25th International Congress on Modelling and Simulation, Darwin, NT, Australia, 9 to 14 July 2023
mssanz.org.au/modsim2023

INVITED

Developing an assessment framework to guide investment in water models

B. Gibbes [a], **D. Botelho** [b], **A. Singh** [c], **D. Rissik** [b,d] **and M. Hipsey** [e]

[a] School of Civil Engineering, The University of Queensland, Brisbane, [b] BMT, Brisbane, Queensland, [c] Queensland Department of Environment and Science, Brisbane, [d] Australian Rivers Institute, Griffith University, Brisbane, Queensland, [e] School of Agriculture and Environment, University of Western Australia, Perth, Western Australia
Email: b.gibbes@uq.edu.au

Abstract: This paper presents the development of an objective, transparent and adaptive process for evaluating water models that identifies key challenges, opportunities and risks for future model development and application. While many frameworks dealing with best practice model development and application exist, there are relatively few attempts to provide more holistic guidance on investment in model development from the perspective of a broader set of stakeholders comprising model developers, model users and decision makers. In response, a model assessment framework (MAF) was developed based on the principles used in the technology readiness levels (TRL) successfully applied in other branches of science and engineering.

Central to the MAF is a generic tool to rank the current state of a model (or set of models) in terms of ability to service the needs of different applications and stakeholders. The MAF comprises a two-part process that first defines the policy drivers or end-use requirements for a given water model. The second phase uses a TRL-style ranking system that considers seven key domains of influence on the development and application of a water model to evaluate the current state of development. This information can then be used to identify the strengths and weaknesses of a model, pinpoint gaps in specific areas, prioritise opportunities and risks for model improvement, and provide a context for model adaptability. The MAF allows a comprehensive analysis of future model development needs and is consistent with the transition in other technology development fields from use of TRLs to system readiness level (SRL) assessments. Application of the MAF is demonstrated via a case study that focuses on pollutant export modelling for the Great Barrier Reef using eWater SOURCE.

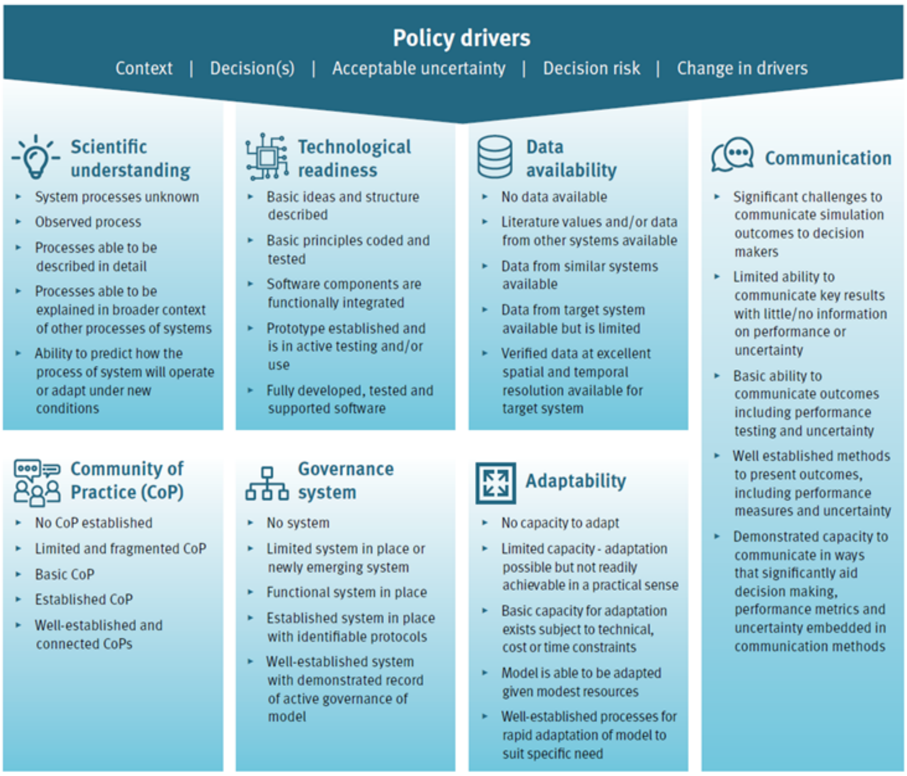

Figure 1. Conceptual representation of the Model Assessment Framework. The framework consists of eight components with two main parts. The first part – policy drivers – is an overarching frame to which all the other components in part 2 are related.

Keywords: *Water model, model assessment frameworks, technology readiness level, eWater SOURCE*

1. INTRODUCTION

Water models support a variety of studies related to policy development, regulatory compliance, development approvals, risk assessments, planning scenarios and overall water management, amongst other uses. A range of different water modelling tools are used by water management agencies to aid and improve critical decision making related to management of environmental and other assets. Widespread use of water models to support decision making has led to an increased awareness of their importance to environmental management activities and as such, there is a pressing need to objectively prioritise investment in water modelling tools to enable their continued use in these critical support roles. While there are numerous guidelines and frameworks for best practice water model development and implementation (e.g., Barnett et al. 2012; Jakeman et al. 2006, 2018; Yu et al. 2022 among others), there appears to be a lack of attention placed on processes for maintenance and ongoing development of water modelling systems. This includes both the updating of modelling approaches to incorporate new science and workforce succession planning to ensure appropriate expertise is available to commission, run and critically assess water modelling projects.

Recent work by the Queensland Water Modelling Network (QWMN) (Gibbes & McIntosh, 2019) highlighted the complex and interconnected nature of the water modelling community which relies on the interactions between different types of organisations and individuals to facilitate water modelling activities. In particular, the commissioning of water modelling projects is often undertaken by different parts of the community to those that develop water modelling systems and those that apply them. Furthermore, resources available to invest in model maintenance and improvement often come from a range of different sources. With an increasing trend towards open source modelling platforms, the capacity for traditional model providers to fund ongoing maintenance and upgrades from licence fee revenues is likely to decrease. This trend suggests that end-users of modelling services will more actively invest in improved capacity and user-friendliness of water models. In many instances, these investors are likely to be in the form of government or collective not-for-profit organisations, which will typically require an increased level of transparency in relation to funding decisions.

While various frameworks for guiding investment in scientific and research activities have been developed and applied in different sectors including aerospace (e.g., Mankins, 1995), nuclear fuels, defence, oil and gas, heavy equipment, power systems and electronics (Olechowski et al. 2015) there are relatively few examples specific to the water modelling sector. These frameworks are generally based on the technology readiness level (TRL) scale introduced by NASA in the 1970s and have subsequently been expanded to consider systems via system readiness level (SRL) approaches (Olechowski et al. 2015). They appear to be readily adaptable and useful to water modelling as, within this sector, there are often a range of models and modelling systems that compete for relatively scarce funding. Given the importance of water models in the decision-making process, the QWMN commissioned a strategic review of water models (Botelho et al. 2021) to identify, substantiate and prioritise investment in water modelling over the next five years. One of the outcomes of this project was the development of an objective, flexible and transparent evaluation framework to support investment allocation across key areas within the water modelling process. The resulting model assessment framework (MAF) is presented and its usefulness demonstrated via application to two case studies. The framework represents a first attempt at adapting investment frameworks from other fields to guide water model development and investment.

2. DEVELOPMENT APPROACH

MAF development comprised the following four steps: i) preliminary design of the framework components; ii) engagement and co-design with the QWMN project team and QWMN community, which is comprised of a variety of Australian-based modelling experts and users; iii) refinement of the preliminary framework design; and iv) application of the framework to selected case studies. Through workshop discussions with users and developers, valuable guidance was obtained for the design of a functional MAF. Some of the key points relating to assessing the performance of a modelling system that were raised at the workshops included:

- Modelling workforce – Identified as a key aspect of the whole water modelling workflow. Participants felt that there was a shortage of new talent emerging and that better training, motivation and mentorship could help attract well-prepared professionals;
- Good quality data – It was recognised that good quality data is important to deliver meaningful outcomes from modelling studies. To alleviate potential issues with data access and sharing, a common platform could be used to bring efficiencies in data collection and improvements in data sharing;
- Communication – A recurring theme that is closely linked to uncertainty is the ability to communicate the results of modelling projects in a way that addresses model uncertainty and enhances uptake and use of models by policy makers, managers and other users. It was widely recognised that communication is a

specialist skill and often modellers may need to work with other experts to make messaging more effective.

Note that these "model performance" issues are related to the broader modelling community and the "business of applying models", rather than focusing on the more technical aspects of a given model. This resulted in the incorporation of wider elements of the modelling workflow into the MAF rather than the application of a more traditional technology focused approach.

3. MODEL ASSESSMENT FRAMEWORK

The framework consists of eight components with two main parts (Figure 1). The first part – policy drivers (§3.1) – is an overarching frame to which all the other components in part 2 (§3.2) are related. The policy drivers set the context for the modelling assessment and can be applied to an already existing model or a new model created specifically for the policy driver purpose. The other components of the framework comprise seven independent areas (with some interconnections) that were identified by the QWMN community as being crucial to assess whether a model is fit-for-purpose and whether future investment in the model is warranted, including whether prioritising different aspects is necessary.

3.1. Policy drivers

Implementation of similar frameworks used to assess technology (e.g. NASA's TRL scale, Mankins, 1995; Olechowski et al. 2015) highlights the importance of understanding end-use requirements. For water models these requirements often inform the development of water management policy or management decisions that are driven by policy. Given the wide scope of model application, it is challenging to establish a universal context or a single driver. Therefore, the inclusion of a process for setting the modelling context is a critical first step in the framework. This step seeks to clarify the following aspects for an objective assessment:

- context – the broad set of socio-economic and political conditions in which a model will be applied
- decision/s – articulation of the decisions that will be made based, in part, on the simulation results
- acceptable uncertainty – an explicit statement about the level of uncertainty (either quantitative or qualitative) that is acceptable in results for the given context and decision/s that will be made
- decision risk – clarification of the implications of making decisions based on erroneous information
- change in drivers – an assessment of the likelihood that the context, decision/s, uncertainty and decision risk/s will change during the lifetime of a given modelling project or model application.

These elements aim to provide information on the overall context and the clear objective for the model, as well as clearly specifying what a model needs to successfully provide the required information to end users.

3.2. Assessing components of model readiness with a five-level scale

The rationale for each of the seven assessment areas is provided below. Each of the other seven areas of model assessment incorporates an expanded set of assessment scales that adopted the NASA TRL concept as an initial template, with a five-level assessment hierarchy (cf. nine levels in the NASA TRL) deemed more appropriate for each assessment area. In each area, a given model can receive a numerical score from 1 to 5.

Scientific understanding
Feedback from stakeholders and experts emphasised the importance of good scientific understanding of the physical, biological and chemical processes being simulated by a water model. The MAF adopts an approach based on the DIKW hierarchy (see Rowley, 2007) which has four levels: i) data; ii) information; iii) knowledge; iv) wisdom. Using this framework, the level of scientific understanding can be rated from low (unknown process, no observations) to high (well understood processes that can be predicted under new conditions).

Technological readiness
Stakeholders reviewed the technological readiness levels (TRLs) commonly used in other technology dominant sectors (Mankins, 1995; Olechowski et al. 2015) and confirmed they had merit for adaption to the assessment of water models. The approach presented here is based on modification of the NASA TRLs as described in Botelho et al. (2021). Technological readiness is rated from low (basic ideas and structure of technology able to be described) to high (fully developed, tested and supported water modelling software).

Data availability
The availability of data for model development, testing, validation and verification is often a limiting factor for successful implementation. This includes input data as well as information on the environmental state variables that the model uses. Therefore, data availability was adopted as one of the core areas in which a

model or modelling system was assessed. The framework ranks data availability from low (no data available) to high (verified data at excellent spatial and temporal resolution available for the target system across a range of system states). In applying the MAF, specific sub-types of data (e.g., rainfall and other meteorological variables, data to inform model process parameters and environmental state variables) can be rated individually.

Communication
Stakeholders consistently highlighted the importance of communicating simulation outcomes in a way that aids decision-making. Effective communication includes factors ranging from well-developed post-processing and data visualisation methods to more complex inclusion of parameter sensitivity and uncertainty quantification techniques. The five-level rating system seeks to capture the range of potential development states in this area, from a low level of communication capability (ability to communicate basic concepts are still under development) to a high level (demonstrated capacity to communicate in ways that significantly aid decision-making, performance metrics and uncertainty embedded in communication methods).

Community of practice
The ability of model practitioners to access support and advice from other professionals who are also involved in the development and application of models was identified as a key factor that can influence the selection of a model. It is therefore important to attempt to quantify the current extent of a given model's community of practice to establish whether additional investment is required. Again, a five-level rating system was applied to rate current state from low (no community established) to well developed (well-established and connected practitioners: a mix of professionals, including notable international links that are internal and external to an organisation, as well as active systems for model development, skill development, recruitment and mentoring).

Governance system
While governance structures are a key part of a good practice approach for modelling projects (Jakeman et al. 2006), model/software governance is also vital. In this context, governance refers to the systems that are in place to manage the model including the underlying model code and the processes and procedures associated with making and communicating model updates. In this area a model is again rated from low (no governance system in place) to well developed (well-established governance process in place, including a structure and set of operating guidelines in place for the ongoing development and custodianship of the model).

Adaptability
The ability of a model to be adapted to meet emerging policy, regulatory or management decision needs is also desirable. Adaptability criteria can be used to identify legacy models that might need investment to remain fit-for-purpose, or alternatively be replaced by other models. Adaptability can incorporate elements of scientific understanding, technological readiness and community of practice presented above. However, in this context, it specifically assesses the adaptability of a model independently from these other areas. The highest possible rating is awarded to models that have demonstrated a well-established process for rapid adaption.

Table 1. Overall five-level assessment scale generated by the MAF

Aggregate score	Categories	Description
≤ 1	Developmental model	Research has identified a pathway for delivering a modelling system that could support decision-making but needs substantial investment in multiple areas to allow a more informed assessment to be made.
1≤ 2	Basic model	A functional model that can generate insights but needs further investment in multiple areas to be considered fit-for-purpose for a given policy driver and setting.
2 ≤ 3	Established model	A model that is considered suitable for providing information in the given policy development or decision-making context.
3 ≤ 4	Mature model	A model that has a demonstrated track record in providing good quality information to inform decision-making across a range of different settings (both environmental settings and policy/ regulatory environments).
4 ≤ 5	State-of-the-art model	A model that achieves excellent to outstanding ranks across all areas and can be readily used to support decision-making across a range of different settings (both environmental settings and policy/regulatory environments).

3.3. Overall assessment scale

When applying the MAF, models are categorised with a numerical value from 1-5 for each component and an overall model assessment score calculated as the sum of all the values awarded, allowing categorisation of the model (Table 1). Different weightings can be applied to each component, if needed. For models that have more discrepant results across different assessment areas, weightings are likely to have larger influence in the overall classification outcome. A neutral weighting is recommended for an initial assessment.

4. CASE STUDIES

Two case studies were used to test the MAF, with both involving application of the SOURCE platform (eWater, 2021), one in the context of water supply planning and another in the context of policy development within the Paddock to Reef program. Inputs for both case studies were from one-on-one interviews with relevant project members. Sessions were recorded, to allow content of discussions to be reviewed and subsequently analysed in more detail. In both case studies, project teams identified that some of the MAF component areas were more important than others. As a result, weightings were applied to each of the component areas according to a qualitative ranking. It was recognised that the choice of weighting could largely influence the overall MAF score. Therefore, three weighting possibilities were then adopted to test the effect of weighting on the overall MAF classification, as follows:

- *Equal weighting* – equal weight was assigned to each of the categories.
- *Heavily biased weighting* – a weight of 1 was assigned to the least important category, a weight of 2 to the second least important category, a weight of 3 to the third least important category and so on.
- *Weakly biased weighting* – a weight of 1 was assigned to the least important category and a weight of 2 was assigned to the most important, with equal subdivisions applied between 1 and 2 for other categories.

4.1. Case study 1: The Border Rivers Model

The Moonie Water Resource Plan (WRP) details the water management arrangements of two surface water and four groundwater sustainable diversion limit resource units identified in the Murray-Darling Basin Plan. The Border Rivers e-Water SOURCE model is used for a variety of applications related to the WRP and the broader water security of the region. The ratings and weighting for each of the assessment categories for Case Study 1 are presented in Table 2. The order of importance for categories was assigned as follows: 1. Data availability; 2. Communication; 3. Governance System; 4. Software Readiness; 5. Scientific Understanding; 6. Community of Practice; 7. Adaptability (as reflected in the weightings).

Table 2. Category and final rating for Case Study 1

Category	Rating	Equal weighting	Heavily biased weighting	Weakly biased weighting
Scientific understanding	4.33	1.00	3.00	1.33
Software readiness	4.00	1.00	4.00	1.50
Data availability	3.75	1.00	7.00	2.00
Communication	3.00	1.00	6.00	1.83
CoP	4.00	1.00	2.00	1.17
Governance system	3.50	1.00	5.00	1.67
Adaptability	5.00	1.00	1.00	1.00
Combined score	N/A	3.94	3.71	3.84

Regardless of the weightings applied, the Moonie WRP model had a combined score between 3.7 and 3.9, indicating that it is an “established” model, with most attributes in place (and relatively little effort required) to be moved to a “mature” model rating. Investment focus should be given to data collection, communication and governance systems. Although given a low weighting and a relatively high score, it was identified that the CoP could benefit from increased model use within universities and other training providers.

4.2. Case study 2: Great Barrier Reef land use change

The Reef 2050 Water Quality Improvement Plan sets a hierarchy of the 2025 water quality targets. The main objective of the e-Water SOURCE model in this context is to identify the impact of changes in land use management on nutrient, sediment and pesticide loads from the catchment. For this case study, the Burdekin catchment model was selected. The ratings and weightings for each of the assessment categories for Case Study 2 are presented in Table 3. Different from Case Study 1, some areas were given equal qualitative importance, in order from most to least important as follows: 1. Communication; 2. Data Availability and Scientific Understanding; 3. Software Readiness; 4. Community of Practice; 5. Governance System; and 6. Adaptability.

Table 3. Category and final rating for Case Study 2

Category	Rating	Equal weighting	Heavily biased weighting	Weakly biased weighting
Scientific understanding	3.00	1.00	3.00	1.67
Software readiness	5.00	1.00	2.00	1.33
Data availability	3.00	1.00	3.00	1.67
Communication	3.00	1.00	4.00	2.00
CoP	3.00	1.00	1.00	1.00
Governance system	4.00	1.00	1.00	1.00
Adaptability	4.00	1.00	1.00	1.00
Combined score	N/A	3.57	3.40	3.48

The Burdekin e-Water SOURCE model received an overall MAF score between 3.40 and 3.57, indicating the model ranks as an “established” model with significant progress towards a “mature” model. Similar to Case Study 1, the overall MAF scores were not sensitive to the weightings. Identified areas for investment (i.e. lower ratings) were data collection, scientific understanding, CoP, and communication.

5. DISCUSSION AND RECOMMENDATIONS

The framework was developed to determine the current status of models, identify how they are being applied and help guide future improvements and developments. In its most simple form, it provides a generic tool to rank the current state of a model in servicing the needs of different applications. The MAF provides a process to identify strengths and weaknesses of a model, prioritise opportunities and risks for model improvement, and provide a context for model adaptability. While the case studies illustrated the application of the MAF, they are not exhaustive or exclusive. Although the outcomes were based on consultation with stakeholders, a sense of priority should not be assigned to the cases presented. Rather, the MAF as applied to the study cases should be seen as how to adopt the approach for structuring future strategic investments (or disinvestments) in models.

With this in mind, it is important to understand the lessons learned with the application of the MAF to the case studies. Specifically, the MAF clearly identified priority areas where (time and financial) investment should be allocated. In the authors’ opinion, the MAF provided valuable insights into how models are applied, areas where the modelling process is successful and the reasons why the modelling process in a particular area is deficient. As such, the MAF provides an opportunity to establish benchmarks or archetypes that can be transferred to other modelling applications. Conversely, archetypes in a MAF assessment area identified as deficient can be borrowed from successful cases elsewhere and used as a starting point for model improvement.

There is, however, a peculiarity of the MAF, as exemplified in these study cases, that would warrant more testing. In the examples provided, the different weightings applied in the case studies had little bearing on the overall MAF classification outcome. The range of ratings assigned to each of the assessment areas did not diverge greatly in either case study. For models that have more discrepant results across different assessment areas, weightings are likely to have larger influence in the overall classification. For instances in which assignment of weightings to different categories are applicable, a sensitivity test of the final MAF result in response to different weight derivations (as applied to the case studies above) is recommended. Regardless of whether weightings are assigned or not, the ratings, and their rationale, are available to understand how the model is performing.

It is recommended that where possible, no bias (i.e., equal weight) should be assigned to all categories, as they all carry inherent risks. It is also advised that, as much as practicable, a collection of modellers, model users, and even additional stakeholders should be involved throughout the whole MAF process. This would harness a collective (and more complete) view with different risk perceptions and would also create shared ownership and responsibility of MAF-derived investment decisions. A transparent and agreed rating/scoring system can minimize potential conflicts once investment decisions are made as a result of the MAF scoring.

To the authors' knowledge, this is the first attempt to create a comprehensive, systematic, and sufficiently generic assessment, with objective ratings, for water models. It is hoped that the application of the MAF will help consistent, efficient and effective identification and prioritisation of investments in water modelling, particularly for organisations with a range of modelling programs and restricted investment capability. Additional testing of the scoring system with a wider group of models and stakeholders may provide pathways for further improvement of the MAF.

ACKNOWLEDGEMENTS

This work was made possible by the QWMN, specifically through the Strategic Review of Queensland Water Models project. The QWMN is an initiative of the Queensland Government that aims to improve the state's capacity to model its surface water and groundwater resources and their quality. The QWMN is led by the Department of Environment and Science (QDES) with key links across industry, research and government, see Queensland Water Modelling Network | Science | Department of Environment and Science, Queensland (des.qld.gov.au) for information. Contributions from all members of the QWMN program team within QDES are gratefully acknowledged. Comments from anonymous reviewers on this paper were greatly appreciated.

REFERENCES

Barnett, B., Townley, L., Post, V., Evans, R., Hunt, R., Peeters, L., Richardson, S., Werner, A., Knapton, A., Boronkay, A., 2012. Australian groundwater modelling guidelines. Waterlines report No. 82, National Water Commission, csiro:EP117658. http://hdl.handle.net/102.100.100/100470?index=1

Botelho, D., Singh, A., Rissik, D., Hipsey, M., Gibbes, B., 2021. Strategic Review of Models – Model Assessment Framework. Report prepared for the Queensland Water Modelling Network, Department of Environment and Science, Queensland Government, Brisbane. https://science.des.qld.gov.au/__data/assets/pdf_file/0019/247114/QWMN-Strategic-Review-of-Models_Assessment-Framework.pdf

eWater. 2021. Source User Guide 5.0. Retrieved 8-Mar-2023 from wiki.ewater.org.au.

Gibbes, B., McIntosh, B. 2019. Queensland Water Modelling Network External Engagement Program Skills and Knowledge Audit. Queensland Department of Environment and Science, available: https://watermodelling.org/wp-content/uploads/2019/09/QWMN-SkillsKnowledgeAudit-July2019.pdf

Jakeman, A. J., Letcher, R. A., Norton, J. P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling & Software, 21(5), 602-614.

Jakeman, A.J., El Sawah, S., Cuddy, S., Robson, B., McIntyre, N., Cook, F. 2018. QWMN Good Modelling Practice Principles, Queensland Department of Environment and Science. https://science.des.qld.gov.au/__data/assets/pdf_file/0011/81110/qwmn-good-modelling-practice-principles.pdf

Mankins, J.C., 1995. Technology Readiness Levels: A White Paper. www.hq.nasa.gov/office/codeq/trl/trl.pdf.

Olechowski, A., Eppinger, S., Joglekar, N., 2015. Technology readiness levels at 40: A study of state-of-the-art use, challenges, and opportunities. 2015 Portland International Conference on Management of Engineering and Technology (PICMET), 2-6 August 2015, Institute of Electrical and Electronics Engineers (IEEE), http://dx.doi.org/10.1109/PICMET.2015.7273196

Rowley, J., 2007. The wisdom hierarchy: representations of the DIKW hierarchy. Journal of information science, 33(2), 163-180.

Yu, S., Sturm, K., Gibbes, B., Kennard, M., Veal, C.J., Middleton, D., Fisher, P.L., Rotherham, S., Hamilton, D.P., 2022. Developing best practice guidelines for lake modelling to inform quantitative microbial risk assessment. Environmental Modelling & Software, 150, 105334, https://doi.org/10.1016/j.envsoft.2022.105334