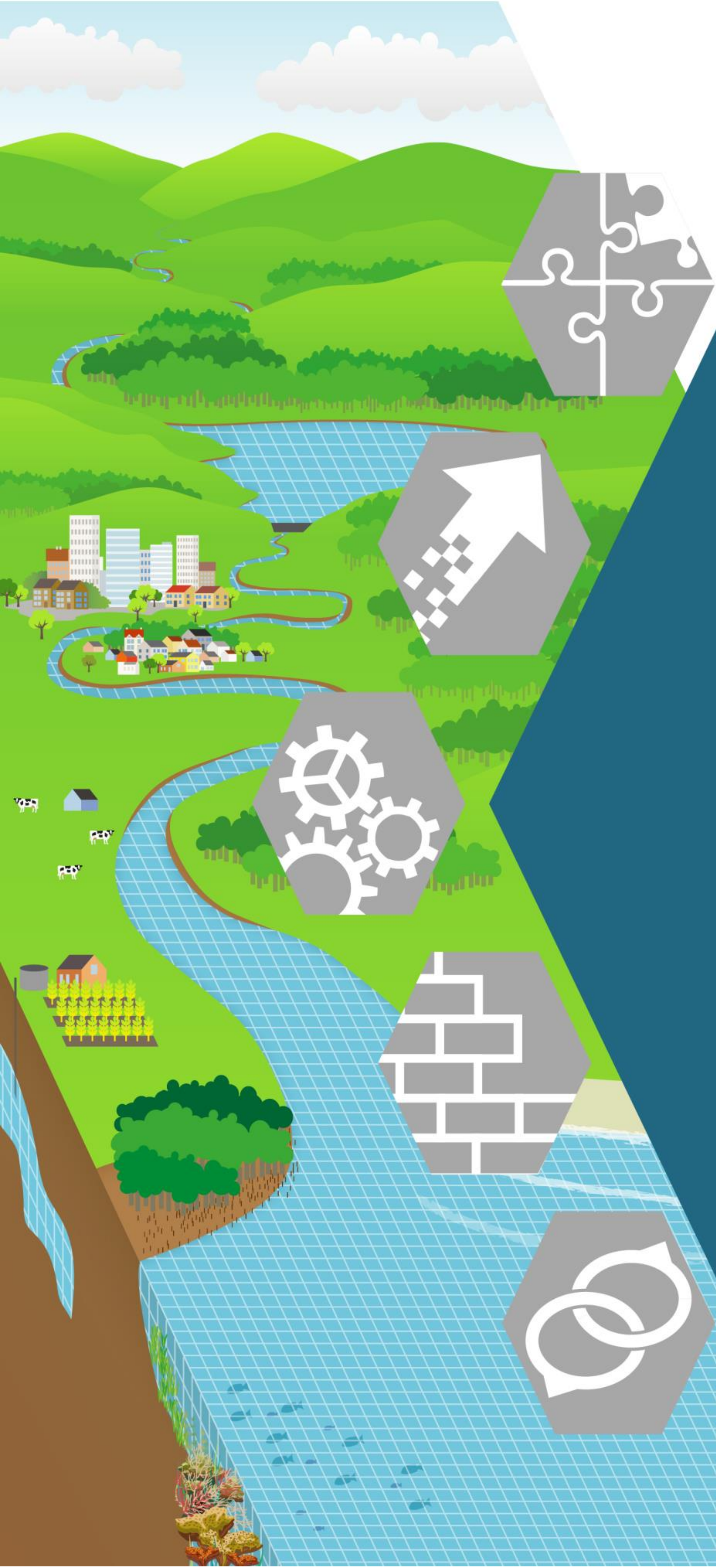
Queensland Government

Australian National University



Queensland Water Modelling Network

Good Modelling Practice

A discussion paper

August 2017

Good Modelling Practice

A discussion paper

Draft Version

22 August 2017

Prepared by

A.J. Jakeman, Australian National University

S. Elsawah, University of New South Wales, Canberra

S. Cuddy, CSIRO Land and Water

B.J. Robson, CSIRO Land and Water

N. McIntyre, The University of Queensland

F.J. Cook, Freeman Cook and Associates

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

August 2017

Table of Contents

1. Introduction

Context

The Department of Science, Information Technology and Innovation (DSITI) has been allocated funding to help address the critical strategic gaps and weaknesses in water models created through the long-term focus on operational issues driven by resource constraints. The objectives are to develop greater capacity and collaborations through engagement with universities, scientific providers and external consultants to position the government for modelling requirements for the future. The four-year funding is to improve the integration of hydrology, groundwater and water quality models State-wide (not just Reef) and across the different scales (paddock, catchment, estuary and marine). The 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 that context, the purpose of this discussion paper (referred to as the paper) is to synthesise existing knowledge and experience on good modelling practices and principles. The paper was developed based on the findings gathered from literature review followed by an expert workshop. The workshop was conducted on June 21-22, 2017, and covered the following:

- Capturing and synthesising the foundations and principles of 'best practice' modelling procedures and model management
- Reflecting on current approaches to modelling and model management
- Developing complementary presentations on R&D modelling principles targeted at modellers and for policy makers
- Recommending 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)

Scope and focus

The scope considered in this paper is with respect to the use of water resource models to investigate impacts on the environmental system in question (e.g. paddock, catchment, reef), both under status quo conditions and in response to management alternatives, climate variations and other uncontrollable forces; as well as to adaptively manage the system such as through additional monitoring and informative studies. The paper's focus is set 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 they aim to predict as outputs one, or usually more, *Quantities of Interest* (*QoIs*) such as some index of water quantity, quality or ecological response as a function of time and/or space. Inevitably the nature and role of these models evolve over time to cope with changes in the policy context and scientific knowledge. The paper covers modelling practices at the different stages of the model development and use lifecycle. The lens through which practices are examined here takes into account the whole modelling process and the sources of uncertainty to be recognised and managed in that process.

Paper organization

The paper is organized as follows: in Section 2, we present an introduction to the concept of good practice modelling, and give an overview of research geared towards identifying those practices. It includes a discussion of uncertainty types and issues in managing them holistically. In Section 3, we present good modelling practices throughout the modelling process with an emphasis on where uncertainties arise and considerations in dealing with them. We wrap up with concluding remarks and recommendations in Section 4.

2. Best Practices

In this section, we define what we mean by best practice modelling. Next, we present views on two essential and complementary components of the modelling process: model development and uncertainty assessment. These views together provide the basis and rationale (i.e. how and why) for identifying the modelling activities, and therefore best modelling practices.

Defining 'best practice'

The quality and outcomes of a modelling process largely depend on the modelling practices that are undertaken at every step. Building quality (i.e. relevance, credibility and validity) into the modelling (process and outputs) is now very much emphasized in the literature. The search for ways to improve the way modelling is conducted is not new. Several attempts have been made to investigate and identify those practices (referred to as best/good/core). Best practices should be proven-to-work practices for managing common problems encountered throughout the modelling process. Identifying best practices helps to provide guidelines for improved modelling practice. Such improvements will ultimately lead to more accurate, credible and useful models, more insightful model-based recommendations, better-informed model adoption, and more importantly improved decision-making.

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 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 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. The aim is to develop purposeful, credible models from data and prior knowledge, in consort with end-users, with every stage open to critical review and revision. Best practice entails identifying clearly the clients and objectives of the modelling exercise; documenting the nature (quantity, quality, limitations) of the data used to construct and test the model; providing a strong rationale for the choice of model family and features (encompassing review of alternative approaches); justifying the techniques used to calibrate the model; serious analysis, testing and discussion of model performance; and making a resultant statement of model assumptions, utility, accuracy, limitations, and scope for improvement. In natural resource management applications, these steps will be a learning process, even a partnership, between model developers, clients and other interested parties."

Table 1 provides a (non-exhaustive) list of research geared towards developing guidance into good/core/best modelling practices.

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

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

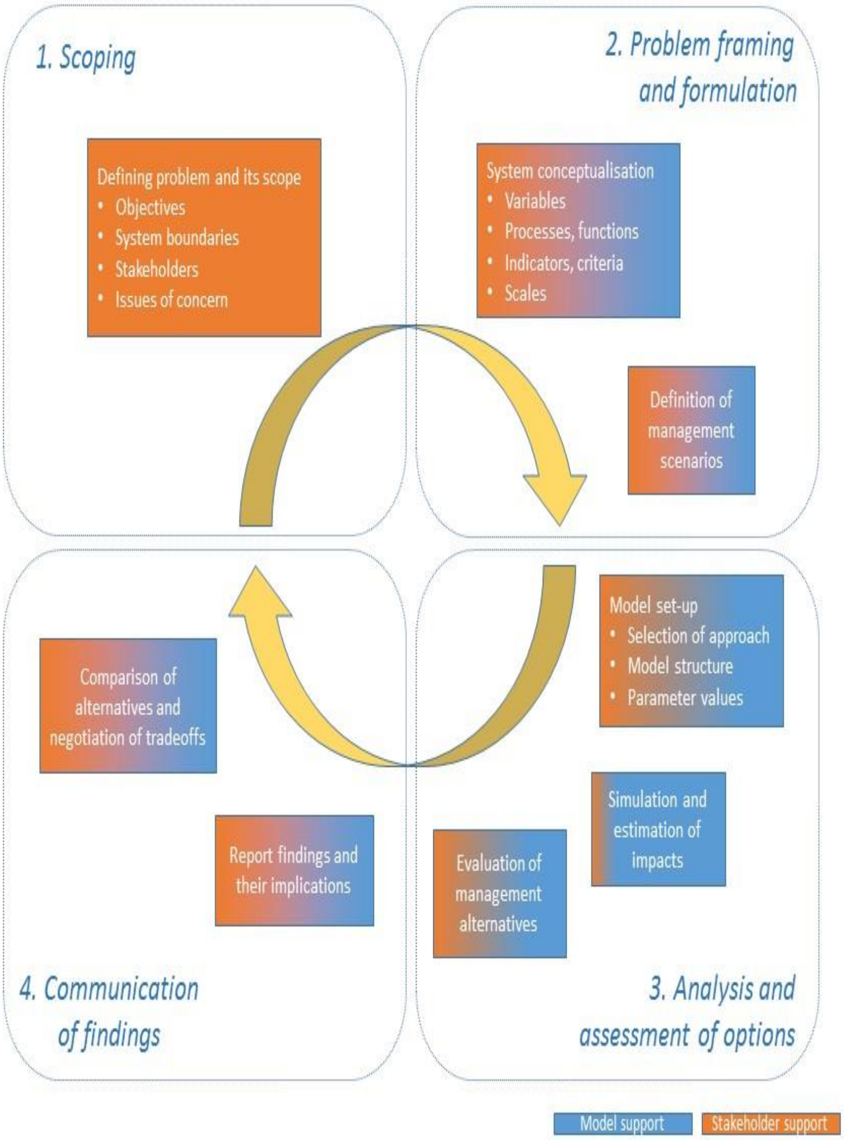
The modelling process

From a development perspective, there are essentially four phases in the modelling and assessment process (Hamilton et al., 2015) and these are reproduced in Figure 1. The phases tend to be iterative and can be described as

- Scoping (Model study plan including identifying model purpose and study objectives)
- Problem framing and formulation (including conceptualisation)
- Analysis and assessment of options (Model Setup, and Calibration and Validation) and
- Communicating of findings (Simulation and Evaluation).

Each phase has several steps as indicated in Jakeman et al.'s (2006) ten steps and in Figure 1 -- see Hamilton et al. (2015). Refsgaard et al. (2007) provide a very similar delineation of the modelling process as indicated by the terms in brackets in the phases above. The main difference is that Hamilton et al. combine the most technical aspects of the process, i.e. model setup, calibration and validation, into one phase. Most of these steps need expert and/or stakeholder engagement, for which there is now much guidance regarding the why and how (e.g. Voinov and Bousquet, 2010). The two colours in Figure 1 indicate notionally the proportional emphasis on stakeholder engagement (versus technical model support) usually needed in each phase.

Figure 1: Phases and steps in the integrated modelling and assessment process



Uncertainty concepts and its management

Being approximations of the real system of interest, models only represent our partial knowledge and views about that system. And because the real system is far more complex to understand and capture in a model representation, uncertainty is inevitably associated with use of the model, and indeed it arises in various ways throughout the modelling process. Uncertainty management is now recognised as an essential part of the modelling process and has quantitative and qualitative aspects aimed at establishing as far as necessary what we know and do not know in terms of predictions (*QoIs*) required for the problem of interest. Therefore the various steps of the modelling process must be paid interdependent attention in order that the addressing of uncertainty sources be complete. We refer to this more holistic attention here as *Uncertainty Assessment* (UA).

Traditionally, attention has focussed predominantly on the quantitative aspects of UA, often known as *Uncertainty Quantification* (UQ). In many situations UQ has been applied almost as an afterthought once the model has been built. Increasingly, guidelines for modelling consider quantitative and qualitative aspects as being complementary and to be addressed throughout the modelling phases. The term *Uncertainty Management* (UM) can be thought of as combining UA (which may well include UQ) with broader aspects of the identification, prioritising, reduction, propagation and communication of uncertainties (see later in this section). Inevitably this should always have a qualitative component because some assumptions cannot be quantified but must be recognised for their relative effect on *QoI* predictions. Thus good modelling practice will attempt to list assumptions and choices made in the modelling process and characterise their effect on those *QoIs*. Qualitative efforts to deal with uncertainty would ideally also include validating the modelling process. For example, Kloprogge et al. (2011) use a so-called pedigree approach to value-ladenness. Van der Sluijs et al. (2004; 2005) combine qualitative and quantitative measures of uncertainty assessment in their so-called NUSAP system.

Several authors have developed typologies of uncertainty in the modelling process. Walker et al. (2003) refer to the nature of uncertainty as either *epistemic* (due to imperfect knowledge) or *stochastic/aleatory* (due to inherent variability that needs to be characterized for its effect on predictions). Epistemic uncertainty is notionally reducible from further studies or data collection. The authors also categorise uncertainty by level and source. Another differentiation is between *ontological* uncertainty and *semantic* uncertainty. Fox (2008) regards ontological uncertainty as that due to participants in a process having different conceptualisations of the system to be modelled while semantic uncertainty derives from participants giving different meanings to the same concepts or terminology. Ontological and semantic uncertainty are best handled through stakeholder engagement activities.

Guillaume et al. (2011, 2012) take this further by proposing an *Uncertainty Management Framework* (see Table 1 which uses the examples of uncertainties ion data, model structure and parameters to suggest methods for dealing with those uncertainties). The framework categorises methods for dealing with uncertainty by source and task according to seven iterative steps.

- identifying the uncertainties
- prioritising resources to address them
- reducing the uncertainties critical to the problem purpose
- describing the uncertainties
- propagating them through the model
- communicating uncertainty to model users and clients, and
- anticipating and managing residual uncertainty.

The rationale for such a framework is that, when considering the different sources of uncertainty (as they arise throughout the steps of the modelling process) and when integrating several different model or system components, it may not be efficient to invest considerable effort and resources toward reducing uncertainty in one source if the results are dominated by uncertainty in another. Thus all sources of uncertainties need to be considered for their criticality in addressing objectives. In other words, to be effective and efficient, uncertainty management should be prioritised toward uncertainties that are most relevant to the task. In the next section we indicate where uncertainties occur in the modelling process and the choices and considerations that must be made to realise the benefits of good practice. Refer to Jakeman and Jakeman (in press) for a more explicit list and technical discussion of the sources of uncertainties as well as new methodological opportunities for addressing them.

Table 2: Categorisation of uncertainty methods by task and source. Categories identified by Matott et al. (2009) are in bold, while qualitative approaches are in italics.

Task	**Modelled outcomes**	**Data**	**Model Structure**	**Model parameters**
Identify	*Expert elicitation, stakeholder methods*	*Quality Assurance* NUSAP	*Identifying assumptions*	(clear from structure)
Prioritise	*Expert opinion*	**Sensitivity analysis**	*Expert opinion*	**Sensitivity analysis, Identifiability analysis**
Reduce	(reducing uncertainty at their source)	Data acquisition planning	*Model verification*	**Parameter estimation, Bayesian Networks**
Describe	Model validation. *Extended peer review*	**Data analysis**, NUSAP (Van der Sluijs *et al.* 2004), DUE (Brown and Heuvelink 2007)	Info-gap theory (Ben-Haim 2006)	**Parameter estimation** GLUE (Beven and Freer 2001), BATEA (Thyer *et al.* 2009)
Propagate	Methods for combined uncertainties. e.g. Meta-models	**Uncertainty analysis, Bayesian networks**	**Multimodel analysis,** Exploratory modelling and analysis (EMA)	**Uncertainty analysis**, **Bayesian networks**, EMA, Error propagation equations
Communicate	Confidence intervals, risk, consequences	Consequences for model	Listing of assumptions and limitations	Consequences for model
Manage		*In modelling, addressed in communication with stakeholders*		

3. Achieving best practice: considerations

Model purpose and objectives

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, and managing client's 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. These considerations need to be clearly communicated to the client, along with setting clear agreements on the format at which model results will be delivered (e.g. reports, raw computer runs, analysis results) and ways to communicate about uncertainties of results (e.g. probability distribution functions, ranges, qualitative or categorical descriptions) and their visualizations and tabulations. In some situations, the client may have a very clear understanding of the modelling objectives. Otherwise, the modeller(s) and the client need to work closely to formulate those requirements. Of course 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.

Actors: modeller, client and other stakeholders

In essence, model development and use is a social communication process, throughout the steps, with stakeholders to build confidence and trust. Successful management of this process is as important as the technical model development aspects because major uncertainties can emanate from such basic aspects as working on a poorly formulated problem, neglecting to include valuable knowledge and perspectives from key interest groups and experts, or poor communication in general. However, this process can be challenging, especially in a multi-agency context, with multiple intended 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, with good soft skills (i.e. communication and interpersonal skills). Some of the useful principles to put into practise are:

- Engage with stakeholders from the very early stages of the modelling process. This includes explicitly accounting and planning for the time and resources required
- Ascertain and communicate the model's value in the problem context, 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 peddling 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 communicate language (avoiding technical and academic jargon), filter and synthesise large amounts of information to communicate the most useful insights, and understand people's cognitive biases and build their understanding step-by-step
- 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.

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 are to enable more rapid model development and use but they 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 (i.e. from a modeller's perspective, the model produces sufficiently accurate results for the purpose at hand) and credibility (same as validity from a client and user perspective, and potentially from a 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 Elsawah, 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 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 involved key stakeholders appropriately?
- Is the process of producing the conceptual model sufficiently credible, for example involved 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 as such then it is provisionally accepted as a possible model and new observations are recommended to further test the model. This scientific viewpoint is different from the more engineering-orientated 'fitness-for-purpose' viewpoint, whereby the model may be accepted as being fit for purpose if it can predict the *QoIs* of interest; however both require the recognition that the model will 'fail' if subjected to a wider range of stress tests and that the model is only one of potentially many acceptable hypotheses.

Data collection, cleansing and preliminary 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 when using the model in a predictive or simulation mode. Appropriate infilling of missing data depends on circumstance (simple interpolation is often inappropriate, for instance between a sample taken during a 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.

There is still much however that can be done to improve such situations. Simple text-book analysis 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 visualization should not be ignored.

There is also much to be gained from more attention to the optimal design of experiments for data collection in the future. 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, however, provide the same amount of information about the processes they are helping inform. Consequently, it is important to design experiments in an optimal way, i.e., to choose some limited number of experimental data to maximize the value of each experiment. Optimal experimental design (OED), that is using physical models to guide experiment selection, 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). When model observables are linear with respect to the model parameters, alphabetic optimality criteria are often used.

Model selection

Models can be categorized in different ways (Kelly et al., 2013; Balci, 2007), including:

- 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)
- 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 the right 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 (physical, chemical and biological) mechanisms of change. 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 (e.g. an Arrhenious equation (Goldman, 1979)). In reality, some rates, such as phytoplankton growth rates, will decline above some optimum temperature (typically around 30°C; e.g. 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 discretization 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 in reducing the numerical error. The other pitfall to recognise and address is that available data do not match the temporal and spatial resolution of the model. This can mean changing the model resolution, implementing methods to interpolate the data, and/or acknowledging the influence of granularity chosen of *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 that at high spatial resolution. It may be valuable to seek available knowledge, including that of 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 consideration 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.

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 (particularly the *QoIs*) to a change in input or parameter. The purpose of SA, often used as a step prior to model calibration, is understanding and quantifying: (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. Thus it 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, for example to revising or indeed simplifying the conceptual model.

SA can also direct additional measurement efforts, whether to improve the prior information used to inform specification of sensitive parameters, to 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.

Some commonly used SA techniques include: local sensitivity methods; variance based techniques; and regional sensitivity analysis. Local SA methods, such as automatic differentiation (Wengert, 1964) and the Morris method (1991), characterize sensitivity by partial derivatives or gradients at the local point. These methods are generally very simple and easy to implement and work well for linear models. However, 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) method involve decomposing the output variance into parts attributed to individual variables and interactions between variables. Regional Sensitivity Analysis (RSA) (Hornberger and Spear, 1981) partitions model realizations into behavioural sets and non-behavioural sets; that is the set of input factors that satisfy the problem constraints and those that do not.

Recently a new technique known as active-subspaces has become popular for identifying lower-dimensional structure. Unlike the aforementioned methods, active subspaces can 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-*QoI* map (Jefferson et al., 2016). Related to sensitivity analysis is break-even analysis. It 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 (Guillaume et al., 2016).

Model emulation (also known as surrogate or meta-modelling) is the practice of developing a simpler (usually statistical) model that is fitted to and approximates the outputs of a more complex model. 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. Also see Castelletti et al. (2012) and Razavi et al. (2012) for reviews and examples of relevant emulation applications. Moreover, model emulation methods promote computational efficiencies by replacing models with slow runtimes, as often occurs with integrated, multi-component models, and mesh-based physical representations such as groundwater and hydrodynamic models.

Calibration and model structure

Parameters may be either calibrated from data, and/or specified from prior knowledge (such as may be assumed 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 parameterization can have significant effect on whether a model can reproduce experimental observations. This is particularly true for parameterization of spatially or temporally varying fields such as conductivity. The complexity of the parameterization of conductivity can range from a single parameter for a homogeneous aquifer, to multiple parameters for a regional conductivity, 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 overfitting 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, then the purpose of the calibration becomes a decision-gate at which the modelling process loops back to one of the earlier stages. 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 *QoIs*. While this principle seems straightforward and certainly obvious, its practice is weak. In surface water hydrology for example there is undue attention to a measure of mean squared error, known as Nash-Sutcliffe efficiency, which places most focus on fitting high flows. Careful attention to what are the precise objective functyions (i.e. error functions of the *QoIs*) is a sure way to reduce uncertainties that would otherwise be manifested.

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 optimizing the estimation of parameters. In particular, it is some *function*(s) of the quantities of predictive interest (*QoIs*) that must be deliberated and specified, requiring knowledge of the natural and human setting, which is often best realized through an appropriate participatory process (Hamilton et al. 2015). An example in hydrology of a more exacting purpose (than say prediction of quantity fluxes) would be where surface and/or groundwater modelling need to predict *QoIs* that relate to ecological needs. Hence the *QoIs* in that case could relate to surface and groundwater levels; but it might be a function of those that is of 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*. And experts might further confirm how accurate in either quantitative or categorical terms the associated prediction of the targets need be for the modelling to be useful. But sometimes a failing can be just that the modeller does not relate the aims of the modelling to either the objective functions used to optimize model parameters or relevant performance measures (Bennett et al., 2013). As an 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. But 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 (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 minimize. In general, it is a good practice to examine the effects of different/multiple objective functions, and to perform sensitivity analysis (for uncertain input and parameters as well as presumed certain parameters).

Selection of calibration periods, and 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 against average 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 different behaviours and parameter values, substantially so when the region of application suffers strong climate variability. Wherever possible one should calibrate a model on different periods and assess the performance of each on all other periods. This so-called cross-validation (see next subsection) 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 how large is the dataset. In some circumstances, the hydrologist for instance 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 and Dichmont, 2004) the objective function and the weighting given to the data in different ranges needs to be carefully considered. Another way is to bin 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.

A common glaring deficiency is the omission of a cross-correlation analysis between model residuals (predictions minus corresponding observations) and model inputs to assess if there seems to be something missing in the model's explanation of outputs. Verification and indeed 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.

Prior knowledge may be used to constrain parameters in the formulated model structure. Inappropriate constraints may underestimate or overestimate uncertainty such as the way priors are 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 *QoIs* 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 with time from a sink (river), where management interventions will be most effective and the fact that it could take millennia before the consequences of interventions have effect. Similarly, for interflow and surface flow, analytical models (Cook et al. 2009; Cook etal. 2011) can offer considerable insight when assessing model output.

Importantly, calibration can be defined to include identification of model structure, inputs and boundary conditions, not just 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 levels of spatial and temporal discretization.

Formal statistical tests for differentiating among different model structures are well developed. They provide criteria which trade the number of parameters against the improvement in model fit to observations. Because of their reliance on statistical assumptions, they are best treated as guides, checking 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 against complexity of representation. A key question not often asked is whether some system descriptors, for instance dimensionality, discretization and processes, can be aggregated to make the aggregation more efficient, worrying only about what dominates the system response indicators at the scales of concern. Allowing more degrees of freedom than warranted in system representation can lead to overfitting (to errors) and unrealistic model behaviours and predictions.

Working with multiple models are 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 instance, 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, for instance, or does not allow for resuspension of diatoms settled to the sediments).

In process-based models, 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. It can be used in a variety of ways, for example:

- Informing Bayesian parameter estimation or uncertainty quantification approaches;
- Setting appropriate initial values and value bounds for optimisation schemes;
- 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.

Whereas there are solutions to support the automated development of model calibration, which can improve the efficiency of the process, some care needs to be taken in using them. Blind 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.

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." Of course validation must be considered through the lens of uncertainty. There are several ways that validation can be approached and a combination of methods is typically appropriate.

Crash or stress-testing the model is an obvious 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 the 'sensibility testing' discussed earlier.

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.

With water quality models the drivers for the transport of the solutes and particulates is the velocity of the flow (advection) and dispersion. Whereas the drivers for water flow are due to potential energy or pressure head differences. This means that calibrating a model for water flow does not necessarily mean that this will 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 bounds. For example, in the case of streamflow, we can consider the upper bound to be the rainfall, i.e. all the water runs off and appears as streamflow during an event. This means that the rainfall times 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 - potential evapotranspiration) times 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 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 m 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 (e.g. Borsuk et al., 2001; Webb et al., 2010; 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 (about expected parameter values as well as confidence in observational data used to calibrate the model) to be taken into account and it 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

Scenario analysis

In its broadest sense, scenario analysis involves exploring multiple, plausible assumptions about future conditions, 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 those 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 data sets with models provides three significant advantages: 1) it facilitates 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 social-environmental model); 2) it facilitates cross-comparison of results between models and ensemble scenario analysis; and 3) it facilitates comparison between scenario predictions from an existing model and from proposed new versions of the same model.

Rather than attempting to develop a priori a minimal sufficient set of scenarios for stakeholders to contemplate, an alternative approach to scenario development is to utilize a model to simulate a large space of possible futures and then allow stakeholders a posteriori to visualize the entire future space and to articulate preferences. This approach is known as exploratory modelling, and is attracting growing attention in the scenario analysis literature (Bankes, 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 (i.e. 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 & Lempert, 2010), Dynamic Adaptive Policy Pathways (Haasnoot et al., 2013; Kwakkel et al., 2016b), and Objective Robust Decision Making (MORDM) (Watson & Kasprzyk, 2017). In principle, these techniques share the idea of open exploration and searching for robust solutions. However, technically, they vary in how the scenario generation process is conducted, and the type of insights to be generated (Moallemi et al., 2017). There is limited understanding of the fundamental differences between these techniques, their relative strengths and limitations, and implications into how uncertainty is treated and solutions identified (Haasnoot et al., 2013, Trutnevyte et al., 2016). Comparative and evaluation studies to investigate differences and complementariness are still needed. To support practice, research into good practices for conducting exploratory modelling is also needed.

Communication

Selecting indicators to communicate about model results

One crucial issue on communication is selecting the appropriate set of indicators to report the modelling results. At a more fundamental level, indicators reflect the objectives/values incorporated in the model. Indicators vary according to multiple aspects, including: level (whole system vs sub-system), purpose (i.e. communicate about average performance versus variability in performance, communicate about snapshot vs pathway), type (e.g. absolute value vs proportional, descriptive vs normative such as difference between hypothesized best value and the calculated value), as well as 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 has full visibility of the effects of different decision options 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. Note that the number of indicators presented to decision makers must be managed. Balancing succinctness and informativeness is desirable. Thus we should be educating decision makers of the need to go beyond single numbers to indicate uncertainty but also realize that we can be too complex and the amount of data overwhelming.

Communicating uncertainty in written reports

The language used in science-policy reports is often very measured and calibrated (McInerny, 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 play on communicating 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.

Visualization

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

A key challenge in model visualization is communication of large datasets, especially in problems with multiple objectives and trade-off solutions. Due to the curse of dimensionality, traditional visualization (e.g. the scatter plot) is no longer an appropriate tool in visualization of a high dimensional objective space. He and Yen (2017) identified three criteria for high quality visualization of high-dimensional multi-objective space. First, it should give accurate information of the Pareto front. Second, it should provide decision makers with a clear indication of trade-off solutions. Third, the tool must be scalable to higher dimensions and larger datasets. The authors reviewed the available approaches, and evaluated their performance on 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 relates to uncertainty communication, especially when incorporating spatio-temporal heterogeneity. 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). See Bonneau et al. (2014) for a review of methods for uncertainty visualization, and Kinkeldey et al. (2017) for a review of effectiveness of some of the methods.

An important concern in developing and using visualization is understanding and mitigating the possible biases in audience's interpretations, which may ultimately lead to over or less confidence in the results (McInerny, et al., 2014; Sacha et al., 2016). For example, rescaling results through visualization 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.

Data and workflow management

Scientific workflows

The limited use of well-thought and transparent experimental design inhibits reproducibility of results, effective reporting of results, and therefore the credibility of models (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). Tools, such as Truii, are readily available for modellers.

Provenance, governance, and meta-data

Management of input, intermediate and output data is one of the more difficult aspects of modelling – 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; how to ensure that the modeller knows what data they are using. Governance of model data requires implementation of strong, internal QA/QC procedures that respect in-house work culture while improving practice. Management of observed data within a specialized 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 low 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.

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, allowing for use and reuse of data and models between states and partners. Automation of parts of the data management workflow can improve its governance – however this requires investment of time and resources, and clear sight of the benefits to offset perceived dis-benefits (e.g. loss of transparency).

An effective data management program requires a strategic investment of effort, with the vision, and the steps to achieve it, clearly articulated and shared with users and practitioners. The goal would be 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. The “gold-standard” is not necessarily appropriate.

4. Concluding remarks and recommendations

Uncertainty Assessment is increasingly being seen as a holistic process that should be a major consideration throughout the whole lifecycle of the modelling process. It should be viewed as standard modelling practice in the water sector, enabled by its comprehensive and central attention in project specifications documents and workflows.

To be complete, uncertainty assessment and management in the water sector will almost always necessitate the use of qualitative assessments, with quantitative assessments wherever possible. The qualitative aspects of uncertainty management include close attention to the problem definition step (in the first instance), going beyond the research questions to: specifying the specific quantities of interest, conceptual modelling to relate cause and effect appropriately, and explicit consideration of the effects of all assumptions and limitations (modelling choices and other sources of uncertainty) on predictions.

Predictive uncertainties should be evaluated, qualitatively and quantitatively, in terms of the effects on precisely specified and argued functions of the quantities of (decision) interest, i.e. including the desirable spatio-temporal function of those quantities (e.g. a mean value, set of moments, probability distribution, some property or pattern of a time series).

Recommendations that warrant specific attention in the water modelling domain are the following:

- Characterise sources, and try to rank the criticality of, uncertainties through such means as expert elicitation, stakeholder engagement, sensitivity and more formal uncertainty analyses.
- Emphasize effective simplification over undiscerning and unnecessary model complexity, especially where complexity reduces transparency, increases uncertainty and/or hinders its assessment.
- Educate users of model results about the dangers of being provided only a single number upon which to base decisions - but also address their needs, by providing uncertainty information in a format that fits within their workflows.
- Factor in the appropriate costs of holistic uncertainty assessment in project budgeting. It will be worth it in the longer term.
- Communicating uncertainty is an area of emerging attention that could be advanced through focussing on meeting its challenges in the water sector. Visualization of indicators of concern are an aspect in such an endeavour. Their design should pay special attention to possible interpretation biases and ways to control them.
- Pay explicit attention to the way model results and uncertainty are communicated in written reports and publications.
- Make effective use of user-centred design for visualization development early in the modelling process, and leverage different visualization tools to engage different audiences (e.g. academics, 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.

- One of the old mantras re measurement bears repeating, that is pay careful attention to the collected data, including measuring the right variables, at the right locations and with the right frequency.

5. References

Alcamo, Joseph. 2001. Scenarios as Tools for International Environmental Assessments. Environmental Issue Report 24. Luxembourg: Office for Official Publications of the European Communities.

Argent, R. M., Sojda, R. S., Guipponi, C., McIntosh, B., Voinov, A. A., and Maier, H. R. (2016) Best practices for conceptual modelling in environmental planning and management. Environmental Modelling and Software, 80: 113-121.

Atkinson, A.C. and Donev, A.N. (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): 435-449.

Bauler, T. (2012) An analytical framework to discuss the usability of (environmental) indicators for policy. Ecological Indicators, 17: 38-45.

Bennett N.D., Croke B.F.W., Guariso G., Guillaume J.H.A., Jakeman A.J., Newham, L.T.H., Norton J.P., Perrin, C., Pierce, S.A., Robson B., Seppelt R., Voinov, A.A., Fath, B. and Andreassian, V. (2013) Characterising Performance of Environmental Models. Environmental Modelling and Software, 40: 1-20

Beven, K. and J. Freer (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology, 249(1-4): 11-29.

Black, D., Wallbrink, P., Jordan, P., Waters, D., Carroll, C. and Blackmore, J. (2011) Guidelines for water management Modelling: towards best practice model application. eWater Cooperative Research Centre, ISBN 978-1-921543-46-3.

Black, D.C., Wallbrink, P.J. and Jordan, P.W. (2014) Towards best practice implementation and application of models for analysis of water resource management scenarios. Environmental Modelling and Software, 52:136-148.

Bryant, B. P., & Lempert, R. J. (2010) Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. Technological Forecasting and Social Change, 77(1): 34-49.

Bonneau, G.-P., Hege, H.-C., Johnson, C.R., Oliveira, M.M., Potter, K., Rheingans, P. and 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, pages 3-27. Springer London, London.

Borsuk M.E., Higdon D., Stow C.A., Reckhow K.H. (2001) A Bayesian hierarchical model to predict benthic oxygen demand from organic matter loading in estuaries and coastal zones. Ecological Modelling, 143(3): 165-81.

Brown, J.D. and Heuvelink G.B.M. (2007) The Data Uncertainty Engine (DUE): A software tool for assessing and simulating uncertain environmental variables. Computers & Geosciences, 33(2): 172-190.

Castelletti, A., Galelli, S., Ratto, M., Soncini-Sessa, R. and Young, P.C. (2012) A general framework for dynamic emulation modelling in environmental problems. Environmental Modelling and Software, 34: 5–8.

Chakladar, S. (2016) A model driven engineering framework for simulation experiment management. PhD diss., Auburn University.

Chen, S. H., and Pollino, C. A. (2012) Good practice in Bayesian network modelling. Environmental Modelling and Software, 37: 134-145.

Coles, J.F. and Jones, R.C. (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): 7-16.

Cook, F.J. and Jayawardane, N.S. (2008) Modelling of water and solute transport at Yanggao County, Shanxi, Peoples Republic of China. Unpublished.

Cook, F.J. Knight, J.H. and Wooding, R.A. (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: 357-372.

Cook, F.J., Neumann, L.N., Siriwardena, L. and Western, A.W. (2011) Does where you plant trees make a difference in hydrologic response. In Chan, F., Marinova, D. and Anderssen, R.S. (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, C. E., McIntyre, N., Jackson, B. M. and Wheater, H. S. (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. and Scarpa, R. (2017) A note on communicating environmental change for non-market valuation. Ecological Indicators, 72: 165-172.

Gaber, N., Foley, G., Pascual, P., Stiber, N., Sunderland, E., Cope, B., Nold, A. and Saleem. Z. (2009) Guidance on the Development, Evaluation, and Application of Environmental Models (EPA/100/K-09/003). https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P1003E4R.PDF

Goldman, J. C. (1979) Temperature effects on steady-state growth, phosphate uptake, and the chemical composition of a marine phytoplankter. Microbial Ecol., 5: 153-166.

Groves, D. G., and Lempert, R. J. (2007) A new analytic method for finding policy-relevant scenarios. Global Environmental Change, 17(1): 73-85.

Guillaume, J.H.A., Croke, B.F.W, El Sawah, S. and Jakeman, A.J. (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, J.H.A., Arshad, M., Jakeman, A.J., 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, J. H., 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, J.H.A., Arshad, M., Jakeman, A.J., 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, J.H.A., Helgeson, C., ElSawah, S., Jakeman, A. J. 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, J. H., Walker, W. E. and Maat, J. T. (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, J.H.A. and Jakeman, A.J. (2015) Integrated assessment and modelling: a review and synthesis of salient dimensions. Environmental Modelling and Software, 64: 215-229.

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

Hornberger. G.M. and Spear, R.C. (1981) An approach to the preliminary analysis of environmental systems. J. Environ. Management, 12: 8-18.

Horsburgh, J. S., Tarboton, D. G., Hooper, R. P. and Zaslavsky, I. (2014) Managing a community shared vocabulary for hydrologic observations. Environmental Modelling and Software, 52: 62-73.

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

Jakeman, A.J. and Letcher, R.A. (2003) Integrated assessment and modelling: features, principles and examples for catchment management. Environmental Modelling and Software, 18(6):491 - 501.

Jakeman, A.J., Letcher, R.A. and Norton, J.P. (2006) Ten iterative steps in development and evaluation of environmental models. Environmental Modelling and Software, 21: 602-614.

Jefferson, J.L., Gilbert, J.M., Constantine, P.G. and Maxwell, R.M. (2016). Reprint of: Active subspaces for sensitivity analysis and dimension reduction of an integrated hydrologic model. Computers and Geosciences, 90: 78-89.

Kelly, R. A., Jakeman, A. J., Barreteau, O., Borsuk, M. E., 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, J.P. and Petersen, A.C. (2011) A method for the analysis of assumptions on in model-based environmental assessments. Environmental Modelling and Software, 26: 289-301.

Larssen, T., Huseby, R.B., Cosby, B.J., 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, R. J., Popper, S. W., and Bankes, S. C. (2003) Shaping the next one hundred years: new methods for quantitative, long-term policy analysis: Rand Corporation.

Maier, H. R., Guillaume, J. H., van Delden, H., Riddell, G. A., Haasnoot, M., and Kwakkel, J. H. (2016) An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together? Environmental Modelling & Software, 81: 154-164.

Matott, L.S., Babendreier, J.E. and Purucker, S.T. (2009) Evaluating uncertainty in integrated environmental models: A review of concepts and tools. Water Resources Research, 45 DOI: 10.1029/2008wr007301.

McVicar T. R., Li LingTao, Van Niel T. G., Hutchinson M. F., 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, B. S., Ascough, J. C., 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, G J., Chen, M., Freeman, R., Gavaghan, D., Meyer, M., Rowland, F. , Spiegelhalter, D.J., 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. M.J. (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, M.D. (1991) Factorial sampling plans for preliminary computational experiments. Technometrics, 33(2): 161-174.

Obenour, D.R., Gronewold, A.D., Stow, C.A. 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 K.J. (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, K.J., Hall, J. and Meadowcroft, I. (2006) Decision tree for choosing an uncertainty analysis methodology: a wiki experiment. Hydrological Processes, 20, no. 17: 3793-3798. DOI: 10.1002/hyp.6541

Poole, D. and Raftery, A.E. (2000) Inference for deterministic simulation models: the Bayesian melding approach. Journal of the American Statistical Association, 95(452): 1244-1255.

Raats, P.A.C. (1978a) Convective transport of solutes by steady flows I. General theory. Agricultural Water Management, 1: 201-232.

Raats, P.A.C. (1978b). Convective transport of solutes by steady flows II. Specific flow problems. Agricultural Water Management, 1: 219-218.

Razavi, S., Tolson, B. A., and Burn, D. H. (2012) Review of surrogate modeling in water resources. Water Resources Research, 48, W07401, doi:10.1029/2011WR011527.

Refsgaard, J.C. and Henriksen, H.J. (2004) Modelling guidelines –terminology and guiding principles. Advances in Water Resources, 27: 71-82.

Refsgaard, J.C., van der Sluijs, J.P., Hojberg, A.L. and Vanrolleghem, P.A. (2007) Uncertainty in the environmental modelling process – a framework and guidance. Environmental Modelling and Software, 22: 1543-1556.

Rietveld, L. C., Van der Helm, A. W. C., Van Schagen, K. M., and Van der Aa, L. T. J. (2010) Good modelling practice in drinking water treatment, applied to Weesperkarspel plant of Waternet. Environmental Modelling and Software, 25(5): 661-669.

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, B.J., 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, J.H.A., Croke, B.F.W. and Jakeman, A.J. (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, B. C., Ellis, G. and Keim, D. A. (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, R. G. (2013). Verification and validation of simulation models. Journal of Simulation, 7(1): 12-24.

Sobol, I.M. (1993) Sensitivity estimates for nonlinear mathematical models. Math. Model. Comput. Exp, 1(4): 407-414.

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.

van Delden, H., van Vliet, J., Rutledge, D. T., and Kirkby, M. J. (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, A. K., Brown, D. G., van Delden, H., Heckbert, S., and Verburg, P. H. (2016) A review of current calibration and validation practices in land-change modeling. Environmental Modelling and Software, 82: 174-182.

Venables, W.N. and Dichmont, C.M. (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, W.E., Harremoes, P., Rotmans, J., van der Sluijs, J.P., van Asselt, M.B.A., Janssen, P. and Krayer von Kraus, M.P. (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, A. A., and Kasprzyk, J. R. (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, J.J., Janssen, J.A.E.B., Booij, M.J. and Krol, M.S. (2010) Identification and classification of uncertainties in the application of environmental models. Environmental Modelling and Software 25(12): 1518-1527.

Webb, J.A., Stewardson, M.J. and Koster, W.M. (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, 940 7(8): 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.