



Queensland
Government



Australian
National
University

Queensland Water Modelling Network

QWMN Good Modelling Practice Principles

April 2018

MODEL
MANAGEMENT



QWMN

Good Modelling Practice Principles

April 2018

Prepared by A J Jakeman, S El Sawah, S Cuddy, B Robson, N McIntyre and F Cook

© The State of Queensland (Department of Environment and Science) 2018

The Queensland Government supports and encourages the dissemination and exchange of its information. The copyright in this publication is licensed under a Creative Commons Attribution 3.0 Australia (CC BY) licence



Under this licence you are free, without having to seek permission from the Department of Environment and Science, to use this publication in accordance with the licence terms.

You must keep intact the copyright notice and attribute the State of Queensland, Department of Environment and Science as the source of the publication.

For more information on this licence visit <http://creativecommons.org/licenses/by/3.0/au/deed.en>

Disclaimer

This document has been prepared with all due diligence and care, based on the best available information at the time of publication. The department holds no responsibility for any errors or omissions within this document. Any decisions made by other parties based on this document are solely the responsibility of those parties. Information contained in this document is from a number of sources and, as such, does not necessarily represent government or departmental policy.

If you need to access this document in a language other than English, please call the Translating and Interpreting Service (TIS National) on 131 450 and ask them to telephone Library Services on +61 7 3170 5725

Executive summary

Best practice modelling reduces model uncertainties and quantitatively and qualitatively documents any uncertainties and assumptions for user transparency. Conversely, poor modelling practices contribute to uncertainties and increase user distrust in the value of modelling.

This paper synthesises existing knowledge and experience on good water modelling practices and principles. Specifically, it provides guidance for new and existing water model development efforts, and informs end users and decision makers about what distinguishes good modelling practices from poor ones. The paper covers the following:

1. An introduction to water models and their role in decision making;
2. An overview of water modelling practices, and the role of best practice modelling in improving model quality and results;
3. A characterisation of best practice modelling in relation to each of the phases and steps in the modelling process, including checklists of things that modelling practice should explicitly address.

This paper's scope covers the use of water resource models to investigate impacts on the environmental system in question, such as a paddock, catchment, or estuary. It includes model use under both status quo conditions and in response to management actions, climate variations or other uncontrollable forces. It also includes model uses to adaptively manage a system, such as through additional monitoring and informative studies.

The following best practice water modelling recommendations warrant specific attention by water model developers and users:

- Specify the objectives, clients and stakeholders for the modelling exercise clearly.
- Document the nature of the data used to build and test the model.
- Justify the selection of model type and calibration method.
- Undertake extensive model testing and report on results, model limitations and assumptions.
- List and characterise information and data sources and try to rank the criticality of uncertainties arising in the entire modelling process by means including expert elicitation, stakeholder engagement, sensitivity and other more quantitative uncertainty analyses.
- Carefully consider appropriate model complexity, taking into account uncertainty, data support and system behaviours. This is likely to include effective simplification with good documentation of the assumptions made and their implications.
- Inform the users of model results about the dangers of being provided only a single number upon which to base decisions. Also address their needs by providing uncertainty information in a format that fits within their workflows.
- Factor in the appropriate costs of a holistic uncertainty assessment in project budgeting. It will be worth it in the long-term.
- Place due emphasis on communicating uncertainty. Visualisation of indicators of concern is one aspect that can be used. The design of such visualisations should pay special attention to possible interpretation biases and techniques to manage them.
- Pay explicit attention to the way model results and uncertainty are communicated in written reports and publications.

- Ensure model visualisations employ user-centred designs early in the modelling process, and leverage different visualisation tools to engage different audiences (e.g. researchers, policy makers, stakeholders).
- Embrace the use of automated methodologies that can both support transparent experimental workflows and allow for systematic understanding of the impacts of the various relationships and factors that influence the model's results.
- Pay careful attention to the collected data, including ensuring that they measure the appropriate variables, at the correct locations and with the required frequency.
- All modelling practices should address the checklists of questions specified in this paper for each phase of the modelling process.
- Help instil good modelling practice within the community by adopting these recommendations and sharing and collaborating on best practice cases in the major water modelling domains.

Table of Contents

Executive Summary	3
1 Introduction	6
2 Background.....	6
2.1 Context	6
2.2 Scope and focus	7
2.3 Paper organisation.....	7
3 Models and modelling	8
4 Modelling practices	10
4.1 Reporting modelling practices	12
5 Identifying and defining good modelling practices	13
5.1 Working with stakeholders	13
5.2 Scoping Phase.....	14
5.3 Problem Framing and Formulation Phase	16
5.3.1 Model conceptualisation	16
5.3.2 Data collection, cleansing and pre-modelling data analysis.....	17
5.4 Analysis and Assessment of Options Phase	19
5.4.1 Model selection.....	19
5.4.2 Sensitivity analysis.....	21
5.4.3 Calibration and model structure	23
5.4.4 Validation and testing.....	27
5.4.5 Scenario analysis.....	29
5.4.6 Data and model management.....	31
5.5 Communication of Findings Phase.....	34
5.5.1 Selecting indicators to communicate model results	34
5.5.2 Communicating uncertainty in written reports.....	34
5.5.3 Visualisation	35
6 Conclusions	37
7 References	38

1 Introduction

Best practice modelling reduces model uncertainties and quantitatively and qualitatively documents any uncertainties and assumptions for user transparency. Conversely, poor modelling practices contribute to uncertainties and increase user distrust in the value of modelling.

This paper synthesises existing knowledge and experience on good water modelling practices and principles. Specifically, it provides guidance for new and existing water model development efforts, and informs end users and decision makers about what distinguishes good modelling practices from poor ones. The paper covers the following:

1. An introduction to water models and their role in decision making;
2. An overview of water modelling practices, and the role of best practice modelling in improving model quality and results;
3. A characterisation of best practice modelling in relation to each of the phases and steps in the modelling process, including checklists of things that modelling practice should explicitly address.

This paper's scope covers the use of water resource models to investigate impacts on the environmental system in question, such as a paddock, catchment, or estuary. It includes model use under both status quo conditions and in response to management actions, climate variations or other uncontrollable forces. It also includes model uses to adaptively manage a system, such as through additional monitoring and informative studies.

2 Background

2.1 Context

We, the Queensland Government Department of Science, Information Technology and Innovation (DSITI), are working to address the critical strategic gaps and weaknesses in water models that have arisen due to the resource constraint driven focus on operational issues. Our objectives are to develop greater stakeholder capacity and collaboration by engaging with universities, scientific providers and external consultants to provide modelling resources for the future. Over four years we aim to improve the integration of all Queensland hydrology, groundwater and water quality models—not just those focused on the Great Barrier Reef—across multiple scales including paddock, catchment, estuary and marine. Our intention is to drive consistency in models and modelling practices across Queensland and, in the longer term, develop a 'community of practice' in model development and model application to better inform decision making.

In this end, the purpose of this paper is to synthesise existing knowledge and experience on good modelling practices and principles. We have developed it based on the findings gathered from a literature review and an expert workshop conducted on June 21-22, 2017. The expert workshop:

- Captured and synthesised the foundations and principles of best practice modelling procedures and model management;
- Reflected on current approaches to modelling and model management;

- Developed complementary presentations on research and development (R&D) modelling principles targeted at modellers and for policy makers;
- Recommended methods to retain currency of the best practice and principles of water modelling and links to a catalogue of government models (in a Stage 2)

2.2 Scope and focus

This paper's scope covers the use of water resource models to investigate impacts on the environmental system in question, such as a paddock, catchment, or estuary. It includes model use under both status quo conditions and in response to management actions, climate variations or other uncontrollable forces. It also includes model uses to adaptively manage a system, such as through additional monitoring and informative studies.

We focus on water models developed for ongoing, regular and operational use. These models are built (or being built) to answer a variety of policy and management questions. Typically, their aim is to predict one, or usually more, Quantities of Interest (QoIs) such as indices of water quantity, quality or ecological response as a function of time and/or space. Inevitably, the nature and role of these models evolves over time to cope with changes in policy and scientific knowledge. We outline modelling practices at all stages of the model development and use lifecycle, taking into account the entire modelling process and the sources of uncertainty that need to be recognised and managed in that process.

2.3 Paper organisation

In this paper we:

1. Introduce the field of modelling and outline the role of models play in decision making in water resource management, specifically within the context of evidence-based policy making (Section 3);
2. Present the concept of best practice modelling, and discuss the need to adopt good modelling practices to improve models and ensure the quality of the decisions that are made based on their results (Section 4); and
3. Identify and define good modelling practice in relation to the various phases and steps in the modelling process, including providing a checklist of the identified best modelling practices (Section 5).

3 Models and modelling

Objective: to introduce the field of modelling and the role models play in decision making, specifically within the context of evidence-based policy making in the water sector.

Models are approximations or simplified representations of a system of interest that link its state to its drivers (inputs) and responses (outputs). Models play a powerful role in assessing water management policies and improving the quality of decision making. Models can be used to:

- Support a methodology for capturing, mapping, and consolidating the various sources of knowledge and data required for understanding water resource systems in a systematic way.
- Provide a platform for bringing together stakeholders and scientists to engage in a science-informed dialogue that helps develop a shared understanding of the problem and possible solutions. Assembling opposing parties around a model of the problem can be a powerful means of sharing concerns and testing ideas with a view to negotiating common grounds and reaching consensus. For example, models can help farmers and natural resource managers come to agreement on water allocation rules.
- Set up experiments, including ‘what-if’ questions, that examine system outcomes (QoIs) under a range of scenarios, where each scenario represents a plausible pathway for managing the system.
- Perform quantitative assessments of decision options and provide objective evidence for linking these options to long-term benefits and costs. This information is useful for devising trade-off decisions in short and longer terms.

Different types of decision models are appropriate for different types of water management and policy questions. However, they are all used for the same purpose: to assist in decision making under uncertainty. According to their use in the decision analysis process, models can be broadly classified into the following (Robinson, 2004):

- Throwing away models: used only for the duration of a modelling study. Such models are developed to investigate one or more issues of concern for decision-making. At the end of the study, when the learning outcomes expected from the study are served, the model is no longer relevant or needed.
- Regular use or operational models: used to support decision making and management on a regular basis, such as surface water models for runoff prediction and groundwater models for sustainable yields of an aquifer. These models often require a high level of accuracy, detail, and resolution to ensure the quality of information on which decisions are made. Most of the development effort occurs upfront, with continuous updates and validation required to keep the model up-to-speed with changes in the real system, technology advances and knowledge accrual.
- Ongoing use models: used as a part of an ongoing effort to investigate a particular problem and answer a variety of policy questions. The model's scope and role may change over time to support the evolving requirements for decision analyses.

A single model can serve more than one use type, and the model use type can evolve over time and across projects. For example, a model that is used regularly can also be used as a throwaway model for a modelling project focused on promoting social learning and stakeholder engagement.

A typical modelling and assessment process has four key iterative phases, each with inherent steps (Hamilton et al., 2015):

1. Scoping, including a model study plan that identifies model purpose, study objectives, stakeholders and issues of concern;
2. Problem framing and formulation, including conceptualisation, linking drivers and responses;
3. Analysis and assessment of options, including model setup, calibration and validation; and
4. Communication of findings, including simulation and evaluation.

Most of these steps require expert and/or stakeholder engagement, for which there is now much guidance regarding why and how they can be achieved (e.g. Voinov and Bousquet, 2010).

Being approximations of the real system of interest, models only represent our partial knowledge and views about that system. And, because a real water resource system has many more issues and is far too complex to completely understand and capture in a model representation, uncertainty is inevitably associated with use of the model and arises in various ways throughout the modelling process.

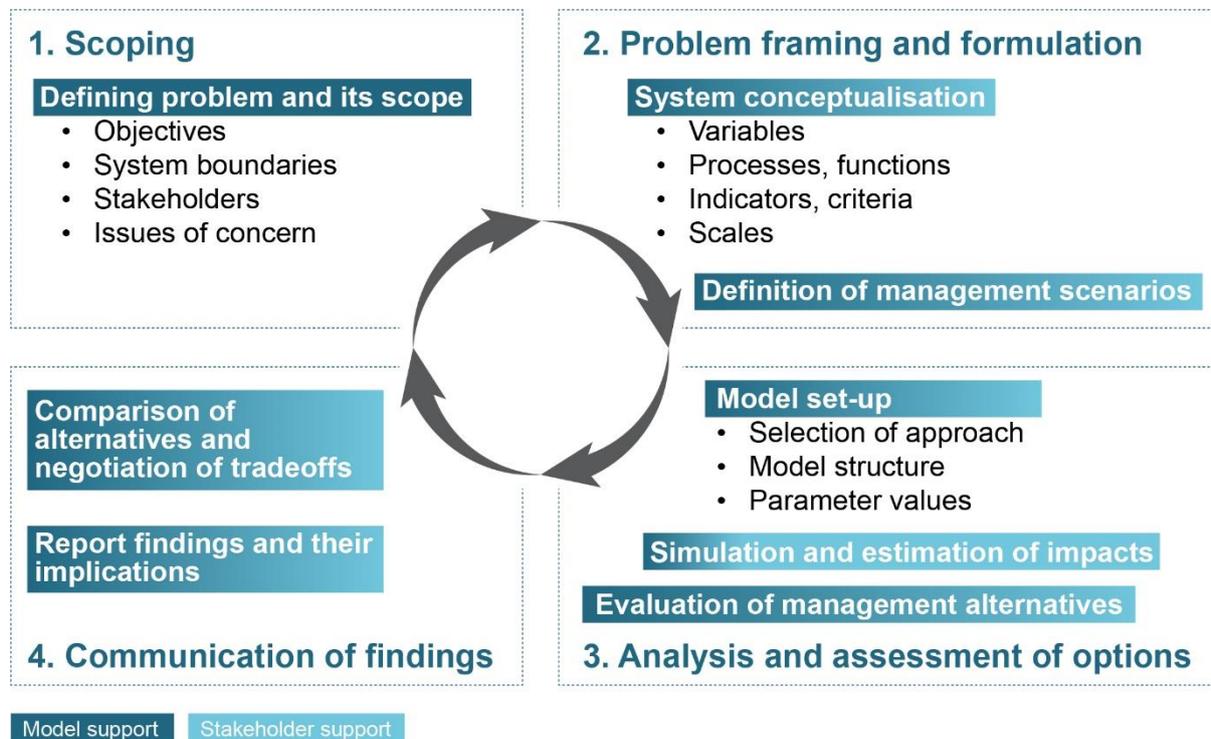


Figure 1: Phases and steps in the integrated modelling and assessment process (adapted from Hamilton et al., 2015)

Uncertainty management is now recognised as an essential part of the modelling process (e.g. Guillaume et al., 2011) and has quantitative and qualitative aspects (e.g. Uuisitalo et al., 2015; Refsgaard et al., 2007) aimed at establishing what we do and do not know about the predictions (QoIs) required for the problem of interest. Therefore, a model user must pay attention to the various steps of the modelling process in order for all interdependent uncertainty sources be addressed.

Key Points - Modelling practices

- Models are simplified representations of the problem of interest and used to answer specific management and policy questions.
- According to their uses, models can be classified into: throwaway, regular, and ongoing models.
- Uncertainty management is an essential part of the whole modelling process and involves quantitative and qualitative aspects for complex water resource issues.

4 Modelling practices

Objective: to present the concept of modelling practices, and the role good modelling practices play in improving models and assuring the quality of model-based decision support.

In this section, we define best practice modelling and present an overview of literature to offer guidance on identifying and defining modelling practices. We also propose the use of tools, such as checklists and templates, as useful means to report and monitor the uptake of good practices.

The quality and outcomes of a modelling process largely depend on the modelling practices that are undertaken at every step. Modelling literature emphasises that it is vital to build quality (i.e. relevance, credibility and validity) into the modelling process and its outputs. The literature also reinforces that best practice modelling should use 'proven-to-work' practices for managing common problems encountered throughout the modelling process. Identifying best practices helps to provide guidelines for improved modelling practice and will ultimately lead to more accurate, credible and useful models; more insightful model-based recommendations; better-informed model adoption; and, most importantly, informed decision-making.

The search for ways to improve the way modelling is conducted is not new. Several attempts have been made to investigate and identify 'best', 'good' and 'core' practices. (Table 1 provides a non-exhaustive list of research geared towards developing guidance into best, good and core modelling practices). In the Australian water industry context, two guideline documents are noteworthy. Black et al. (2011) developed a set of guidelines with support from the eWater Cooperative Research Centre (CRC). The other is the Australian

Groundwater Modelling Guidelines (2012) prepared for the National Water Commission by Sinclair Knight Merz and the National Centre for Groundwater Research and Training. Both go into more technical detail than is presented here and are valuable resources to be used in conjunction with this document. This paper synthesises these two existing guidelines, together with other key literature, for non-modellers and provides checklists of questions that policy makers might refer to in order to understand if a modelling proposal or report has adhered to good practices.

Table 1: A list (non-exhaustive) of literature offering guidance into best, good and core modelling practices

Publication	Scope, focus
Jakeman et al. (2006)	Steps in development and evaluation of environmental models
Refsgaard et al. (2007)	Uncertainty in the modelling process
Robinson (2007, 2008)	General modelling and simulation, conceptual modelling
Gaber et al. (2009)	US EPA Guidance on the development, evaluation, and application of environmental models
Rietveld et al. (2010)	Drinking water treatment, whole modelling process
McIntosh et al. (2011)	Environmental modelling, design for improved use and adoption
Black et al. (2011)	Guidelines for water management modelling: towards best practice model application
Australian Groundwater Modelling Guidelines (2012)	Model calibration and uncertainty, groundwater
Chen, S. H. and Pollino, C. A. (2012)	Environmental modelling, model set up and formulation, Bayesian network modelling
Kelly et al. (2013)	Environmental modelling, model selection
Horsburgh et al. (2014)	Hydrological modelling, data sharing
Black et al. (2014)	Water management, whole modelling process, scenario-based models
Black et al. (2015)	Guidelines on implementing a risk-based approach to water resources planning (companion to Black et al. 2011)
Argent et al. (2016)	Environmental modelling, conceptual modelling
van Vliet et al. (2016)	Land use change, model calibration and validation
Elsawah et al. (2017)	Environmental modelling, whole modelling process, system dynamics

According to Black et al. (2011), “best practice modelling can be defined as quality assured model implementation to deliver a credible, robust model that is fit for purpose, and its application to deliver results, using methodology that is transparent, defensible and repeatable.” Modellers working on environmental problems not only build and use models according to strict fundamental disciplinary principles, such as mathematics, statistics, hydrology, computer science and ecology, they are also faced with the ongoing challenge of juggling cost, time, and other resource constraints while producing quality products and managing stakeholder expectations and interactions. Therefore, best practice means the best achievable procedures and outcomes taking into account intended modelling purpose, and trade-offs in knowledge, data, resource and time constraints.

Another corroborating viewpoint of good model development and evaluation practice is in Jakeman et al. (2006) who outline ten basic steps of good, disciplined model practice towards building credible models. The authors state that best practice modelling must:

- clearly identify the clients and objectives of the modelling exercise;
- document the nature (quantity, quality, limitations) of the data used to construct and test the model;
- provide a strong rationale for the choice of model family and features (encompassing review of alternative approaches);
- justify the techniques used to calibrate the model and conduct detailed analysis, testing and discussion of model performance; and
- make a resultant statement of model assumptions, utility, accuracy, limitations, and scope for improvement.

These steps, expanded upon further in Section 5, should be carried out through a learning process, or even in partnership, between model developers, clients and other interested parties.

4.1 Reporting modelling practices

It is crucial to document the practices employed throughout the modelling process in a systematic and transparent way that helps decision-makers form their own judgment about the model’s results and improves their confidence in using the model as a decision-aid. We propose that a checklist of modelling practices can be an effective means to document and report the modelling efforts and help distinguish between good and poor modelling practices. Such a checklist is a minimum requirement for evaluating the modelling process and informs sources of model uncertainties. A ‘practices description template’ (e.g. Alwazae et al., 2015) is another useful tool for describing the detail of the practices implemented, including items including the rationale for practice use, the steps carried out in the implementation of the practice, and references to data sources used to carry out the practice.

Key points - Modelling practices

- Good modelling practice comprises a 'proven-to-work' routine that improves a model's quality and results.
- Good modelling practices can be employed at every phase and step in the modelling process (Figure 1).
- We suggest the use of practice documentation tools such as checklists and templates to document the modelling efforts and help distinguish between good and poor modelling practices.

5 Identifying and defining good modelling practices

Objective: to identify and define good modelling in relation to the various aspects and phases in the modelling process.

Given that stakeholder participation spans across the modelling process (Figure 1), we start by discussing the general practices underpinning stakeholder engagement. Next, we present and discuss modelling practices related to the different modelling activities, including the stakeholder engagement practices specific to each phase. We support the discussion using examples from the literature and the findings of the 2017 expert workshop.

5.1 Working with stakeholders

In essence, model development and use involve social communication processes used throughout the modelling process to build confidence and trust with stakeholders. Successful management of these processes is as important as the technical model development because major uncertainties can emerge from basic aspects such as working on a poorly formulated problem, neglecting to include valuable knowledge and perspectives from key interest groups and experts, or by poor communication in general. However, stakeholder engagement can be challenging, especially in a multi-agency context with multiple intended model uses and end users, each with slightly different needs (e.g. government agencies and farmers, or operational river managers and policy planners). The effectiveness of this process requires well-rounded modelling competencies including good communication and interpersonal skills. Some useful principles to put into practice include:

- Engage with stakeholders from the very early stages of the modelling process. This includes explicitly accounting and planning for the time and resources required;
- Determine and communicate the model's value within the context of the problem and develop realistic expectations about what the model and the modelling process can and cannot do;
- Agree on the underlying conceptual models of the system with stakeholders;
- Approach the process from a position of humility and goodwill, embracing relationship building, rather than a position of selling expertise;
- Work with stakeholders to design communication products and model interrogation tools (e.g. end-user interfaces, visualisation methods) that suit their needs;
- Adopt effective science communication practices, such as using: easy to understand language that avoids technical and academic jargon, filtering and synthesising large

- amounts of information to communicate the most useful insights, and understanding stakeholders' cognitive biases, and building their understanding step-by-step; and
- Document and peer review both the model itself (including its scientific basis and practical implementation) and, just as importantly, the model development process to establish credibility and legitimacy.

Checklist – Working with stakeholders

Aspect	Questions to consider
Project actors and roles	<ul style="list-style-type: none"> • Who will be engaged (i.e. stakeholder and interest groups, group size)? • What is the level at which each group will participate in the modelling process (i.e. information extraction, consultation, co-building of the model)? • Are the roles of the various stakeholders (including clients and model developers) clearly delineated and understood? • Does the project have a steering committee? If not, how is this decision justified?
Process design	<ul style="list-style-type: none"> • What participatory methods will be used to engage stakeholders? • When will stakeholders be engaged through the modelling process?
Communication and expectation management	<ul style="list-style-type: none"> • Is the information about the model's capabilities, limitations, and assumptions well-communicated to stakeholders? • Have the expectations of stakeholders about the model been obtained and appropriately managed? • Is this information communicated in a way that is easily accessible to stakeholders so they can fully understand the model's purpose?

5.2 Scoping Phase

In the scoping phase, the modelling team works with the client and primary stakeholders to establish the project's objectives and modelling purpose; and chart a plan of how to realise these.

The purpose and objectives of a model should include a clearly articulated set of user data requirements, processes to be represented, questions, functionalities, system boundaries and predictive quantities of interest (QoIs). The model's purpose and objectives need to be considered within the project's constraints, such as available time and resources, whilst managing client and stakeholder expectations and avoiding over-sell. This includes determining whether QoIs are absolute values or are relative to a baseline. It also includes functionality in terms of what input variables or model parameters may need to be varied as part of model application. The strength of evidence sought from the model, in terms of supporting decisions, should be agreed. For example, is it making broad generalisations to support state land management policy among other sources of evidence, or is it intended to be the main line of evidence in assessing the impacts of a local project?

1. Scoping

Defining problem and its scope

- Objectives
- System boundaries
- Stakeholders
- Issues of concern

These considerations need to be clearly communicated to the client, along with establishing clear agreements on the format in which model results will be delivered (e.g. reports, raw computer runs, analysis results) and ways to communicate about result uncertainties (e.g. probability distribution functions, ranges, qualitative or categorical descriptions) and their visualisations and tabulations. In some situations, the client may have a very clear understanding of the modelling objectives. In others, the modeller and the client must work closely to formulate those objectives and associated requirements. Ongoing projects may have already specified their modelling purpose but even then there may be some advantage in revisiting the specification to make it more exact, and/or to simplify the problem in a way that still answers valuable questions but does so with more certainty.

Checklist – Scoping phase

Aspect	Questions to consider
Considering the use of models	<ul style="list-style-type: none"> • Is there sufficient evidence (e.g. similar case studies, literature) that modelling can be useful for the problem? • Is there sufficient leadership buy-in into the use of models? If not, can leadership be rallied? • Are the primary stakeholders on board? If not, can they be brought on board? • What is the organisational history around the use of models?
Defining the model's purpose and role	<ul style="list-style-type: none"> • Is the model's use clearly stated (e.g. provide quantitative evidence, facilitate stakeholder engagement)? • Are the questions that the model is intended to answer clearly defined? • Can these questions be adequately answered within the available time and resources? • Is the model's end user(s) clearly identified?
Project resources	<p>Is the project sufficiently resourced to realise the defined objectives, including resources allocated for the following:</p> <ul style="list-style-type: none"> • model validation and testing; • stakeholder engagement; and • data collection, analysis, and storage.
Stakeholder engagement	<ul style="list-style-type: none"> • How will the model users interact with the model (e.g. through interactive experimentation, through predefined runs to be conducted by the modellers)? • Is there a clear agreement among the modelling team and the client about the model's outcomes and outputs (e.g. visualisations)? • How will the model users determine (or measure) what the model does and whether it is successful?
Uncertainty communication	<ul style="list-style-type: none"> • Is there a clear agreement among the modelling team and the client about how modelling uncertainty will be communicated?
Documentation	<ul style="list-style-type: none"> • Are the options and scenarios to be evaluated clearly documented? • Are the methods and practices used to define the model objectives clearly documented?

5.3 Problem Framing and Formulation Phase

5.3.1 Model conceptualisation

Conceptual models are qualitative representations of the model content, its components and relationships.

Developing conceptual models involves making assumptions and simplifications. Assumptions are made when there are uncertainties or beliefs about the real world being modelled. Simplifications incorporated in the model enable more rapid model development and use, including recognising that there might not be knowledge or data on some processes. Simplifications can also be used to reduce uncertainties that would be associated with an overly complex model.

Risk-focused validation of the conceptual model is needed to improve the model validity and credibility. This includes, from a modeller's perspective, ensuring that the model produces sufficiently accurate results and is valid for the purpose at hand, and making sure that it is credible from a client, user and legal perspective. The risk associated with each assumption can be assessed (quantitatively and qualitatively as relevant) according to the level of confidence and impact (Guillaume and Elsworth, 2014), along with transparent documentation of the methods and data used to conclude the risks (Sargent, 2013). Validating and testing the conceptual model should not only be limited to the conceptual model itself, but needs to also include the process used to produce the conceptual model, raising questions such as:

- Is the process of producing the conceptual model sufficiently legitimate, for example did it involve key stakeholders appropriately?
- Is the process of producing the conceptual model sufficiently credible, for example did it involve relevant expertise and independent peer review?
- Is the conceptual model credible and defensible by virtue of the fact that suitable calibration and validation procedures have been followed?

If the strength of evidence provided by the model is examined in court, the legitimacy and credibility of the conceptual model are likely to be closely scrutinised. From a scientific perspective, a conceptual model should be treated as a hypothesis, whereby if the model output can be confidently concluded to be inconsistent with the observed behaviour of the system then the conceptual model is rejected; and if this cannot be concluded then it is provisionally accepted as a possible model and new observations are recommended to further test the model.

2. Problem framing and formulation

System conceptualisation

- Variables
- Processes, functions
- Indicators, criteria
- Scales

Definition of management scenarios



Checklist - Model conceptualisation

Aspect	Questions to consider
Model assumptions	<ul style="list-style-type: none"> • Are the model's assumptions transparent and supported by theoretical and/or empirical evidence? • Are the model's assumptions reasonable given the overall objective and perspective of the model? • Are all causal relationships explained and supported by theoretical and/or empirical evidence? • Are assumptions sufficiently validated by expert opinion and independent peer review?
Stakeholders engagement	<ul style="list-style-type: none"> • Have key stakeholders been involved in the process of developing the conceptual model? • Is the conceptual model communicated to stakeholders in an easily accessible way so they can understand the model's logic, and the implications to the model's results? • Is this information communicated in a way that is easily accessible to stakeholders so they can understand the conceptual model?
Documentation	<ul style="list-style-type: none"> • Is the conceptual model clearly documented, including reporting all the underpinning assumptions and simplifications? • Are the methods and practices to develop and validate the conceptual model clearly documented?

5.3.2 Data collection, cleansing and pre-modelling data analysis

Data with respect to environmental model development are typically imprecise, often sparse in space and/or time, with systematic and/or random errors, and/or inadequate coverage of conditions, rendering them insufficiently informative for model calibration. Their errors affect calibration of the model, while errors in data inputs also affect outputs (QoIs) when using the model in a predictive or simulation mode. Appropriate infilling of missing data depends on circumstance and simple interpolation is often inappropriate (for instance, between a sample taken during a river's low-flow period and a little after the start of a flow event). Inadequacies in data, both from errors and non-informativeness, need to be taken into account in the method for calibrating a model, and appropriate limitations on its subsequent use reported and communicated.

However, there is still a lot that can be done to improve such situations. Simple text-book analyses of data to reveal their signals and uncertainties before modelling is under-practised, or at least under-reported. A wealth of tools is available to detect outliers, trends, implausible correlations, timing errors in model response, and generally to extract information from data. The value of simple plotting and visualisation should not be ignored.

There is also much to be gained from paying more attention to the optimal design of experiments for data collection in the future. However, because collection of experimental data is expensive and only a limited amount of experimental data can be obtained, it must be recognised that not all experiments provide the same amount of information about the processes they are helping inform. Consequently, it is important to design experiments in an optimal way, such as by choosing a limited number of experimental data to maximise the

value of each experiment. Optimal experimental design (OED) uses models to guide experiment selection and has been shown to drastically improve the cost effectiveness of experimental designs for a variety of models, including those based on ordinary differential equations, partial differential equations and differential algebraic equations. OED has been developed in both Bayesian and non-Bayesian settings (Atkinson and Donev, 1992).

Data and modelling are complementary and considerable benefits can be gained by data collectors and managers working in close partnership with modellers. Those involved on the data side are in fact stakeholders in the modelling process and can thereby be shown appreciation of their role in the whole decision process. The partnership can be a synergistic one where the most appropriate data are monitored, their strengths and weaknesses known and opportunities sought for optimising model use by improvements in monitoring network design.

While most good modelling practice guidelines include sections on data management, they rarely include sections on implementing data management policies that are targeted at those responsible for formulating and implementing government strategy to improve water information. 'Good Practice Guidelines for Water Data Management Policy', recently released by the Australian Bureau of Meteorology under the World Water Data Initiative, provides valuable guidance on standards and protocols, including licensing and open access.

Checklist - Data collection, cleansing and pre-modelling data analysis

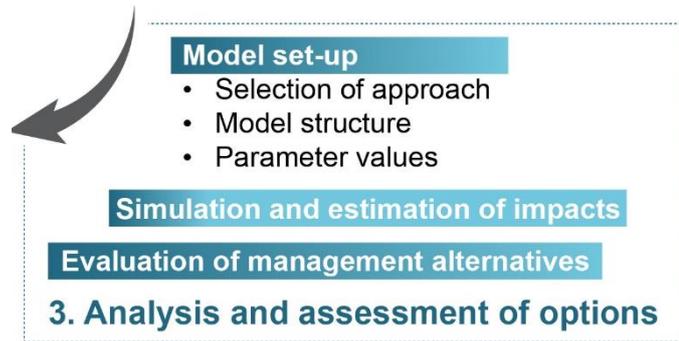
Aspect	Questions to consider
Data collection and cleansing	<ul style="list-style-type: none"> • Is there a clear data collection plan in place (i.e. what data, collected by whom, why, when and where)? • What procedures have been used to identify and select data sources? • How have these decisions been justified? • Have modellers been included in making these decisions?
Data analysis and handling	<ul style="list-style-type: none"> • Does the available data match the model's temporal and spatial resolution? If not, how have mismatches been handled? • Have the methods used for handling mismatches been justified and reported in sufficient detail? • Has special attention been placed on data used for estimating crucial model parameters? Is the preliminary data analysis methodology based on justifiable and sound statistical grounds? • Have the methods used for handling missing data been justified and reported in sufficient detail?
Stakeholder engagement	<ul style="list-style-type: none"> • Have key stakeholders been involved in the process of identifying datasets and sources? • Is the data analysis communicated to stakeholders in an easily accessible way so they can understand the implications to the model's results?
Documentation	<ul style="list-style-type: none"> • Has the quality of data been appropriately checked, assessed and documented? • Have all data used in the model been described and referenced in sufficient detail?

5.4 Analysis and Assessment of Options Phase

5.4.1 Model selection

Models can be categorised in a number of different ways (Jakeman et al. 2006; Balci, 2007; Kelly et al., 2013), including by:

- type (e.g. empirical, conceptual, physical, numerical, analytical);
- treatment of space (e.g. non-spatial models, lumped spatial models, grid spatial models);
- treatment of time (e.g. non-temporal, steady state, lumped discrete, dynamic);
- composition (e.g. coupled, integrated); and
- execution (e.g. distributed, web-based).



Various considerations influence the modeller's choice of the most appropriate model. First of all, the model needs to have the ability to estimate the parameters/variables of interest for the study at an appropriate scale and resolution (i.e. temporal, spatial, and thematic) which matches the rate of change in the system of interest (van Delden et al., 2011). Empirical and statistical models are appropriate only for predicting responses within the range of existing observational datasets (Robson, 2014). Observational datasets may consist of historical observations for a particular system, or observations from a range of similar systems with varying characteristics (e.g. similar catchments with varying land uses). If a model is to be used outside this range, for instance to predict effects of long-term climate change or to predict results for a region in a different climatic zone, then it is necessary to use a process-based model that reflects what is known of the mechanisms of change (physical, chemical and biological).

Even when using a process-based model, it is important to evaluate the assumptions underlying the model to verify that they still apply in the changed circumstances. For example, representations of the effects of variations in temperature in most aquatic ecosystem models assume increasing biogeochemical process rates with increasing temperature, such as an Arrhenius equation (Goldman, 1979). In reality, some rates, such as phytoplankton growth rates, will decline above an optimum temperature of around 30°C (Coles and Jones, 2000) so these models may need modification if applied to tropical regions or for climate change scenarios. Another example is the case of a hydrological model that is calibrated using flow estimates derived from a rating curve. Large flood events may take the system beyond the valid range of the rating curve, where the actual relationship between water level and flow may be quite different from that predicted (e.g. due to overbank flow). Long-term hydrological simulations may need to take into account changing river morphology, while for short-term simulations this is usually not necessary.

Another crucial consideration is the ability to scale up results from the model (temporally and spatially). Special attention needs to be given to the spatial and temporal discretisation used

in the model, and how these may influence the output accuracy. Very finely grained models in time and/or space do not necessarily lead to more accurate results. Detailed models can be mistakenly perceived as highly accurate, while the benefit of using fine time steps and grid sizes can be over-stated in reducing the numerical error. The other pitfall to recognise and address is that of the available data not matching the temporal and spatial resolution of the model. This can result in the need to change the model resolution, implement methods to interpolate the data, and/or acknowledge the influence of the granularity chosen on QoIs.

A further consideration is whether data are available as input to drive the model of choice and, more importantly, whether we have means to validate the model output, especially models of high spatial resolution. It may be valuable to seek existing knowledge, including from both scientific experts and land/water managers, about how the system functions, and how observations from other contexts (e.g. other catchments or paddocks), can be generalised or adjusted to be useable in the model.

A model's flexibility, including the ability to update code and functionalities, can be an important concern in model selection, especially for models whose basics are likely to have a long shelf life.

Finally, there are contextual factors (e.g. past experience of the modelling team, previous investments in modelling platforms) and constraints (e.g. the requirement to use the same model across the region for consistency) that can be influential in model selection.

In the next two sections we discuss sensitivity analysis and calibration issues. These can be valuable for helping decide between models of the same type but of different complexities such as in level of process description and/or parameterisation.

Checklist – Model selection

Aspect	Questions to consider
Time scale and horizon	<ul style="list-style-type: none"> • Is the model's time horizon sufficiently long to reflect all important differences between decision options? • Is the model's temporal scale sufficient to capture the transient interaction among modelled processes?
Model's purpose and objectives	<ul style="list-style-type: none"> • Is the selected model type appropriate to address the defined modelling questions?
Data	<ul style="list-style-type: none"> • Is the selected model type appropriate considering its data requirements with respect to available data? • Have the costs of obtaining, updating and storing this data been considered?
Reuse and adaptation	<p>In case of selecting an already developed model or modelling components:</p> <ul style="list-style-type: none"> • Is the model sourced from an authoritative source? • Is the model flexible in terms of the ease of updating code and functionalities?
Documentation	<ul style="list-style-type: none"> • Are the methods and practices used to select the model clearly documented?

5.4.2 Sensitivity analysis

Sensitivity analysis (SA) comprises a formal, quantitative set of methods used to identify the sources of uncertainty arising from model parameters and inputs, and their relative influence on outputs (Saltelli et al., 2004). A sensitivity index measures the ratio of a change in a model output (in particular QoIs) to a change in input or parameter. SA is often used as a step prior to model calibration and its purpose is to understand and quantify: (a) how each model parameter and potentially other model inputs, such as initial conditions and forcing variables like climate, affect relevant model outputs; and (b) how any parameter interactions contribute in strength to model outputs. Therefore, SA is of assistance in determining which parameters and parameter combinations should be prioritised in calibration, and more generally, which model inputs should be prioritised for uncertainty reduction. Results may suggest looping the modelling process back to an earlier step, such as revising or simplifying the conceptual model.

SA can also direct additional measurement efforts, including whether or not to improve the prior information used to inform specification of sensitive parameters; improve measurement of inputs to which the model is particularly sensitive; or to improve monitoring in ways that will better constrain calibration of sensitive parameters and other model inputs. SA may also be used post-validation, in application of the model, to test how outputs vary over different management options. Identifying sensitive inputs allows future research to focus on increasing knowledge of the behaviour of the inputs in order to constrain the input variability and hence reduce the output uncertainty. Good and robust SA can save a lot of time and effort. Identifying insignificant inputs can also help refine model structure through the combining or removal of parameters that have negligible effect on the behaviour of the model. Table 2 provides an overview of some techniques that can be used for sensitivity analysis, their strengths and limitations.

Methods	Description	Strengths and limitations
Local SA methods, such as: automatic differentiation (Wengert, 1964) and the Morris method (1991).	Characterise sensitivity by partial derivatives or gradients at the local point.	<ul style="list-style-type: none"> • Very simple and easy to implement and work well for linear models. • When the model is non-linear, the results obtained at a nominal point are in general not representative of the entire space.
Variance based techniques, such as the Fourier Amplitude Sensitivity Test (FAST) (Saltelli and Bolado, 1998) and the Sobol (1993).	Decompose the output variance into parts attributed to individual variables and interactions between variables.	<ul style="list-style-type: none"> • Sensitivity is an average over the range selected and so varies with the range specified.

Methods	Description	Strengths and limitations
Regional Sensitivity Analysis (RSA) (Hornberger and Spear, 1981).	Partitions a model output realisation for a given set of assumptions into either a behavioural set or non-behavioural set, that is the set of input factors (assumptions, parameter values, model inputs) that satisfy the problem constraints and those that do not.	<ul style="list-style-type: none"> • Gives insights into the input factors which yield an acceptable result.
Active-subspaces (Jefferson et al., 2016).	Identify directions in parameter space which may not be aligned with the parameter axes that significantly influence a QoI. These directions are the eigenvectors of a matrix derived from the gradient of the parameter-to-QoI map.	<ul style="list-style-type: none"> • Can be more computationally efficient than variance-based techniques.
Break-even analysis (Guillaume et al., 2016).	Identifies model variables at tipping points where one is considering management options two at a time; and conditions and uncertainties can be generated to define at which points one option is as good as another.	<ul style="list-style-type: none"> • Requires astute handling of multiple model variables.

Model emulation (also known as surrogate or meta-modelling) is the practice of developing a simpler, usually abstract statistical, model that is fitted to a more complex model and approximates its outputs. The surrogate model can be used to facilitate a more thorough sensitivity or uncertainty analysis than would be possible with the more complex model, or it can be used to allow simulation of a wider range of scenarios. Fraser et al. (2013) review the types of emulator models that are relevant to predicting time-series of environmental variables and examine the errors that arise when this approach is used to upscale complex field-scale models into a catchment scale model. Castelletti et al. (2012) and Razavi et al. (2012) also review and provide examples of relevant emulation applications. Model emulation methods promote computational efficiencies by replacing models that have slow runtimes, such as integrated, multi-component models, and mesh-based physical representations such as in most groundwater and hydrodynamic models.

Checklist – Sensitivity analysis

Aspect	Questions to consider
SA methods selection	<ul style="list-style-type: none"> • Has sensitivity analysis been undertaken? If not, how is this justified? • Are the methods used for sensitivity analysis appropriate and justified? • Are the ranges used for sensitivity analysis stated clearly and justified?
Use of SA results	<ul style="list-style-type: none"> • Have the results from the sensitivity analysis been used to determine which parameters and parameter combinations should be prioritised in calibration? • Have the results from the sensitivity analysis been used to investigate possibilities for simplifying the model's structure to improve computational efficiency and reduce modelling complexity?
Documentation	<ul style="list-style-type: none"> • Are the methods and practices used for sensitivity analysis clearly documented?

5.4.3 Calibration and model structure

Finding parameter values

Model parameters may be either calibrated from data, and/or specified from prior knowledge such as from expert opinion or measurement. Estimated parameters will always have uncertainty but so will parameters that are considered known or can be measured. For example, in the latter case aquifer properties vary across very small scales yet a parameter value for conductivity obtained from a groundwater pump test at a specific location may be used or adjusted to represent them at some specified larger scale. The chosen level of model parameterisation can have significant effect on whether or not a model can reproduce experimental observations. This is particularly true for parameterisation of spatially or temporally varying fields such as conductivity. The complexity of the parameterisation of conductivity can range from a single parameter for a homogeneous aquifer, to multiple parameters for a regional conductivity, through to thousands or even millions of parameters for a fully spatially distributed conductivity. A single parameter may be easily estimated from data, however may result in poor fitting to data, whereas a highly distributed conductivity may lead to 'over-fitting' and only a subset of parameters being informed by data.

While the purpose of model calibration is to identify the parameter sets that may be considered 'optimal' in terms of the selected objective function, often the focus is on finding the single best parameter set. As the optimal value of the objective function may be below a pre-specified threshold for the model to be considered potentially fit-for-purpose, the purpose of the calibration becomes a decision-point at which the modelling process loops back to one of the earlier stages.

In process-based models, which are often used in groundwater, hydrodynamic and biogeochemical modelling studies, parameters usually represent rates and traits that are, at least in principle, measurable. In this case, it is often possible to derive considerable prior information about the expected values of parameters from the literature, or from local measurements. This information should not be ignored in calibration, rather can be used in a variety of ways, including:

- informing Bayesian parameter estimation or uncertainty quantification approaches;

- setting appropriate initial values and value bounds for optimisation schemes; and
- guiding the selection of parameters that can be assigned fixed values to reduce the scope of the calibration exercise and reduce the risk of ‘over-fitting’ the model.

Whilst assuming prior knowledge may be used to constrain parameters in the formulated model structure, inappropriate constraints may underestimate or overestimate uncertainty such as the way prior information (known as ‘priors’) is selected for estimating aquifer parameters for conductivity and storativity. For example, an under-estimation of the variance in model priors will lead to a misleading under-estimation in the uncertainty of outputs of a groundwater model. Similarly, an over-estimation of prior uncertainty can lead to overly conservative estimates of uncertainty in predictions. Thus, prior knowledge should be assigned its own level of uncertainty and the effect of that on predicted Qols and associated indicators determined. For groundwater flow, simplified models based on analytical solutions such as those of Raats (1978a,b;) can offer insights that can help with understanding, for example, the likely distance that solutes can travel from a sink, such as a river, over time; the location where management interventions will be most effective; and accounting for the fact that it could take millennia before the consequences of any interventions have effect. Similarly, for interflow and surface flow, analytical models (Cook et al., 2009; Cook et al., 2011) can offer considerable insight when assessing model output.

Selecting an appropriate objective function

A fundamental principle for sensitivity analysis and calibration, in the context of developing a fit-for-purpose model, is that the target objective function should be a relevant error function or metric of the Qols. While this principle may seem straightforward, its practice is weak, and caution is advised. In surface water hydrology, for example, there is undue attention to a measure of mean squared error, known as Nash-Sutcliffe efficiency, which places emphasis on fitting high flows more closely than low flows. Careful attention to determining the precise objective functions (e.g. error functions of the Qols) is a sure way to reduce uncertainties that would otherwise manifest.

Bennett et al. (2013) present a wide range of performance and objective function measures and methods, including visual plots, which should be considered as objective function metrics for optimising parameter estimation. In particular, a function(s) of the quantities of predictive interest (Qols) must be deliberated and specified, requiring knowledge of the natural and human setting, which is often best realised through an appropriate participatory process (Hamilton et al. 2015). An example in hydrology that has a more specific purpose than the prediction of quantity fluxes, is where surface and/or groundwater modelling need to predict Qols relating to ecological requirements. The Qols in this case could relate to surface and groundwater levels, but might be a function of those with a more specific interest, such as the timing and pattern of surface and/or groundwater flows, which, in turn, might need to be specified in numerical terms as targets or indicators. Moreover, experts might further confirm how accurate, in quantitative or categorical terms, the associated predictions of the targets need be for the modelling to be useful.

Sometimes a model weakness may simply be that the modeller does not effectively relate the aims of the modelling to either the objective functions used to optimise model parameters, or the relevant performance measures (Bennett et al., 2013). For example, one may wish to accurately predict the level of an aquifer at a set of specific locations. In this

case, a very fine scale spatial model will be important for capturing the desired quantities. However, a lumped model, which may be good at predicting total water volume in the aquifer, would not be capable of predicting local quantities accurately. However, the fine scale model may need to be most accurate at certain locations, for example, where interactions with surface water occur, and/or that the model may need to be most accurate at times when the stream is losing (or gaining) water to (or from) the aquifer. The objective function for model calibration, therefore, needs to take into account the QoIs and the type of predictive error in them that is appropriate to effectively minimise. In general, it is a good practice to examine the effects of different/multiple objective functions, and to perform a sensitivity analysis for uncertain input and parameters as well as for presumed certain parameters.

Selection of calibration periods

Examining the effects of different calibration periods is crucial. The period of calibration should be determined in the context of the model's purpose and use. For example, a model that is calibrated under average climatic conditions, and assessed only in these conditions, should not be used directly to predict quantities associated with extremely wet or dry states. Water models calibrated on different periods will have substantially different behaviours and parameter values when the region of application experiences strong climate variability. Wherever possible, one should calibrate a model on different periods and assess the performance of each calibration on all other periods. This is called cross-validation and is an empirical integrator of uncertainties and provides a valuable assessment of the minimum uncertainties to be expected when making predictions.

One consideration is the extent to which the datasets for calibration cover the potential range of inputs to the model rather than the size of the dataset. In some circumstances, the hydrologist may want the model to fit best to low flows or high flows. Since the model will fit best to the mean of the data set (Venables & Dichmont, 2004) the objective function and the weighting given to the data in different ranges needs to be carefully considered. Another way is to exclude data in the range where the model does not need to fit as well as a means of lowering the weight given to this data.

Assessing calibration performance

There are many methods available to assess the performance of a calibrated model, as presented in Bennett et al. (2013). A common deficiency in assessing calibration performance is the omission of a cross-correlation analysis between model residuals (predictions minus corresponding observations) and model inputs to assess if there is anything missing in the model's explanation of outputs. Verification and validation must not be carried out deterministically but rather executed to account for the model uncertainty (e.g. variation in convergence rates of mesh refinement studies, due to parameter uncertainties).

Identification of model structure and embracing multiple models

Importantly, calibration can be defined to include the identification of model structure, inputs and boundary conditions, and not just an estimation of a model's parameters. Model structure in the water domain will relate predominantly to the complexity of process (types and detail), assumptions considered, as well as the levels of spatial and temporal discretisation.

Formal statistical tests for differentiating among different model structures are well developed. They provide criteria which compare the number of parameters against the improvement in model fit to observations. Because of their reliance on statistical assumptions, formal statistical tests are best treated as guides that can be used to check the results of the structure recommended on other grounds such as: predictive performance on independent data sets, credibility of parameter estimates, and consistency with prior knowledge. The underlying aim is to balance sensitivity to system variables with complexity of representation. A key question not often asked is whether some system descriptors, for instance dimensionality, discretisation and processes, can be aggregated to make the model representation more efficient, concentrating only on what dominates the system response indicators at the scales of concern. Allowing more degrees of freedom than warranted in system representation can lead to 'over-fitting' (to errors) and unrealistic model behaviours and predictions.

Working with multiple models is also a useful way to explore uncertainties in model formulations. Different model structure candidates or perspectives can be used with tools like sensitivity analysis to understand sources of uncertainty. Various techniques such as Bayesian model selection can then be used to assess the strengths and weaknesses of each, and under which conditions each model is more suitable. Calibrated parameter values can also provide clues about the structural accuracy of models. If a model provides a better fit to the observational data when one or more parameters are calibrated to unexpectedly or unreasonably high or low values, it suggests either a systematic bias in measurements or an error in model structure (or in values of other parameters). For example, a model that requires an unrealistically high parameter value for phytoplankton growth rate may be missing a source term (seeding of phytoplankton from weir-pool blooms, for instance, or from germination of akinetes) or over-estimating a loss term (perhaps it does not allow for the unpalatability of some phytoplankton species to grazers or does not allow for resuspension of diatoms settled to the sediments).

On automated calibration tools

There are solutions to support the automated development of model calibration, which can improve the efficiency of the process, however, care needs to be taken in using them. Over reliance on these tools may bring the risk of losing data insights that can help interpret model results and help understand the system. The use of automated tools should not preclude the use of sensibility (or common-sense) testing. It is helpful to the reliability of modelling if these sensibility tests are built into the models so that all model runs can be conveniently benchmarked.

Checklist - Calibration and model structure

Aspect	Questions to consider
Methods selection and setup	<ul style="list-style-type: none"> • Are the methods used for model calibration appropriate and justified? • Is the method used for selecting the calibration period appropriate and justified given the model's purpose and use? Has this involved examining the effects of different calibration periods? • Has a cross-correlation analysis between model residuals and model inputs been conducted? • In the case of using automated model calibration tools, has sensibility (common-sense) testing still been conducted? • Has special attention been paid to accounting for calibration issues related to spatial heterogeneity?
Calibration objective function	<ul style="list-style-type: none"> • Is the target objective function(s) used for model calibration the relevant error function or metric of the QoI? • Have different/multiple objective functions been examined and compared in terms of their effects on the model's fitness for purpose? If not, how is this justified?
Data	<ul style="list-style-type: none"> • Do calibration datasets cover the potential range of input conditions to the model?
Use of calibration results	<ul style="list-style-type: none"> • Have the results from the sensitivity analysis and calibration efforts been used to address the structural uncertainties? If not, how is this justified?
Documentation	<ul style="list-style-type: none"> • Are the methods and practices used for model calibration clearly documented?

5.4.4 Validation and testing

Model validation is defined by Refsgaard and Henriksen (2004) as “Substantiation that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model.” Validation must be considered through the lens of uncertainty and there are several ways that validation can be approached and a combination of methods is typically appropriate.

‘Crash’ or ‘stress-testing’ the model is a simple but under-practised exercise to explore model strengths and weaknesses. It can be similar to scenario modelling (see Section 3.9) but with a different purpose in that attempts are made to see what model parameter sets, observation periods and other assumptions and conditions establish limitations or invalidate the model. This should include examining the performance of the model through time and/or space to assess inadequate performance. Stress-testing should be applied as much as resources can allow.

Hipsey et al. (in prep.) propose a four-level evaluation framework for process-based models such as hydrodynamic-biogeochemical models:

- Level 0: Is the model's behaviour plausible in light of existing theory and system understanding? This can be evaluated in consultation with disciplinary experts and/or

stakeholders and equates to sensibility testing (see section 5.4.3).

- Level 1: Traditional model evaluation of model performance against monitoring data, such as time series of nutrient, sediment and chlorophyll concentrations. Metrics should include measures of correlation, measures of bias and other measures of error.
- Level 2: Evaluation of predicted process rates, such as comparing observed versus simulated nitrification and denitrification rates, zooplankton grazing rates, and net ecosystem metabolism.
- Level 3: Evaluation of the model's ability to reproduce system-scale emergent properties that are not built into the model's structure and were not considered during calibration. Examples may include phytoplankton community structure (the relationship between percent nano- or pico-phytoplankton and chlorophyll a concentration), length scales of eddies, or the statistical distributions of nutrient concentrations in different parts of flood plumes.

In water quality models, the drivers for the transport of solutes and particulates are the velocity of the flow (advection) and dispersion, whereas the drivers for water flow (quantity) are determined by the potential energy or pressure head differences. This means that calibrating a model for water flow does not necessarily work well for solutes and particulates.

A crucial part of testing is placing physical bounds on the uncertainty that can exist. These physical bounds can help in reducing what would otherwise be unrealistic uncertainties and also help with understanding whether the model is giving sensible answers, as the results should always occur within the set bounds. For example, in the case of streamflow volume, the upper bound can be considered to be the rainfall volume (i.e. all the water runs off and appears as streamflow during an event). This means that the rainfall multiplied by the area of the catchment should be the upper bound for the cumulative streamflow for a rainfall event. The lower bound for streamflow can also be defined as the larger of zero and rainfall minus potential evapotranspiration multiplied by the area of the catchment, as it is unlikely that all of the potential evaporation will be realised.

Similarly, limits for water quality constituents can be estimated based on sensible limits and used to assess if the model is giving sensible results. Defining these limits is more difficult, but plausible upper limits based on observed extreme values of quantities like sediment concentration and other water quality parameters are available. In addition, there are physical constraints on volumetric sediment concentration, and relationships among some water quality constituents that are based on stoichiometric principles. Zero can be taken as the lower limit of constituent concentrations.

When using evaporation estimates from countries outside Australia, it is necessary to check where the data come from. In China, they often use a 0.20 metre diameter pan, so the pan evaporation figures are much greater than what would be found with a class A pan (McVicar et al., 2005). Because of this Cook and Jayawardane (2008 unpublished) found the pan evaporation had to be multiplied by 0.44 to get the reference evapotranspiration. Thus, when calculating bounds, it is essential to check that the data used make sense first.

Close investigation of issues related to uncertainty propagation through coupled and integrated models is a promising topic for research and practice. Several groups, including

Borsuk et al. (2001), Webb et al. (2010) and Obenouer et al. (2014), have applied Bayesian Hierarchical Modelling approaches to uncertainty quantification and parameter estimation. Key advantages of this approach are that it allows prior information on the expected parameter values as well as confidence in observational data used to calibrate the model to be taken into account and provides both calibrated parameter values and model predictions in a probabilistic framework. The probability distributions that arise as outputs from one component of an integrated model can be used as prior distributions for input to another component of the integrated model system. In this way, uncertainty can be propagated through the model system without the exaggeration that occurs if propagating confidence intervals (e.g. Larssen et al., 2006). The related approach, Bayesian Melding (Poole et al., 2000), can also be used to consider uncertainty in model structure.

Checklist – Validation and testing

Aspect	Questions to consider
Structure validation	<ul style="list-style-type: none"> • Has the mathematical logic of the model been tested thoroughly? • Has expert opinion been used to validate the model's structure?
Behaviour validation	<ul style="list-style-type: none"> • Has the model's behaviour been tested against actual data especially that not used for calibration? • How does the model's testing account for the physical bounds on possible uncertainties? • Have the model's results been compared with those of other models and any differences in results explained? • Has the model been stress-tested? Are the methods used for stress-testing appropriate and justified given the model's type and use?
Uncertainty propagation	<ul style="list-style-type: none"> • Have issues related to uncertainty propagation been thoroughly investigated?
Documentation	<ul style="list-style-type: none"> • Are the methods and practices used for model testing and validation clearly documented?

5.4.5 Scenario analysis

In its broadest sense, scenario analysis involves exploring multiple plausible assumptions about future conditions (largely model inputs), model structure and parameter values (Alcamo, 2001). For example, in an aquifer context, future climate will affect the amount of recharge of precipitation to the groundwater, making predictions uncertain. Cross-sectoral issues creating future uncertainties may relate to the interactions of proposed energy extraction projects with existing groundwater uses for agriculture, or a government policy to issue more groundwater access to increase food production. Scenario analysis can be used for many purposes (Maier et al., 2016), such as to promote discussion and sharing of knowledge and perspectives and/or to search for such scenarios that lead to good, intermediate and poor outcomes. At its core is simulation of model drivers and parameter samples, and analysis of the model's QoI functions (i.e. target indicators).

The use of well-defined, standard and consistent scenario sets (i.e. scenario library) that are

packaged as a part of the model is a good practice. In addition to preserving replicability, packaging scenario datasets with models provides three significant advantages. Packaging facilitates:

1. extension of the scenarios to related domains (e.g. running the same or similar scenarios used with a hydrological model, but for a water quality model, or an integrated socio-environmental model);
2. cross-comparison of results between models and ensemble scenario analysis; and
3. comparison between scenario predictions from an existing model and from proposed new versions of the same model.

Rather than attempting to develop a minimal set of presumptive scenarios for stakeholders to contemplate, an alternative approach to scenario development is to utilise a model to simulate a large number of possible outcomes and then allow stakeholders to deductively visualise the entire suite of outcomes and then articulate preferences. This approach is known as exploratory modelling and is attracting growing attention in the scenario analysis literature (Banks, 1993; Walker et al., 2013). Exploratory modelling represents a family of techniques whose aim is to explore robust solutions under various future possibilities as captured in different model assumptions and parameter values (referred to as cases, scenarios, ensembles, and eras). Some of these techniques include: Robust Decision Making (RDM) (Groves and Lempert, 2007; Lempert et al., 2003), Scenario Discovery (Bryant and Lempert, 2010), Dynamic Adaptive Policy Pathways (Haasnoot et al., 2013; Kwakkel et al., 2016b), and Objective Robust Decision Making (MORDM) (Watson and Kasprzyk, 2017). In principle, these techniques share the idea of open exploration and searching for robust solutions. However, they vary technically in how the scenario generation process is conducted and the type of insights that are generated (Moallemi et al., 2017). There is limited understanding of the fundamental differences between these techniques, their relative strengths and limitations, and the implications of how uncertainty is treated, and solutions identified (Haasnoot et al., 2013; Trutnevyte et al., 2016). Comparative and evaluation studies to investigate differences and complementarity are still needed. To support practice, research into good practices for conducting exploratory modelling is also needed.

Checklist - Scenario analysis

Aspect	Questions to consider
Scenario selection	<ul style="list-style-type: none"> • Is there a clearly stated scientific approach used for defining and formulating scenarios in a transparent way? If not, how is this justified? • Has some consideration been given to using automated scenario generation tools to promote scenario consistency and diversity? If not, how is this justified?
Data	<ul style="list-style-type: none"> • Is the choice of scenario datasets described and justified? • Is there a proper protocol around storing and packaging the scenario datasets with the model?
Documentation	<ul style="list-style-type: none"> • Are the methods and practices used for scenario analysis clearly documented?

5.4.6 Data and model management

5.4.6.1 Provenance, governance, and meta-data

Management of input, intermediate and output data is one of the more difficult aspects of modelling. Some of the key challenges are:

- what and how much data to store from a model run and how many model runs to store;
- how to manage updates to input data and record its provenance;
- how to manage updates to the model executable itself; and
- how to ensure that the modeller knows what data they are using.

Governance of model data requires implementation of strong internal Quality Assurance and Quality Control (QA/QC) procedures that respect in-house work culture while improving practice. Management of observed data within a specialised database (e.g. Hydstra for hydrological data) is an industry norm that is rarely extended to modelled data. Adoption of new technologies such as scientific workflows and data and model service brokering services is limited, for instance, in the hydrology modelling community—perhaps a reflection of the level of control in the overall modelling lifecycle required of modellers. There can also be tension between corporate IT and its data governance practices and the requirements of the modelling community to manage exploratory testing and production environments.

An effective data management program requires a strategic investment in effort with a clearly articulated vision and steps to achieve it, that is shared with users and practitioners. The goal may be to achieve a shift in culture, supported by in-house infrastructure and management. Tools such as the Data Management Maturity model (adapted from the Carnegie Capability Maturity Model) can be used to assist in identifying the level of data management that is required, and achievable. In that model, there are five levels for managing data. Table 3 summarises these levels in the order of highest to lowest maturity level. This is a simple example as it contains only three capabilities: safety, versioning and lineage, as extracted from the Provenance Maturity Model developed by Taylor et al. (2015). In this example 'safety' is defined as resilience to failures in hardware and/or processing errors; 'versioning' is whether there is a scheme to identify versions; and 'lineage' is whether information is kept to identify the source of the data/model.

Table 3 Maturity levels for three data/model capability (from Taylor et al., 2015)

Level	Data/model safety	Data/model versioning	Data/model lineage
Optimising	Highly valued business process	Integral part of model management	Intrinsic part of modelling practice
Managed	Automated and widely used	Full adoption, and partially automated	Standard practice and partially automated
Defined	Defined standard process	Defined standards, mostly adopted	Agreed metadata standards, mostly adopted
Tactical	Some backups, but not rigorous	Individuals have their own process but not standard	Some metadata recorded but very patchy
Initial (Chaotic)	No or ad-hoc backup strategy	No versioning or naming standards	No idea of source data or its state

Capability Maturity Models have been developed for many domains, including provenance management (e.g. Taylor et al. 2015) and the software development process (the original Carnegie Capability Maturity Model). Both these models have multiple capability matrices. e.g. the Provenance Maturity Model of Taylor et al. (2015) has six components (one being Data management), each of which has multiple capabilities described at five levels of maturity. The use of Maturity models is recommended as they provide a 'simple' mechanism for capturing aspiration and operational reality and provides a mechanism for moving between levels and agreeing on what is achievable, noting that it is not necessary for each capability to be at the same level.

Data and model sharing between collaborators and with the wider data provisioning community is improving with the increasing adoption of creative commons and data sharing licensing, which allows for use and reuse of data and models between jurisdictions and partners. The Australian Government supports research data management through its National Collaborative Research Infrastructure Strategy (NCRIS) funded Australian National Data Service (ANDS), which Australian universities and research agencies link to through tailored data access portals. The requirement to lodge research data (inputs and outputs) with ANDS as part of a modelling project proposal and planning has proved to be an effective way of fast-tracking cultural change towards best practice data management and governance.

Parts of the data management process can be automated, but this requires a significant investment in time and resources and a clear understanding of the benefits gained and how they offset any perceived disadvantages, such as the loss of transparency. Many tools now exist to automate parts of the data and model management life cycles, ranging from relatively unsophisticated tools that manage code versioning and extract metadata from that code or datasets, through to ones that fully automate the model execution process.

5.4.6.2 Automating the data management and modelling process

A modelling process that lacks well-thought-out and transparent experimental design inhibits the reproducibility of results, the effective reporting of results, and therefore, the credibility of the model (Teran-Somohano et al., 2014). Modellers usually go through an iterative process of ad-hoc experimentation and adaptation till they land on the final set of results on which to base recommendations. In many cases, model results are presented as a ‘bunch of results’ without much explanation as to why those experiments/results have been cherry-picked, and how they are driven from an experimental design that logically flows from the model’s objectives and research/policy questions. This is poor practice especially when interrogating large complex models, where many possible interactions among factors and outcomes play out. Instead, modellers need to embrace the use of automated methodologies that can support transparent experimental workflow and allow for systematic understanding of the impacts of the various relationships and factors that influence the model’s results (Chakladar, 2016). Methodologies, such as model-driven engineering (MDE) and model-driven science (MDS), provide principles, techniques and tools that meet these needs (Yilmaz et al., 2016).

Workflows are a useful concept and construct for designing and managing the modelling process itself. They provide an intuitive mechanism for the composition of the modelling and provide a repeatable, automated way of running and documenting all the elements that constitute a model execution. Sophisticated workflow engines support the ability to ‘plug-and-play’, either through changing input datasets, managing a range of parameter sets that can be switched in and out, and even switching model components. They track the provenance of data used in the model run and the provenance of the model (at least its version number) and manage the storage of outputs from multiple runs.

Workflows have been around for a long time in the business world. For example, an online shopping cart is a workflow with a set number of sequenced steps and prescribed logic for moving forward or backwards through those steps. The use of workflows to manage data and model life cycles is more recent. Data management tools that manage the data workflow are readily available, such as Truui. Scientific workflow management systems to manage the modelling workflow have grown out of the need to deal with big data, increased model complexity, and collaborative research. They are designed to support team/individual-based scientific experimentation and the sharing of data, models and workflows. Examples of such systems include workflows that manage DNA sequencing, the construction and composition of remotely sensed imagery, and the running of complex coupled modelling suites. There are many scientific workflow products available on the market, though most have come out of specific scientific disciplines such as bioinformatics, astronomy, chemistry and require significant investment in their technical back-ends, implementation and deployment.

Electronic notebooks, such as Jupyter, can be used as ‘low-level’ workflow entry points. They support the writing of ‘documents’ that contain and execute code, store results, allow for intervention in the model execution process, and can be deployed to provide unified software management and data access within organisations.

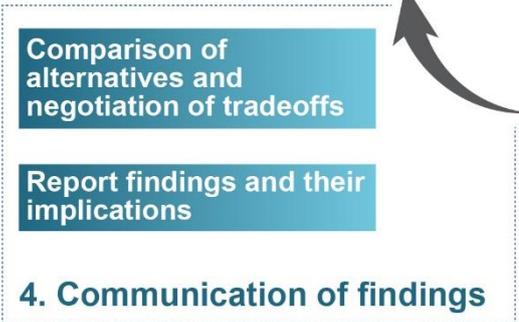
Checklist - Automating the data management and modelling process

Aspect	Questions to consider
Planning	<ul style="list-style-type: none"> • Is there a clearly stated data management plan in place (i.e. storage, sharing, security, cataloguing, publishing, archiving, maintenance)? • Does the data management plan follow common standards for data management? • Does the modelling team have the skills and resources needed to manage the data? • Is there serious thought into the data and model management maturity levels, and justification for the level to be adopted? • Is there an agreement about whose responsibility it is to maintain and update this data?
Documentation	<ul style="list-style-type: none"> • Are the methods and practices used for data management clearly documented?

5.5 Communication of findings phase

5.5.1 Selecting indicators to communicate model results

One of the most crucial issues when it comes to the communication of model results is the selection of the appropriate set of indicators to report the modelling results. Fundamentally, indicators reflect the objectives/values incorporated in the model. They vary according to multiple aspects, including: level (whole system or sub-system), purpose (average performance or variability in performance, and ‘snapshot’ or ‘pathway’), type (absolute value or proportional, descriptive or normative such as difference between hypothesized best value and the calculated value), and formulation. Different indicators can be used to diagnose different system characteristics. Identifying and selecting a suite of integrated and balanced indicators is important to ensure that the decision maker can clearly understand the effects different decision options have on the system over its lifetime (Bauler, 2012). For example, Fu et al. (2017) examined a suite of mathematical indicators used for evaluating the non-market value of environmental change. They concluded that all indicators have limitations and stressed the need for contextual information to mitigate possible biases. Also note that the number of indicators presented to decision makers must be managed. Balancing succinctness and informativeness is desirable. Thus, modellers should strive to educate decision makers on the need to go beyond single numbers to indicate uncertainty, but also realise that modellers can be too complex and overwhelming in their communication and the amount of data presented.



5.5.2 Communicating uncertainty in written reports

The language used in science policy reports is often very measured and calibrated (McInerney, et al., 2014), especially when acknowledging uncertainties and knowledge gaps.

However, this does not consider how the reader interprets these findings and the uncertainty implications. Special attention needs to be paid to the way uncertainty is communicated in written reports. The way language is used to communicate uncertainty (i.e. uncertainty framing) plays a significant role in how uncertainty is interpreted by the reader (Guillaume et al., 2017). Towards the development of best practices around framing uncertainty, Guillaume et al. (2017) have developed a typology of eighteen uncertainty frames. The typology has both a descriptive and prescriptive function to assist modellers communicate uncertainty. In its descriptive role, the typology can be used to describe the existing uncertainty frames (at least in abstracts) employed. The outcome of the descriptive function is to evaluate how the selection of a particular uncertainty frame influences the way the reader interprets the findings. In its prescriptive role, the typology gives users conceptual guidance into how to think and select uncertainty frames that best communicate their intended message (i.e. fits the purpose). The availability of a range of frames helps to raise awareness about multiple ways of delivering the message, which ultimately leads to more critical thinking about this when writing a publication or report.

5.5.3 Visualisation

Effective visualisation tools are needed to provide intuitive descriptions of complex and large volumes of simulation data. The importance of this task has been recognised, including by the US Department of Energy (DOE) which has funded the SciDAC Institute of Scalable Data Management, Analysis and Visualization (SDAV). Model visualisation is not just aesthetic, but effective visualisation tools can facilitate better understanding of the processes that produce the data and reveal interesting characteristics of the datasets. For decision makers, visualisation helps distil the key information without overwhelming the user with all the modelling details. The effectiveness of a visualisation technique depends on the problem on hand and consideration of factors such as audience, the intent of the message to be communicated (e.g. trade-offs and uncertainty), as well as the data types.

A key challenge in model visualisation is the communication of large datasets, especially in problems with multiple objectives and trade-off solutions. Traditional visualisation techniques, such as scatter plots, are no longer appropriate tools in the visualisation of a high-dimensional objective space. He and Yen (2017) identified three criteria for high quality visualisation of high-dimensional multi-objective space. They state that a visualisation tool should, firstly, provide accurate information of the 'Pareto Front' (see below for details). Secondly, it should provide decision makers with a clear indication of cost-benefit and/or trade-off solutions. Thirdly, the tool must be scalable to higher dimensions and larger datasets. The authors reviewed the available approaches and evaluated their performance in meeting these criteria. They concluded that the reviewed techniques can satisfy one or two of these three criteria to some degree. However, none can fully satisfy them all, which leaves the door open for integrated approaches that can leverage the strengths of existing techniques.

Another challenge of effective visualisation relates to uncertainty communication, especially when incorporating spatio-temporal heterogeneity. As mentioned above, a key tool now used in portraying uncertainty is the 'Pareto Front'. Its portrayal of a prediction versus degradation of model fit underscores the fact that multiple models might be considered reasonable and provides a view of how much model fit would need to be lost in order to meet a specific model outcome (Australian Groundwater Modelling Guidelines, 2012). Bonneau et al. (2014)

present a review of methods for uncertainty visualisation, and Kinkeldey et al. (2017) a review of effectiveness of some of the methods.

An important consideration when developing and using visualisation tools is understanding and mitigating the possible biases in audience's interpretations, which may ultimately lead to over or less confidence in the results (McInerney, et al., 2014; Sacha et al., 2016). For example, rescaling results through visualisation can invite systematic biases. McMahon et al. (2015) found that a group of novice readers, who were shown a graph of climate change projections, misinterpreted the intended message about the role of socio-economic factors in the IPCC scenarios.

Checklist - Visualisation

Aspect	Questions to consider
End users	<ul style="list-style-type: none">• What are the assumptions (implicit and explicit) underpinning the methods and tools used to communicate model results (e.g. audience's technical and mathematical knowledge)?• Is there evidence that the model visualisation is useful to the user and distils the message to be communicated?
Uncertainty communication	<ul style="list-style-type: none">• Is there special attention given to how the user may interpret a model's results especially in regard to reported uncertainty?

6 Conclusions

The most notable observation about this paper is that, although some attention has been given to defining good modelling practices in natural resource management, there is still little progress across the globe in putting these into *actual and* routine practice—despite the critical need for such practices in coupling and integrating models. Much of the discourse around adopting good modelling practices is still conducted at a high level, such as general advice around comprehensive testing, without much drilling down into the details of implementing these recommendations.

It is clear that there is still a strong need for an ongoing pursuit to identify, define, and document modelling practices. This presents an opportunity for us, the Queensland Water Modelling Network (QWMN), to lead the way in establishing and demonstrating good practices within and across our water modelling domains. The QWMN intends to focus on the transparent description of modelling practices at an appropriate level of detail to allow knowledge sharing and effective communication not only among the modelling community, but also among modellers, end users and decision makers. To this end, we have provided a series of checklists comprising questions that should be addressed in the pursuit of good modelling practices.

Because modelling is as much an art as a science, and involves many choices at every step, progress in achieving best practice is best facilitated by undertaking case studies and sharing the lessons. Indeed, for the water sector, cases could be undertaken and experiences shared in each of the major domains (surface water, groundwater, estuary, marine etc.) where some of the experiences tend to differ markedly. An even more desirable outcome would be to achieve best practice across an issue that encompasses all the major domains that would lead to cultural change in modelling practice and its associated multiple benefits.

7 References

- Alcamo, J 2001, 'Scenarios as tools for international environmental assessments'. *Environmental Issue report 24*, Luxembourg: Office for Official Publications of the European Communities.
- Alwazae, M Perjons, E. and Johannesson, P 2015, 'Applying a template for best practice documentation', *Procedia Computer Science*, pp.72; 252-60.
- Argent, RM, Sojda, RS, Guipponi, C, McIntosh B, Voinov AA, and Maier HR 2016, 'Best practices for conceptual modelling in environmental planning and management', *Environmental Modelling and Software*, pp.80; 113-21.
- Atkinson, AC, Donev, AN 1992, *Optimum experimental designs*, Oxford University Press.
- Australian Groundwater Modelling Guidelines (2012)*, Sinclair Knight Merz and National Centre for Groundwater Research and Training, Waterlines Report Series No. 82 June 2012.
- Bankes, S 1993, 'Exploratory modeling for policy analysis', *Operations Research*, 41(3); pp.435-49.
- Bauler, T 2012, 'An analytical framework to discuss the usability of (environmental) indicators for policy', *Ecological Indicators*, 17: 38-45.
- Bennett, ND, Croke, BFW, Guariso, G, Guillaume, JHA, Jakeman, AJ, Newham, LTH, Norton, JP, Perrin, C, Pierce, SA, Robson B, Seppelt R, Voinov, AA, Fath, B, Andreassian, V 2013, 'Characterising Performance of Environmental Models', *Environmental Modelling and Software*, 40 pp.1-20.
- Beven, K, Freer J 2001, 'Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology', *Journal of Hydrology*, 249 (1-4): pp.11-29.
- Black, D, Wallbrink, P, Jordan, P, Waters, D, Carroll, C, Blackmore, J 2011, *Guidelines for water management Modelling: towards best practice model application*. eWater Cooperative Research Centre, ISBN 978-1-921543-46-3.
- Black, DC, Wallbrink, PJ, Jordan, PW 2014, 'Towards best practice implementation and application of models for analysis of water resource management scenarios', *Environmental Modelling and Software*, 52: pp.136-148
- Bryant, BP, Lempert, RJ 2010, 'Thinking inside the box: A participatory, computer-assisted approach to scenario discovery', *Technological Forecasting and Social Change*, 77(1): pp.34-49.
- Bonneau, GP, Hege, HC, Johnson, CR, Oliveira, M M, Potter, K, Rheingans, P, Schultz, T 2014, 'Overview and state-of-the-art of uncertainty visualization', In Charles D. Hansen, Min Chen, Christopher R. Johnson, Arie E. Kaufman, and Hans Hagen, editors, *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*, pp. 3-27. Springer London, London.

- Borsuk ME, Higdon D, Stow CA, Reckhow KH 2001, 'A Bayesian hierarchical model to predict benthic oxygen demand from organic matter loading in estuaries and coastal zones', *Ecological Modelling*, 143(3): pp.165-81.
- Brown, J D, Heuvelink, GBM 2007, 'The Data Uncertainty Engine (DUE): A software tool for assessing and simulating uncertain environmental variables', *Computers & Geosciences*, 33(2): pp.172-190.
- Castelletti, A, Galelli, S, Ratto, M, Soncini-Sessa, R, Young, PC 2012, 'A general framework for dynamic emulation modelling in environmental problems', *Environmental Modelling and Software*, 34: pp. 5–8.
- Chakladar, S 2016, *A model driven engineering framework for simulation experiment management*, PhD diss., Auburn University.
- Chen, SH, Pollino, CA 2012, 'Good practice in Bayesian network modelling' *Environmental Modelling and Software*, 37: pp. 134-145.
- Coles, JF, Jones, RC 2000, 'Effect of temperature on photosynthesis-light response and growth of four phytoplankton species isolated from a tidal freshwater river', *Journal of Phycology*, 36(1): pp.7-16.
- Cook, FJ, Jayawardane, NS 2008, Modelling of water and solute transport at Yanggao County, Shanxi, Peoples Republic of China. Unpublished.
- Cook, FJ, Knight, JH, Wooding, RA 2009, 'Steady groundwater flow to drains on a sloping bed: Comparison of solutions based on Bossinesq equation and Richards equation', *Transport in Porous Media*, 77: pp. 357-372.
- Cook, FJ, Neumann, LN, Siriwardena, L, Western, AW 2011, 'Does where you plant trees make a difference in hydrologic response'. In Chan, F, Marinova, D and Anderssen, RS (eds) *MODSIM2011, 19th International Congress on Modelling and Simulation*, Modelling and Simulation Society of Australia and New Zealand, December 2011, pp. 2317-2323. ISBN: 978-0-9872143-1-7.
- Fraser, CE, McIntyre, N, Jackson, BM, Wheeler, HS 2013, 'Upscaling hydrological processes and land management change impacts using a metamodeling procedure', *Water Resources Research*, 49, doi:10.1002/wrcr.20432.
- Fu, B, Dyer, F, Kravchenko, A, Dyack, B, Merritt, W, Scarpa, R 2017, 'A note on communicating environmental change for non-market valuation', *Ecological Indicators*, 72: pp.165-172.
- Gaber, N, Foley, G, Pascual, P, Stiber, N, Sunderland, E, Cope, B, Nold, A, Saleem, Z 2009, (EPA/100/K-09/003). *Guidance on the Development, Evaluation, and Application of Environmental Models* <https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P1003E4R.PDF>
- Goldman, JC 1979, 'Temperature effects on steady-state growth, phosphate uptake, and the chemical composition of a marine phytoplankter', *Microbial Ecol.*, 5: pp.153-166.
- Groves, DG, Lempert, RJ 2007, 'A new analytic method for finding policy-relevant scenarios' *Global Environmental Change*, 17(1): pp.73-85.
- Guillaume, JHA, Croke, BFW, El Sawah, S, Jakeman, AJ 2011, *Implementing a framework for managing uncertainty holistically*. Paper presented at the 8th IWA

Symposium on Systems Analysis and Integrated Assessment Watermatex2011, San Sebastian, Spain, 20-22 June 2011.

Guillaume, JHA, Arshad, M, Jakeman, AJ, Jalava, M and Kummu, M 2016, 'Robust discrimination between uncertain management alternatives by iterative reflection on crossover point scenarios: Principles, design and implementations', *Environmental Modelling and Software*, 83:326-343.

Guillaume, JH, and ElSawah, S 2014, 'Fostering assumption-based stress-test thinking in managing groundwater systems: learning to avoid failures due to basic dynamics', *Hydrogeology Journal*, 22(7): 1507-1523.

Guillaume, JHA, Arshad, M, Jakeman, AJ, Jalava, M and Kummu, M 2016, 'Robust discrimination between uncertain management alternatives by iterative reflection on crossover point scenarios: Principles, design and implementations', *Environmental Modelling and Software*, 83:326 -343.

Guillaume, JHA, Helgeson, C, ElSawah, S, Jakeman, AJ and Kummu, M 2017, *Towards best practice framing of uncertainty in scientific publications: a review of Water Resources Research abstracts*. Water Resource Research (in press).

Haasnoot, M, Kwakkel, JH, Walker, WE and Maat, JT 2013, 'Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world', *Global Environ. Change*, 23: 485.

Hamilton, S, El Sawah, S, Guillaume, JHA. and Jakeman, AJ 2015, 'Integrated assessment and modelling: a review and synthesis of salient dimensions', *Environmental Modelling and Software*, 64: 215-229.

He, Z, and Yen, GG 2017, 'Comparison of visualization approaches in many-objective optimization', In *Evolutionary Computation (CEC)*, 2017 IEEE Congress, pp. 357-363, IEEE.

Hornberger, GM and Spear, RC 1981, 'An approach to the preliminary analysis of environmental systems'. *J. Environ. Management*, 12: 8-18.

Horsburgh, JS, Tarboton, DG, Hooper, RP and Zaslavsky, I 2014, 'Managing a community shared vocabulary for hydrologic observations', *Environmental Modelling and Software*, 52: 62-73.

Jakeman, AJ. and Jakeman, JD (in press) *An overview of methods to identify and manage uncertainty for modelling problems in the water-environment-agriculture cross-sector*. Springer.

Jakeman, AJ and Letcher, RA 2003, 'Integrated assessment and modelling: features, principles and examples for catchment management', *Environmental Modelling and Software*, 18(6):491 - 501.

Jakeman, AJ, Letcher, RA and Norton, JP 2006, 'Ten iterative steps in development and evaluation of environmental models', *Environmental Modelling and Software*, 21: 602-614.

Jefferson, JL, Gilbert, JM, Constantine, PG and Maxwell, RM 2016, 'Reprint of: Active subspaces for sensitivity analysis and dimension reduction of an integrated hydrologic model', *Computers and Geosciences*, 90: 78-89.

Kelly, RA, Jakeman, AJ, Barreteau, O, Borsuk, ME, ElSawah, S, Hamilton, S and van Delden, H 2013, 'Selecting among five common modelling approaches for integrated environmental assessment and management', *Environmental Modelling and Software*, 47: 159-181.

Kinkeldey, C, MacEachren, AM, Riveiro M & Schiewe J 2017, 'Evaluating the effect of visually represented geodata uncertainty on decision-making: systematic review, lessons learned, and recommendations', *Cartography and Geographic Information Science*, Vol. 44, Iss. 1, 2017.

Kloprogge, P, van der Sliujs, JP and Petersen, AC 2011, 'A method for the analysis of assumptions on in model-based environmental assessments', *Environmental Modelling and Software*, 26: 289-301.

Larssen, T, Huseby, RB, Cosby, BJ, Høst, G, Høgåsen, T and Aldrin, M 2006, 'Forecasting acidification effects using a Bayesian calibration and uncertainty propagation approach', *Environmental Science & Technology*, 40(24): 7841-7847.

Lempert, RJ, Popper, SW, and Bankes, SC 2003, *Shaping the next one hundred years: new methods for quantitative, long-term policy analysis*: Rand Corporation.

Maier, HR, Guillaume, JH, van Delden, H, Riddell, GA, Haasnoot, M, and Kwakkel, JH 2016, 'An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together?', *Environmental Modelling & Software*, 81: 154-164.

Matott, LS, Babendreier, JE and Purucker, ST 2009, 'Evaluating uncertainty in integrated environmental models: A review of concepts and tools', *Water Resources Research*, 45 DOI: 10.1029/2008wr007301.

McVicar TR, Li LingTao, Van Niel TG, Hutchinson MF, Mu XingMin, and Liu ZhiHong 2005, *Spatially Distributing 21 Years of Monthly Hydrometeorological Data in China: Spatio-Temporal Analysis of FAO-56 Crop Reference Evapotranspiration and Pan evaporation in the Context of Climate Change*. CSIRO Technical Report, CSIRO Land and Water, 8/05.

McIntosh, BS, Ascough, JC, Twery, M, Chew, J, Elmahdi, A, Haase, D and Chen, S 2011, 'Environmental decision support systems (EDSS) development—challenges and best practices', *Environmental Modelling and Software*, 26(12): 1389-1402.

McInerny, GJ, Chen, M, Freeman, R, Gavaghan, D, Meyer, M, Rowland, F, Spiegelhalter, DJ, Stefaner, M, Tessarolo, G and Hortal, J 2014, 'Information visualisation for science and policy: engaging users and avoiding bias', *Trends in Ecology & Evolution*, 29, no. 3: 148-157.

Moallemi, E, ElSawah, S, Ryan, MJ 2017, 'Robust Decision Making and Epoch-Era Analysis: Comparing two model-based approaches for decision making under uncertainty', *European Journal of Operations Research* (under review).

Morris, MD 1991, 'Factorial sampling plans for preliminary computational experiments', *Technometrics*, 33(2): 161-174.

- Obenour, DR, Gronewold, AD, Stow, CA and Scavia, D 2014, 'Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts', *Water Resources Research*, 50(10): 7847-7860.
- Pappenberger F and Beven KJ 2006, 'Ignorance is bliss—or 7 reasons not to use uncertainty analysis', *Water Resources Research*, 42(5): W05302. DOI: 10.1029/2005WR004820.
- Pappenberger, F, Harvey, H, Beven, KJ, Hall, J and Meadowcroft, I 2006, 'Decision tree for choosing an uncertainty analysis methodology: a wiki experiment', *Hydrological Processes*, 20:17: 3793-3798. DOI: 10.1002/hyp.6541
- Poole, D and Raftery, AE 2000, 'Inference for deterministic simulation models: the Bayesian melding approach', *Journal of the American Statistical Association*, 95(452): 1244-1255.
- Raats, PAC 1978a, 'Convective transport of solutes by steady flows I. General theory', *Agricultural Water Management*, 1: 201-232.
- Raats, PAC 1978b, 'Convective transport of solutes by steady flows II. Specific flow problems', *Agricultural Water Management*, 1: 219-218.
- Razavi, S, Tolson, BA, and Burn, DH 2012, 'Review of surrogate modeling in water resources', *Water Resources Research*, 48, W07401, DOI:10.1029/2011WR011527.
- Refsgaard, JC and Henriksen, HJ 2004, 'Modelling guidelines –terminology and guiding principles', *Advances in Water Resources*, 27: 71-82.
- Refsgaard, JC, van der Sluijs, JP, Hojberg, AL and Vanrolleghem, PA 2007, 'Uncertainty in the environmental modelling process – a framework and guidance', *Environmental Modelling and Software*, 22: 1543-1556.
- Rietveld, LC, Van der Helm, AWC, Van Schagen, KM, and Van der Aa, LTJ 2010, 'Good modelling practice in drinking water treatment, applied to Weesperkarspel plant of Waternet', *Environmental Modelling and Software*, 25(5): 661-669.
- Robinson, S 2004, *Simulation: the practice of model development and use* (p. 336). Chichester: Wiley.
- Robinson, S 2008, 'Conceptual modelling for simulation Part I: definition and requirements', *Journal of the Operational Research Society* 59(3): 278-290.
- Robinson, S 2008, 'Conceptual modelling for simulation Part II: a framework for conceptual modelling', *Journal of the Operational Research Society*, 59(3): 291-304.
- Robson, BJ 2014, 'When do aquatic systems models provide useful predictions, what is changing, and what is next?', *Environmental Modelling & Software*, 61, pp.287-296.
- Shin, M-J, Guillaume, JHA, Croke, BFW. and Jakeman, AJ 2015, 'A review of foundational methods for checking the structural identifiability of models: results for rainfall-runoff', *J Hydrology*, 510: 1-16.
- Sacha, D, Senaratne, H, Kwon, BC, Ellis, G and Keim, DA 2016, 'The role of uncertainty, awareness, and trust in visual analytics', *IEEE Transactions on Visualization and Computer Graphics*, 22(1): 240-249.

- Saltelli, A and Bolado, R 1998, 'An alternative way to compute Fourier amplitude sensitivity test (fast)', *Comput. Stat. Data Anal.*, 26(4): 445-460.
- Saltelli, A, Chan, K and Scott, E 2004, *Sensitivity Analysis*. New York, Wiley.
- Sargent, RG 2013, 'Verification and validation of simulation models', *Journal of Simulation*, 7(1): 12-24.
- Sobol, IM 1993, 'Sensitivity estimates for nonlinear mathematical models', *Math. Model. Comput. Exp*, 1(4): 407-414.
- Taylor, K, Woodcock, R, Cuddy, S, Thew, P, and Lemon, D 2015, 'A Provenance Maturity Model',. In Denzer, R. et al. (eds) *ISESS 2015*, IFIP AICT 448, pp1-18.
- Trutnevyte, E, Guivarch, C, Lempert, R, and Strachan, N 2016, 'Reinvigorating the scenario technique to expand uncertainty consideration', *Climatic Change*, 135(3): 373-379. DOI: 10.1007/s10584-015-1585-x.
- Teran-Somohano, A, Dayıbaş, O, Yilmaz, L, and Smith, A 2014, 'Toward a model-driven engineering framework for reproducible simulation experiment lifecycle management', In *Proceedings of the 2014 Winter Simulation Conference*, pp. 2726-2737, IEEE Press.
- Uusitalo, L, Lehtikoinen, A, Helle, I And Myrberg, K 2015, 'An overview of methods to evaluate uncertainty of deterministic models in decision support', *Environmental Modelling and Software*, 63: 24-31.
- van Delden, H, van Vliet, J, Rutledge, DT, and Kirkby, MJ 2011, 'Comparison of scale and scaling issues in integrated land-use models for policy support', *Agriculture, Ecosystems and Environment*, 142(1): 18-28.
- Van der Sluijs, J, Janssen, P, Petersen, A, Kloprogge, P, Risbey, J, Tuinstra, W and Ravetz, J 2004, *RIVM/MNP guidance for uncertainty assessment and communication: Tool catalogue for uncertainty assessment*. Utrecht University Retrieved 27th April, 2011, from <http://www.nusap.net/sections.php>.
- van Vliet, J, Bregt, AK, Brown, DG, van Delden, H, Heckbert, S, and Verburg, PH 2016, 'A review of current calibration and validation practices in land-change modeling', *Environmental Modelling and Software*, 82: 174-182.
- Venables, WN and Dichmont, CM 2004, 'GLMs, GAMs, and GLMMs: an overview of theory for applications in fisheries research', *Fisheries Research*, 70: 319-337.
- Voinov, A and Bousquet, F 2010, 'Modelling with stakeholders', *Environmental Modelling and Software*, 25: 1268-1281.
- Walker, WE, Harremoes, P, Rotmans, J, van der Sluijs, JP, van Asselt, MBA, Janssen, P and Krayen von Kraus, MP 2003, 'Defining uncertainty: a conceptual basis for uncertainty management in model based decision support', *Integrated Assessment*, 4(1): 5-17.
- Walker, W, Haasnoot, M, and Kwakkel, J 2013, 'Adapt or perish: A review of planning approaches for adaptation under deep uncertainty', *Sustainability*, 5: 955.

Watson, AA, and Kasprzyk, JR 2017,. 'Incorporating deeply uncertain factors into the many objective search process', *Environmental Modelling and Software*, 89: 159-171. DOI: <http://dx.doi.org/10.1016/j.envsoft.2016.12.001>

Warmink, JJ, Janssen, JAEB, Booij, MJ and Krol, MS 2010, 'Identification and classification of uncertainties in the application of environmental models', *Environmental Modelling and Software*, 25(12): 1518-1527.

Webb, JA, Stewardson, MJ and Koster, WM 2010, 'Detecting ecological responses to flow variation using Bayesian hierarchical models', *Freshwater Biology*, 55(1): 108-126.

Wengert, R.E. (1964) 'A simple automatic derivative evaluation program', *Commun. ACM*, 9(4): 463-464.

Yilmaz, L, Chakladar, S and Doud, K 2016, 'The goal-hypothesis-experiment framework: a generative cognitive domain architecture for simulation experiment management', *Proceedings of the 2016 Winter Simulation Conference*, Roeder et al. (eds), pp.1001-1012. IEEE Press.