models also provides a useful metric for predictive uncertainty.
- **Scenario error:** We don't know the future, so plausible scenarios will include errors.

Potential for scaling the effort in uncertainty analyses

Uncertainty analyses have sometimes been presented as 'all or nothing'. By this, we mean that the calibrated 'fit-for-purpose' models are then used in an uncertainty analysis. This can be an expensive exercise and hence might not scale according to budgets and need. In this section, we consider three ways this might be addressed.

One of the methods used by risk frameworks in other domains is embedding the assessment process in an iterative or hierarchical process. This is shown conceptually in Figure 7. Initially, a preliminary risk assessment is done, possible risk mitigations are considered, and the model is conceptualised to meet the objectives. This is discussed in more detail for groundwater modelling by Guillaume et al. (2016) and in discussion paper 2 (Peeters 2017). In going through more iterations or steps down in the hierarchy, there is a winnowing of the objectives according to risk, and complexity is added as necessary. In the preliminary stages, there may not be any need for numerical modelling. If risks are not high, nothing more is required, and resourcing is stopped. The process could, in some cases, take several years or it could take a few months.

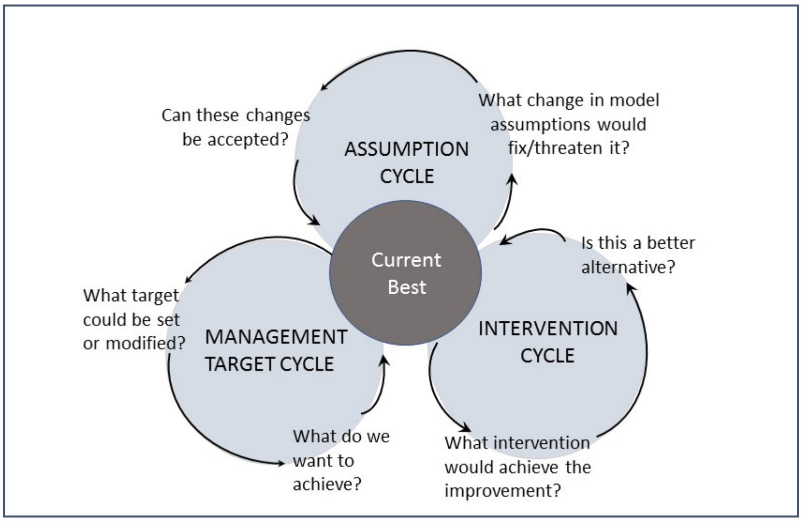


Figure 7: An iterative approach for groundwater modelling involving setting objectives, risk mitigation options and modelling conceptualisation (from Guillaume et al., 2016)

To make such a framework able to be implemented, it is necessary to discuss two issues in more detail i.e. complexity and the use of more than one model. Complexity refers to the representation of processes in the model relative to the information available. The adage that 'models should be as simple as possible, but no simpler' applies here. Additional complexity leads to greater effort required to set up and calibrate models, analyse outputs and revise the models, and can be less transparent. They are also more difficult to stabilise and automate, thus making uncertainty analyses difficult and expensive. Model structure can lead to significant bias. While they are often the source of dispute, they are not often analysed due to the time required to alter and recalibrate models.

Additional complexity may better represent real processes, but does not necessarily result in better predictions. The workshop discussion paper 3 (Doherty and Moore 2017) shows that due to the limitations in the information content in historical data, there is an optimal number of parameters (or combinations) that can be fitted to that data to reduce predictive uncertainty. Trying to fit more parameters ('overfitting') leads to increased predictive uncertainty. Sensitivity analyses may be used on the prior distributions of non-fitted parameters as part of developing a metric for predictive uncertainty. Complex models may provide an educational role in investigating sensitivities and feedbacks. However, because of the predictive uncertainty, they do not necessarily provide better projections despite the additional overheads in using them. Greenfield sites, where there has been little previous development, may have little information relevant to parameters. However, such data and information should become available quickly after development. If development is agreed upon after assessment of risks, regular groundwater updates are important, as the ability to constrain groundwater parameters should improve in parallel with monitoring.

One of the arguments for complex models is that they appear to be more defensible to criticism. It is relatively easy to identify seemingly important processes that are omitted in the model, but less obvious to identify the 'sweet spot' in regard to complexity. While the notion of 'parsimony' is well-ingrained in groundwater modelling, there are various worldviews as to where the 'sweet spot' is. The workshop discussion paper 3 (Doherty and Moore 2017) shows an objective way to view this issue. There may well be a need to review the concept of parsimony for uncertainty analyses to be sensibly done.

Monte Carlo methods produce many plausible model 'implementations'. The culture of having one 'fit-for-purpose' model is well-embedded. Custodians of many of the groundwater models are often subject to diminishing resources. The notion of one model that is used to support decisions such as allocations has been important in defending decisions. A single model could be chosen based upon the percentiles (e.g. 50th or 95th) against one objective, but the actual choice of model is likely to vary for different objectives. A model based on optimisation with historical data may not be the best for predictions. It should not be surprising that the one model may not be suitable or ideal for all objectives. Trying to achieve one model suitable for use for all modelling objectives may lead to unwanted complexity. To make robust decisions, alternative conceptualisations may need to be trialled or different weighting on different objectives needs to be considered. Some disciplines, such as climate and hydrology, use ensembles of models built on different platforms to provide increased confidence in outputs. Some (e.g. Ferré 2017) have argued that this is the way that groundwater modelling should proceed. The notion of more than one model being used should therefore not be surprising. The flexibility afforded by multiple models may be required to implement the iterative process described above. The issue of multiple models is further discussed in the workshop discussion paper 3.

While these three issues are important in ensuring resources are justified, it is likely that including uncertainty analyses will sometimes be more resource intensive. Common sense would dictate that adding iterations or using more models will not necessarily reduce required resources. In a world of finite funding, this means uncertainty analyses will be dependent on priority.

Changing the groundwater culture

Some of the inertia on the introduction of uncertainty analysis is caused by non-technical issues. The groundwater community has been using numerical groundwater models for some time without uncertainty analyses. They have become familiar with the notion of relatively complex fit-for-purpose models. Administrative processes have been built around these. Some models support legislation, others form the basis of inter-jurisdictional agreements. Understanding uncertainty, and how it can be used – rather than be abused – will take some time. Some of these models may be expensive and slow to change over to models that incorporate uncertainty. Communication with a range of stakeholders for them to understand any changes in modelling processes and how this may impact on decisions will be even slower. It is therefore likely to be a staging of the introduction of uncertainty techniques, starting with high risk, contentious new developments. Workshops targeted at non-modellers may be useful in supporting this.

Some other processes, such as peer review and modelling guidelines checklists may unduly ingrain poor practices with respect to complexity and unnecessarily constrain practices. Thus, while the guidelines have undoubtedly raised the general standard of modelling, quality assurance and model documentation, this may have been at the cost of constraining innovation and implementing some of the changes that uncertainty analysis requires. It may be that most gains in future might be made by tailoring the modelling guidelines to meet decision-making frameworks and focusing on the outcomes of the modelling rather than refining and prescribing modelling processes. By doing this, the framing of objectives, prioritisation of resources, iteration of the risk management system and tailoring of outputs may be more naturally addressed. It may also encourage different approaches while doing away with superfluous processes that do not improve decision making.

Despite the modelling guidelines requiring uncertainty analyses and the availability of courses on the topic, it is clear that there is a need to continue to raise skills across the groundwater community. There is no textbook on uncertainty methods, software can be difficult to use and terminology difficult for the novice. Targeted workshops, mentoring as part of on-the-job work and other approaches will continue to be required.

Finally, there is one issue where perhaps the pendulum has swung too far. Adaptive management is justifiably used to address environmental issues. However, the large time lags and lack of risk treatments for groundwater actions can mean that once an action is taken, it may be difficult to reverse the impacts. Hence, by the time monitoring shows that a significant ecological asset will be affected, it may be too late to act. Turning off the pumps will not solve the problem. In such situations, there needs to be very careful understanding of the system and a significant sensitivity analysis before proceeding. Any lack of details of risk treatments in the initial environmental impact statements can be misleading in regard to the residual risk and in the ability to adaptively manage. In such cases, the likely adaptive management strategies should be included in the scenario modelling and trialled as part of the initial assessment.

References

AS/NZS. 2009. *AS/NZS ISO 31000:2009 Risk management – Principle and Guidelines*. Jointly published by Standards Australia, Sydney and Standards New Zealand, Wellington.

Barnett B, Townley L, Post V, Evans R, Hunt R, Peeters L, Richardson S, Werner A, Knapton A and Boronkay A. 2012, *Australian groundwater modelling guidelines.* Technical report, Waterlines Report no. 82, National Water Commission, Canberra.

Cui T, Peeters L, Rassam D, Raiber M, Crosbie R, Gilfedder M, Pickett T, Hartcher M, Marvanek S, Bruce C and Davies P. 2017. *Groundwater numerical modelling for the Clarence-Moreton bioregion*. Product 2.6.2 from the Clarence-Moreton Bioregional Assessment. Department of the Environment and Energy, Bureau of Meteorology, CSIRO and Geoscience Australia, Australia.

DEWNR. 2012. *Risk management for water planning and management*. SA Department of Environment, Water and Natural Resources, Adelaide.

Doherty J and Moore C. 2017. 'Simple is beautiful'. Discussion paper no.3 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. NCGRT.

Ferré, TPA. 2017. 'Revisiting the relationship between data, models and decision-making'. *Groundwater*, 55.5, 604–614

Guillaume JHA, Hunt RJ, Comunian A, Blakers RS, and Fu B. 2016. 'Methods for exploring uncertainty in groundwater management predictions'. AJ Jakeman et al. (eds), *Integrated Groundwater Management*, Chapter 28, 711–737.

Hayes P and Nicol C. 2017. 'Uncertainty by Design: Sufficient model simplification to make predictive uncertainty analysis tractable'. Discussion paper no.4 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. NCGRT.

McIntyre N and Wood C. 2011. *Developing adaptive groundwater management arrangements in the South-East Australia: a pilot study*. Department of Water, South Australia, 47pp.

Peeters L. 2017. 'Uncertainty analysis in groundwater modelling: where to start?' Discussion paper no.2 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. NCGRT.

QWC. 2012. *Underground water impact for the Surat Cumulative Management Area*. Queensland Water Commission, Queensland.

Richardson S, McMillan M and Currie D. 2017. 'Communicating uncertainty to decision makers and stakeholders'. Discussion paper no.5 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. NCGRT.

Walker G. 2017. 'Predictive uncertainty in groundwater modelling: how is it best used?' Discussion paper no.1 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. NCGRT.

Uncertainty analysis in groundwater modelling: Where to start?

Luk Peeters, CSIRO Land and Water, Adelaide

Introduction

As discussed in Walker (2017), the goal of an uncertainty analysis is to build confidence with clients and stakeholders by communicating what we do know and can predict. This means that we have to honestly and transparently communicate that there can exist a wide range of model predictions that are all consistent with our current understanding of the system, even though gaps limit our system understanding. Part of an uncertainty analysis is therefore to assess how these knowledge gaps might affect the model predictions.

The methodologies for evaluating which predictions are consistent with the observations and system understanding are the focus of Doherty and Moore (2017), while the communication aspect is thoroughly discussed in Richardson et al. (2017), and the overview describing all these elements is presented in Middlemis et al. (2017).

This paper is about the groundwork required before predictive uncertainty can be evaluated, and how quantitative and qualitative uncertainty analysis fits in the modelling and project workflow. Workflow elements are discussed in terms of practical guidance issues, such as 'Where to start?', 'Which assumptions matter?', 'What changes to make to the model?', 'How to bring in observations?' and 'Do initial parameter values matter?'

There are several papers that provide workflows and approaches for uncertainty analysis for environmental modelling in general (Jakeman et al., 2006; Refsgaard et al., 2007; Liu et al., 2008; Bennett et al., 2013) and specifically for groundwater flow and transport modelling (Refsgaard et al., 2012; Guillaume et al., 2016).

The common principles that emerge from examining these workflows, which also can be found in the *Australian groundwater modelling guidelines* (Barnett et al., 2012), are:

1. Explicitly define project objectives and what the model needs to predict.
2. Constrain the model predictions with all available observations and information.
3. Discuss model assumptions and choices and how they affect predictions.
4. Iterate and revisit model objectives, assumptions and predictions during the project in consultation with clients and stakeholders.

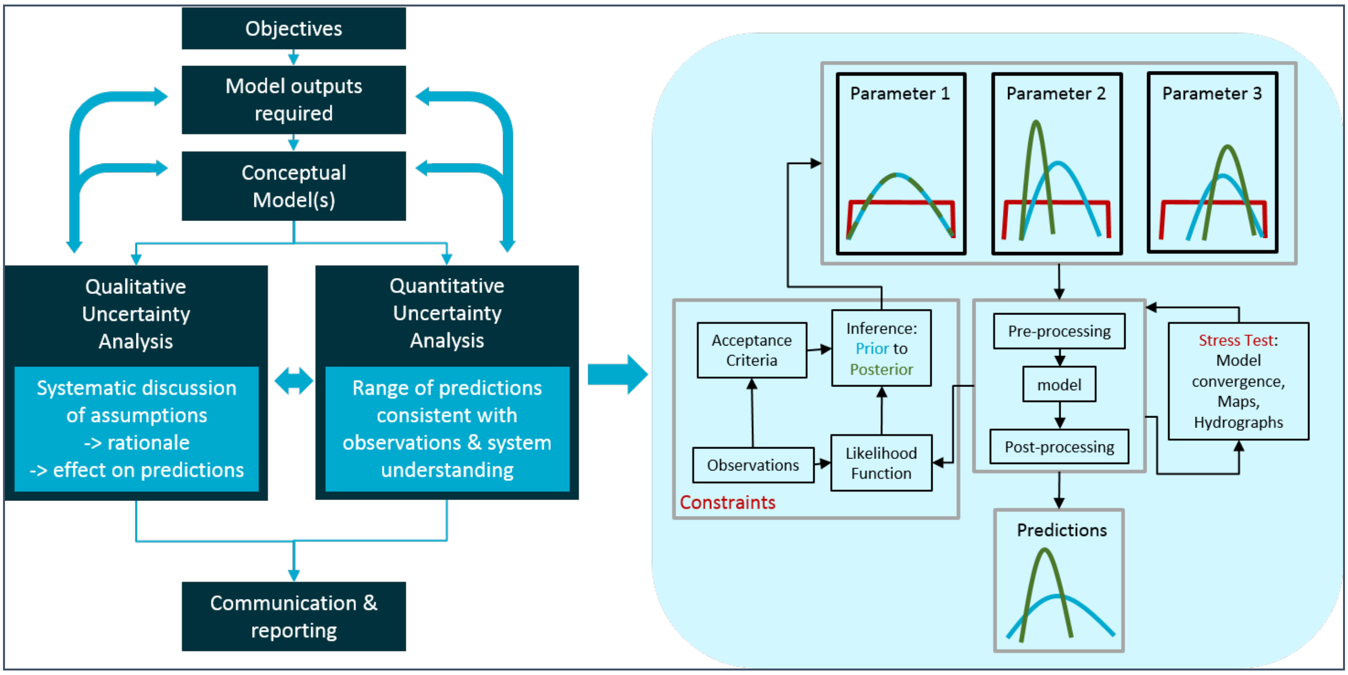


Figure 1: Uncertainty analysis workflow

These concepts form the basis of the workflow depicted in Figure 1. The left-hand side provides a high-level workflow, emphasising the need to explicitly define project objectives and the types of information to be provided by the predictions so that these can be taken into consideration when developing conceptual models.

From that stage, there are two parallel processes: a qualitative and quantitative uncertainty analysis. The qualitative uncertainty is a formal discussion of all model assumptions and choices and how they affect model predictions. The quantitative uncertainty analysis seeks to find all model predictions that are consistent with (or constrained by) the observations. The workflow depicted on the right provides more detail on the aspects of the quantitative uncertainty analysis, such as including what to change to an existing model, how to define initial parameter ranges and how to bring in observations. The final step in the workflow is the communication and reporting. The thick blue arrows indicate the need to iterate between all the components of the workflow during the model project.

In this paper various aspects of this workflow are discussed based on questions pertinent to the modelling process, such as ‘Where to start?’, ‘Which assumptions matter?’, ‘What changes to make to the model?’, ‘How to bring in observations?’ and ‘Do initial parameter values matter?’

Where to start?

The obvious starting point is stated in any modelling textbook or model guidelines; a modelling project starts with the objectives of the project, the research question that needs to be addressed. While obvious, it is not trivial to formulate the research question, let alone to explicitly define which model outcomes will answer the question. Morrison-Saunders et al. (2014) claim that one of the largest issues in environmental impact assessments is a lack of focus due to vaguely formulated research objectives.

Let me illustrate this with a very simple example (Figure 2). In an unconfined, alluvial aquifer a new irrigation development is proposed (D), using both surface water and groundwater. The river is regulated, considered to be losing-disconnected and supports a Ramsar wetland (R). In the aquifer is an existing stock and domestic bore (B) and a spring (S), both up-gradient from a geological fault that may or may not act as a barrier to flow.

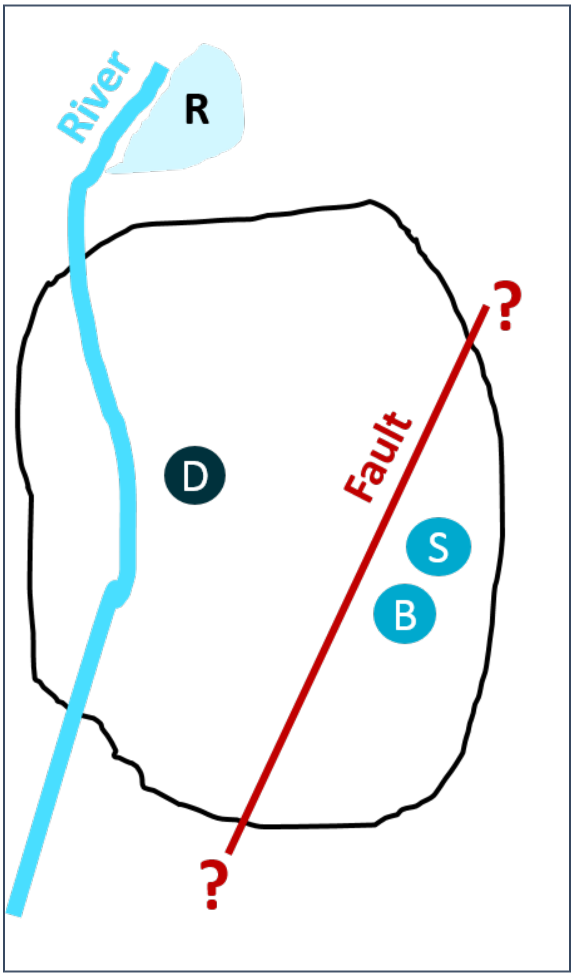


Figure 2: Hypothetical example to illustrate uncertainty analysis workflow concepts. In an unconfined, alluvial aquifer a new irrigation development (D), using both surface and groundwater is proposed. The river is regulated, considered to be losing-disconnected and supports a Ramsar wetland (R). In the aquifer is an existing stock and domestic bore (B) and a spring (S), both up-gradient from a geological fault that may or may not act as a barrier to flow.

The objective of a model study in this case, the reason to build a groundwater model, could be 'Is this proposed development sustainable?' While this is a very pertinent research question from a water resources management point of view, it is impossible for a groundwater model to directly answer. For a groundwater model to be able to inform this issue, the details of what is be sustained must be detailed, and then predictions can be specified explicitly in those terms. In this example that can look like:

In the next 15 years, the irrigation development should not cause:

- More than 20 cm drawdown at the spring S;
- More than 2 m drawdown at the stock and domestic bore B;
- An increase in the number of no flow days at the wetland R to more than 14 days per year;
- A decrease in the flood inundation event frequency at the wetland R to less than 1 in 3 years.

This short statement describes very precisely the future scenario to consider (15 years of irrigation development), the kind of hydrological impacts we seek to avoid (drawdown, no flow days, flood frequency), the maximum tolerable impact (20 cm / 2 m drawdown, 14 no flow days) and the location where these impacts need to be evaluated (at a spring, at a stock and domestic bore and at a wetland).

Defining such predictions explicitly is much more challenging than formulating the research question; it requires a thorough and detailed local understanding of not just the hydrogeology and hydrology of the system, but also of the social, economic and ecological functioning of the study area. Participatory modelling, in which clients and stakeholders are actively engaged in the modelling process (Castilla-Rho et al., 2015; Castilla-Rho, 2017), illustrates the importance of iterating between the modelling and model predictions to identify which model aspects are relevant for groundwater management.

The emphasis on defining model predictions is not just part of good modelling practice in general, it is crucial for uncertainty analysis. It is not possible to carry out an uncertainty analysis of a groundwater model. It is only possible to carry out an uncertainty analysis on the predictions of a groundwater model. Without explicitly defining model predictions, it is impossible to do an uncertainty analysis.

Which assumptions matter?

The qualitative uncertainty analysis of the workflow (Fig. 1) is a formal and structured discussion of all model choices and assumptions and their effect on predictions (Kloprogge et al., 2011). The discussion is organised by answering following four questions with 'low', 'medium' or 'high' (Peeters, 2017a; Peeters et al. 2017b):

- What is the likelihood that I would have made the same choice if I had more or different ***data***?
- What is the likelihood that I would have made the same choice if I had more ***time*** and ***budget***?
- What is the likelihood that I would have made the same choice if I had a better ***model/software*?**
- What is the likelihood that the model ***predictions*** are very different if I change the ***assumption***?

Assume in the example in Figure 2 that the choice is made to not represent the fault in the groundwater model; for example, because a lot of extra data is required, such as the geometry of the fault, the displacement and the hydraulic properties of the fault. Assume that such data is not available in this example, so the first question is scored 'high', Likewise, adding a fault to a groundwater model requires additional time and resources especially since the added complexity will require model runs during model calibration and uncertainty analysis. If the budget for the project is limited and if the timelines are strict, the second question also receives a 'high' score. Implementing faults in existing groundwater modelling software is possible, but not trivial. The third question therefore is scored 'medium'. The final question, however, is the most important one as it pertains to the effect on predictions. For the drawdown predictions, it would be scored 'high' as an impermeable fault might completely prevent drawdown from propagating to the spring and stock and domestic bore. For the hydrological change at the wetland however, the scoring would be 'low' as the river is losing-disconnected, and the fault is up-gradient from the irrigation development. Despite the 'high' score for drawdowns, the choice might still be justified as, in this example, not representing the fault would be conservative as it overestimates drawdown.

In this illustration, the effect on predictions is discussed from first principles, considering a fault as a barrier to flow. In Peeters (2017) a number of additional approaches are discussed, including sensitivity analysis, hypothesis testing with custom-made models, analogues from literature and expert opinion. An example of hypothesis testing, salient to this discussion, can be found in Knapton and Middlemis (2017), where a fault was modelled both as a barrier and as a conduit of flow. It was only possible to match the observed groundwater levels in the latter conceptualisation, which provided an argument against the fault acting as a barrier to flow.

Elements of this scoring and discussion of assumptions and limitations are already part of most groundwater model reports, as recommended in Barnett et al. (2012). Rather than relying on a 'confidence level' classification prone to misinterpretation, an open and transparent plain English discussion is a great way to engage with clients and stakeholders who do not necessarily have the same in depth knowledge of the system or groundwater modelling as a technical reviewer might have. In addition to that, critically evaluating assumptions has since long been shown to reduce overconfidence in predictions, an affliction experts in many scientific disciplines suffer from (Koriat et al., 1980).

The qualitative uncertainty ideally complements a quantitative uncertainty analysis, but the concept can also be applied to a deterministic groundwater model. It serves as a first evaluation of conceptual model uncertainty and the results identify the most crucial knowledge and data gaps.

What to change in the model?

In essence, a quantitative uncertainty analysis consists of evaluating different parameter combinations with a groundwater model. This is not materially different to calibration or sensitivity analysis. Calibration and sensitivity analysis are often done manually, by changing parameter values in a graphical user interface and running the model again. However, automated analysis is becoming more common, as a manual approach becomes very impractical if the number of model runs required exceeds a couple of dozen.

To systematically explore a wide parameter range, the model needs to be run, and the parameters changed in an automated way. This means that any pre -or post-processing step needs to be captured as part of the model variables. Recharge is for instance often computed through a separate process. To account for uncertainty in this recharge estimation, the recharge estimation and processing script needs to be integrated in the model batch file that is run from the command line.

Similarly, model predictions are often not direct outputs of the model, but need to be post-processed. Think for instance of the maximum difference in groundwater levels between two scenarios or a mass-balance over a part of the model domain. To compute these, additional programs or scripts need to be executed and, if they are the target of an uncertainty analysis, these programs need to be included in the batch file. Recent advances, such as Python packages Flopy (Bakker et al., 2016) and PyEmu (White et al., 2016), combine the flexibility and numerical analysis capability of a scripting language (Python) with traditional groundwater modelling (MODFLOW) and uncertainty analysis (PEST) packages. In addition to the practical advantages, a scripted workflow increases transparency and reproducibility of the modelling process greatly (Fienen et al., 2016).

There is however a large drawback to automating the running of groundwater models. Everyone who has ever created and run a groundwater model knows that they are not unconditionally stable. A large part of model development is often devoted on ensuring that a model converges to a solution. This is hard enough when dealing with a deterministic model and a single parameter combination, let alone when considering a wide range of parameter combinations. To make matters even more challenging, it is quite likely that custom-made scripts have bugs or coding errors.

Solving these issues during calibration or uncertainty analysis is very time-consuming and frustrating. A stress test, shown in the detailed workflow in Figure 1, can help to identify and avoid potential issues before they become a problem. In a stress test, a limited number of extreme parameter combinations (coloured red in Figure 1) are evaluated with the batch file that contains the pre-processing scripts, the model executable and the post-processing scripts. The goal is to break the model; to find out for which parameter combinations the model no longer converges.

Non-convergence for some parameter combinations is not always an issue. It is not hard to imagine that if you combine a low hydraulic conductivity value with a high storage value that a model will struggle to converge. As it is quite unrealistic that such a parameter combination exists in reality, the model does not need change, but in the uncertainty analysis such parameter combinations need to be avoided, for instance by specifying a covariance between hydraulic conductivity and storage. It is different when a model fails to converge for a realistic parameter combination as it points to a structural issue with the model. While it is often hard to fix these structural issues, a stress test will at least make the modeller aware of their existence before starting a computationally expensive quantitative uncertainty analysis. Stress-testing also allows verification that all the processing scripts work as intended and that the model produces sensible results over a wide range of parameter values.

How to bring in observations?

Observations of state variables, for instance groundwater level observations or stream flow data, come into a calibration and uncertainty analysis through an objective function or likelihood function. In calibration, the minimum of the objective function is sought, while in uncertainty analysis, a range of parameter values is sought consistent with observations. In the detailed workflow in Figure 1 we emphasise that it is therefore not enough to specify a likelihood function to bring in observations; one needs to specify how much mismatch between model and observations can be tolerated before one considers a parameter set to be not consistent with observations.

A common approach in calibration and uncertainty analysis in groundwater is to assign the weights of each observation as the inverse of the standard deviation of observation error (Hill et al., 2007). This approach assumes that observation error is normally distributed. The distribution of parameter combinations that result from an uncertainty analysis, the posterior distribution, will then produce a distribution of model residuals that is comparable with the observation error. As an example, consider the standard deviation of observation uncertainty of a groundwater level observation to be 1 m. If this value of observation uncertainty is used in an uncertainty analysis, it implies that 5% of posterior parameter combinations will result in a residual larger than 2 m.

The result of the uncertainty analysis, the width of the uncertainty bounds, will largely depend on what one defines as observation uncertainty or acceptance criteria. Unfortunately to date, there is not a lot of research available to provide guidance on this matter. An acceptance criterion can be defined by unpacking the observation process, such as in the case of groundwater level observations, via the accuracy of the data logger and the survey level on the observation bore, and by accounting for scale issues related to the resolution of the groundwater model. Such a first principle approach is relatively straightforward for groundwater level observations, but becomes very challenging when dealing, for instance, with stream flow fluxes or salinity concentrations. It is however crucial when combining different sources of data in the likelihood function as it determines the weight each data type receives.

The approximate Bayesian computation (ABC) approach, applied in groundwater by Vrugt et al. (2013) and Peeters et al. (2017), uses a rejection algorithm in which a set of criteria to evaluate a model is defined and only those parameter combinations are accepted that are within predefined thresholds. The approach misses the statistical rigor a formal Bayesian likelihood function provides, but has the great advantage that it makes the uncertainty analysis process less opaque and more flexible. The goal of the ABC approach is, in discussion with local experts, clients and stakeholders, to establish a rule set of what is an acceptable model. This can be related to the mismatch with observations (e.g. groundwater levels must be matched to within 2 m) but can also include conditions based on the system understanding (e.g. conductivity of layer 2 is always lower than conductivity of layer 3, the groundwater contribution to a stream cannot be more than stream flow). Many of these conditions may appear self-evident at first sight, but they are invaluable in formally constraining a groundwater model.

Do initial parameter values matter?

The quantitative uncertainty analysis workflow shown in Figure 1 is a generic representation of the Bayesian inference approach that underpins most quantitative uncertainty analysis methodologies. It starts with defining an initial range of values for each parameter, which is referred to as the prior parameter distribution (blue in Figure 1). The quantitative uncertainty analysis evaluates this initial parameter range with the model. By comparing the model outcomes with the objective function or likelihood function, a range of parameter combinations is obtained that is consistent with the observations. This is referred to as the posterior parameter distribution (green in Figure 1). The ensemble of predictions that correspond to the posterior parameter combinations are referred to as the predictive posterior distribution.

If the observations are informative, the posterior parameter distribution can be quite different to the prior distribution (parameters 2 and 3). Some parameters, however, will not be constrained by the observations (parameter 1). An obvious example is that it is not possible to infer storage values from a steady state model. For complex models, it is often not that straightforward to work out which parameters can be constrained by the observations. A systematic sensitivity analysis, often a by-product of an uncertainty analysis, can provide valuable insight into this aspect, for instance by quantifying parameter identifiability (White et al., 2014).

This implies, however, that, when a prediction strongly depends on a parameter that cannot be constrained by observations, the definition of the prior parameter distribution is crucial. Even in the case where a parameter can be constrained by observations, it is possible that the parameter is compensating for structural issues with the model. River-bed hydraulic conductivity, for instance, can compensate for an incorrectly specified river stage. A prior parameter distribution informed by local knowledge can safeguard against parameters assuming unrealistic values.

Unfortunately, we are rarely in a position that it is possible to describe a prior distribution of a parameter at the scale of model discretisation, based on field measurements, and the range applied in the modelling is mostly informed by literature values and expert knowledge. It is, however, important to be aware that parameter values that fit the observations are not necessarily correct.

Conclusion

The goal of an uncertainty analysis is to build confidence with clients and stakeholders in what we do know and can predict (Walker, 2017). The generic workflow presented and discussed in this paper considers uncertainty analysis as an integral and crucial component of a modelling project. It highlights the need to explicitly specify what predictions are required of the model and to justify all assumptions in the context of these predictions. This paper emphasises the need for a stable model and well-defined acceptance criteria based on the available observations and informative priors, especially for parameters that cannot be constrained by parameters or are prone to compensate for structural model issues. This is closely related to the complexity of models discussed in Doherty and Moore (2017).

Transparently documenting and honestly reporting the results of an uncertainty analysis, together with its limitations and assumptions, is crucial to instil confidence in the model predictions with stakeholders and clients. Richardson et al. (2017) elaborate more on this communication aspect of uncertainty analysis.

References

Bakker M, Post V, Langevin CD, Hughes JD, White JT, Starn JJ and Fienen MN. 2016. ‘Scripting MODFLOW Model Development Using Python and FloPy’. *Groundwater*, 54.5, 733–739.

Barnett B, Townley L, Post V, Evans R, Hunt R, Peeters L, Richardson S, Werner A, Knapton A and Boronkay A. 2012, *Australian groundwater modelling guidelines.* Technical report, Waterlines Report no. 82, National Water Commission, Canberra.

Bennett ND, Croke BFW, Guariso G, Guillaume JHA, Hamilton SH, Jakeman AJ, Marsili-Libelli S, Newham LTH, Norton JP, Perrin C, Pierce SA, Robson B, Seppelt R, Voinov AA, Fath BD, and Andreassian V. 2013. ‘Characterising performance of environmental models’. *Environmental Modelling & Software*, 40, 1–20.

Castilla-Rho J. 2017. ‘Modelling with stakeholders: Finding the complexity that matters’. Submitted to *Ground water*, 55.

Castilla-Rho J, Mariethoz G, Rojas R, Andersen M, and Kelly B. 2015. ‘An agent-based platform for simulating complex human–aquifer interactions in managed groundwater systems’. *Environmental Modelling & Software*, 73, 305–323.

Fienen MN and Bakker M. 2016. ‘HESS Opinions: Repeatable research: what hydrologists can learn from the Duke cancer research scandal’. *Hydrology and Earth System Sciences,* 20.9, 3739–3743.

Guillaume JHA, Hunt RJ, Comunian A, Blakers RS and Fu B. 2016. ‘Methods for Exploring Uncertainty in Groundwater Management Predictions’. *Integrated Groundwater Management*. Springer Nature, 711–737.

Hill MC and Tiedeman CR. 2007. *Effective groundwater model calibration*. Wiley.

Jakeman A, Letcher R and Norton J. 2006. ‘Ten iterative steps in development and evaluation of environmental models’. *Environmental Modelling & Software*, 21.5, 602–614.

Kloprogge P, van der Sluijs JP, and Petersen AC. 2011. ‘A method for the analysis of assumptions in model-based environmental assessments’. *Environmental Modelling & Software*, 26.3, 289–301.

Knapton A and Middlemis H. 2017. *Integrated modelling and alternative conceptual model testing of catchment with tailings storage in Victoria, Australia*. In preparation.

Koriat A, Lichtenstein S, and Fischhoff B. 1980. ‘Reasons for confidence’. *Journal of Experimental Psychology: Human Learning & Memory*, 6.2, 107–118.

Liu Y, Gupta H, Springer E, and Wagener T. 2008. ‘Linking science with environmental decision making: Experiences from an integrated modeling approach to supporting sustainable water resources management’. *Environmental Modelling & Software*, 23(7), 846–858.

Middlemis H, Walker G, Peeters L, Hayes P, Moore C and Richardson S. 2017. *Overview on improved evaluation of the effect of hydrogeological uncertainty on project objectives*. Overview paper for NCGRT national groundwater modelling uncertainty workshop. Sydney, 10 July 2017.

Doherty J and Moore C. 2017. 'Simple is beautiful'. Discussion paper no.3 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. NCGRT.

Morrison-Saunders A, Pope J, Gunn JA, Bond A and Retief F. 2014. 'Strengthening impact assessment: a call for integration and focus'. *Impact Assessment and Project Appraisal*, 32.1, 2–8.

Peeters LJM. 2017a, 'Assumption hunting in groundwater modelling: find assumptions before they find you'. Accepted for publication in *Groundwater* 55.5.

Peeters LJM, Pagendam D, Crosbie RS, Viney NR, Rachakonda PK, Dawes WR, L. Gao, M. S., Zhang YQ and McVicar TR. 2017b. 'Building confidence in environmental impact models through quantitative and qualitative uncertainty analysis'. Submitted to *Environmental Modelling and Software*.

Refsgaard JC, Christensen S, Sonnenborg TO, Seifert D, Højberg AL and Troldborg L. 2012. 'Review of strategies for handling geological uncertainty in groundwater flow and transport modelling.' *Advances in Water Resources*, 36, 36–50.

Refsgaard JC, van der Sluijs JP, Højberg AL and Vanrolleghem PA. 2007. 'Uncertainty in the environmental modelling process - A framework and guidance'. *Environmental Modelling & Software*, 22.11, 1543–1556.

Richardson S, McMillan M and Currie D. 2017. 'Communicating uncertainty to decision makers and stakeholders'. Discussion paper no.5 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. NCGRT.

Vrugt JA and Sadegh M. 2013. 'Toward diagnostic model calibration and evaluation: Approximate Bayesian computation'. *Water Resources Research*. 49.7, 4335–4345.

Walker G. 2017. 'Predictive uncertainty in groundwater modelling: how is it best used?' Discussion paper no.1 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. NCGRT.

White JT, Doherty JE and Hughes JD. 2014. 'Quantifying the predictive consequences of model error with linear subspace analysis'. *Water Resources Research*. 50.2, 1152–1173.

White JT, Fienen MN and Doherty JE. 2016. 'A python framework for environmental model uncertainty analysis'. *Environmental Modelling & Software*, 85, 217–228.

Simple is beautiful

John Doherty, Watermark Numerical Computing, and Catherine Moore, CSIRO Land and Water, Brisbane.

Preface

This short text summarises a much larger document delivered through the 'Smart Models for Aquifer Management' project undertaken by GNS, New Zealand. That document, Doherty and Moore (2017), provides a theoretical basis for the discussion that is presented herein. It can be downloaded from:

https://www.gns.cri.nz/Home/Our-Science/Environment-and-Materials/Groundwater/Research-Programmes/Smart-Aquifer-Models-for-Aquifer-Management-SAM/SAM-discussion-paper

Introduction

We commence with an apology to EF Schumacher, the author of the much-acclaimed book *Small is Beautiful*. The book's subtitle is 'a study of economics as if people mattered'. To cement our plagiarism we should subtitle the present text 'a study of environmental modelling as if decisions mattered'.

The subject of this paper is the construction, calibration and deployment of models that are built to support environmental and resource management. Nowadays, it is common practice to build models for this purpose. We submit that many of these models fail to deliver the support which they promise because they are too complex and insufficiently agile to use in the decision-making process. But first we discuss the role that models can play as an environmental decision-support tool, and then we explain why they so often fail to provide that support.

Models and decisions

The contribution that numerical modelling should make to the environmental decision-making process is succinctly described by Freeze et al. (1990). Though often cited, the arguments presented in this landmark paper are mostly ignored by the modelling community. Freeze et al. (1990) point out that modelling introduces to the decision-making process the vital ingredient of risk. Risk can be roughly equated to the probability of a bad thing happening as a consequence of a particular decision, multiplied by the cost associated with its occurrence. It follows immediately that if a model is deployed to support environmental decision making, its predictions of environmental behaviour under different management options must be accompanied by estimates of the uncertainties associated with those predictions.

Moore and Doherty (2005) show that model predictions of groundwater behaviour can be accompanied by large uncertainties, even where these predictions are made by a 'calibrated model' i.e. by a model whose parameters have been adjusted to ensure that its outputs match historical measurements of system state (for example heads in wells and flows in streams). They show that, to the extent that a prediction is sensitive to components of the calibration null space, the uncertainty of that prediction is not reduced through the calibration process at all. (The 'calibration null space' refers to model parameters, and/or combinations of parameters, which are uninformed by historical measurements that comprise the calibration dataset.) Such predictions, therefore, tend to be those that are sensitive to hydraulic property detail and hydraulic property variability that exists at small spatial scales. They include (but are not limited to) movement of a contaminant through a heterogeneous sub-surface, nuances of groundwater/surface-water interaction, and the response of a natural system to extreme climatic events, and/or to large changes in its management regime. Unfortunately, many decision-critical predictions fall into one or more of these categories.

Doherty and Simmons (2013) and Doherty and Vogwill (2015) extend the concepts introduced by Freeze et al. (1990) to define failure of a modelling enterprise in lending support to the decision-making process. Based on the premise that decisions are taken to avoid the occurrence of an unwanted outcome (i.e. a 'bad thing'), a model fails in its decision-support role if its predictive uncertainty margins underestimate the probability of occurrence of that bad thing.

Examined from a frequentist perspective, decision making requires exploration of the hypothesis that a bad thing will happen if a certain management strategy is adopted. (This strategy can include so-called 'adaptive management', where mitigation actions are prescribed if certain monitoring thresholds are crossed.) Ideally, an environmental model should be employed to implement the scientific method, whereby rejection of the 'bad thing hypothesis' is attempted by demonstrating incompatibility of its occurrence with information about the system that is encapsulated in the model. This information is comprised of expert knowledge of system properties and processes on the one hand, and measurements of the historical behaviour of the system on the other hand. A type 2 statistical error occurs if the hypothesis of a bad thing happening is falsely rejected; we define this as failure.

Armed with a definition of failure, a firm conceptual reference point exists for setting specifications for any model that is built to support environmental decision making. In the present discussion, specifications of most interest are those pertaining to the level of model complexity.

Another criterion of relevance to model design is that of usefulness. If the uncertainty bounds calculated by a model for a decision-critical prediction are too broad, then the bad thing hypothesis can never be rejected. Failure to reject a false hypothesis is referred to as a type 1 statistical error. A model's usefulness increases to the extent that it can reduce the uncertainty of a decision-critical prediction to the lower limit set by availability of current information pertaining to that prediction.

Models as receptacles for information

It is apparent from the above discussion that a model's support for the decision-making process rests on its ability to provide receptacles, or containers, for information about the system that is undergoing management. As stated above, this information falls into two broad categories, namely expert knowledge on the one hand, and the historical behaviour of the system on the other hand. Expert knowledge is expressed through the construction details of the model, its boundary conditions, and its parameterisation (i.e. how hydraulic properties of the simulated system are represented in the model).

Conceptually, all of these (especially its parameterisation) must be stochastic, as that is the nature of expert knowledge as it pertains to environmental systems. Information that is encapsulated in the historical behaviour of the system is introduced to the model through the history-matching process. This information constrains parameters to a narrower stochastic range than that based on expert knowledge alone. This is because parameters employed by a model must be such as to allow the model to replicate that behaviour. These two types of information constitute the prior probability and likelihood terms that appear on the right side of Bayes equation; see any statistical text for more details.

Bayes' equation states that the outcomes of the history-matching process are probabilistic, this pertaining to parameters employed by a model, and to predictions made by a model. At first glance it would appear that the concept of the 'calibrated model' which forms the basis for widespread, model-based decision making is at odds with this fundamental premise. This is indeed the case. However, if model calibration is considered as the first step in a two-step process of:

- Finding a solution of minimum error variance to the inverse problem posed by fitting model outputs to the calibration dataset; and then
- Quantifying the error in predictions made by the thus calibrated model;

then the precepts of Bayes equation can be followed in a way that is far easier to implement than Bayes' equation itself, with outcomes that reduce the probability of occurrence of a type 2 statistical error to an acceptable level. See Doherty (2015) for further details.

Instead of being considered as devices through which the scientific method can be introduced to the decision-making context, models are often construed (especially by non-modellers who pay for them) as (presumably accurate) simulators of environmental behaviour at a particular study site. However even the most complex model can only be approximate in its attempted simulation of local environmental processes. Nevertheless, the ability to simulate, even in an approximate fashion, processes that are operative at a location of interest endows a model with the information receptacles that it requires to fulfil its decision-support role. Importantly, it is these receptacles which are of primary importance.

In contrast, exact (and elusive) replication of environmental processes is of secondary importance, for the role of simulation is to serve the greater task of providing receptacles for information that resides in local knowledge of the system and in measurements of the behaviour of that system. Furthermore, exact replication of environmental behaviour may not even be required if the occurrence of a bad thing can be relegated to a low level of probability using only a fraction of the information that is available at a particular study site.

Complex models

Presumably, the more complex is a model, the greater is the amount of information for which it can provide receptacles.

Complex models provide more receptacles for expert knowledge because of their 'realistic' representation of real-world conditions, and because of their 'physically-based' simulation of real-world processes. Local measurements of system attributes, and of hydraulic properties, can thus be transferred directly to the model. These can then exert constraints on properties assigned to parts of the model domain where no such measurements have been made, using correlations that have emerged from local site characterisation.

Complex models provide receptacles for information encapsulated in historical measurements of system state because they can support many adjustable parameters, and because these parameters embody realistic descriptions of system heterogeneity. It follows that a good fit between model outputs and historical measurements should be readily achievable. This does not, of course, result in parameter uniqueness. However, it does result in tighter constraints on parameter variability than that which is permitted on the basis of expert knowledge alone.

In theory, environmental decision making is therefore well supported by a complex model. In practice, this is rarely the case. Complex models have long run times. Often they are plagued by numerical instability. Furthermore, the more detail that a model can express, the greater is the necessity for that detail to be expressed stochastically (i.e. probabilistically). Stochasticity is required because neither direct measurements of system properties at a discrete number of locations, nor inferences of system properties made through history matching, can yield unique estimates of that detail throughout a modelled area. Unfortunately, however, the long run times of complex models preclude running the model the number of times required for proper stochastic analysis. These same run times, accompanied by a penchant for numerical instability, often make the task of history-matching very difficult indeed – if not impossible. A simplistic parameterisation scheme is often therefore draped over the domain of a complex model to ease the burden of history matching. The model then becomes simple again – losing the benefits of complexity while eschewing the quick run times that model simplicity can bring.

It is often the complex model itself, rather than the support that it must provide to management decisions, that constitutes the deliverable of a model construction exercise. This is based on the premise that a complex model can be built to support the making of many different decisions. To avoid failure (as defined above) of the time-consuming and expensive model-construction process, the model that emerges from this process must be capable of quantifying the uncertainties of many different predictions. Furthermore, for each of these predictions a guarantee must be provided that a type 2 statistical error has been avoided. Meanwhile, model usefulness (as defined above) requires that predictive uncertainties be reduced to a level that is commensurate with all available information. This requires inclusion in the model of all parameters that can be informed by the calibration dataset. Meanwhile, avoidance of failure requires inclusion in the model of all parameters that are not informed by the calibration dataset but to which one or more predictions of management interest may be sensitive. Both sets of parameters must be capable of variation under calibration constraints, to support exploration of the uncertainties of all predictions that the complex model may be required to make. Where run times are high and where model numerical stability is questionable, this is not possible.

Meanwhile, the personal difficulties faced by a modeller who has been engaged to build a complex model are considerable. With his/her focus on the model as a deliverable, the modelling process descends into that of making the model 'look good' by adding ever-increasing amounts of non-stochastic detail to the model, to forestall accusations of its inadequacy as an 'accurate simulator' of reality. The decisions that the model must support are forgotten as the modeller nurses the model to a point at which it can be delivered to a manager or client, with the hope that the latter will not use it in ways that expose its numerical deficiencies.

Simple models

Simple models can be built and calibrated more quickly than complex models. However, they must be used with caution. As is discussed extensively by Doherty and Moore (2017), the design, calibration and deployment of a simple model must be tailored to one or a small number of specific predictions to ensure its decision-support role is not corrupted. This is not necessarily a bad thing. The relationship between a simple model and the decision that it is built to support is thus clearly defined, right from the start of the model construction process. Furthermore, in deploying the simple model in support of the decision-making process, the modeller becomes an integral part of that process. The simple model is thus seen as an agent for better decision making rather than as an end in itself.

In general, simple models do not provide good receptacles for the information contained in expert knowledge. They are often abstract in nature; their parameters are often ‘lumped’. It is therefore difficult for point measurements of real-world system properties to inform these parameters. Nor can the prior uncertainties of lumped parameters be readily established through site characterization studies. Hence if a prediction is sensitive to parameters that are not well informed by the calibration dataset, it is difficult to calculate the uncertainty of that prediction because the receptacles for expert knowledge (the only other source of information pertinent to the prediction) are not provided by a simple model. At the same time, the abstract parameter set employed by a simple model may not be capable of representing local heterogeneities in system properties to which a management-critical prediction may be sensitive. The assignment of too little or too much prior uncertainty to the parameters encapsulated in a simple model may, under these circumstances, lead to a type 2 statistical error in the first case or a type 1 statistical error in the second case.

A simple model may include too few parameters to support a good fit with a calibration dataset. Presumably this design defect would be rectified prior to its deployment, or an alternative model would be chosen. However, the ability of a simple model to fit a calibration dataset does not provide a guarantee that its predictions are without bias. Nor does it guarantee that the uncertainty interval ascribed to a prediction reflects the true uncertainty of that prediction, together with any ‘manufactured uncertainty’ (i.e. bias) accrued through use of the simple model itself. Predictive bias can arise in two ways. It may be a direct reflection of simple model inadequacy as it pertains to that prediction. Alternatively, it may be incurred through the very process of attempted uncertainty reduction; that is, through the process of model calibration. Doherty and Christensen (2011), White et al. (2015) and Doherty (2015) show how a simple model may be capable of making relatively unbiased predictions if parameter values are based on expert knowledge alone. However, these same predictions may be accompanied by considerable bias once the model has been calibrated because of the surrogate roles that parameters must play to compensate for model defects as they are adjusted to fit the calibration dataset. These authors show that this applies to some predictions made by a simple model but not to others. For some predictions, model defects can be ‘calibrated out’. For other predictions, the calibration process can actually thwart the model’s ability to make an unbiased calculation.

Because simple models tend to employ fewer parameters than complex models, the null space exposed through the calibration process is generally of lower dimension for a simple model than it is for a complex model. A simple model may be designed in such a way, and calibrated in such a way (see below), as to reduce the chances of predictive bias. Furthermore, the calibration dataset may be rich in information pertaining to the prediction that a simple model is designed to make. However, if the prediction is dependent on some null space parameter components, then these parameter components should be represented in the simple model, not because they are estimable, but precisely because they are inestimable; this avoids a type 2 statistical error when quantifying the uncertainty of the prediction. If the model is too simple to include these parameters, then their lack of representation must be accommodated through strategic inflation of the predictive uncertainty interval calculated by the simple model. Means through which this can be achieved will be context-specific, and will probably involve a degree of subjectivity.

Despite all these shortcomings, there are some types of predictions that a simple model can make with impunity. These are predictions that are wholly dependent on the solution space of the ‘real world model’ of which the simple model is an emulator. These predictions tend to be similar in nature to those which comprise the simple model’s calibration dataset. White et al. (2015) show that for predictions of these types, a sufficient and necessary condition for the making of an unbiased prediction of future system behaviour is that the model is able to replicate past system behaviour to a level that is commensurate with measurement noise. Because the uncertainties of these types of predictions are dependent only on measurement noise that accompanies the calibration dataset, and not on the nature and dimensionality of the calibration null space, a simple model is able to quantify their uncertainties with integrity.

Overcoming simple model defects

The problems that beset complex models when used in the decision-support context are not about to disappear. Hence, despite the difficulties facing simple model usage that have been outlined above, ways must be found to build, calibrate and deploy them with integrity on a widespread basis as fleet-footed, flexible, prediction-specific replacements for the slow, prediction-general, complex models that are currently used on a widespread basis for decision support. Doherty and Moore (2017) outline ways in which this can be achieved. Some of these are now briefly discussed. Refer to the original text for further details.

Specifications of a simple model must be such that it is capable of making, with as little bias as possible, the prediction for which purpose it was built. Here it is worth noting that most models are better at predicting differences than absolutes. Even complex models are compromised in their ability to associate numbers with future system states. However, these compromises tend to be reduced for predictions that are defined as alterations to system state following alterations to current management practice. The same applies to comparative system states emerging from two competing management strategies. It follows that it is preferable to base management decisions on alterations to system states, or on comparative system states, than on the absolute values of system states.

The type of prediction required of a model sets a lower limit on the complexity of the model that is used to make it. This applies particularly to predictions that are sensitive to system and parameterisation detail. The uncertainty associated with such predictions is often large. This follows from the fact that such detail may lie largely within the calibration null space. At the same time, representation and parameterisation of such detail will probably be far more simple in the model than in the real world; this may lead to under-estimation of predictive uncertainty, especially if attempts are made to constrain inappropriately broad scale or lumped parameters through history-matching. In cases like these, the model's decision-support role may be best served by not calibrating the model at all on the basis that the prior uncertainty of the prediction can serve as a useful (and not overly-conservative) surrogate for its posterior uncertainty. Evaluation of prior uncertainty is numerically easier, and far cheaper, than evaluation of posterior uncertainty. Methodologies such as multiple point geostatistics can be easily used to generate 'realistic' random parameter fields that pay maximum respect to expert knowledge. Meanwhile, history-matching can be undertaken in a 'stochastic' sense, by ensuring that model-generated counterparts to the calibration dataset calculated using these random parameter fields collectively encompass measurements comprising the calibration dataset.

The most difficult predictions to make with a simplified model are those whose uncertainties are partially reduced through history-matching, but which still retain a significant amount of null space sensitivity. These are the predictions that are most likely to incur bias as attempts are made to reduce their uncertainties through calibration; see White et al. (2015) for details. Nevertheless, because of their partial solution space dependency, considerable reduction in predictive uncertainty can be achieved through the history matching process; hence calibration must be undertaken to promulgate model usefulness. Doherty (2015) and White et al. (2015) show that it is possible to reduce the potential for calibration-induced predictive bias through formulation of a multi-component objective function that incudes observations and corresponding model outputs that have been specially processed prior to matching. This processing is designed to isolate or 'orthogonalise out' aspects of the calibration dataset that are likely to entice parameters to adopt surrogate roles in order that model outputs can match that dataset. In many cases, it is not a difficult matter to design appropriate model and data processing schema to achieve this aim. It may only require that vertical, horizontal and temporal differences, as well as the individual measurements themselves, be included in the calibration dataset. The objective function components that emerge from introduction of these 'differences observations' must then be weighted for visibility in the overall calibration objective function.

Finally, the prior uncertainties that are ascribed to the lumped and averaged parameters with which a simple model is endowed may require special consideration to avoid underestimation of posterior predictive uncertainties. These prior lumped parameter uncertainties may need to be wider than the uncertainties associated with their spatially distributed real world counterparts. This ensures that any prediction-salient details (for example continuous zones of high or low hydraulic conductivity) that may be 'hidden' in these parameters because they are undetectable through the calibration process, can nevertheless find expression when evaluating the uncertainties of these predictions.

Concluding remarks

The history of model-based decision support is chequered. Too often, specifications for model construction and deployment have been based on an illusion that human beings can create a numerical surrogate for a complex environmental system. It is assumed that this surrogate can be used to predict the future of that system under any proposed management strategy before that strategy is implemented in the real world. Millions of dollars have been spent on models that were built in pursuit of this goal. Frequently, those who paid for models such as these were disappointed with what they received.

The authors of this paper suggest that models can provide better support for environmental decision making if the premise of their construction is altered from that of a simulator of complex environmental processes to that of a tool for implementation of the scientific method. This method requires that hypotheses be tested, and maybe rejected, through testing their compatibility with information pertaining to the nature and properties of the system whose management is being decided. In the decision-making context, hypotheses that require testing are clearly defined. They are the bad things that we seek to avoid if a certain course of management action is adopted. For each such bad thing, it should be possible to construct a model, specific to that bad thing, which can provide receptacles for information against which the probability of its occurrence can be tested. In all likelihood, such a model will be 'simple', for its performance in carrying out the task for which it was built will be enhanced if it is unencumbered by unnecessary complexity, and is not required to provide receptacles for unnecessary information.

Reducing the probability of decision-support failure to an acceptable level requires a guarantee that the hypothesis of occurrence of a bad thing is not falsely rejected. Model-calculated uncertainty intervals must therefore accommodate contributions to predictive uncertainty that are an unavoidable outcome of the necessarily defective nature of any model that attempts to simulate a real-world system. Sadly, defects are the price that must be paid for an ability to quantify uncertainty. Pursuit of modelling 'perfection' inevitably results in a level of numerical complexity that makes uncertainty analysis difficult, if not impossible.

With acceptance of the benefits of prediction-specific simplicity in simulation of real-world systems comes abandonment of the premise that a single model can be used as a basis for all environmental management within a certain study area. So too follows abandonment of the notion that model development should be separated from the decisions that a model is intended to support. Instead, construction, calibration and deployment of a model must all be done as part of a single, prediction-specific enterprise. This enterprise must result in demonstrably conservative quantification of the uncertainty of that prediction (thus avoiding a type 2 statistical error) while reducing that uncertainty to a level that reflects all available information (thus avoiding a type 1 statistical error).

Our industry is yet to acquire the skillset required for widespread construction of simple, prediction-specific models that maximise the support that model-based data-processing can provide to environmental decision making. This is an area in which research is urgently needed. Paradoxically, it is often easier to build a complex model than a simple model. Construction of the latter requires a deep understanding of the system that is being simulated, founded on detailed data analysis. It also requires a deep understanding of the numerical strategies that are required for design of a tool that can provide receptacles for information from disparate sources through which specific hypotheses of management failure can be tested, and maybe rejected. Development of this skillset is fundamental to the future of model-based decision making.

References

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

Doherty J and Moore CR. 2017. *A Theoretical Analysis of Model Simplification.* GNS. New Zealand. https://www.gns.cri.nz/Home/Our-Science/Environment-and-Materials/Groundwater/Research-Programmes/Smart-Aquifer-Models-for-Aquifer-Management-SAM/SAM-discussion-paper

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

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

Doherty J and Vogwill R. 2015. 'Models, Decision-Making and Science'. Vogwill R (ed). *Solving the Groundwater Challenges of the 21st Century*. CRC Press.

Freeze RA, Massmann J, Smith L, Sperling T and James B. 1990. 'Hydrogeological decision analysis: 1 A framework'. *Ground Water*, 28.5,738–766.

Moore C and Doherty J. 2005. 'The role of the calibration process in reducing model predictive error'. *Water Resources Research*. 41.5. W05050.

White JT, Doherty JE and Hughes JD. 2014. 'Quantifying the predictive consequences of model error with linear subspace analysis'. *Water Resources Research,* 50.2, 1152–1173. DOI: 10.1002/2013WR014767

Uncertainty by design: sufficient model simplification to make predictive uncertainty analysis tractable

Phil Hayes, Jacobs (now University of Queensland) and Chris Nicol Groundwater Logic

Introduction

This paper outlines methods to make groundwater model uncertainty analysis tractable. It looks at the Menindee groundwater model as a practical example of focus and simplification that enabled robust assessment of the uncertainty of key model predictions. It is intended to complement Doherty and Moore (2017), which makes the case for model simplicity and uncertainty analysis, and to show real world ways in which uncertainty can be incorporated within the modelling workflow, from setting aims and objectives through to providing communication of confidence to decision makers.

The Menindee groundwater model developed for WaterNSW in 2016 (Jacobs, 2016) is unusual in the size of the dataset available, and the detailed conceptualisation completed previously by Geoscience Australia (GA) during the Broken Hill Managed Aquifer Recharge (BHMAR) in 2012. The example uses PEST and the Null Space Monte-Carlo (NSMC) method of uncertainty analysis (Doherty, 2016). Many of the techniques used to make predictive uncertainty analysis tractable for the NSMC method and the Menindee model would apply to other predictive uncertainty analysis techniques, where computational effort is a limiting factor.

Uncertainty by design

Planning for uncertainty analysis started with the overall project aim and model objectives. These were defined iteratively by WaterNSW, who procured the model, and the modelling team, in discussion with the regulator, DPI Water. The overall project aim was assessment of impacts from planned time-limited pumping of two new wellfields installed to provide emergency water supply to Broken Hill, and the subsequent recovery of groundwater systems. The impact assessment was to support abstraction licence applications.

From that overall aim, the following specific study objectives (Walker, 2017; Peeters, 2017) were set as:

- To assess the extent, magnitude and timing of drawdown impacts on groundwater-dependent ecosystems and other water users.
- To predict the impact on river flows and the timing of impacts, and whether primarily at periods of high or low flow.
- To assess the extent, magnitude and timing at which confined aquifers are made unconfined.
- To assess the likely mechanisms of aquifer recharge and the rate and frequency of recharge.
- To predict the recovery of groundwater levels after pumping and their rate of recovery.
- To predict whether and over what timescale groundwater resources will be replenished.
- The new wellfields had not been pumped beyond individual well pumping tests lasting a few days. Model calibration to the required type of prediction – the response to sustained pumping – was therefore not possible. A further objective was therefore set: to define the uncertainty, or error in predictions using the Null-Space Monte-Carlo method (Doherty, 2016).

The specified objectives are targeted on impact assessment; they are specific enough to describe required results, and they are realistic. They are not general; no mention is made of developing a ‘water resource management model’. And they are flexible enough to trust the modelling team to plan effectively, in that they do not prescribe the aquifers to be simulated or the domain. And they include predictive uncertainty analysis from the outset.

At first glance, these objectives appear to fall short of specifying ‘bad things we don’t want to happen’, as set-out by Doherty and Moore (2017). They do not explicitly set out hypotheses to be tested. However, the model scenarios and results naturally require a greater level of detail than the objectives list and those details were developed as the project progressed. Once this is done, the testing of relevant hypotheses is straightforward, and can be undertaken in a number of ways, including using plots shown later in this paper.

From the objectives, the factors that influenced model design, the modelling workflow and overall modelling philosophy included:

- That the model should focus on the required key predictions: assessment of environmental impacts of wellfield pumping in terms of drawdown, and the recovery of the groundwater system. It must capture sufficient complexity to represent the key aspects of the system for predictions.
- There was no experience of historic long-term wellfield operations and no experience of the drawdown and recovery impacts – the key modelling predictions sought. Therefore, the model needed to use as many different types of pertinent field data as possible to inform model calibration. Use of different data types in calibration assisted in narrowing down estimates of unknown parameters, such as vertical hydraulic conductivity, that strongly influence the drawdown and recovery responses.
- Known hydrogeological complexity, the number of observations available and estimates of distributions of parameters from GA made the use of parameter estimation software such as PEST essential. To best inform the model whilst ensuring a tractable solution, the time variant calibration covered as long a period of historical record as practical, over a period of intense hydrological drying followed by a significant flood. In the absence of long term pumping, this hydrological variability is critical to understanding groundwater recovery post-pumping.
- The groundwater model used and honoured the significant GA datasets and conceptual model as far as possible, unless evidence suggested they were not applicable.
- For key predictions, uncertainty analysis defined the likely range of impacts of wellfield pumping. With software such as PEST required for calibration, it made sense to use PEST and proven workflows for uncertainty analysis.
- Model design needed to take account that PEST was to be used for calibration and uncertainty analysis. A balance was required between model runtime and the complexity represented, such that the model could achieve its key objectives, with runtimes that were sufficiently short to make use of the full capability of PEST. The number of model cells, layers and estimated parameters, in addition to stress period numbers and durations, required careful judgment, and some model stress-testing (Peeters, 2017).
- Finally, the modelling was to follow good practice and be broadly consistent with the *Australian groundwater modelling guidelines* (Barnett et al., 2012).

The design of the model was discussed during the development stage. The following aspects were agreed or influenced by discussion between the modelling team and the project review group consisting of client, reviewers and the regulator:

- For predictions, the model considered a 10-year period. The regulator, DPI Water, assesses sustainability over a decadal timescale, so predictions of recovery were considered over this timeframe.
- The predictive model scenarios considered a range of borefield pumping rates, regimes, time frames and bores.
- The model used predictive uncertainty analysis on two main pumping and recovery scenarios. These were then used to inform a further eight scenarios run deterministically.
- The types of predictive model outputs were designed to provide objective information on the effects of pumping and the system response/recovery post-pumping.
- The model calibration period covered the range of drought to flood conditions from 1985 to 2016, and predictive scenarios were run using the latest ten years of hydrological data to 2016 as a surrogate for likely future variability.
- The modellers honoured GA-estimated hydraulic conductivity ranges and spatial patterns (which were derived from a range of methods, as described in the next section) to an agreed extent.
- The approach to incorporating faults into the model (to test elements of alternative conceptualisations) was also agreed upon.

A surfeit of data

Maintaining sufficient simplicity for the Menindee groundwater model was challenging in part due to the breadth and magnitude of the available data (Geoscience Australia, 2012). The content and scale of the dataset is described briefly below to set in context the model design focus and simplifications.

The primary source of information was the intensive investigation and research of the Broken Hill Managed Aquifer Recharge Study (BHMAR). The BHMAR study was undertaken by Geoscience Australia between 2008 and 2012 with funding from the Commonwealth Department of the Environment. The study comprised a wide suite of scientific and technical investigations with the objective of ‘assessing the viability of managed aquifer recharge (MAR) and/or groundwater extraction options to provide improved drought security for Broken Hill’

(GA, 2012). Subsequent to the BHMAR study completion in 2012 and reporting in 2013, further work by WaterNSW concentrated on the drilling of the Lake Menindee and Talyawalka borefields and associated wellfield yield assessments, see Figure 1.

The BHMAR study comprised a comprehensive range of scientific and technical investigations over four years that included conventional hydrogeological techniques such as drilling, piezometer installation, pumping tests, remote sensing data processing and analysis, hydrochemical studies and laboratory testing. This was supplemented by an extensive airborne and downhole geophysics program that included flying more than 30,000 km of airborne electro-magnetic (AEM) surveys with helicopters, ground resistivity surveys and downhole geophysics including nuclear magnetic resonance (NMR) profiling.

The scale of investigations and the outputs produced can be illustrated by the five main reports from the study that in total cover over 3,000 printed pages, and the digital GIS and other datasets produced that comprise more than 42 GB (the GIS component is 29 GB). The scope of the GIS dataset is illustrated in Table 1 by the section headings used within ArcGIS.

One unusual output to be integrated into the model design was spatial estimates of the range of hydraulic conductivity by aquifer.

Table 1: Section headings from BHMAR spatial datasets

Section Headings in GA ArcGIS Dataset			
Boundaries and Misc	Bores	Hydrochemistry	Ground Geophysics
Borefields and Targets	Topography	River	Geology
Land	Physiography	Geomorphology	Hydrogeology
AEM Data	AEM Sections	AEM Data	AEM Interpreted Products
Surface Materials	Flood Extent	Vegetation	Elevation
Basemaps			

Model design and simplification

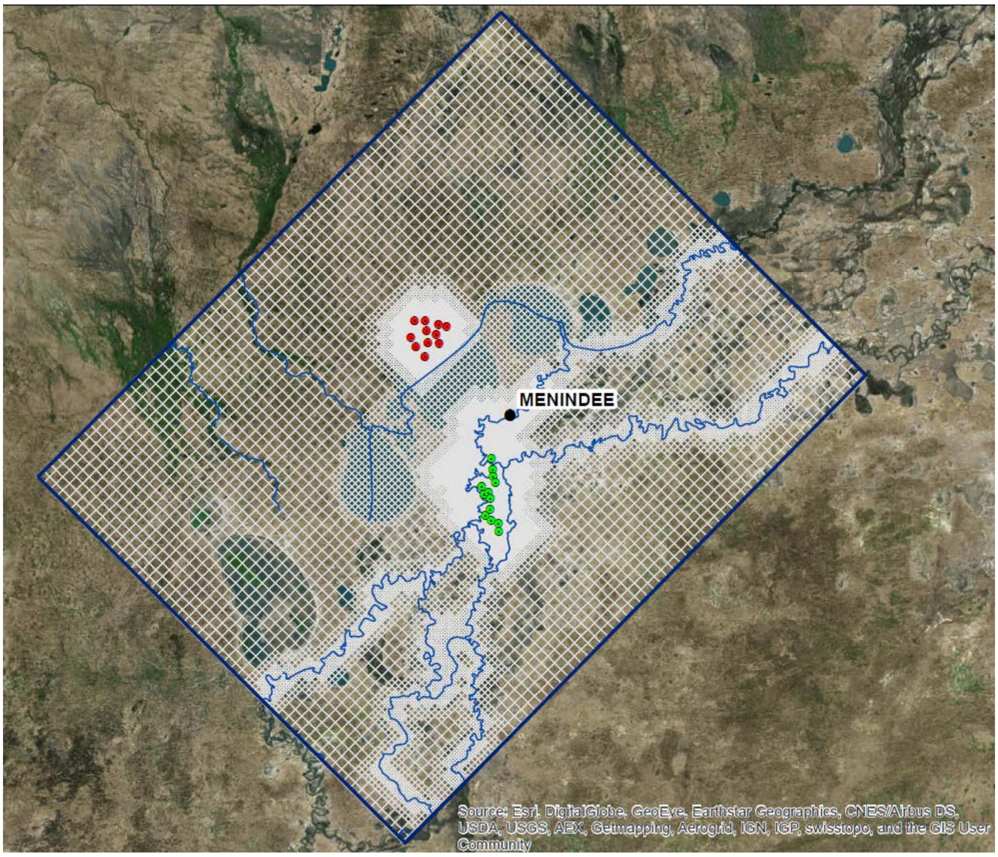
The key areas where model design was kept simple, or where potentially numerically intensive processes were streamlined, often using detailed data analysis, included:

- **Model code, domain and grid**
  - MODFLOW-USG was selected due to flexible mesh and integration with PEST.
  - A simulation of pumping using a rapidly developed 1-layer prototype model of the main aquifer helped define the required model domain.
  - Grid design focussed on predictions with refinement at wellfields, and where the system recharges, along the rivers and lakes, see Figure 1.
- **Model layering**
  - Three model layers were used for detailed modelling. A shallow unconfined aquifer layer, comprising three or more geological units including a clay drape that effectively excludes areal recharge across most of the floodplain, a confining clay layer (aquitard) with holes/gaps in certain localised areas, and the Calivil Formation, the main productive aquifer horizon.
  - Underlying units were excluded due to low vertical connectivity with the main aquifer.
  - Holes or gaps in the clay aquitard were represented in a simple manner, replacing aquitard properties with those of the overlying layer.
- **Aquifer properties**
  - The best available information from pumping tests at wellfields and the GA datasets was used.
  - Workflow was developed in conjunction with John Doherty to use GA spatial estimates and ranges within PEST calibration.
- **Model domain boundary conditions**
  - Boundaries were placed at sufficient distance to not unduly influence results.
  - They were represented in a simple manner by non-varying general head boundaries.

- **Rivers and lakes**
  - Surface water is the predominant source of recharge. The aim was to ensure sufficient representation of the surface-water system, without adding a large numerical burden.
  - With little evidence of active recharge from anabranches and other channels, only the main channel of the River Darling was included.
  - The MODFLOW river package was used as it is simpler to configure than the stream flow routing and lake packages, and because accurately simulating river and lake stages and their variability was critical to paying due respect to the GA conceptualisation of recharge processes via rivers and lakes. A PEST 'observation' constrained infiltration to prevent unrealistic river infiltration.
  - Hydrographs were combined with river-bed surveys and LiDAR topography to set river bed elevations (to establish physical/geometric constraints on the recharge process).
  - The ephemeral Menindee Lakes were included from historical lake stage data and bathymetry / topography. As the lakes dry, the boundary condition was switched from a surface-water boundary to evapotranspiration to take account of losses.
  - The surface-water time series was used to define variable stress periods of between 7 and 90 days. This ensured focus during the calibration, and predictive models, on periods of rapid change to the groundwater system, and enabled the models to reflect GA's conceptualisation of recharge processes from surface waters.
  - Part of GA's conceptual model was implemented by changing river-bed conductance depending on river stage (high/dynamic river flows erode the within-channel clay drape and allow for greater recharge than during low/stable flows). This was incorporated within the Python program that wrote the MODFLOW river and evapotranspiration packages and included parameters to enable PEST to optimise the degree of conductance change. Suitable allowable conductance ranges were defined through simple preliminary 1-dimensional spreadsheet modelling of observed groundwater levels and river stage.

  Evapotranspiration
  - Classification and mapping of vegetation by GA was simplified to three main classes: forest, woodland, and shrubs and grasses.
  - Extinction depths were set from literature on vegetation rooting depths.
  - PEST was able to optimise extinction depth.
  - Potential evaporation was set simply from average monthly open pan at Menindee. As actual evaporation is typically much less than potential, there was little to be gained from inclusion of the full historical timeseries.
  - A soft calibration target was included in PEST: modelled evapotranspiration should be higher in areas mapped by GA as being GDEs than in other areas.
- **Abstraction**
  - Prior multi-layer modelling of pumping indicated that production well yields should be sustainable for at least two years; this was also confirmed by source reliable output analysis using pumping test and installed capacity data.
  - Abstraction wells, shown on Figure 1, were represented using the 'standard' WEL package ahead of multi-node wells, or connected linear networks in MODFLOW-USG to reduce the chances of instability and convergence issues with PEST.
- **Neotectonic faults**
  - GA mapped over 500 small fault features with sufficient offset to juxtapose the shallow and Calivil sand aquifers and to open potential pathways to recharge.
  - Faults were incorporated using the simple method of modifying vertical hydraulic conductivity in the layer 2 aquitard, taking account of grid size.
  - More complex methods of representation, such as connected linear networks, J-array modification or grid offsets were rejected as taking too long to implement and / or increasing the possibility of instability and convergence issues.
  - Parameters were included in the fault processing Python script to allow PEST to optimise fault connectivity with constraints that faults with larger overlap were modelled with higher hydraulic conductivities that those with smaller overlaps.
- **Model simulation period**
  - The calibration was run over the period 1985–2016. Earlier data was available, but temporal data density is much lower prior to 1985.
  - Predictive scenarios were run as a hind cast of the most recent decade of hydrological data, with pumping starting in 2006 with recovery through to 2016. This shorter period was used for all predictive pumping scenarios. The computational burden of the stochastic realisations made creation of a shorter predictive simulation period worthwhile.

Figure 1: Menindee model domain showing mesh refinement at wellfields and along surface-water features



Optimising calibration

Calibration proceeded iteratively, starting from mean (but spatially variable) values of aquifer parameters estimated by GA. Initially, aquifer parameters were constrained to vary in a manner that respected both the spatial pattern developed by GA and the hydraulic conductivity range. Subsequently, the requirement to maintain the spatial pattern was relaxed, when it became clear that the conceptualisation warranted some refinement. Steps to make the model calibration process as efficient as possible included:

- The observed groundwater level data were resampled where necessary – from sub-daily time intervals to as long as necessary to adequately reflect the observed temporal groundwater level variability. A total of 8314 groundwater level observations were used in model calibration, having been re-sampled down from a data set of around 220,000 data points.
- The groundwater level data were incorporated into calibration in several ways:
    - As groundwater levels (8,314 observations);
    - As vertical head differences at nested bore sites (2,191 observations);
    - As lateral head differences between the river and/or lakes and nearby groundwater observation bores (4,410 observations);
    - As temporal head differences per bore – from each observation to the next (8,220 observations);
    - As temporal head differences per bore (drawdown) – from the first observation to all subsequent (8,220 observations).
- The number of parameters to be calibrated was controlled by careful placement of pilot points used to estimate aquifer parameters, and global factors in PEST were used to control other aspects such as EVT, faults, and the change in river conductance with stage.
- Additional ‘observations’ were specified to stop PEST from exploring nonsensical solutions, such as infiltrating more water than was flowing in the river. These observations assisted in constraining the calibration with additional ‘expert knowledge’.
- A total of 409 calibratable parameters and 31,953 calibration targets were ultimately used.

For model calibration, each run took approximately three hours to complete including pre- and post-processing. Calibration required approximately 5,000 hours of compute time, using PEST-HP, multiple cores and machines and a combination of local and cloud computing resources. Predictive scenarios consumed a similar level of resources.

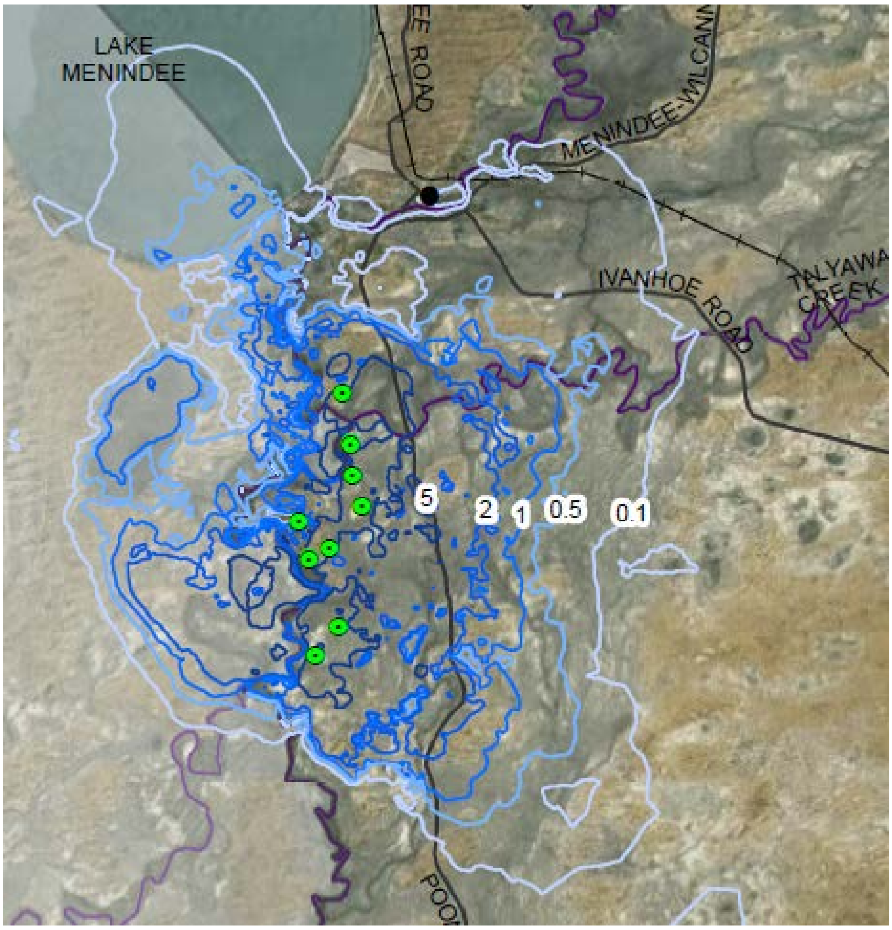


Figure 2: Example output showing predicted 50th percentile unconfined aquifer drawdown after two years of pumping

Predictive scenarios

Predictive uncertainty analysis was completed following the null-space Monte-Carlo method with 1,000 realisations created by PEST from the calibrated model datasets and run through the calibration period from 1985–2016. From these runs a sub-set of 153 realisations were selected where the objective function had not changed by more than 5% from the calibrated model. No recalibration of the stochastic runs was undertaken, saving significant run time.

Predictive uncertainty analysis was then completed using the ensemble of 153 realisations. Full predictive analysis was restricted to two primary scenarios: one operational base case for each wellfield. Each scenario was run from 2006-2016, using starting heads from the calibration run. As the stress period setup was changed to accommodate pumping, a base case of 153 non-pumping realisations was also run. The computational saving from a reduced duration for the predictive scenarios was worth the additional pre-processing as flexible utility program design simplified the task.

A final efficiency step was achieved with pre-agreed model outputs. This meant that post-processing for each stochastic realisation was automated via batch files to be executed upon completion of the model run.

Example results showing predicted 50th percentile drawdown is shown in Figure 2 with a stochastic representation of abstract impact and recovery shown in Figure 3.

Discussion

The Menindee model demonstrates that predictive uncertainty analysis is tractable, even with the largest of datasets. It is regarded as a successful modelling project that met all objectives. Upon completion, WaterNSW noted that the:

> *'capacity to manage potential complexities from the modelling brief ensures that the model is as simple as needed to offer a rigorous response to the predictions sought and the time constraints. At the same time, it satisfies the expectations of peer reviewers and our regulator.'*

WaterNSW also commented on the clear communication of model technicalities and its results (echoing principles outlined in Richardson et al., 2017). What made for that success? From our perspective, the foundations to success were:

- Uncertainty analysis was planned from the outset, and included within project budgets.
- Objectives were focussed (see also Walker, 2017; Peeters, 2017). Unhelpful objectives, such as defining a maximum RMS error for the model calibration, were omitted.
- The client contributed by engaging with the modelling team regarding scope, by not insisting on lump sum costs, and facilitating the project review group.
- Model design decisions were left to the informed judgement of the modelling team, with the project review group commenting.
- A knowledgeable modelling team was assembled, with members capable of making informed decisions on simplification and with sufficient programming skill to create tools to shorten model development and runtimes, and to facilitate efficient incorporation of the extensive data sets.
- Reviewers were accessible throughout, and also acted as trusted advisors. Reviewers were constructive and helpful, rather than negative and point scoring.
- The Menindee model described here was a second phase of modelling that followed simple wellfield yield modelling completed in 2015. The experience from that work assisted significantly when deciding on key aspects of model design such as layering.

In terms of the simplifications made to make the predictive uncertainty analysis workflow tractable the areas we highlight are:

- Many of the simplifications and approximations made to the conceptual model helped reduce model runtime. A standard calibration MODFLOW run was kept to 60–90 minutes maximum.
- Other design decisions, such as adoption of standard wells and avoidance of connected linear networks or unsaturated zone representation, were made to promote model stability such that PEST could calculate continuous derivatives. Some decisions, such as adoption of MODFLOW-USG and grid refinement helped both runtime and stability.
- PEST was applied as efficiently as possible by controlling the number of calibration parameters. In hindsight, we should have included more spatial variability than we did (to better explore alternative conceptualisations and parameterisations); however, time was a limiting factor. Predictive scenarios were shortened to reduce computational burdens.

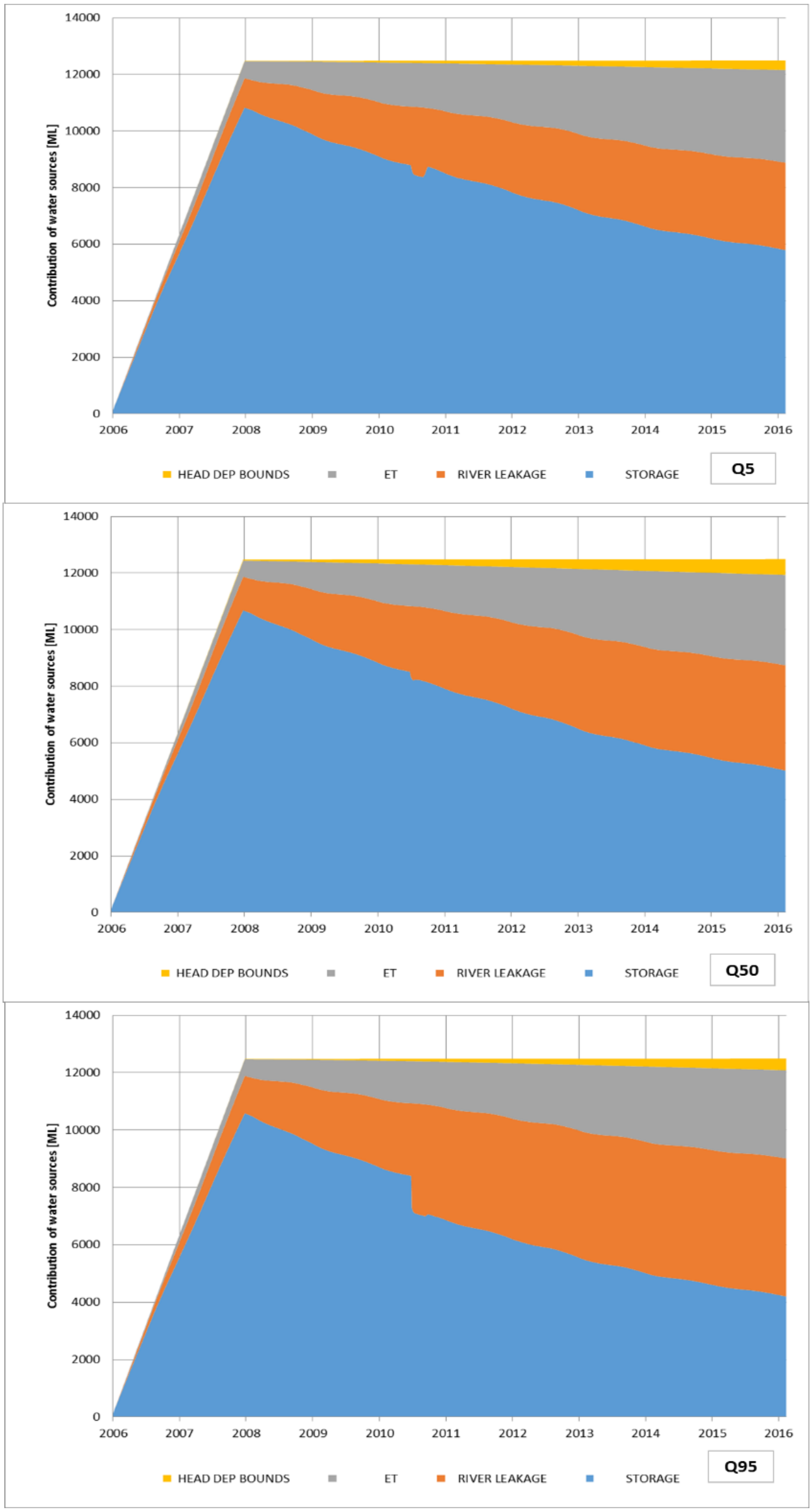


Figure 3: Example probabilistic output showing aquifer depletion and recovery using water balances

Conclusions

We conclude that the Menindee model example demonstrates that predictive uncertainty analysis can be rendered tractable, even with large datasets. To do so requires careful planning from the outset, both by those requisitioning a model, and the modellers, with tightly defined objectives. Focussing the model and numerical effort on the key drivers of the hydrogeological system and those that control predictions enables runtimes to be limited. And judicious application and use of PEST enables calibration and uncertainty analysis to be completed without supercomputing power.

Acknowledgements

We acknowledge work, assistance and review by Vincent Puech and Brian Barnett of Jacobs, WaterNSW, DPI Water, the external reviewers and advisors, and staff of the GA Groundwater Group.

We are very grateful to WaterNSW for permission to publish this paper.

References

Barnett B, Townley L, Post V, Evans R, Hunt R, Peeters L, Richardson S, Werner A, Knapton A and Boronkay A. 2012, *Australian groundwater modelling guidelines.* Technical report, Waterlines Report no. 82, National Water Commission, Canberra.

Doherty J and Moore C. 2017. ‘Simple is beautiful’. Discussion paper no.3 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. NCGRT.

Doherty J. 2016. *PEST, Model-Independent Parameter Estimation User Manual, v6*. Watermark Numerical Computing, Brisbane.

Geoscience Australia. 2012. *BHMAR Project: Assessment of Potential Groundwater Resources and Underground Storage Options in the Darling Floodplain near Menindee*. Geoscience Australia, Canberra.

Jacobs. 2016. *Menindee Regional Groundwater Model*. IH087100_Rev1. Unpublished report prepared for WaterNSW. 10 November 2016.

Peeters L. 2017. ‘Uncertainty analysis in groundwater modelling: where to start?’ Discussion paper no.2 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. NCGRT.

Richardson S, McMillan M and Currie D. 2017. ‘Communicating uncertainty to decision makers and stakeholders’. Discussion paper no.5 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. NCGRT.

Walker G. 2017. ‘Predictive uncertainty in groundwater modelling: how is it best used?’ Discussion paper no.1 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. NCGRT.

Communicating uncertainty to decision makers and stakeholders

Stuart Richardson, CDM Smith; Dr Miriam McMillan, Office of Water Science; and Dr Dougal Currie, CDM Smith

Introduction

Models cannot predict the future with complete (100%) confidence and they cannot be used to make decisions for us. This is the job of decision makers and stakeholders who must exercise their judgement to decide what level of risk is acceptable for a specific context. Modelling aids the process of determining the potential outcomes for specified courses of action, and uncertainty analysis can quantify or qualify the confidence we have in the modelled outcomes. Together, the model results and uncertainty analysis should be used by decision makers as a guide to the likelihood of consequences eventuating (be they beneficial or adverse) for any chosen course of action.

A key part of uncertainty analysis is, however, often overlooked or not adequately addressed: its communication. When modelling is intended to support decision making it is essential that uncertainty is communicated in a manner that is useful for decision making. Poor communication of uncertainty often leads to the impression that model outputs are unreliable; though in fact, 'sound decision making requires a full understanding of the full range of possible consequences and associated possibilities' (Mastrandrea et al., 2010). Peeters (2017) provides an overview of assessment workflow that includes uncertainty analysis and highlights the need to have transparency in communication and reporting of uncertainty.

The key to successful communication of uncertainty is to present the information in a way that is most likely to aid decision making. To achieve this, analysis of uncertainty information in model output needs to be: (i) adequately tailored to decision-makers' needs, (ii) focussed on the messages that are most likely to be relevant to their decisions, and (iii) presented in plain and clear (precise, non-technical) language. These three key factors require a meaningful two-way dialogue between modellers and decision makers from the early stages of a project. Transparency about the modelling objectives is also necessary and these need to be discussed early in the project workflow with the decision makers. This can help to make sure that modellers and decision makers agree on the objectives.

This paper

In this paper, we outline some key principles that underpin sound communication of uncertainty analysis with a discussion of:

- How uncertainty analysis is perceived and why it is important to decision makers;
- How analysis of uncertainty helps with decision making and what should be communicated; and
- Some best practice guidance for communication of uncertainty analysis.

This discussion demonstrates that it is possible to improve how information about uncertainty analysis is communicated to decision makers by using some existing examples of good practice as a guide. Taking the perspective of the decision maker rather than the modeller we aim to provoke discussion on how to reach an acceptable trade-off between what is needed and what can be provided.

Understanding of roles

Effective communication of uncertainty requires an understanding of the role of the decision maker and their needs. It is imperative to also understand how they interact with other parties informing and responding to the decision making process.

In the context of water resource planning, Refsgaard et al. (2007) talk about four different types of actors in the process:

- **Water manager**
  The person or organisation responsible for the management or protection of the water resources, and thus of the modelling study and the outcome (the problem owner).
- **The modeller**
  A person or an organisation that develops the model and works it, conducting the modelling study (if the modeller and the water manager belong to different organisations).
- **The reviewer**
  A person who is conducting some kind of external review of a modelling study. The review may be more

or less comprehensive depending on the requirements of the particular case. The reviewer is typically appointed by the water manager to support her/him to match the modelling capability of the modeller.

- **The stakeholders/public**
  Any interested party with a stake in the water management issue, either in exploiting or protecting the resource. Stakeholders include the following categories: (i) competent water resource authority (typically the water manager, cf. above), (ii) interest groups, and (iii.) general public.
- **The project proponent**
  The project proponent (an additional 'actor' where the analysis involves a development such as a mine or infrastructure) is the person or organisation that owns the project or development (e.g. a mine). The project proponent is asking the water manager to make a decision related to impacts to a water resource. The project proponent commissions studies by outside professionals such as consulting hydrogeologists.

The water manager may be the primary decision maker. This person should interact with all parties including the technicians and the public (e.g. irrigators). The decision maker will require the modeller to generate outputs from models that can be understood by all stakeholder groups – requiring clarity in communication with stakeholders.

Stakeholders still need to understand what the range of possible predictions means for their interests in the decision-making process. This is especially true when the range of predictions is likely to be relevant for particular concerns (e.g. a clear view of the likelihood of a possible reduction in groundwater flux to a river), and indicates how it can be modified or managed. The decision maker may want to understand the range of possible outcomes including whether a set of circumstances is possible that will likely result in unacceptable outcomes.

The modeller is in a good position to describe the range of uncertainty but may not be as well placed to know what it means for the decision. Therefore, the modeller's focus should be on making sure the estimates of uncertainty are accurately and clearly described using language and statistical measures that non-experts can understand, and framed in a manner that supports the decision-making process.

The decision maker will not only require information about the overall uncertainty of model predictions, but will need to understand the sources of this uncertainty and whether there are biases inherent in the modelling undertaken. Knowledge about what is contributing to the uncertainty and its relative influence enables a decision maker to prioritise efforts to reduce uncertainty or manage its effects on decisions and management plans. Knowledge about biases in the modelling undertaken can assist the decision maker in determining whether the model predictions are conservative, balanced or optimistic.

The format of the uncertainty information provided to the decision maker is important and should ideally be quantitative given that most people: '(i) like receiving explicit quantitative expressions of uncertainty (such as credible intervals), (ii) can interpret them well enough to extract their main message, and (iii) are liable to misinterpret qualitative expressions of uncertainty (e.g., 'good' evidence, 'rare' side effect) (Fischoff and Davis, 2014).

Principles for communication of uncertainty – framed to support the decision-making process

The type of information about uncertainty that needs to be communicated with decision makers and the level of detail required to inform their decision making is context-specific and dependent on how sensitive the risk (event) is to the magnitude of uncertainty present (Flage and Aven, 2009). In general, where information about uncertainty analysis is relevant to a particular decision, the decision maker needs to understand the types, sources and levels of uncertainty. This information should be presented quantitatively where the magnitude of the impact and uncertainty is significant or otherwise be presented more simply as a qualitative description of biases and data gaps that impact on the estimated magnitude of the event (noting that qualitative descriptions of uncertainty need to be precisely defined to avoid misinterpretation).

The most critical aspect is tailoring the uncertainty analysis to what is needed to inform the decision (Walker, 2017) before the analysis begins. Achieving this necessitates careful design and meaningful review of the uncertainty analysis from an early stage in the project. Communicating end-products to decision makers is a last step in what (ideally) should be multiple communication steps beforehand. However, there are practical limitations to the extent of engagement with stakeholders during a technical process. Middlemis et al. (2017) suggest that engagement doesn't have to be 'intensive', but rather targeted at critical points in the workflow such as conceptualisation of causal pathways.

Ultimately, the nature of the decision will guide how the uncertainty analysis should be framed. Fischhoff and Davis (2014) discuss three types of situations where a decision is required:

- For situations where a threshold or trigger for action needs to be defined (e.g. whether a mitigation measure needs to be activated or not), uncertainty analysis should be framed around the confidence in the model predictions (for example) of the defined threshold being breached.

- For a choice between fixed options (e.g. setting different pumping limits), uncertainty analysis should be framed around the confidence in predicted outcomes (either adverse or beneficial) that result for all options being evaluated (e.g. the rate of groundwater discharge to a stream that results from different pumping rates).
- For decisions about potential options testing what may be possible (e.g. pumping optimisation or water supply investigations), uncertainty analysis should be framed around the confidence in the processes being simulated which shape the outcomes (e.g. does the model include all the processes which influence water availability, such as inter-aquifer leakage in response to drawdown, and how well is each of these processes known).

In many cases the decision maker needs to know the 'most likely' outcome and to understand whether there are circumstances or possibilities that result in an unacceptable outcome – will the option fail under conditions z, y or z? This could relate to assumptions on parameters set within a model prediction, or some other component of the conceptual model. The uncertainty analysis should be framed around understanding the confidence level to predict the likelihood of failure. That is, what is the level of confidence in the quantification of the likelihood of failure? In this regard, uncertainty analysis can support quantitative risk analysis (see for example Flage and Aven, 2009).

Principles for communication of uncertainty – systematic documentation and presentation of sources of error and potential bias

Thorough uncertainty analysis systematically documents all sources of uncertainty and potential bias which contribute to the overall uncertainty in the model predictions. Bias is an important component of model uncertainty, but is often not made explicit in uncertainty analysis. The diagram shown in Figure 1 illustrates how errors and bias contribute to uncertainty.

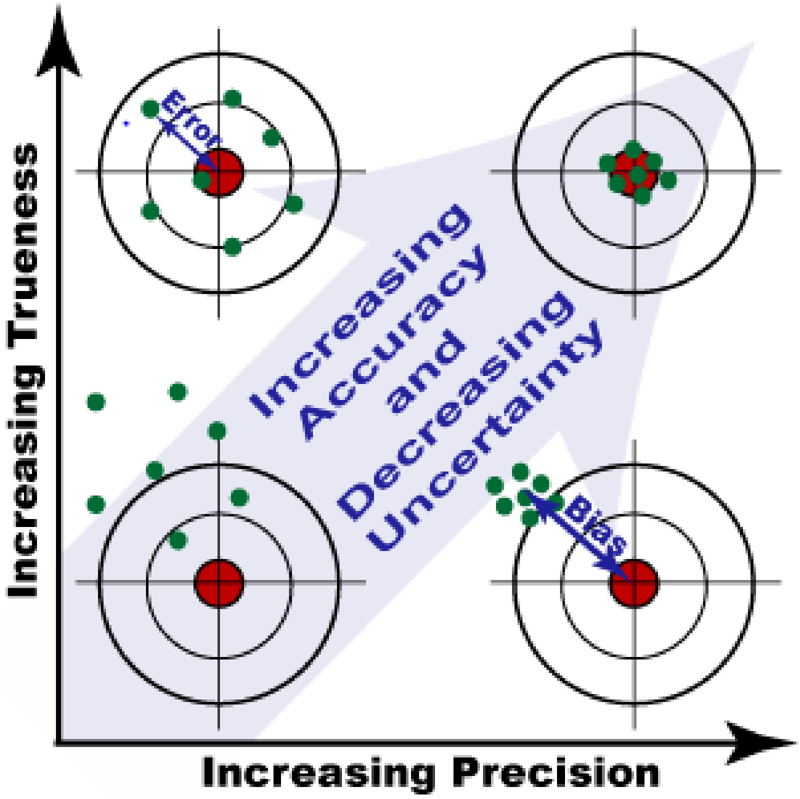


Figure 1: Errors, biases and their influence on uncertainty (NDT Resource Center, 2017)

Random error is one component of uncertainty that may result from a variety of factors (e.g. natural variability, measurement noise), influencing the precision in the model output. Bias refers to systematic error, which displaces the model outputs in a predictable way. This influences the trueness in the model output, where trueness is the difference between the average value obtained from model predictions and an accepted true value. Being over committed to one conceptualisation over others (bias), perhaps the wrong one, could lead to predictions that overestimate or underestimate an impact. If uncertainty analysis only focusses on errors and neglects to account for biases, incomplete and distorted evidence of the modelling accuracy will be provided. It is therefore crucial that biases are documented.

When communicating uncertainty, the sources of uncertainty and their relative importance to overall uncertainty should be clearly outlined. An example of how this can be applied is provided in Table 1. A table such as this facilitates decision making by readily identifying components driving overall uncertainty so efforts to reduce uncertainty can be prioritised. While the approach uses straightforward language, each of the terms would need to be clearly defined if it is to be used effectively, to avoid misinterpretation.

Table 1: An example of an uncertainty matrix showing 'Sources of uncertainty and their importance in a specific project context' (Refsgaard et al. 2007)

Source of uncertainty	Type of uncertainty				Importance	
	Statistical uncertainty	Scenario uncertainty	Qualitative uncertainty	Recognised ignorance	Weighting	(Uncertainty × weight)
Problem context						
– Future agricultural practice		Medium	Medium	Medium	Large	Medium
– Future climate		Medium	Medium	Large	Medium	Medium
Input data						
– Catchment data	Medium			Small	Large	Medium
– Nitrate load from agriculture	Small			Small	Large	Small
Parameter uncertainty						
– Water quantity	Small			Small	Medium	Small
– Water quality	Medium			Medium	Medium	Small
Model structure (conceptual)						
– Geology		Large	Large	Medium	Large	Large
– Nitrate reduction in underground		Medium	Medium	Large	Large	Large
Model technical uncertainty						
– Numerical approximation	Small			Small	Medium	Small
– Bugs in software				Medium	Medium	Small
					SUM:	

Being aware of subjective biases

Having uncertainty information that fully informs risk assessments has been shown to result in better decision making (e.g. Joslyn and LeClerc, 2012). However, many decisions involving risk are complex and require a lot of mental effort involving multiple concepts that often conflict with each other. To reduce this mental effort, our brains often use shortcuts that allow us to make rapid judgements with less mental effort (Kahneman, 2011). These shortcuts are hard-wired into human brains and result in consistent cultural and personal biases in how we make decisions (New Zealand Government, 2014). These biases can affect how different groups of people (including experts and non-experts) interpret information about uncertainty (Kahneman, 2011; Kloprogge et al., 2007). This is shown to be true in particular for interpretation of information about probabilities (New Zealand Government, 2014).

It is not possible to completely eliminate biases from limit-setting decision making because the decisions require value judgements to be made regarding what is important and what is an acceptable level of risk (New Zealand Government, 2014).

When seeking to effectively communicate information around uncertainty to stakeholders involved in groundwater modelling processes, professionals benefit when subjective biases are identified and communicated to the greatest extent possible (New Zealand Government, 2014). Minimising the influence of these possible biases should come from early acknowledgement of possible biases (for greater transparency), being able to present output in a non-emotive way and by being data-driven in the analysis.

Relevant and common biases

Four of the most relevant and common biases that can affect communication of uncertainty are (following New Zealand Government, 2014):

- **Availability bias**
  'People tend to judge events that are easily recalled as more risky or more likely to occur than events that are not readily available to memory. An event may have more availability if it occurred recently, if it was a high-profile event, or if it has some other significance for an individual or group.'
- **Confirmation bias**
  'Confirmation bias refers to the filtering of new information to fit previously formed views. In particular, it is the tendency to accept as reliable new information that supports existing views, but to see as unreliable or erroneous and filter out new information that is contrary to current views. People may ignore or dismiss uncertainty information if it contradicts their current beliefs.'

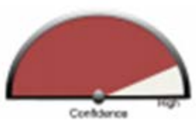
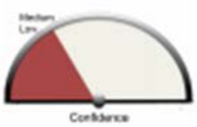
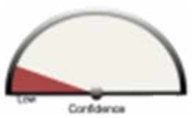
- **Confidence bias**
  'People typically have too much confidence in their own judgements. This appears to affect almost all professions, as well as the lay public. The few exceptions are people who receive constant feedback on the accuracy of their predictions, such as weather forecasters. The psychological basis for this unwarranted certainty seems to be insensitivity to the weaknesses in assumptions on which judgements are based.'
- **Framing bias**
  'How probabilistic information is framed can influence how that information is understood as well as the confidence that people have in the information. 'Priming' the brain with a particular stimulus can affect how it responds to a later stimulus. Using expressions that take advantage of this priming (i.e. the direction and expression are consistent) can reduce cognitive strain, which makes it easier for stakeholders to understand the idea presented without requiring further analysis. For example, the phrase 'there is a 5% chance the drawdown in the groundwater level will be greater than 0.2 m' may leave a different impression than the phrase 'there is a 95% chance the water drawdown level will be less than 0.2 m'; even though the two phrases contain the same information (New Zealand government 2014). The latter requires less mental workload because your brain is already 'primed' to think about being 'down' when it hears 'less than'. This is particularly effective when paired with explicit advice about whether precautionary action is advised.'

Concise, consistent and easily understood

For the decision maker it is important to have a clear description of the confidence in the model's ability to provide accurate predictions; for example, discussion of evidence to support knowledge of a particular process. Being concise is an integral part of this. If modellers and technicians include excessive detail and complexity in the communication of uncertainty, decision makers often become overwhelmed and could be less inclined to rely on the model results due to a lack of understanding. This is particularly important when model results are to be communicated to a wide audience of stakeholders, with varying levels of scientific understanding.

Consistency and precision in language is required to help avoid the subjective decision-making biases by the water manager or the project proponent. It is critical to not distort the relative importance of the findings presented. Moss and Yohe (2011) provide some guidance on the use of consistent language in communicating uncertainty and their definition of confidence levels is outlined in Table 2 which could be modified to fit a groundwater-related context. Other examples on the use of qualitative terms to describe uncertainty can be found in Regan et al. (2002) and Uusitalo et al. (2015).

Table 2: Confidence levels as defined by Moss and Yohe (2011)

	Confidence Level	Example combinations of factors that could contribute to this confidence evaluation
	High	Strong evidence (established theory, multiple sources, consistent results, well documented and accepted methods, etc.), high consensus
	Medium High	Moderate evidence (several sources, some consistency, methods vary and/or documentation limited, etc.), medium consensus
	Medium Low	Suggestive evidence (a few sources, limited consistency, models incomplete, methods emerging, etc.), competing schools of thought
	Low	Inconclusive evidence (limited sources, extrapolations, inconsistent findings, poor documentation and/or methods not tested, etc.), disagreement or lack of opinions among experts

The New Zealand government (2014) builds on the guidance of the IPCC (Mastrandrea et al. 2010) to suggest a consistent approach to combining the narrative descriptors of likelihood, of a given outcome, with quantitative ranges in probabilities (Table 3). This table provides descriptors of likelihood from other sources/contexts.

Table 3: Suggested narrative descriptors of ranges in likelihood (New Zealand Government, 2014)

Probability	Intergovernmental Panel on Climate Change (IPCC) scale[4]	Scale based on legal standards of proof[5]	Environmental risk management authority (ERMA) scale[6]
100%	–	Beyond any doubt	–
>99%	Virtually certain	Beyond a reasonable doubt	Highly likely
90–99%	Very likely	Clear and convincing evidence	Highly likely
80–90%	Likely	Clear showing	Highly likely
67–80%	Likely	Substantial and credible evidence	Likely
50–67%	About as likely as not	Preponderance of evidence	Likely
33–50%	About as likely as not	Clear indication	Unlikely (occasional)
10–33%	Unlikely	Probable cause, reasonable belief	Very unlikely
1–10%	Very unlikely	Reasonable grounds for suspicion	Highly improbable
<1%	Exceptionally unlikely	No reasonable grounds for suspicion	Highly improbable
0%	–	Impossible	–

Note: Dashes (–) indicate that no equivalent point is provided in the IPCC and ERMA scales.

This same work by the Ministry of the Environment (2016) describes the strengths and weaknesses in numeric, narrative and visual methods for communication. A simple example of combining these approaches is provided in Table 4 where the ranges of IPCC probability classes are simplified and combined with a visual colour coded cue. This concept is explored further within the summary report on Groundwater Modelling Uncertainty by tailoring the narrative descriptor, so it relates to the likelihood of exceedance of a threshold condition.

While the numeric approach allows for ‘precision’ in the description, the presentation of numbers does not always appeal to people who look for more contextual descriptions. A key advantage of the visual is that it can allow a person to quickly see patterns in outputs, which can be easily related to the numeric and narrative descriptors.

The concept of the use of simple ‘calibrated’ language is used by the IPCC (Mastrandrea et al, 2010) to rank confidence in analysis based on combinations of agreement and evidence (Table 5). In this example, there is deemed to be greater confidence where ‘high’ agreement and ‘robust’ evidence intersect. Agreement is a qualitative term that could come from a technical reference group.

4 Mastrandrea et al. (2010).

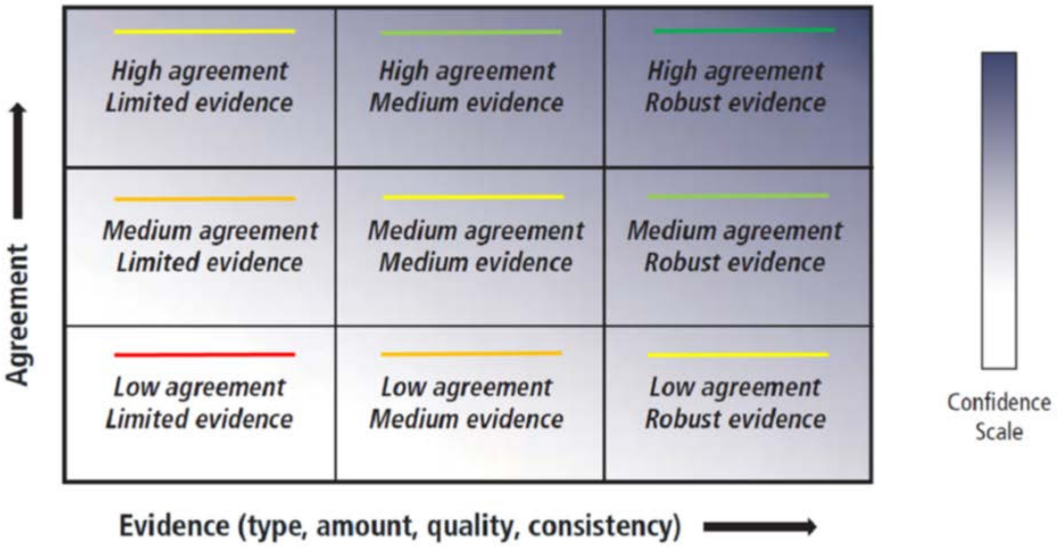
5 Weiss (2003).

6 Environmental Risk Management Authority (2009).

Table 4: Example approach to combined numeric, narrative and visual approach to describing likelihood (Ministry of the Environment, 2016)

Narrative descriptor	Probability class	Description	Colour code
Very likely	90–100%	Likely to occur even in extreme conditions	
Likely	67–90%	Expected to occur in normal conditions	
About as likely as not	33–67%	About an equal chance of occurring as not	
Unlikely	10–33%	Not expected to occur in normal conditions	
Very unlikely	0–10%	Not likely to occur even in extreme conditions	

Table 5: An example of using calibrated language to define (and rank) confidence (from Mastrandrea et al, 2010)



Best practice for communicating uncertainty

The following guiding principles are suggested for sound communication of uncertainty:

1. Identify the stakeholders (audience) and engage with them early in the process to decide on whether the evaluation of uncertainty is important for the decisions the modelling is intended to inform, and if so, why.
2. Communicate the context of the uncertainty analysis. Early conversations should convey the positive aspects of being clear about uncertainty, i.e. there is greater transparency and credibility so that analysis can be seen to be unbiased.
3. Contextualise the uncertainty by describing the risk, that is, the consequences and boundaries within which the uncertainty is applicable.
4. Related to point 2, align the type and presentation of uncertainty to objectives, i.e. if the analysis is to inform an approval for a new development then uncertainty should relate to the range of possible impacts associated with that development.
5. Check that the analysis is relevant to the context of the issue at hand, (for example, if the analysis relates to a water planning issue then it should be policy-relevant) i.e. support the iterative science-based development of water policy with an initial range in predicted outcomes that are assessed and used to refine policy settings.
6. Indicate the type of uncertainty present (and from what source/s) and how it was characterised. Describe the relative contribution of each source of uncertainty (following Kloprogge et al., 2007).
7. Match how the uncertainty is expressed (positive or negative framing) to the expressions used in the decision context. Using expressions that take advantage of this priming (i.e. the direction and expression are consistent) can reduce cognitive strain, making it easier to understand ideas presented

without further analysis (Joslyn and LeClerc, 2012). A good way to frame uncertainty output is to present threshold probability estimates (at the threshold relevant to the user's decision) where possible – these are often the most usable kind of uncertainty information (Joslyn and LeClerc, 2012). If the fact is presented in isolation, also present the uncertainty values in terms of odds (e.g. 1 in 100 as well as 1% chance) (Fischhoff et al., 2002).

8. Provide a clear description of the scale of uncertainty using combinations of numeric, narrative and visual descriptors. Use plain and clear, precise non-technical language. Present uncertainties in multiple formats, including probabilities and associated qualitative (categorical) descriptions (e.g. likely, meaning greater than 66%). Use consistent (preferably established, e.g. IPCC) calibrated terminology to define quantitative and qualitative descriptors of uncertainty. This calibrated terminology should be clearly defined so decision makers understand what the terms mean and are aware of biases.
9. Discuss how uncertainty may influence decisions, i.e. are the methods used to estimate uncertainty known to have biases making them likely to either over or under estimate the potential impacts? And if so, what is the magnitude?
10. Indicate which uncertainties (if any) have been or may be reduced and how that is achieved. Also indicate a general structure of the cost–benefit to reducing it, i.e. time/$/person hours etc. vs how much the uncertainties can be reduced and whether it's likely to impact the decision.
11. Use positive and balanced language. Too often, it seems, a narrative description of uncertainty focusses on those aspects 'we don't know'. This sometimes has the effect of creating confusion and usually leads to perceptions the work is of low value. A more positive approach is suggested in this paper where the reporting framework is one of probability or likelihood of an event occurring. A narrative of uncertainty can also be balanced by describing those aspects 'we do know about'.

References

Environmental Risk Management Authority (ERMA). 2009. *Decision making: A technical guide to identifying, assessing and evaluating risks, costs and benefits.* ERMA New Zealand, Wellington.

Fischhoff B, Bostrom A, and Jacobs-Quadrel M. 2002. 'Risk perception and communication'. *Oxford Textbook of Public Health*, *4th ed*. Eds. Detels R, McEwen J, Reaglenhole R and Tanaka H). Oxford University Press, New York, 1105–1123

Fischhoff B and Davis A. 2014. 'Communicating scientific uncertainty'. *PNAS* 111.4, 13664–3671, 16 September, 2014.

Flage R and Aven T. 2009. 'Expressing and communicating uncertainty in relation to quantitative risk analysis'. *Reliability & Risk Analysis: Theory & Application*, 2.13, 9–18.

Joslyn SL and LeClerc JE. 2012. 'Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error'. *Journal of experimental psychology: applied*, 181, 126.

Kahneman D. 2011. *Thinking, fast and slow.* Macmillan.

Kloprogge P, van der Sluijs J and Wardekker A. 2007. *Uncertainty communication*. Report NWS-E-2007-199. Copernicus Institute for Sustainable Development and Innovation. http://www.nusap.net/downloads/reports/issuesandgoodpracticefinalmanuscript.pdf

Mastrandrea MD, Field CB, Stocker TF, Edenhofer O, Ebi KL, Frame DJ, Held H, Kriegler E, Mach KJ, Matschoss PR and Plattner GK. 2010. *Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties*.

Middlemis H, Walker G, Peeters L, Hayes P, Moore C and Richardson S. 2017. *Overview on improved evaluation of the effect of hydrogeological uncertainty on project objectives*. Overview paper for NCGRT national groundwater modelling uncertainty workshop. Sydney, 10 July 2017.

NDT Resource Center. 2017. https://www.nde-ed.org/GeneralResources/ErrorAnalysis/UncertaintyTerms.htm

Ministry for the Environment and Ministry for Primary Industries. 2016. *A draft guide to communicating and managing uncertainty when implementing the National Policy Statement for Freshwater Management*. New Zealand Government, New Zealand. http://www.mfe.govt.nz/publications/fresh-water/draft-guide-communicating-and-managing-uncertainty-when-implementing

Moss RH and Yohe G. 2011. *Assessing and Communicating Confidence Levels and Uncertainties in the Main Conclusions of the NCA 2013 Report: Guidance for Authors and Contributors*. National Climate Assessment Development and Advisory Committee (NCADAC),. www.globalchange.gov/

NDT Resource Center. 2017. www.nde-ed.org/GeneralResources/ErrorAnalysis/UncertaintyTerms.htm

Office of Best Practice Regulation, 2014. *Research Report: Environmental valuation and uncertainty*. Department of Prime Minister & Cabinet, Australia. www.dpmc.gov.au/sites/default/files/publications/012a_Research_Report_Environment.pdf

Peeters L. 2017. 'Uncertainty analysis in groundwater modelling: where to start?' Discussion paper no.2 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. NCGRT.

Petersen AC, Janssen PHM, van der Sluijs JP, Risbey JS and Ravetz JR. 2003. *RIVM/MNP Guidance for Uncertainty Assessment and Communication: Mini-Checklist and quickscan questionnaire*. Netherlands Environmental Assessment Agency, Netherlands. http://www.rivm.nl/bibliotheek/digitaaldepot/Guidance_MC_QS-Q.pdf

Refsgaard JC, van der Sluijs JP, Højberg, AL and Vanrolleghem PA. 2007. 'Uncertainty in the environmental modelling process–a framework and guidance'. *Environmental Modelling & Software*, 22.11, 1543–1556.

Regan HM, Colyvan M and Burgman MA. 2002. 'A taxonomy and treatment of uncertainty for ecology and conservation biology'. *Ecological applications*, 12.2, 618–628

Uusitalo L, Lehikoinen A, Helle I and Myrberg K. 2015. 'An overview of methods to evaluate uncertainty of deterministic models in decision support'. *Environmental Modelling & Software*, 63, 24–31

Walker G. 2017. 'Predictive uncertainty in groundwater modelling: how is it best used?' Discussion paper no.1 in Middlemis et al. 2019. *Groundwater modelling uncertainty – implications for decision making*. NCGRT.

Weiss W. 2003. 'Scientific uncertainty and science-based precaution'. *International Environmental Agreements: Politics, Law and Economics*, 3, 137–166.

APPENDIX B

Glossary of key terms used in groundwater modelling uncertainty analysis

Bias	Bias refers to systematic error, which displaces the model outputs in a predictable way. This influences the trueness in the model output, where trueness is the difference between the average value obtained from model simulations and an accepted true value. (Richardson, McMillan and Currie, 2017). Four of the most relevant and common biases that can affect communication of uncertainty are listed below (New Zealand Government, 2014).
Bias – Availability	People tend to judge events that are easily recalled as more risky or more likely to occur than events that are not readily available to memory. An event may have more availability if it occurred recently, if it was a high-profile event, or if it has some other significance for an individual or group.
Bias – Confirmation	Refers to the filtering of new information to fit previously formed views. In particular, it is the tendency to accept as reliable new information that supports existing views, but to see as unreliable or erroneous and filter out new information that is contrary to current views. People may ignore or dismiss uncertainty information if it contradicts their current beliefs.
Bias – Confidence	People typically have too much confidence in their own judgements. This appears to affect almost all professions, as well as the lay public. The few exceptions are people who receive constant feedback on the accuracy of their predictions, such as weather forecasters. The psychological basis for this unwarranted certainty seems to be insensitivity to the weaknesses in assumptions on which judgements are based.
Bias – Framing	How probabilistic information is framed can influence how that information is understood as well as the confidence that people have in the information. 'Priming' the brain with a particular stimulus can affect how it responds to a later stimulus. Using expressions that take advantage of this priming (i.e. the direction and expression are consistent) can reduce cognitive strain, which makes it easier for stakeholders to understand the idea presented without requiring further analysis. For example, the phrase 'there is a 5% chance the drawdown in the groundwater level will be greater than 0.2m' may leave a different impression than the phrase 'there is a 95% chance the water drawdown level will be less than 0.2m'; even though the two phrases contain the same information (New Zealand Government, 2014). The latter requires less mental workload because your brain is already 'primed' to think about being 'down' when it hears 'less than'. This is particularly effective when paired with explicit advice about whether precautionary action is advised.
Calibration – Conditional	Conditional calibration is a process by which parameters are adjusted until model simulations fit historical measurements or observations, indicating that the model has not yet been falsified by tests against observational data, and that the model is accepted as a good representation of (or receptacle of knowledge about) the physical system of interest.
Calibration null space	Model parameters/combinations not informed by historical measurements. To the extent that a simulation is sensitive to individual parameters, and/or to combinations of parameters, that lie within the 'calibration null space', the uncertainty of that simulation is not reduced via the calibration process at all. (Doherty and Moore, 2017)

Communication / engagement	The key to successful communication is to present the information about uncertainty in a way that is most likely to aid decision making. To achieve this, analysis of uncertainty information in model output needs to be: (i) adequately tailored to decision-makers' needs; (ii) focussed on the messages that are most likely to be relevant to their decisions; and (iii) presented in plain and clear (precise, non-technical) language. (Richardson et al., 2017)
Complexity (Middlemis et al., 2001)	The degree to which a model application resembles, or is designed to resemble, the physical hydrogeological system (adapted from the model fidelity definition given in Ritchey and Rumbaugh, 1996, cited in Middlemis et al 2001). A hierarchical classification of three main complexities in order of increasing complexity: basic, impact assessment and aquifer simulator. Higher complexity models have a capability to provide for more complex simulations of hydrogeological processes and/or address resource management issues more comprehensively. In the 2001 modelling guideline (Middlemis et al 2001, the term complexity is used in preference to fidelity (to assuage community concerns at the time).
Effective and attractive option (risk treatment)	Able to reduce risk and be implemented in a timely manner, and economically and socially acceptable risk treatment.
Equifinality	Explicit recognition that there may be multiple model representations that provide acceptable simulations for any environmental systems. The concept of equifinality is distinguished from non-identifiability or non-uniqueness: non-identifiability can be described as a poorly defined optimum of the calibration objective function, while non-uniqueness can be described as multiple local optima (Beven, 2002).
Fit-for-purpose See *italic* text opposite for proposed new definition for uncertainty context. (see also Middlemis et al 2001 guideline 'definition' below).	This term is not actually defined in the 2001 guidelines (nor in the AGMG) , but 'purpose' is used in the guiding principle for defining the model study objectives, complexity and resources (it is always linked to complexity and objectives). The term 'fit-for-purpose' is only ever used in the context of evaluating the model performance (i.e. review of whether performance is adequate in relation to the stated purpose, complexity etc). Suggested new definition in the context of uncertainty analysis: *The purpose of a modelling study is to provide information about uncertainties in the conceptualisations and model simulation outputs in a way that allows decision makers to understand the effects of uncertainty on project objectives (echoing the ISO 31000 risk definition) and the effects of potential bias.*
Hypothesis	In the environmental risk management context where groundwater modelling is applied, the hypothesis to be tested typically comprises the conjecture of an unwanted outcome or consequence associated with a particular development and/or management strategy. In practice, the hypothesis should be clearly stated in terms of threshold impacts (preferably regulatory-based) and/or resource condition indicators, and should be closely linked with the specified modelling objectives. The hypothesis of an unwanted outcome can never be completely rejected (a 'known unknowns' issue).
High-risk systems	Systems where environmental or economic risks are high, that lack attractive and effective risk treatments (see definitions) and where long time lags apply require detailed uncertainty analysis.
Low-risk systems	Systems with low environmental or economic risks and/or with effective and attractive risk treatments (see definition of 'Effective and attractive option') may not need detailed uncertainty analysis.
Model – Coupled – Externally	Solve for surface flow and sub-surface flow separately but without iteration within a time step (i.e. solve surface flow first then groundwater, then advance to next time step; usually applies short time steps for dynamic surface-water system, and longer time steps for groundwater system). Example: Mike-SHE.

Model – Coupled – Fully	Solve for surface flow and sub-surface flow and dynamic exchanges simultaneously within time step (i.e. iteration proceeds at same time step for all processes, usually constrained by surface water as that is most dynamic; this can cause significant computational overhead). Examples: HydroGeoSphere, MODHMS.
Model – Coupled – Iteratively	Solve for surface flow and sub-surface flow separately, but iteratively within time step (i.e. solve surface flow first then groundwater, iterate within time step to convergence before advancing to next time step; short time steps for surface water and longer time steps for groundwater). Example: MODFLOW.
Model – Deterministic	Same output for same input – most regional models developed for impact assessment are deterministic.
Model – Distributed	Spatial variability of parameter distributions across domain, and local-scale processes also represented, such as recharge/discharge zones, rivers, wells.
Model – Empirical	Algorithms or mathematical relationships that are based on observations or evidence (empiricism) but do not necessarily have a physical basis (e.g. regressions that do not necessarily establish a causal relationship).
Model – Integrated	Integrated solution of surface and groundwater flow and dynamic exchanges via coupling techniques.
Model – Lumped	Hydrological processes lumped to catchment-scale (no spatial variability within catchment/domain).
Model – Physically-based	Algorithms designed to realistically represent physical processes (e.g. depth-dependent ET).
Model – Stochastic	Different output for same input (element of random). Can invoke stochastic via PEST on deterministic model.
Modeller (actor in engagement process)	A person or an organisation that develops the model and conducts the modelling study.
Model failure	Model has failed if the predictive uncertainty margins underestimate probability of a bad thing happening. Or, if there is sufficient bias for a poor decision to be made on the basis of the bias, especially if the consequence of this is large.
Model purpose (after Middlemis et al,. 2001, Table 2.1)	A guideline for defining modelling study objectives, complexity and resources: (i) The modelling study objective and purpose must be clearly stated in specific and measurable terms, along with the resource management objectives that the model will be required to address. (ii) The overall management constraints should be outlined in terms of budget, schedule, staged development and long-term maintenance, and eventual ownership and use of the model. (iii) The model complexity must be assessed and defined to suit the study purpose, objectives and resources available for each model study (iv) The model complexity assessment must involve negotiation between a client/end-user and the modelling team, including the model reviewer, and relevant government agency representatives.
Non-identifiable (Barnett et al., 2012; s.5.4.1)	Model parameters can be non-identifiable or non-unique if the mathematical equations that describe a situation of interest depend on parameters in combination (e.g. R/T, R/Sy or T/S), rather than individually, in such a way that the product or ratio of parameters may be identifiable, but not the individual parameters themselves (Barnett et al., 2012).

Non-uniqueness (Middlemis et al., 2001; s.3.2.2)	The principle that many different possible sets of model inputs can produce nearly identical computed aquifer head distributions for any given model.
Precautionary principle (ESD context)	Where there are threats of serious or irreversible environmental damage, lack of full scientific certainty should not be used as a reason for postponing measures to prevent environmental degradation; www.environment.gov.au/about-us/esd/publications/national-esd-strategy-part1#GoalsEtc. PP has four central components: (i) Taking preventive action in the face of uncertainty; (ii) Shifting the burden of proof to the proponents of an activity; (iii) Exploring a wide range of alternatives to possibly harmful actions; and (iv) Increasing public participation in decision making (Kriebel et al., 2001). It has been suggested that ecologically sustainable development (ESD) is not a factor to be balanced against other considerations; rather, ESD is the balance between development and environment/social imperatives.
Probability density function Probability distribution function (PDF)	The probability distribution of a random variable specifies the chance that the variable takes a value in any subset of the real numbers. For example; 'there is a probability of p that the variable is between x and y'.
Proponent (actor in engagement process)	The person or organisation that owns the project or development (e.g. a mine). The project proponent is asking the water manager to make a decision related to impacts to a water resource. The project proponent commissions studies by outside professionals such as consulting hydrogeologists.
Reviewer (actor in engagement process)	A person conducting an external review of a modelling study. The review may be more or less comprehensive depending on the requirements of the particular case. The reviewer is typically appointed by the water manager to support her/him to match the modelling capability of the modeller.
Risk (calculation)	Combination of consequence and likelihood (AS/NZS ISO 31000:2009).
Risk (definition)	The effect of uncertainty on management objectives (AS/NZS ISO 31000:2009). Effect can be positive or negative deviation from the expected.
Risk (descriptive)	Can be roughly equated to the probability of a bad thing happening as a consequence of a particular decision multiplied by the cost associated with its occurrence (Doherty and Moore, 2017).
Sensitivity analysis (Middlemis et al, 2001)	The measurement of the uncertainty in a calibrated model as a function of uncertainty in estimates of aquifer parameters and boundary conditions.
Sensitivity analysis (Peeters, 2017)	The study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input.
Simplicity (effective) (Middlemis et al., 2001)	The simplicity (or parsimony) principle implies that a conceptual model has been simplified, yet it retains enough complexity that it adequately represents the physical system and its behaviour for the specified purpose of the model. (The term 'effective model simplicity' was discussed by Voss, 2011. Model simplification involves testing and removing all redundant elements of the model that the prediction is insensitive to.)
Simplicity (optimal)	Optimal model simplicity is achieved by testing and removing all elements of the model that the prediction is *not* sensitive to (optimal means no redundant elements). However, where the prediction is sensitive to parameters that are *not* informed by the calibration dataset, uncertainty reduction via calibration can be minimal, even if a model is perfectly calibrated. Furthermore, the parameters that *cannot* be estimated uniquely are just as important as those that *can* be estimated uniquely when exploring predictive uncertainty. (Doherty and Moore, 2017)

Stakeholders/public (actors in engagement process)	An interested party with a stake in the water management issue, either in exploiting or protecting the resource. Stakeholders include the following categories: (i) competent water resource authority (typically the water manager, cf. above); (ii) interest groups; and (iii) general public.
Type 1 error (statistical)	'False Positive' – failure to correctly reject, or incorrect or false acceptance of the hypothesis (e.g. accepting a hypothesis of a bad thing happening when it is indeed unlikely).
Type 2 error (statistical)	'False Negative' – falsely rejecting the hypothesis (e.g. wrongly rejecting a hypothesis of a bad thing happening when it can indeed eventuate).
Type 3 error (generic)	Not definitive, but put simply: the right answer to the wrong question.
Uncertainty	Alternates: probability of threshold impacts. Confidence of predictions.
Uncertainty (definition)	Uncertainty is the state, even partial, of deficiency of information related to the understanding or knowledge of an event, its consequence, or likelihood (AS/NZS ISO 31000:2009).
Uncertainty (source/type)	Any deficiency in information relating to understanding or knowledge in four main classes/sources of uncertainty: 1. Structural/Conceptual uncertainty 2. Parameter/input uncertainty 3. Measurement error 4. Scenario uncertainties
Uncertainty – Measurement error	Combination of uncertainties associated with the measurement of complex aquifer system states (heads, discharges), parameters and variability (3D spatial and temporal) with those induced by upscaling or downscaling (site-specific data, climate data).
Uncertainty – Parameterisation	Hydrogeological property values and assumptions applied to represent complex reality in space and time (any system aspect that can be changed in an automated way in a model via parameterisation).
Uncertainty – Predictive	The quantification of uncertainty in predictions. The bias and spread associated with model predictions that are made via a model that is consistent with the conceptual understanding of the system and associated measurements.
Uncertainty – Scenario	Guessing future stresses, dynamics and boundary condition changes (e.g. mining, climate variability; land and water use change).
Uncertainty – Structural/Conceptual	Geological structure and hydrogeological conceptualisation assumptions applied to derive a simplified view of a complex hydrogeological reality (any system aspect that cannot be changed in an automated way in a model). To test alternative conceptualisations (structural uncertainty), one needs to parameterise the conceptualisation issue (e.g. faults) or apply Bayes theorem to combine/evaluate the known and unknown conceptual models. Faults may be included as specific model features only where explicit evidence exists. Where some minor/inconclusive evidence exists, faults could be considered as part of a sensitivity/uncertainty analysis involving parameterisation of the fault features, and consideration of probabilities.
Uncertainty analysis (Predictive uncertainty)	The quantification of uncertainty in predictions.

Uncertainty analysis (Qualitative)	A formal and structured discussion of all model choices and assumptions and their effect on simulations. The discussion is organised by answering following four questions with 'low', 'medium' or 'high' (Peeters, 2017): • What is the likelihood that I would have made the same choice if I had more or different data? • What is the likelihood that I would have made the same choice if I had more time and budget? • What is the likelihood that I would have made the same choice if I had a better model / software? • What is the likelihood that the model simulations are very different if I change the assumption?
Uncertainty analysis (Quantitative)	Quantitative uncertainty analysis seeks to find all model predictions that are consistent with (or constrained by) the observations.
Water manager (actor in engagement process)	The person or organisation responsible for the management or protection of the water resources, and thus of the modelling study and the outcome (the problem owner).

APPENDIX C

NCGRT national workshop on groundwater modelling uncertainty

AGC2017 Uncertainty Workshop (Monday 10 July 2017)		
Role	**Name**	**Organisation**
Convenor	Hugh Middlemis	Hydrogeologic Pty Ltd
Convenor	Glen Walker	Grounded in Water
Convenor	Luk Peeters	CSIRO Land and Water
Convenor	Stuart Richardson	CDM Smith
Convenor	Phil Hayes	Jacobs
Convenor	Catherine Moore	GNS Science
Convenor	Prof Craig Simmons	NCGRT
Federal Agency	Peter Baker	Office of Water Science
Federal Agency	John Higgins	Office of Water Science
Federal Agency	Peter Hyde	Murray-Darling Basin Authority
Federal Agency	Olga Barron	CSIRO Land and Water
Federal Agency	Hashim Carey	Geoscience Australia
State Agency	Joel Hall	WA Department of Water
State Agency	Dale Cobban	NT Dept Land Resource Mgt, Water Resource Division
State Agency	Juliette Woods	SA Dept Environment, Water & Natural Resources
State Agency	Andrew Druzynski	NSW Department of Primary Industries - Water
State Agency	Keith Phillipson	Office of Groundwater Impact Assessment (QLD)
State Agency	Sanjeev Pandey	Office of Groundwater Impact Assessment (QLD)
State Agency	Ashley Bleakley	QLD Department of Natural Resources and Mines
Consultant	Noel Merrick	HydroSimulations
Consultant	Anthony Knapton	CloudGMS
Consultant	Kevin Hayley	Groundwater Solutions
Consultant	Brian Barnett	Jacobs
Consultant	Aine Patterson	Pells Sullivan Meynink, Engineering Consultants
Academic	Tony Jakeman	Australian National University
Academic	Wendy Timms	University of New South Wales
Academic	Tim Peterson	University of Melbourne
Academic	Doug Anderson	University of New South Wales
Academic	Martin Andersen	University of New South Wales
Industry	Keith Brown	Rio Tinto
Industry	Shawan Dogramaci	Rio Tinto
Industry	Ryan Morris	Origin Energy
Internationals	Steve Berg	Aquanty
Internationals	Prof Fabien Cornaton	DHI Water & Environment

1: Glen Walker – Predictive uncertainty in groundwater modelling: How is it best used? (2017)

- **Q1.1 – Can a complex model answer all questions?**
    - A1.1.1: Complex models can be built but don't expect that they can answer all questions, especially answers that require uncertainty analysis (UA). Engagement is common theme in papers, confirming need to set objectives then iterate through process to set specific questions to be answered, then design the model complexity and method required to provide the answers. Sometimes more than one model is needed; one model does not suit all questions.
- **Q1.2 – Should we run (many) multiple conceptualisations?**
    - A1.2.1: Competitive market and budget constraints mean that (many) multiple conceptualisations are not possible (unless budgets increase).
    - A1.2.2: Some examples of two alternate conceptual models in practice (e.g. IAH SA seminar 2016), but not many.
    - A1.2.3: Menindee project used preliminary model study to help set boundaries and explore key process, and final detailed uncertainty-driven modelling also used simple sub-models to explore certain processes, and that is a form of testing multiple alternate conceptual models.
- **Q1.3 – Suggestion that main source of error is estimating recharge, which means we should use physics-based process models integrated with groundwater model (e.g. Mike-SHE).** Taken as comment.
    - A1.3.1: Questions about the important hydrological processes need to consider all contexts, not just recharge (e.g. subsidence in coal mining jobs, fracture effects, aquitards etc.).

- **Q1.4.– Where should we concentrate our efforts, given that prioritisation of uncertainty is not done in environmental science? Sources of uncertainty should include 'asking the "right" question'**, **which is as important as recharge.** Taken as statement.
- **Q1.5 – Need more engagement with economists etc. regarding the value of groundwater and the effect of uncertainty in terms of risks and cost/investment to guide decision making.** General agreement and noted Glen Walker's paper regarding embedding UA within risk framework.
- **Q1.6 – Do we have cascading uncertainty (structural on top of parameterisation on top of scenario uncertainty etc.) and how do we grapple with that?**
    - A1.6.1: Question taken on notice, and later addressed during presentations by Luk Peeters and Cath Moore, who demonstrated that UA can help reject certain hypotheses and thus reduce the problem of cascading uncertainty.
    - A1.6.2: Subsequently also addressed during presentation by Phil Hayes, in that scenario uncertainty can sometimes be reduced almost as a by-product of careful discussion with stakeholders about the specific questions to be answered by the model (e.g. Menindee Lakes study).

Notes from breakout session on where to from here (Walker)

- **Note 1.1 – Good discussion involving industry, government and consultants**
    - Demand-driven need for uncertainty analysis
    - Discussion focussed about EIS rather than intra-government processes
    - Change in terms of references for EIS should be focussed and based on tech issues
    - Matching ToRs with risk, need indication of $$ (resources required)
    - This involves government and industry, with consultation with others
    - Robustness of management plans
- **Note 1.2 – Training, training, training**
    - Practitioners: high priority; logical framework with respect to different approaches
    - Also for agencies, community, model procurers

- **Note 1.3 – Conceptual uncertainty needs to be addressed, where risk requires it**
  - Currently too focussed around parameter uncertainty
- **Note 1.4 – Guidelines were seen to be important and useful**
  - Often are high level, so perhaps consider more case studies and more discussion of methods and computing and other technical requirements
- **Note 1.5 – Language – use of IPCC-style calibrated language recommended and consider different terminology that does not put too much focus on uncertainty, which can be mistaken for lack of knowledge or understanding. Echoed later with suggestions for focus on confidence, not uncertainty.**

2: Luk Peeters – Uncertainty analysis in groundwater modelling: Where to start? (2017)

- **Luk also addressed Q1.3**
  - About recharge – the model must be efficient for uncertainty analysis, which would usually preclude complicated recharge models.
  - About which is largest source of uncertainty: it is different for different contexts and the UA (if done properly) will help identify which are the largest sources of uncertainty for any particular case.
- **Q2.1 – Could the answer be that one is better off not doing a model? Can it be difficult to communicate to stakeholders that a model may not be required?**
  - A2.1.1: Straw poll – about half the workshop attendees put up their to agree that they had ever said that a model is not required.
  - A2.1.2: Sometimes not doing a model can be the answer, and that has been argued during the Bioregional Assessments. Professionals involved in the Bioregional Assessments indicated that, yes, it can be difficult to suggest that a model is not strictly required, even within an informed community. In general, the public usually expects a model and experience suggests that, if folks think/feel there is an impact, then a model is likely to be needed to predict/quantify impacts (whether benign or not).
  - A2.1.3: If the question is who we are doing the models for, then the answer is often that models are used to provide objective evidence to the community (rather than 'trust me'). Suggested that Naomi Oreskes' seminal paper provides an answer.
- **Q2.2 – What is the model being built for? While the fit-for-purpose catch-all is often used to suggest the model is suited to multiple purposes, models are usually built for one main purpose (e.g. mine dewatering) and may not be that applicable for other purposes, such as assessing environmental impacts (e.g. fine grid at mine pit rather than detail at a GDE).**
  - A2.2.1: Should not be misled by fine grid at the mine suggesting a dewatering focus; fine grid is needed at pit because that is where water table curvature is greatest (not simply because that is where the mine is). Models can be designed with an equal focus on mine dewatering and on environmental impacts. Furthermore, NSW Aquifer Interference Policy requires a certain design for the model to provide information on environmental impacts, so any mining impact assessment models are definitely designed mainly for environmental assessment purposes and yet they are also appropriate for mining.
  - A2.2.2: Suggestion that model should be developed to provide answers to the questions posed, including where to get more information to reduce uncertainty and help inform community and improve confidence. One can't build a model if one does not know what question needs answering.
  - A2.2.3: Starting to have success in SA in having discussions up front with stakeholders and then building the right sort of model and designing the modelling methodology to answer the well-considered/defined specific questions.

  - A2.2.4: Other delegates suggested that they would like regulators to be less prescriptive (e.g. less insistent on a single complex model), so this initiative about uncertainty analysis is worthwhile.

- **Q2.3 – Should we be using a Pareto front rather than one single solution?**
  - A2.3.1: If a Pareto front is a required outcome, the you need to build a model specifically to answer a specific (single) question, and then build a Pareto front from the results from a number of such models.
- **Q2.4 – Modflow is often specified by regulators; how do we encourage change and allow for diversity and innovation in approaches and software? Other questions about whether PEST is the 'best' software for UA?**
  - A2.4.1: Essentially, there is no one best software solution for modelling or for uncertainty. Question taken on notice on how to allow for diversity in tools and methods.
- **Q2.5 – How do we upskill to make the UA process efficient? Often, we fall back to what we know, and the problem is the learning curve on scripting and numerical methods.**
  - A2.5.1: Departments often work with grads a lot, so the learning curve issue is a challenge and training initiatives are required. General support from delegates for initiatives for training and for hosting tools.
  - A2.5.2: Scripting is becoming more available, but there is a learning curve even if the scripts are available, and then specialist computer help is needed. USGS quite a good source of scripts etc., as is John Doherty (https://water.usgs.gov/ogw/modflow/utilities.html; https://water.usgs.gov/ogw/flopy/; http://pesthomepage.org/Utility_Support_Software.php).
- **Q2.6 – What is the difference between sensitivity analysis and uncertainty analysis?**
  - A2.6.1: Sensitivity analysis (SA) is done in addition to uncertainty analysis (UA); in definitions below, SA is done after UA for insight on what drives prediction uncertainty.
    - Uncertainty analysis is the quantification of uncertainty in predictions.
    - Sensitivity analysis is the study of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input.
  - A2.6.2: Can do SA as a screening exercise.
- **Q2.7 – Parameter uncertainty is a common UA method, but how to test for conceptual uncertainty? Perhaps hypothesis testing may be where we need to go forward?**
  - A2.7.1: To efficiently test alternative conceptualisations, one needs to parameterise the conceptualisation issue (e.g. fault parameters).
  - A2.7.2: Often small modelling exercises are used to test different conceptualisations (i.e. not all equally developed conceptual realisations) – example of Menindee Lakes model (Hayes and Nicol, 2017).
  - A2.7.3: Example of experienced modeller recalling surprise at how two very different models were developed in the 1990s from the same datasets (one by modeller one, by QLD govt); now happy to admit that both are probably equally wrong.
  - A2.7.4: There is a question about how to deal with questions about alternate (hypothetical) conceptualisations (e.g. unmapped faults). This was considered in the breakout session: see below Breakout Note 4.1.
  - A2.7.5: There is a formal way to use Bayes to combine the known and unknown conceptual models – see Doherty and Moore (2017) and related references.
- **Q2.8 – Can we specify a set of minimum requirements for UA? Discussed at breakout session.**
  - A2.8.1: Maybe, but only if the minimum requirements are couched in risk terms, and in terms of issues to be considered and then addressed by the study and documented in the report, which should be reviewed.

- A2.8.2 Luk Peeters suggested these five minimum requirement points need to be considered during a UA study, and objective evidence that they have been addressed needs to be provided in the report, in a transparent way, with all assumptions discussed, so it is amenable to scrutiny:

1. What are the objectives? How do they link to model outcomes?
2. Which parameters or model aspects are included in the uncertainty analysis? What is the reason to include a parameter? Is there a prioritisation done beforehand, during conceptualisation, of what are the main sources of uncertainty, how they can affect the predictions/objectives? (This is not just for the modeller to answer, this is as much a question for the client/stakeholder.)
3. How is the parameterisation done? How are the parameters varied? How is prior knowledge incorporated? What are the prior distributions for parameters? (Literature and expert knowledge are valid ways to get to prior distributions, it just needs to be spelled out.)
4. What kind of likelihood function / objective function is used? Are there regularisation terms (i.e. soft constraints based on expert knowledge of the system)? What is considered an acceptable mismatch? How is observation uncertainty accounted for? How are observations weighted?
5. Uncertainty analysis method: are the assumptions underpinning the method documented and discussed in function of the model and objectives (model linearity, sampling design, likelihood function)? (This information is not readily available, and we need clear guidance.)

Notes from breakout session on uncertainty analysis – Where to start (Peeters)

- **Note 2.1 – Terms and language used**
  - 'Uncertainty' not helpful in building confidence
  - Probability of threshold impacts, and confidence of predictions would be more helpful
  - Effective model simplicity – optimal simplicity rather than simple models
- **Note 2.2 – Training, training, training – suggested courses**
  - Modelling and Uncertainty
    - for managers
    - for regulators
    - for model procurers
    - for community?
  - Communicating the main concepts in modelling and uncertainty analysis
  - Modelling workflows and options for uncertainty analysis
  - Data visualisation
  - Objective and evidence-based reviewing of modelling studies
  - Data processing and mathematical tools for uncertainty analysis
    - scripting, weighting functions, cloud computing, Algomesh
  - Designing specific questions for models to answer to meet objectives
  - Streamlining models for efficient uncertainty analysis
- **Note 2.3 – In-house coaching of junior staff; start slowly/simply and build slowly, with formal training**
- **Note 2.4 – Why and when do we do uncertainty analysis?**
  - Make uncertainty analysis an integral part of a risk framework
  - Low risk – do at least qualitative UA (e.g. Table 2.1 of the AGMG)
  - High risk – comprehensive and detailed UA required
  - Can do early with simple models, to explore alternative options/processes

    - Can do later with more mature models, as a comprehensive uncertainty analysis
    - Can use independent reviewers to advise during a project (e.g. as in Menindee Lakes model)
- **Note 2.5 – Question regarding the ethics of modellers who claim that uncertainty has been addressed, citing the 2012 AGMG but presenting no objective evidence**
    - Model review criteria are often used as a bureaucratic checklist in an attempt to avert future accusations of policy error or wrongdoing by deflecting responsibility in advance; however, this uncertainty initiative requires that objective evidence be presented showing how uncertainty issues have been addressed and the degree of stakeholder engagement.
    - Corrective Action suggested – 2012 AGMG model review criteria ('Checklist') should be updated to require objective evidence of conformance/compliance.
- Note 2.6 – Engagement is not always easy, especially when third party proponents are not comfortable discussing with agencies what may be sensitive issues for them, and they may not wish to be seen to be possibly over-committing. Similarly, there are sensitivities for agencies, who like to see final study report before committing.
    - Transparency, honesty and 'without prejudice' would seem to be the only way to conduct engagement (as suggested in Richardson et al., 2017).

3: Catherine Moore – Simple is beautiful (Doherty and Moore, 2017)

- **Key Points regarding presentation on discussion paper no.3**
    - If calibration is sensitive to parameters that are not informed by dataset, then suggested to not bother with calibration and just go to uncertainty assessment. Parameters that cannot be estimated uniquely should be included in the process of uncertainty analysis.
    - Uncertainty Methods tips: Rejection sampling is difficult with highly parameterised model. Markov Chain Monte Carlo (iterative update of prior distro) is more efficient than rejection sampling but still inefficient. Ensemble or Kalman Filters/Smoothers (adjusting parameters so that match to calibration dataset) show significant improvements recently by Jeremy White (the 'Ensemble Smoother').
    - Pragmatic alternatives to Bayesian include Approximate Uncertainty Analysis (Doherty). Non-Linear: foundation of Null Space Monte Carlo, which uses a linearised form of Bayes (i.e. not strictly Bayesian, but it is numerically tractable). Linear: sensitivity assessment (Jacobean) then linearised form of Bayes used. Quite fast methods and allows assessment of influence of data and parameters on prediction uncertainties.
    - If we can calculate uncertainty, we can estimate how much uncertainty would be reduced by more data in key areas of uncertainty.
    - Simplification: numerical stability and long run time issues mean that we need simplification. Requires building models that remove details that the prediction is not sensitive to. The cost is the error incurred due to simplification, but sometimes get a systematic or bias term, so simplify carefully.
    - Art of model simplification involves the use of expert knowledge. Success can be judged when simplification-induced uncertainty is small compared to the predictive uncertainty.
- **Q3.1 – What do you mean by model simplification?**
    - A3.1.1: See above key points. In simple terms, remove all elements of the model that the prediction is insensitive to (all redundant elements).
    - A3.1.2: There is tension between building a complex model that can answer many/all questions but ends up so complex that it can't efficiently answer the questions posed for uncertainty analysis; hence the need for some simplification.
    - A3.1.3: Delegate suggested rejecting the notion that complex models are difficult to calibrate where one knows the recharge very well.
        - If you assume that you know the recharge perfectly then you can remove that uncertainty from the calibration. But most hydrogeologists believe that recharge

estimations are fraught with uncertainty, so that issue should be considered very carefully indeed.

- **Q3.2 – What about cost of UA?**
    - A3.2.1: Senior agency delegate suggested that regulators will drive the need for better UA and industry may need to wear the cost, noting the value of improved decision making.
    - A3.2.2: It is expensive, so multiple conceptualisations are not possible in competitive consulting, but parameter UA is possible and is happening slowly. There is a timeline issue, in that groundwater is already the slowest element of EIS, and the mining industry for example has not allowed in their project timelines for groundwater to become slower.
- **Q3.3 – Should PEST be specified as a requirement (a bit like Modflow often is)?**
    - A3.3.1: No. There is no one software solution for all modelling or for all uncertainty analysis.
- **Q3.4 – Can we learn from surface water and annual allocation setting?**
    - A3.4.1: SA is moving towards that with changes in legislation. Other states?
- **Q3.5 – 'Simple' model might not be the best term? And use of the term 'uncertainty' is not helpful in building stakeholder confidence.**
    - A3.5.1: Note that the words 'simple' and 'complex' are hard-wired into NSW Aquifer Interference Policy.
    - A3.5.2: Suggested use of terms like 'confidence' rather than uncertainty.
    - A3.5.3: Suggested use of terms like 'streamlined' or 'optimal'.
    - A3.5.4: Would be useful to see more examples of Cath Moore's work on comparing simple versus complex (steady state versus complex).

4: Phil Hayes – Uncertainty by design (Hayes and Nicol, 2017)

- **Key points regarding presentation on discussion paper no.4:**
    - Broken Hill MAR background briefly presented then model development (not in paper, but in report). Comprehensive dataset (42 GB) based on $30 million plus of GA investigations.
    - Objectives (licensing): drawdown impact assessment, timing of impacts (low/high flows), recharge mechanisms, recovery (rates, levels, sustainability).
    - Discussion of approach and model design as per paper. Null Space Monte Carlo method.
    - Worked closely with stakeholders.
    - Successful project: focussed objectives, planned UA from outset with appropriate budget; client contributions with review (HM and JD); key decisions left to modellers, with inputs from others; modelling team capability, assisted by initial simple modelling.
    - Conclusion: predictive uncertainty analysis is tractable, even with large datasets.
- **Q4.1 – Regarding the Menindee study, how much data was redundant, and would early engagement optimise that requirement?**
    - A4.1: Did not need all the data to do the model, but it was used as an example where maximum data was used to optimise the model for UA. Early engagement may have optimised the requirement.
- **Q4.2 – What makes a good model?**
    - Balance between conceptual and numerical features and capability.
    - The right software tool and careful design with review and discussion with stakeholders.
    - Fit-for-purpose is not a dirty word when the purpose is commensurate with objectives.

Notes from breakout session on effective simplicity (Moore) and uncertainty by design (Hayes)

- **Note 4.1 – Conservatism: modelling with worst case has helped modelling progress a long way**
  - Faults, for example; should include only where explicit evidence exists.
  - Where other minor evidence exists, include faults in sensitivity / uncertainty analysis (involves parameterisation and should consider probability).
- **Note 4.2 – Petroleum Engineering**
  - Standard workflows involve uncertainty analysis and using results for hard economic decisions.
  - What can be learnt from their workflows?
  - What about communication? (Tornado plots for example).
- **Note 4.3 – OGIA Surat cumulative impact model lessons learned**
  - Started simply, but with added complexity over time.
  - Now moving to simpler, smaller scale models to investigate processes.
- **Note 4.4 – Lack of skills is currently a constraint and training initiatives required**
  - Maths / programming / stats and uncertainty.
  - Uncertainty training for regulators, decision makers and model procurers.
  - Do we want or need accreditation, or is a suite of examples enough?
- **Note 4.5 – Objectives must be translated to specific targeted questions**
  - 'No significant impact' is not specific enough.
- **Note 4.6 – Grid / scale independent modelling is a developing area**
  - Rapid moving from large scale to small, complex to streamlined, and feeding back.
  - Hydro-Algorithmics are developing a tool: www.hydroalgorithmics.com/software/algomesh/
- **Note 4.7 – Emulators (models of models) is a developing area**
  - Could speed up and allow integration to other models, economic/ecological.
  - But similar issues of training, communication etc.

5. Stuart Richardson – Communicating uncertainty (Richardson et al., 2017)

- **Key points regarding presentation on discussion paper no.5**
  - Embed uncertainty within risk framework and link modelling objectives.
  - Results need to be tailored to needs, focussed on key messages, plain and clear.
  - Need to move into the mindset of a regulator. Communication/engagement.
  - *Do we overplay the concept of uncertainty and underplay the value of UA?*
  - Actors: proponent, water manager, modeller, reviewer, stakeholders/community
    - Would be useful to include in glossary/definitions – done!
    - Need careful identification of stakeholders.
  - Undertake meaningful co-design and co-production on UA. Documentation of results is the end product that needs to be designed at the start via engagement process.
  - Reduce cognitive strain by matching how uncertainty is expressed to align with expressions in the decision making (+ve or -ve).
  - Guidance from IPCC on calibrated language (discipline required on report writing).

  - Provide clear description of scale of uncertainty via multiple outputs and use numeric, visual and commentary elements, with consistent and precise language (IPCC calibrated language).
  - Indicate type and source of uncertainty present and relative contributions.
  - Describe how biases could skew the results.
  - Indicate whether/how uncertainty has been or can be reduced (with extra info).
- **Q5.1 – Do we need to allow for extra runs to test alternative conceptual models?**
  - A5.1: A sound communication process with engagement should ensure that the model and the overall approach are designed appropriately, allowing for testing alternate conceptual models if that is needed.

Notes from breakout session on communicating uncertainty (Richardson)

- **Note 5.1** – Cross-section of stakeholders in group.
- **Note 5.2** – Definitions: uncertainty, certainty, confidence. Negative. We need positive language.
- **Note 5.2** – 'Engagement' may be a better title for the paper, to highlight the need for communication with stakeholders throughout the uncertainty analysis (engagement theme runs through all the papers).
- **Note 5.3** – Different requirements from different decision makers (e.g. stakeholders involved in policy-driven settings require qualitative assessments while stakeholders making investment decisions require quantitative analysis like probability of occurrence).
- **Note 5.4** – Communication is made easier where it can be shown that certainty is improved – 'building confidence' through the modelling workflow.
- **Note 5.5** – Focus on telling people what we do know rather than what we don't know – likelihood/probability of occurrence is a useful communication tool.
- **Note 5.6** – Support to develop 'calibrated language' for groundwater modelling – types of uncertainty, similar to IPCC.