QUEENSLAND WATER MODELLING NETWORK

Strategic Review of Models

MODEL ASSESSMENT FRAMEWORK

Report prepared by
BMT, The University of Queensland and
The University of Western Australia
For Queensland Water Modelling Network

Queensland Government

The Queensland Water Modelling Network (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 with key links across industry, research and government.

Prepared by: BMT, The University of Queensland and The University of Western Australia

August 2021

Citation

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.

Table of contents

Introduction

Water models are used to inform a wide range of decision-making processes including policy, planning and management decisions, and to assess likely impacts from external drivers, such as rainfall change, sea level rise or population growth (QWMN, 2021).

The Queensland Water Modelling Network (QWMN) was established in 2017 to build the capacity of the water modelling sector, encouraging engagement between modellers, end-users and stakeholders.

A range of collaborative projects have been initiated to improve the state's capacity to model its surface and ground water resources, from cataloguing major water models used by the Queensland Government, through to improving integration between agricultural and water catchment models.

The QWMN's 2018-2020 Research, Development and Innovation Strategy raised the need to conduct a strategic review of water models to identify, substantiate and prioritise investment in water modelling over the next five years (2020-2025). The Strategy noted that investment in water modelling would benefit from an objective, transparent and adaptive process for evaluating water models and identifying key challenges, opportunities and risks for future model development and application.

The QWMN commissioned BMT, The University of Queensland and The University of Western Australia to undertake a strategic review of Queensland water models, including developing an approach to classify models and a framework for assessment. This report provides a summary of the Model Assessment Framework that:

- presents a generic tool to rank the current state of a model (or sets of models)
- reflects modellers and end-users' feedback on existing and future model capabilities required to support water policies in Queensland
- builds the basis for an open dialogue between model users and model software owners (and developers) to help make models as efficient and effective as possible.

The Strategic Review of Models – Model Classification report provides more information about the classification approach and together with the Great Barrier Reef eWater SOURCE case study is recommended reading to understand how all the elements work together in practice.

Framework methodology

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 was designed as a generic tool to rank the current state of a model (or set of models) in servicing the needs of different applications.

The framework provides a process 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 authors recommend reporting on the assessment of individual models (or set of models) within the context of its application to ensure the results are not only objective and transparent but can also support an ongoing exchange between model owners (or developers), modellers and model end-users about future modelling requirements.

The framework development process consisted of the following steps:

- preliminary design of the framework components
- engagement with the QWMN project team
- engagement and co-design with the QWMN community
- refinement of the preliminary framework design
- application of the framework to selected model application case studies.

Model assessment framework

Development considerations

Two key features underpin the framework.

- Systematic approach – as the modelling process can be quite diverse, the framework is compartmentalised to provide a structure that can be logically followed and used to direct specific areas and actions for future focus and investment.
- Generic application – the framework is universal in its approach to ensure it can be applied within multiple modelling contexts and be conducted by different assessors without loss of its core elements. In this sense, the framework is not only applicable to Queensland Government modelling programs, but also more generically to any modelling program seeking to focus its efforts and investments more strategically.

In addition to the key features identified above, there are also three other considerations informing the framework's operations. First, as a model can be used for different purposes, the application context needs to be clearly articulated so ratings can be objectively obtained, and future assessments can verify whether the model achieved the specified improvements.

Second, most water models have been used for some time and therefore carry a legacy from which policy development and investment decisions were (at least in part) derived. Any future modifications should be supported by proactive engagement with relevant stakeholder groups, including a transition plan for a new model or modification to an existing model.

Third, the definition for each of the ratings needs to be clear with distinct step changes between them. This will enable an assessor to collect relevant information about the current state of a model and make recommendations that not only lead to an improved rating, but more importantly, to more effective and accurate model components (i.e. underlying science, output uncertainty, as per model components described below) to inform the decision-making process.

Framework components

The framework consists of eight components (see the diagram below). The first – policy drivers – is an overarching frame to which all the other components are related. It sets 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.

Component assessment criteria

Policy drivers

Implementation of similar frameworks used to assess technology (e.g. NASA's TRL scale, Heder, 2017) highlights the importance of understanding end-use requirements. For water models these requirements are often to 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 policy drivers or 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 specific decisions that will be made based, in part, on the model's simulation results
- acceptable uncertainty: an explicit statement about the level of uncertainty (either quantitative or qualitative) that is acceptable in simulation results for the given context and decision/s that will be made
- decision risk: clarification of the implications of making decisions based on erroneous modelling information
- change in drivers: an assessment of the likelihood 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 that will lead to the development of a clear objective for the model. In particular these elements intend to specify what a model needs (in terms of capability and fitness for purpose) to provide the required information to policy developers, regulators or decision-makers.

Model Assessment Framework

Conceptual representation of the Model Assessment Framework illustrating the overarching role of defining the policy drivers in relations to each of the other components.

Policy drivers

Context | Decision(s) | Acceptable uncertainty | Decision risk | Change in drivers

Scientific understanding

- System processes unknown
- Observed process
- Processes able to be described in detail
- Processes able to be explained in broader context of other processes of systems
- Ability to predict how the process of system will operate or adapt under new conditions

Technological readiness

- Basic ideas and structure described
- Basic principles coded and tested
- Software components are functionally integrated
- Prototype established and is in active testing and/or use
- Fully developed, tested and supported software

Data availability

- No data available
- Literature values and/or data from other systems available
- Data from similar systems available
- Data from target system available but is limited
- Verified data at excellent spatial and temporal resolution available for target system

Communication

- Significant challenges to communicate simulation outcomes to decision makers
- Limited ability to communicate key results with little/no information on performance or uncertainty
- Basic ability to communicate outcomes including performance testing and uncertainty
- Well established methods to present outcomes, including performance measures and uncertainty
- Demonstrated capacity to communicate in ways that significantly aid decision making, performance metrics and uncertainty embedded in communication methods

Community of Practice (CoP)

- No CoP established
- Limited and fragmented CoP
- Basic CoP
- Established CoP
- Well-established and connected CoPs

Governance system

- No system
- Limited system in place or newly emerging system
- Functional system in place
- Established system in place with identifiable protocols
- Well-established system with demonstrated record of active governance of model

Adaptability

- No capacity to adapt
- Limited capacity - adaptation possible but not readily achievable in a practical sense
- Basic capacity for adaptation exists subject to technical, cost or time constraints
- Model is able to be adapted given modest resources
- Well-established processes for rapid adaptation of model to suit specific need

Scientific understanding

Feedback from stakeholders and experts emphasised the importance of good scientific understanding of the underlying physical, biological and chemical processes being simulated by a model. For clarity, scientific understanding refers to process understanding which is distinctly separate from the availability of data for a given water modelling problem.

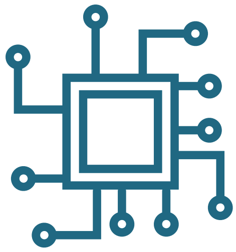
Scientific understanding of water modelling processes can be categorised in a number of ways. The framework adopts an approach based on the DIKW hierarchy (for background see Rowley, 2007) which has four levels:

- data: discrete objective facts or observations in an unorganised or unprocessed form -- generally considered of no use without context or interpretation
- information: organised or structured data that has been processed so that it has relevance for a particular purpose or context (also referred to as functional data)
- knowledge: a fluid mix of framed experience, values, contextual information, expert insight and grounded intuition that provides an environment and framework for evaluating and incorporating new experiences and information (Dewey and Bentley, 1949)
- wisdom: the ability to apply knowledge to make good decisions, particularly in relation to future events or in circumstances that are different from those in which knowledge was developed.

These levels have been modified into a five-level system based on scientific understanding of the processes or system that is under investigation, as described below:

SU1	Unknown process: no observations or data are available – the process can be hypothesised to exist but data has not yet been collected to test.
SU2	Observed process: data exists that suggests a process or relationship between different environmental factors within a given system.
SU3	Described process: a theory has been developed, based on data, to explain the relationship between different environmental factors within a given system.
SU4	Explained process: the process is able to be described for a range of different systems and settings.
SU5	Predictable process: it is possible to predict how the process will operate or adapt in under new conditions.

Technological readiness



Stakeholder feedback reviewed the technological readiness levels (TRLs) commonly used in the defence and aerospace industries (Banke, 2020; Heder, 2017) 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 to shorten the list to a five-level system as described below.

TR1	Basic ideas and structure described with links to potential applications hypothesised.
TR2	Basic principles coded: properties of algorithms, representations and concepts are defined, and basic principles coded (experiments performed with synthetic data).
TR3	Software components functionally integrated: end-to-end software elements implanted and validated for a limited set of conditions against existing systems and/or data.
TR4	Prototype established and in active testing and/or use: functional software available but not yet fully de-bugged or applied to a wide variety of environmental settings (documentation under development but not yet complete).
TR5	Fully developed, tested and supported software available: software has been de-bugged and successfully applied to a range of different environmental settings/problems (all documentation has been completed, functionality has been successfully demonstrated in operational scenarios, verification and validation complete).

Data availability

The availably of data for model development, testing, validation and verification is often a limiting factor for successful model implementation. This includes input data as well as information on the environmental state variables that the model uses – specifically when quantitatively assessing model performance (i.e. the application of model skill scores in calibration).

The framework adopts the five-level rating system below.

DA1	No data available.
DA2	Literature values and/or data from other systems available but limited in extent, scope or applicability to the target system.
DA3	Data from similar systems available but might be limited in spatial or temporal coverage – data quality might not be verified.
DA4	Data from target system available but is limited in spatial or temporal scope – data quality might not be verified.
DA5	Verified data at excellent spatial and temporal resolution available for target system across a range of system states.

Note that it might be useful to apply this assessment system to specific sub-types of data including input data (e.g. rainfall and other meteorological variables, data to inform model process parameters and environmental state variables).

Communication

Stakeholders consistently highlighted the importance of being able to readily communicate simulation outcomes in a way that aids decision-making. Effective communication includes factors ranging from well-developed post-processing and data visualisation methods and systems to more complex inclusion of parameter sensitivity and uncertainty quantification techniques in the underlying modelling system. All these factors support the effectiveness in communicating modelling outputs and outcomes.

The following five-level rating system seeks to capture the range of potential development states in this area.

C1	Significant challenges to communicate simulation outcomes to decision-makers. Methods to post-process model outputs to communicate basic concepts are still under development or require significant resources or expertise.
C2	Limited ability to communicate key results with little/no information on performance or uncertainty. Basic outcomes can be communicated but significant investment is required to improve the uptake of information in decision-making.
C3	Basic ability to communicate outcomes including performance testing and uncertainty. Outcomes can be communicated with some post-processing and information to support overall concepts/findings can be understood by decision-makers. Performance testing (i.e. calibration/validation) and uncertainty information can be developed for the simulation output).
C4	Well-established methods to present outcomes, including performance measures and uncertainty. Well-supported techniques and systems are in place to post-process simulation outputs into forms that are useful for policy development and regulatory decisions.
C5	Demonstrated capacity to communicate in ways that significantly aid decision-making, performance metrics and uncertainty embedded in communication methods. An excellent suite of post-processing tools that apply well-founded methods are available and extensively used to communicate outcomes, and performance metrics and uncertainty in multiple forms.

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 has been identified as a key factor that can influence the selection of a model in both the short and long-term.

The framework uses the following five-level rating system.

CP1	No community established: a recognisable community of practitioners, either formal or informal is unable to be identified.
CP2	Limited and fragmented: a limited number of model users are known to other professionals.
CP3	Basic community: an identifiable group of practitioners, either internal or external to an organisation, engage in sporadic communication in relation to a given modelling platform.
CP4	Established community: a group of practitioners has self-identified as a community of practice and has established formal and/or informal protocols for exchanging information and ideas (can include groups both internal and external to an organisation).
CP5	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 often a key part of specific modelling projects and form good practice (Black et al, 2011; Jakeman et al, 2006), model/model software governance is vital when assessing the model designed for a given task. 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 changing the code and making users aware of updates.

The following five-level system was developed to assess these characteristics.

GS1	No governance process in place.
GS2	Limited system in place or a newly emerging system in the process of being developed: the system not considered functional (i.e. system is unclear or unable to respond in reasonable timeframes).
GS3	Functional system in place: an identifiable system has been established for managing the model but improvements in communication and or funding are required.
GS4	Established system in place with identifiable protocols and a track-record of successful management including a process for interacting with model users and developers, including well-established benchmark testing applications.
GS5	Well-established governance process in place: a structure and set of operating guidelines are in place for the ongoing development and custodianship of the model (with an established record of delivering model adaptations including funding and making them available to all users).

Adaptability

The ability of a model or software suite to be adapted to provide information to meet emerging policy, regulatory or management decision needs is also essential. An adaptability criteria can therefore be used to identify legacy models that might need substantial investment to remain fit-for-purpose, or alternatively be replaced by other models. It is anticipated adaptability may incorporate elements of the scientific understanding, technological readiness and community of practice components presented above. However in this context it specifically assesses the adaptability of a model independently from these other assessment areas.

The framework uses the following five-level system.

A1	No capacity to adapt.
A2	Limited capacity: adaptation theoretically possible but not readily achievable in a practical sense.
A3	Basic capacity for adaptation exists subject to technical, cost or time constraints.
A4	Model is able to be adapted given modest resources.
A5	Well-established adaptation processes.

Five-level assessment scale

Each component outlined above specifies a five-level assessment hierarchy that incorporates an expanded set of assessment scales that are based on the NASA Technological Readiness Level (TRL) concepts (see also Sauser et al. 2006, Banke 2010, Heder, 2017). It is intended 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. Different weightings could be applied to each component, if needed. It should be noted, that 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.

It is envisaged that the total score would result in an overall categorisation as:

Aggregate score	Model categorisation	Description
≤ 1	**Developmental model**	Research has identified a pathway for delivering a model or simulation information that could support decision-making has been identified 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 is able to be readily used to support decision-making across a range of different settings (both environmental settings and policy/regulatory environments).

Issues and insights

Framework implementation

In developing the framework, the authors believe the process described above offers 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. It also provides an opportunity to establish benchmarks or archetypes that can be transferred to other modelling applications. Conversely, elements in the framework identified as deficient can be borrowed from successful cases elsewhere and used as a starting point for model improvement initiatives. Finally the framework helps identify potential areas for future investment in model development and roll out.

Modellers and end-users

Several other considerations were identified by stakeholders involved in the development process.

- Identification of the individuals (and respective organisations) operating the models, the model end-users and stakeholders (this can be added to the context setting) will help to more readily identify those involved in the implementation of future recommendations.
- Specification of the outputs for end-users and stakeholders (this can be added to the communication) will provide additional context for eventual issues that are identified through the assessment process.
- Specification of the format in which the information/data is provided to end-users and stakeholders (this can be added to the communication) will provide additional context for eventual issues that are identified through the assessment process.

Component weighting

Assessors should be mindful of the assessment purpose in determining whether weighting of components is required. For models that have more discrepant results across different assessment areas, weightings are likely to have larger influence in the overall classification outcome. For instances in which assignment of weightings to different categories are applicable, a sensitivity test of the final classification result in response to different weight derivations is recommended.

It is also recommended that where possible, no bias (i.e. equal weight) should be assigned to the different categories, as they all carry inherent risks in the modelling process. In this case, it is advised that, as much as practicable, modellers, model users, and stakeholders be involved à priori to discuss the assessment areas. This would harness a collective (and more complete) view with different risk perceptions across the assessment process. For example, hypothetically, a modeller may give more weighting to data availability and underpinning science; whereas a manager could give greater weighting to governance and communication of uncertainty. Involving all stakeholders in the assessment process would create shared ownership and responsibility of future improvements. It follows that it is imperative to have a transparent and agreed rating/scoring system for assessments. This way, any unwarranted biases to a given component, and more importantly, potential conflicts once decisions are made as a result of the classification would be avoided.

References

Banke, J. (2010). Technology readiness levels demystified. NASA, https://www.nasa.gov/topics/aeronautics/features/trl_demystified.html.

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, 92p.

Dewey, J. and Bentley, A.F. (1949). Knowing and the Known. The Beacon Press. Boston.

Héder, M. (2017). From NASA to EU: The evolution of the TRL scale in Public Sector Innovation. The Innovation Journal, 22(2), 1-23.

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.

QWMN (2021). Queensland Water Modelling Network (QWMN) information accessed from https://watermodelling.org/water-modelling/how-are-water-models-used

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

Sauser, B., Verma, D., Ramirez-Marquez, J., & Gove, R. (2006). From TRL to SRL: The concept of systems readiness levels. In Conference on Systems Engineering Research, Los Angeles, CA (pp. 1-10).