

Safe to Swim? Rapid Risk Forecasting for Horizons Regional Council Contact Recreation Sites

Utilising rainfall, flow, and site history to predict daily public health risks in-lieu of physical sampling

1. Background Throughout New Zealand, regional councils monitor recreational water quality to assess public health risks from faecal contamination—indicated by *E. coli* in freshwater and enterococci in coastal environments. Traditionally, this monitoring relies on routine physical sampling followed by laboratory analysis. However, laboratory analysis inherently introduces a 24 to 48-hour delay before results are returned. Factoring in transportation and reporting protocols, a sample collected on Monday morning may not have its results communicated to the public until Wednesday or Thursday. Because water quality can change rapidly following heavy rain, this lag can result in the public being given out-of-date information that does not reflect current health risks. To bridge this gap, organisations are shifting toward dynamic "nowcasting" models that synthesise historical sampling data with real-time environmental variables to predict swimmability on a more frequent basis.

Note on Public Reporting: Physical sampling data and real-time predictions generated by these models are reported to the public via the Land, Air, Water Aotearoa (LAWA) website. The LAWA platform serves as the national "one source of truth" for recreational water quality communications, displaying real-time predictions alongside long-term site grades in its "Can I Swim Here?" section so that the public can make informed decisions before swimming.

2. Challenge Overview / Problem Specification The core problem is to determine whether we can accurately predict the real-time risk to swimmers in lieu of physical sampling, using environmental drivers such as rainfall, turbidity, and river flow. A key component of this mathematical challenge is managing the asymmetric, uneven risk associated with the model's predictions:

- **False Positives:** If the model predicts it is unsafe to swim when the water is clean, it creates an opportunity cost and a loss of recreational amenity.
- **False Negatives:** If the model predicts the water is safe when it is contaminated, there is a severe public health risk. For instance, a recent evaluation of the Greater Wellington Regional Council (GWRC) model [2] highlighted that eliminating false negatives must be the primary focus to protect public health.

The study group will need to determine whether a single, generic model can be applied across all swim sites, or if localised models—accounting for complex hydrodynamics and individual catchment topographies—are required.

3. Key Factors and Variables To build a robust predictive framework, the mathematical approach will need to consider several physical and environmental parameters. Current models in New Zealand commonly utilise:

- **Rainfall:** 24-hour and 48-hour observed rainfall.
- **Hydrological Dynamics:** River flow (discharge) rates.
- **Site History:** Long-term microbiological history, often expressed as a 5-year Swimmability grade.
- **Catchment Characteristics:** Upstream land use (e.g., urban vs. pastoral).
- **Turbidity Data:** discrete spot sampling and (less common) continuous data.

4. Data Available The team will be provided with datasets to train and validate the predictive models, potentially including:

- Historical microbiological lab results (concentrations of *E. coli* and enterococci) spanning multiple years.
- Continuous hydrological data, including river flow and rainfall data.
- Upstream land-use characteristics.

5. Desired Outcomes The goal is to develop a computational model (or suite of models) capable of outputting a reliable daily swimmability risk score for recreational sites, which can then be fed directly into public-facing platforms like LAWA. The ideal solution will outline a clear algorithm that successfully penalises false negatives to prioritise public health while remaining sensitive enough not to issue blanket "unsafe" warnings all summer.

Model Transparency and Interpretability: Regulators must be able to defend their decisions to the community, making model transparency critical. The study group should explicitly consider the trade-off between model complexity and usability. An opaque, "black-box" artificial neural network that achieves 99% accuracy may be less useful operationally than a slightly less accurate but highly interpretable, framework like a decision tree. We strongly prefer an interpretable model where the exact drivers of risk (e.g., "the algorithm flagged this site because of the 48-hour flow rate") can be easily audited by staff and clearly communicated to the public.