**Workshop reporting**

**1st Australia and New Zealand Water Quality Modelling Symposium**

7-10 April 2024, Queenstown, NZ.

**Workshop Topic** 2.2 Strategy for uptake and use of new modelling technologies.

# **Vision**

The water quality modelling community is familiar with and makes greater use of new and emerging technologies and methods, enhancing the accessibility, predictive power and impact of our models.

For example:

* The community is skilled in using HPC, GPU, cloud computing, quantum computing and new programming languages, and know how and when to effectively use them;
* New data sources and innovative uses of big data, are discoverable, accessible, usable, trust-worthy and are quickly and ethically adopted as they become available;
* Machine learning is widely used in ways that complement and enhance process-based modelling to improve prediction and decision-making;
* The community is using a broader range of communication methods, including AR, VR, gamification and participatory processes, informed by understanding of stakeholder needs, collaborating with experts in social science, computer science, graphic art, PR and science communication.

Scope and Issues

The strategy considers both technologies that are new (such as large language models and quantum computing) and technologies are methods that are somewhat established but are still emerging with respect to their uptake and use within water quality modelling (such as remote sensing data, high performance computing, and a wider range of communication methods).

Technologies and methods of interest include:

* Machine learning (ML) and AI;
* GPU, HPC and quantum computing;
* New and under-used data sources: satellites, high-frequency sensors, citizen science data, and innovative use of data collected for other purposes, such as Strava and social media data, economic and social data;
* New engagement and communication technologies: VR, AR and new visualisation and graphics approaches, and gamification;
* Digital twins;
* New platforms, bringing in economics.

# Goals and implementation strategy:

From among the above new and emerging technologies, four topics were chosen for further exploration: enhanced uptake and better use of big data; new computing technologies (HPC, GPU, clous computing and quantum computing); machine learning; and new communication and engagement methods.

## New data sources and big data

**By MODSIM 2015**, the SIG should aim to:

* Develop a clear scope for workflows and guidelines to integrate new data sources, ML and simulation models; and
* Collate a collection of examples of workflows and use cases for new and emerging data sources technologies to facilitate learning and uptake. These examples should be specific to water quality modelling so that they are easier to adopt than more generic examples available in existing resources.

Other short-term goals could include:

* Development of clear ethical guidelines for assessment of social and other risks of adopting new technologies, including respect for traditional owners and affected stakeholders;
* Working towards consistency in adoption of existing protocols to ensure that our models, workflows and model outputs are discoverable, accessible, usable, interpretable, well documented, and transparent.
* Ensuring that data management is routinely included in project scoping and planning so that data and metadata management are adequately funded and projects are held accountable.

**Within five years**, we would like to be in a position such that:

* We have good workflows and guidelines to integrate new data sources, ML and simulation models;
* We have new tools to help us evaluate, format and process data from a wide range of emerging sources;
* Observational data is more easily available (discoverable, accessible, free or funded, usable, useful, transparent and trust-worthy; and is provided in a format that is readily translated into something meaningful);
* Our model output data and workflows are also discoverable, accessible, usable, interpretable, reproducible, well documented and transparent. Protocols have been reviewed, improved and unified across the Tasman;
* We have education and training opportunities so that we can handle big data more confidently;
* Modellers consistently work with data collectors and have a clear statement of needs to ensure that it adds value to modelling and decision-making. We also add value to data platforms.

## Machine Learning and AI

ML and AI can be used in water quality modelling in a wider range of ways. Approaches include:

* Using ML *instead* of traditional water quality modelling;
* Using ML *in ensemble with* traditional models;
* Using ML to generate *better input data* and boundary conditions for traditional models (e.g., spatial distribution of observational data, derived variables from observed variables, interpretation of satellite data);
* Using ML *surrogates* for simulation models (i.e., ML models trained on input and output data from traditional models), for example to enhance sensitivity analysis, for parameter estimation, for forecasting, or to create responsive user interfaces to explore scenarios;
* Using ML to *improve* traditional models, for example:
  + For data assimilation;
  + Using genetic algorithms to evolve improved models and ML to generate ensembles of traditional models;
  + Using ML to generate synthetic data to train traditional models.
* Using ML for *visualisation and interpretation* of model output;
* To fill gaps or extend the scope and spatial or temporal domain of simulation models;
* To improve and speed up modelling *workflows* – for example, an AI-guided model set up process.

While all of these approaches are being used in water quality modelling, in most cases uptake is still in its infancy and most water quality modellers are not yet familiar with or confident in their use. To enhance effective and efficient uptake, we should aim to:

* Develop a platform for accessing and automating water quality modelling tools, datasets, workflows and communication tools; and
* Enhance ML capacity in the WQ modelling community through:
  + More accessible training with WQ-specific workflows; and
  + Establishment of a community of practise or seminar series

We can also use our influence to advocate for:

* + Funding for small projects that are opportunities for modellers to try out ML methods;
  + More professional development time for modellers;
  + Funding of specialists to help other modellers to work with new ML techniques.

## Emerging computing technologies

Uptake of new and emerging computing technologies (high performance computing, GPU programming, cloud computing, new programming languages and – in future – quantum computing) by the water quality modelling community is uneven: high performance computing workflows have become common in ocean modelling, but are less common in lake modelling, while newer computing technologies such as cloud computing and GPU computing are less widely used.

We would like to work towards a future in which the WQ modelling community is skilled in using these technologies and know how best to use them. Objectives to this end include:

* Increasing the availability and uptake of relevant training, including training in HPC and emerging programming languages for computation efficiency (e.g., Julia);
* Sharing our skills and knowledge about what training resources are available;
* Undertaking a horizon scan with input from computer scientists;
* Collaboration to improve access to HPC resources and bring down costs.

If appropriately adopted, new computing technologies should lead to better, more informative models through better uncertainty estimates, longer runs, higher resolution, better process representation, wider domains, and greater use of ensembles. To ensure these outcomes, we need to be well informed so that we make appropriate decisions about when to use HPC resources and when to keep things simple.

## New communication and engagement methods

Research in other fields has improved understanding of how to effectively communicate and educate, but these lessons have not yet been widely adopted in communication of water quality modelling work. Through collaboration with computer scientists, social scientists, graphic artists, science communicators and experts in education and public relations, there is an opportunity to develop better ways to communicate water quality information, tailored to the needs of our stakeholders. This may include:

* Better communication and engagement throughout the modelling process, resulting in shared ownership;
* Better use of apps, web tools and emerging tech such as virtual reality, augmented reality, and gaming tools; and
* Following and tailoring protocols for participatory modelling.

A useful first step towards these outcomes would be development of a toolbox to help water quality modellers decide which communication methods are appropriate depending on context (e.g., on the inform --- empower scale).

# Priorities for the Special Interest Group

While some of the goals described above require broader engagement from the wider community, in the short term, it would be helpful and within the resources of the SIG over the next five years to:

1. Collate a collection of examples of workflows and use cases for emerging and under-utilised technologies and methods to facilitate learning and uptake. These examples should be specific to water quality modelling (so that they are easier to adopt than more generic examples available in existing resources) and should include both successes and failures. Examples might include:
   * A workflow for accessing remote sensing chlorophyll data and assimilating it to enhance lake model parameterisation.
   * A workflow for implementing uncertainty analysis for a water quality model using multiple nodes on a HPC machine.
   * A workflow for training a deep learning algorithm on the output from a water quality hindcast model to provide a tool for water quality forecasting.
   * A workflow for using ChatGPT or Gemini to more quickly generate custom visualisations from netCDF formatted model output.
   * Example communication workflows.
2. Establish a community of practise or seminar series to share learnings between groups that have experience in adopting these technologies;
3. Work with experts to develop guidelines for assessing social and other ethical implications of adoption of new technologies in our work;
4. Publish a paper documenting this discussion and the potential for greater use of emerging technologies in this domain;
5. Develop a toolbox to help water quality modellers decide which communication and engagement methods are appropriate in a given context;
6. Conduct a horizon scan with input from computer scientists.

In the longer term, we should also aim to:

1. Establish a platform for accessing and automating water quality modelling tools, datasets, workflows and communication tools