Freshwater Management Tool

August 2021

FWMT Report 2021/5

Report 5
Review of the Freshwater Management Tool Baseline State Assessment (Rivers)

Auckland Council
Te Kaunihera o Tāmaki Makaurau

Freshwater Management Tool: Report 5. Review of the Freshwater Management Tool Baseline State Assessment (Rivers)

August 2021

Contributing authors:

Australian Rivers Institute, Griffith University, Brisbane
David P. Hamilton

National Institute of Water and Atmospheric Research (NIWA), Napier
J. Christopher Rutherford

Taiao Natural Resource Management
Nic Conland

Auckland Council Healthy Waters Department
FWMT Report 2021/5

ISSN 2815-9772

ISBN 978-1-99-100275-4 (PDF)

Recommended citation
Auckland Council (2021). Freshwater management tool: report 5. Review of the freshwater management tool. Baseline state assessment (rivers). FWMT report, 2021/5. Prepared by the Australian Rivers Institute, NIWA and Taiao Natural Resource Management for Auckland Council.

Review of Freshwater Management Tool

Baseline State Assessment (Rivers)

Report to Auckland Council by the Peer Review Team

7 June 2021

David P. Hamilton[1], J. Christopher Rutherford[2] and Nic Conland[3]

1. Australian Rivers Institute, Griffith University, Brisbane
2. National Institute of Water and Atmospheric Research, Napier
3. Taiao Natural Resource Management

Author (reviewer) contact details

Professor David Hamilton
+61 7373 53544; +61 429 395 041
david.p.hamilton@griffith.edu.au

Dr J. Christopher Rutherford
+64-6-833-7712
kit.rutherford@niwa.co.nz

Nic Conland
+64-27-413-5699
Nic.Conland@thestream.nz

Report citation

Hamilton, DP, Rutherford JC, & Conland, N (2021). Review of Freshwater Management Tool Baseline State Assessment. Report to Auckland Council by the Peer Review Team. Australian Rivers Institute, NIWA and Taiao Natural Resource Management.

Preface

This report summarises information from a peer review process undertaken for Auckland Council by the three authors to evaluate the Freshwater Management Tool developed for Auckland Council by Paradigm Environmental Hydraulic Analysis Limited (USA) and Morphum Environmental (NZ).

Acknowledgements

The authors acknowledge the valuable assistance of the Auckland Council team (Dr Theodore Kpodonu, Dr Tom Stephens and Nick Brown), Dustin Bambic, John Riverson and David Rosa (Paradigm Environmental Hydraulic Analysis Limited) and Caleb Clarke (Morphum Environmental).

Review of Freshwater Management Tool Baseline State Assessment (Rivers)

Plain English summary

Auckland Council is responsible for the development and implementation of plans and policies to improve the environment and reduce the risks of contamination of streams and other receiving waters (e.g., lakes, beaches, harbours) in the Auckland region. Monitoring is an important part of this responsibility and is used to report on the state of the environment (SoE) and identify pressures affecting the current environmental condition. The SoE monitoring is expensive and captures only a subset of the regional stream network and the variations in stream flow and contamination concentrations, it is complemented with models that are designed to fill the gaps between monthly measurements and locations across the stream network. Importantly, a model that satisfactorily reproduces measured data can be used to inform plans and policies by testing current performance and generating future scenarios to predict what might occur using stream data generated by the model, with the input data adjusted to represent what is being tested. Such a model is also useful to test different environmental pressures (e.g., land use change, climate change).

Auckland Council has made a significant investment in a stream flow and contaminant model, known as the Freshwater Management Tool (**FWMT**), which includes a lake component. To understand the usefulness of the FWMT to inform plans and policies, the Council commissioned an external peer review of the four draft reports which document the modelling process. The draft reports include (1) data inputs to the FWMT, (2) the performance of the FWMT to assess its suitability for scenario generation, and FWMT simulations of the baseline state of (3) rivers and (4) lakes. This report presents the findings of the peer review team (**PRT**) that was composed of three experienced modellers familiar with the national and regional context for applications of models to inform plans and policies.

Stage 1 of the FWMT sought to satisfactorily simulate the baseline state of streams in the Auckland region based on six contaminants of concern: nitrogen (N) – total and dissolved forms, phosphorus (P) – total and dissolved forms, copper (Cu) and zinc (Zn) – total and dissolved forms, sediment – total suspended solids (TSS) and faecal indicator bacteria – *E. coli*. Other variables simulated by model were also used to assess its baseline state performance and included stream flow conditions, dissolved oxygen (DO) and pH. Future iterations of the model have the potential to simulate conditions for aquatic plants growth and decay cycles (large plants like macrophytes and microscopic plants like planktonic algae).

In the opinion of the PRT, the FWMT is a suitable framework for modelling the baseline state of flow and contaminant generation and delivery to streams at catchment scale but there are some aspects of the model that could be better aligned with input data representing the local conditions, biophysical processes and the management requirements of Auckland Council. Some predictions by the FWMT are classified as 'unsatisfactory' and should be the focus of the next application of the FWMT. The lake FWMT is not currently adequately developed to support scenarios at this stage and requires integration with the stream FWMT to properly quantify streamflow and nutrient loads to lakes.

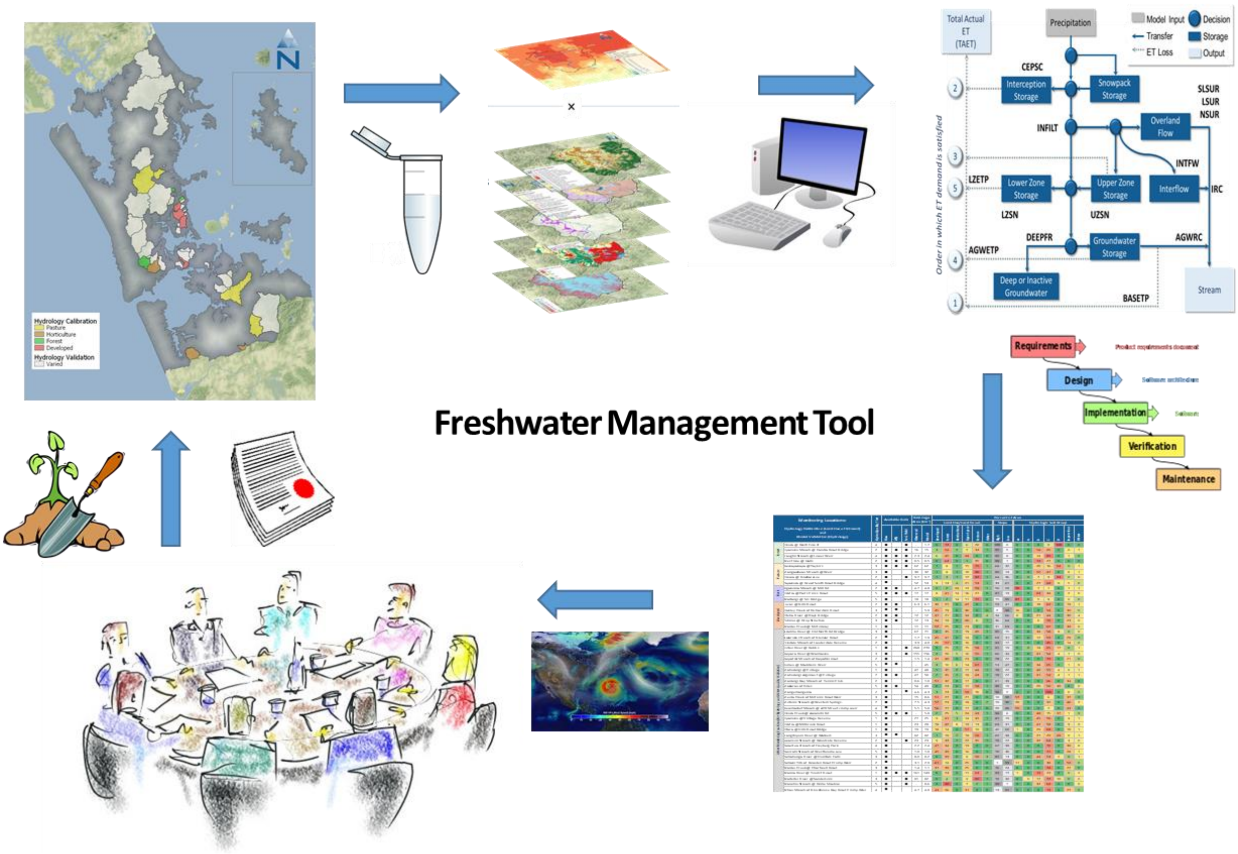
The PRT identified a need to clearly define the way in which the FWMT is used to consider strategic regional planning for infrastructure, rural land use and operations, as well as mitigation activities that include changes in land use practices, wastewater treatment and green infrastructure enhancement. This

forward planning is essential to extend the value of the FWMT beyond the baseline state assessment but also means that consideration needs to be given to complementing the FWMT with other models specifically suited to the management and planning questions being asked.

The PRT also identified a need to improve the assessment and communication of the FWMT performance so that the model can be demonstrated to be a robust and reliable tool for comparing the relative performance of different planning and management scenarios. Following the identification of regional values and community goals, additional work on key processes in the model that affect DO and aquatic plants, informed by local observations and knowledge, are expected to improve the model fit to nutrient observations (nitrogen and phosphorus species). This development step will build confidence in the utility and forecasting capability of the model. In addition, some technical details of FWMT data reporting can be improved, including understanding what leads to uncertainty in the model output, including differences between the model output and the observations of variables.

The PRT supports the modelling strategy that Auckland Council has initiated in developing the FWMT for the planning of urban and rural development, good farm practices, assessment of life cycle costs and optimising investment in catchment mitigations/remediation. The PRT also supports the FWMT in assisting the Council to address the goals of the National Policy Statement for Freshwater Management to set and achieve objectives for water quality in freshwater management units in the region.

Graphical abstract



Auckland Council collects information for the region (top, left) in a variety of formats, including georeferenced data (top, middle), suitable for application in a computer mathematical model (top, right) that is calibrated and validated (middle, right) to a baseline configuration (bottom, right) that can be visualised (bottom, middle) and assessed to develop plans and policies (bottom, left) that are implemented for the region (middle, left).

Executive Summary

Auckland Council is supporting the development of plans and policies, including life cycle costs and remedial activities, with a Freshwater Management Tool (**FWMT**) that uses model simulations to generate 'baseline state' of water quality and future outcomes from potential remedial activities.

As part of a peer review process, Auckland Council commissioned three scientists experienced in the evaluation and application of environmental models, to review the FWMT. The review documents consisted of four reports provided to the Peer Review Team (**PRT**) by Auckland Council, including: (1) FWMT Baseline Data Inputs Report, (2) FWMT Baseline Configuration and Performance Report, (3) FWMT Baseline State (Rivers) Report, and (4) FWMT Baseline State (Lakes) Report. The specifications for the review were contained in the Terms of Reference for the PRT, which asked the PRT to consider 11 main points important to the function, use and performance of the FWMT.

Modelling is a key part of Auckland Council's tools for developing well-informed plans and policies to improve the environment and reduce the risks of contamination of streams and other receiving waters (e.g., lakes, beaches, harbours) in the Auckland region. Monitoring is also used to report on the baseline state of the environment (**SoE**) and identify pressures affecting the current environmental condition. The SoE monitoring captures only a subset of the regional stream network and the variations in stream flow and contaminant concentrations. It is therefore complemented with models that fill the gaps between routine SoE measurements and locations across the stream network, as well as providing data with which to calibrate and validate the model(s). Satisfactory model performance, demonstrated by statistical measures comparing model output with observed data, provide a foundation on which to inform plans and policies, allowing increased confidence in model scenarios of different possible management actions and mitigation strategies. The current application of the model may also be useful in future to test and integrate other environmental pressures (e.g., climate change) that form a basis for well-informed future planning.

In the opinion of the PRT, the FWMT is a suitable framework for modelling aspects of hydrology and contaminant generation and delivery to streams at catchment scale but aspects of the model application can be strengthened for subsequent phases beyond the baseline state performance assessment. Some predictions by the FWMT are classified as 'unsatisfactory' and should be the focus of the next application of the FWMT. The PRT supports the identification of the six priority contaminants (total and dissolved forms of N, P, Cu and Zn, and TSS and *E. coli*) but recommends that greater focus is directed towards DO, linking it to metabolic processes of production and respiration, particularly with relevance to aquatic plant growth in streams and the importance of DO for biogeochemical processes. For lakes, the FWMT catchment hydrology and contaminant loads should be linked to the stream FWMT, to generate streamflow and nutrient loads that improve estimates of water residence time and contaminant attenuation processes in lakes, as well as supporting management actions that can improve the relatively poor current grading of lakes based on the NPS-FM National Objectives Framework (**NOF**).

An opportunity exists to increase the reliability and validity of the FWMT model output using sensitivity/uncertainty analysis of the input data, model structure and calibrated parameters and to better communicate the strengths and weaknesses of the model to stakeholders. This is an important step to finalise the Stage 1 process, including that:

- where there may be uncertainty in the model output, the causal factors are reasonably well understood and can be taken into account when making decisions, and

- clearly communicate where modelled scenarios align sufficiently well with actual outcomes to enable confidence in whether targets can be attained and differences among management scenarios are realised, and if alignment is unsatisfactory seek alternative management approaches.

A wider range of comparative measures to assess model output against observed data will help to deepen understanding of the performance of the model over a range of environmental conditions. This will assist with clarity of communication about uncertainty and risk that is required by Auckland Council, to build confidence in the suitability of the FWMT to inform stakeholders and guide management.

The PRT recommends making use of a pressure-state-response framework in assessing the six major contaminants modelled in the FWMT. This framework would focus on the contaminants of concern and the sources, transport and fate of each contaminant, including attenuation and lag times. Future FWMT reports could then consider how the FWMT predicts concentrations, loads and impacts, and use the recommended risk analysis to determine whether or not proposed policy and mitigation actions will achieve regional and NPS-FM targets.

The PRT recognises that a great deal of work has gone into the Baseline FWMT configuration and that it has the potential to be a valuable tool to support decisions made by Auckland Council. It will also be looked upon as an exemplar for assisting catchment-level decision making by regional councils in New Zealand. The PRT hopes that its key recommendations can lead to realisation of the potential of the FWMT to be used routinely to optimise investments in mitigation actions and assist with planning and policy decisions by Auckland Council.

Table of contents

Introduction

Four reports have been generated as part of this phase of the Freshwater Management Tool (FWMT) development and each has been peer reviewed by the Review Committee. The reports are:

1. FWMT Baseline Data Inputs Report – describes sources of input data used for model configuration, calibration and derivation of the Baseline State Assessment of rivers and lakes;
2. FWMT Baseline Configuration and Performance Report – describes model configuration, calibration and validation;
3. FWMT Baseline Rivers Baseline State Report – describes the current attribute state of Auckland's rivers; and
4. FWMT Baseline Lakes Baseline State Report – describes the current attribute state of Auckland's lakes.

The peer review process commenced with delivery of the four reports to reviewers, followed by a virtual initiation workshop on the FWMT development and review process, as set out in Section 9 below.

The review committee included:

Professor David Hamilton, Australian Rivers Institute, Griffith University, Brisbane, Australia

David Hamilton is the Deputy Director and a Professor at the Australian Rivers Institute, Griffith University. He has held positions in Environmental Engineering at the University of Western Australia and Biological Sciences at the University of Waikato in New Zealand. He was the inaugural Bay of Plenty Regional Council Chair in Lake Restoration at the University of Waikato in 2002 and held this position for 15 years. His research makes use of measurements and models to map the recovery of freshwater ecosystems. He has been closely involved in management and policy implementation for freshwater ecosystems, involving advisory groups with the Ministry for the Environment (NZ) and for regional councils and industry groups in New Zealand and Australia. He is editor-in-chief of the scientific journal Inland Waters.

Dr J. Christopher (Kit) Rutherford, National Institute of Water and Atmospheric Research (NIWA), Hamilton

Kit Rutherford is a Principal Scientist Emeritus with NIWA where he was previously employed as Principal Scientist Freshwater and before that with the Ministry of Works and Development and the Department of Scientific and Industrial Research since 1973. His role is in the development and application of mathematical models of nutrient runoff from catchments, and their effects on water quality and aquatic plants. Kit holds degrees of BE (1970) and PhD in Engineering Science (1975) from the University of Auckland. He has worked internationally on projects where he was based in the Water Research Centre, UK (1976-1977), the University of Exeter, UK (1991), the University of Boulder, USA (1995), the Swiss Federal Institute for Water and Wastewater (2001) and CSIRO Land & Water, Australia (2003-2004).

Nic Conland, Taiao Natural Resource Management (Taiao NRM),

Nic has 25 years' experience in the environmental assessment field being involved in both the model design, preparation and review of water quality effects assessment modelling and in field monitoring, managing, reviewing and reporting on water quality programmes at the farm, enterprise and catchment scale. Since 2013, Nic has provided strategic advice and science communication to the primary sector on their water quality impacts and potential mitigations for catchment scale management within a freshwater quality accounting framework under the National Policy Statement for Freshwater Management (NPS FM 2014; 2017; and 2020) and regional plan developments in the Northland, Auckland, Waikato, Taranaki, Manawatu, Bay of Plenty, Wellington, Hawkes Bay, Canterbury, Otago and Southland.

Auckland Council was represented by:

- Theodore Kpodonu, Tom Stephens, Nick Brown: Auckland Council, Healthy Waters FWMT Project Management Team.

The modelling group (referred to as the consultants) that developed the FWMT included:

- Caleb Clarke, Hana Judd, Andrew Rosaak: Morphum Environmental, Healthy Waters FMWT Consultants, and
- Dustin Bambic, John Riverson, Khalid Alvi, David Rosa, Xiaomin Zhao: Paradigm Environmental Hydraulic Analysis Limited, Healthy Waters FMWT Consultant.

The authors particularly acknowledge the valuable assistance of a smaller group involved in consultation with the reviewers, consisting of Theodore Kpodonu, Tom Stephens and Nick Brown (Auckland Council team, Dustin Bambic, John Riverson and David Rosa (Paradigm Environmental Hydraulic Analysis Limited) and Caleb Clarke (Morphum Environmental).

Additional assistance in the review process was provided by Auckland Council and Council-controlled organisation (CCO) subject matter experts (SMEs).

Purpose and Scope of Review of Baseline State Assessment

The purpose of the review is to provide an independent assessment of whether the Baseline State Assessment, processes leading to the determination of baseline state and its derivation as described in the four reports had been adequately described and met, including whether it provided robust evidence of freshwater quality (contaminant) baseline state to support the wider purpose and objectives of the FWMT.

The Terms of Reference state that the objective of the Baseline State Assessment of the FWMT is '…*to report on the baseline state...of Auckland's streams, rivers and lake water quality…to satisfy freshwater accounting and grading requirements; offer a robust evidence base for decision-making and reporting; and offer a decision-support tool for stakeholder engagement*…' and the purpose of the review is '…*to provide an independent assessment of whether the Baseline State Assessment…provides robust evidence of…baseline state to support the wider purpose and objectives of the FWMT*…'

The Terms of Reference further state that the purpose of the FWMT is to '…*provide a robust evidence base for water management decision making across Auckland Council*…' While this purpose logically follows on from the Baseline State Assessment it encompasses Future State Planning which extends beyond the scope of this review although we are asked to comment on whether the FWMT can provide support for Future State Planning and Assessments (**FSP**).

The Review Committee was tasked to assess the four draft FWMT reports for:

A. Clarity of purpose for the model development; whether the FWMT as described in the reports is adequate to be used in providing robust evidence case for water quality management decision making across Auckland Council.
B. Validity of assumptions on which the model was built (data preparation, data wrangling, model configuration and parameterization).
C. Suitability of input data for the configuration, calibration and validation of the model (whether the 15-minute/1-hour model outputs can be (a) validated with measured data and (b) scaled to annual data suitable for (i) the NPS-FM assessment and (ii) the (approx.) time scales on which land use-contaminant relationships are typically developed.
D. Suitability of the model configuration, including approach taken for contaminant source modelling via Hydrological Response Units (**HRU**s). It will include appropriateness of parameterization, calibration and validation steps at the HRU scale, choice of performance measures and bands therein, as well as narratives on model performance.
E. Reporting of baseline state and associated modelled information (e.g., contaminant sources, events or conditions associated with water quality variation). Additionally, comment may be made on whether FWMT configuration and performance is suitable for scenario assessment to optimized management strategy development (e.g., reporting changes in water quality grading, sources, pathways and events or conditions).
F. How changes in HRUs or boundary conditions (e.g., climate) might affect optimized strategy development.
G. Approach taken to grading or assessing water quality (contaminant) state, including implications of using simulations for grading against National (and Regional) Objective Framework numeric attribute states (i.e., potential for the latter not to have considered modelling applications that provide continuous time-resolved outputs and their effects on model reporting).
H. The appropriateness of the approach used in the lake component of the model (e.g., identification of limitations and recommendations for addressing limitations).
I. Limitations of the FWMT Stage 1 including issues that may arise for its use in strategic (regional planning) and operational (plan implementation) management aligned with the Future State Assessment.
J. Recommendations for improvements to the Baseline State Assessment in FWMT Stage 1 and directions for FWMT Stage 2, including potential expansion of capability.
K. Any other factors that affect the ability of the Model to meet its purpose and objectives (see Section 2).

Review Deliverables

The deliverable for this review is an independent report coordinated by the lead reviewer, that provides reviewers' annotations on the existing FWMT reports, details the findings and recommendations of the Review Committee as outlined in Section 8, and gives any agreed additions from the required final review workshop as set out in Section 9.

Review Timeline

1. All reports sent and received by reviewers – 23 Dec 2020
2. Initiation virtual workshop – around 20 to 29 Jan 2021
3. Review of draft Baseline State Assessment reports – around Jan to Apr 2021
    - 10 Feb-Report 1
    - 28 Feb- Report 2
    - 31 Mar- Report 3
    - 30 Apr- Report 4
4. Finalisation virtual workshop – 5^{th} May 2021
5. Peer review report delivered – 15^{th} June 2021

Layout of Report

The Purpose and Scope of Review of Baseline State Assessment (Rivers) has provided a working framework for this report. Specifically, sections in the review are outlined according to points A to K above.

General comments

The reviewers are grateful to the modellers and Auckland Council for the clear and open dialogue and availability of information that enabled this review. We commend the modellers for work done in collating information, documenting the current calibration of the FWMT and summarising findings to date. The reviewers' aims are, through questioning and commenting, to try to improve the clarity of the reports and add value where possible, so that Auckland Council can make best use of the FWMT in its current form and through ongoing model development. The reviewers appreciate the difficulties faced by those involved and in no way are seeking simply to 'find fault'; rather, as reviewers, we are seeking to 'add value'.

The FWMT is guided by the overarching framework of the National Policy Statement for Freshwater Management (**NPS-FM**), which states that:

"*…the NOF [National Objectives Framework] process requires regional councils to…identify FMUs [Freshwater Management Units]…set values for each FMU…set outcomes for each value…define attributes for each value…define baseline states for those attributes…set target attribute states…set rules and action plans to achieve environmental outcomes*…".

The NPS-FM sets the overarching framework from which plans, policies and processes are developed by Auckland Council.

The objective for Auckland Council in Stage 1 of the FWMT was to develop a modelling framework suitable to:

1. Determine values and environmental outcomes (done by managers and stakeholders);
2. Identify pressures and attributes (engineers and scientists);
3. Set targets (dialogue with managers, stakeholders, engineers, scientists, etc.);
4. Identify a suitable model (or models) (including input, calibration and testing data);
5. Calibrate the model and undertake sensitivity/uncertainty analysis;
6. Make predictions, including risk analysis.

The reviewers interpret that the Loading Simulation Program (**LPSC**) model was selected as the basis for the FWMT because of its widespread and successful use in the USA and because the modellers (primarily Paradigm) have considerable experience and a deep understanding of LPSC and the associated System for Urban Stormwater Treatment and Analysis Integration (SUSTAIN) model. The reports presented to the reviewers and ensuing discussions suggest to us that stormwater management is also the major focus of the FWMT and likely influenced the choice of LSPC and SUSTAIN.

In our opinion the FWMT is a suitable framework for modelling aspects of hydrology and contaminant generation and delivery to streams at catchment scale but there are some aspects of the model parameterisation that could be better aligned with input data representing the local conditions, biophysical processes and the management requirements of Auckland Council, as expanded on in the sections below.

Previous studies in the Auckland region have highlighted that nutrients, sediment, faecal matter and metals (copper and zinc) are likely to be the most widespread and serious risks to coastal and freshwater quality. Stage 1 of the FWMT includes the model variables of six priority contaminants: nitrogen (N) – total and dissolved forms, phosphorus (P) – total and dissolved forms, copper (Cu) and zinc (Zn) – total (directly) and dissolved forms (indirectly), sediment – total suspended solids (TSS) and faecal indicator bacteria – *E. coli*. The purpose of the review (as set out in the Terms of Reference) includes assessing whether the four reports adequately describe the model and whether it adequately predicts baseline state (viz., satisfactorily models the six priority contaminants). We note that there is a well-established framework for modelling these contaminants and a large body of work to support the setting of guidelines and targeted mitigations.

The Terms of Reference also include reviewing the ability of the model to support wider objectives (including additional contaminants and values). While the choice of the six priority contaminants is understandable, the reports indicate that considerably less attention has been directed to other important values, attributes, processes and variables (e.g., dissolved oxygen (**DO**), pH, aquatic plants (planktonic and benthic algae, and macrophytes), macroinvertebrates, fish (index of biotic integrity F-IBI), deposited sediment (percentage fine sediment cover)) that are relevant to the National Objectives Framework (**NOF**) of the National Policy Statement for Freshwater Management 2020 (NPS-FM) or are important ecologically. We provide details below that support our reaching this conclusion.

The reviewers understand and support the basis for the process-based modelling approach adopted in the FWMT, as it relates to the opportunity to increase evidence of causality alongside any correlative statistical relationships. It also related to ability to carry out model runs in response to possible changes in the attributes and their values with refinement of the National Objectives Framework. The overall confidence in the scenario results generated from the FWMT will be tied to several factors that are examined in greater detail in the sections below. Briefly, these include confidence in the model structure and processes, selection of relevant variables, modelling techniques that enable sufficient confidence in parameter values to support scenario generation, model alignment with the adequacy of input data, uncertainty and risk analysis and skilled communication strategies that clearly articulate the relevance and importance of the modelling in decision making.

A. Clarity of purpose of model development

This section addresses whether the FWMT as described in the reports is adequate to be used in providing robust evidence case for water quality management decision making across Auckland Council.

The NPS-FM gives four Compulsory National Values – ecosystem health, human contact, mahinga kai and threatened species (NPS-FM Appendix 1a). The attribute used by the NPS-FM to quantify the risk to human contact for recreation is *E. coli*. The FWMT includes this as one of the six variables modelled in Stage 1 and the Baseline State Assessment – Rivers reports compliance with guidelines in the NPS-FM.

The NPS-FM encourages an integrated approach to recognise the interactions "…*between freshwater, land, water bodies, ecosystems, and receiving environments*…" (NPS-FM section 3.5). Ecosystem health is affected by contaminants, water chemistry, nutrients, algae (including blooms), sediment, temperatures, oxygen, invasive and endemic species, flow regimes, essential habitats and connectivity of waterbodies. The FMWT models five contaminants that affect ecosystem health (N, P, Cu, Zn and TSS) and flow regimes across the model domain. The NPS-FM sets standards for N toxicity (NO_3-N and NH_4-N) and the Stage 1 reports address this issue. We note that standards/guidelines for Cu, Zn, TN, TP and SS are not defined in the Stage 1 FWMT reports – but are needed when undertaking future state analysis.

The FWMT is designed to (a) help quantify the current (baseline) state and (b) assist with adaptive planning. However, the stated scope of the FWMT is somewhat narrow and mention could be made of remedial activities such as infrastructure planning (urban development including earthworks, etc.), rural land use planning (pastoral and horticultural), industrial waste management, domestic wastewater management and stormwater overflows that have the potential to affect freshwater ecosystems and water quality. We believe greater use of the well-accepted 'pressure-state-response' framework would improve the reports; we provide specific examples below where this process can be implemented. We also suggest reviewing what is written in the Introduction to better clarify the purpose of the model. For example, future state scenario testing and analysis is a key objective of model development and application to support planning and policy implementation.

The reviewers acknowledge that Stage 1 deals with six important attributes and consider it to be a useful first step in helping manage Auckland waterways. The FWMT modelling framework aligns with conventional numerical modelling methods for catchment hydrology and water quality. In common with other catchment models the FWMT simulates the rate at which contaminants are generated by activities in the catchment, and how these contaminants accumulate and are transported into waterways (e.g., sediment associated – rainfall driven, soluble – through leaching) or are discharged directly into waterways (e.g., stormwater overflows). Once in the streams, the FWMT routes contaminants to where they cause adverse effects, accounting for transformations and attenuation by instream processes.

The FWMT predicts baseline state for 2,567 sub catchments, each assigned an 'instream grade' for the modelled reach after calibration to data from 36 SoE sites (water quality) and 46 hydrological stations (water quantity) across the regional monitoring network. The baseline state prediction of the 6 freshwater attributes (N, P, Cu, Zn, TSS and *E. coli*) is graded based on performance metrics determined at the SoE monitoring sites. The model performance at each site is assessed on how well it matches SoE monitoring of continuous concentration and load (daily and annual) using standard statistical outputs (PBIAS, NSE and r^2) to reach a summary grade from "Very Good" to "Unsatisfactory". Model predictions are also assessed on how well they match the SoE NPS-FM grades A to D for freshwater outcomes. While all sites have differences in attribute performance, the reports provide sufficient explanation of the predictive ability of the FWMT (based on performance grading), with generally higher grading for loads than concentrations that reflect higher performance of the model output for stream discharge. At several sites the match between observed and predicted concentrations was assessed to be 'unsatisfactory' and should be addressed in future iterations of the FWMT, including identification of the reasons for mismatches. Auckland Council observes that "lower accuracy in A-graded sites is less concerning than lower accuracy in D- graded sites, for FWMT purposes".

The ongoing investment in the FWMT is intended to support its use in communicating how well predicted flows, concentrations, loads and grades (as defined by the NPS-FM) match observations in monitored sub-catchments. Identifying and communicating the confidence that can be placed in the FWMT performance enables it to be useful to determine the likely baseline state in these sub-catchments and in unmonitored sub catchments, and to estimate the effectiveness of proposed mitigations, action plans and policy objectives of Auckland Council.

Recommendation 1: *The ability to clearly communicate the baseline state conditions outside the 36 stream reaches should be supported by a common language, with consistent use of terms for uncertainty, risk, likelihood and consequence, as well as consistent descriptors for likelihood of occurrence (e.g., very likely, likely, about as likely as not, unlikely, very unlikely). The FWMT can therefore predict the likelihood of whether the six main attributes meet designated NPS-FM grades (or some target concentration for metals) at a site.*

Recommendation 2: *Detailed information should be provided to show how the FWMT can support activities undertaken by Auckland Council to manage its waterways according to the goals of the NPS-FM. This may include adaptive planning related to key infrastructure investments (assessments of 'grey' and 'green' infrastructure), policy development and community consultation. We recommend making greater use of the pressure-state-response framework when discussing the TSS, TN, TP,* E. coli*, Zn and Cu sub-models, which would include more detail on the sources, generation/transport/attenuation/modelling of these processes, the available remedial actions and how these were implemented in the model, and how risk analysis will be undertaken to determine the likelihood that a proposed policy or remedial action will achieve NPS-FM targets.*

The NPS-FM sets standards for periphyton biomass and councils must also set limits on dissolved inorganic nitrogen (**DIN**) and dissolved reactive phosphorus (**DRP**) as instream concentration limits to ensure these standards are met. The Stage 1 modelling does not include periphyton or the effects of DIN and DRP concentrations on plants. The LPSC model contains a rudimentary periphyton sub-model and Auckland Council says "…*future FWMT iterations might simulate instream ecological outcomes (e.g., periphyton, macrophytes, macroinvertebrates, fish)*…". Periphyton tends to proliferate in shallow, stony-bed streams (e.g., on the east coast of both the North and South Islands) and may not cause widespread problems in Auckland although the four reports reviewed do not make specific mention of periphyton abundance in Auckland streams. The periphyton sub-model might prove useful as a surrogate for aquatic plants (e.g., macrophytes which proliferate in some Auckland streams where they cause problems with diurnal swings of DO and pH, sedimentation and reduced drainage).

The NPS-FM recognises five components that contribute to ecosystem health (water quantity and quality, habitat, aquatic life and ecological processes). Water temperature, DO, pH, TSS, and toxicants strongly influence ecosystem health and water quality, and also affect rates of nutrient transformations. Although the FWMT includes DO, pH and temperature, these variables are not modelled (viz., calibrated or results analysed) in Stage 1, nor are there yet standards in the NPS-FM for these attributes. The FWMT Baseline C&P report hints that future iterations might add these attributes to the FWMT. If successful, such additions would help assess the impacts of management on ecosystem health but would pose challenges because the underlying processes are complex and information to calibrate and test the model may not be available.

The reports indicate that future FWMT iterations will consider simulations of some of the instream attributes that determine ecosystem health (e.g., DO, pH, periphyton, macrophytes, macroinvertebrates (**MCI**) and fish). We note that Stage 1 does not include assessing the effects of aquatic plants on DO and pH, nor the diurnal and seasonal variations in DO and pH that have direct impacts on stream organisms and can be important for nutrient regeneration from bottom sediments and plant uptake/release of nutrients.

The FWMT may not have the capability to predict some NOF attributes (e.g., macrophytes, deposited sediment, macroinvertebrates, fish). We note that DO, pH and temperature may be useful when considering macroinvertebrates and fish – 'safe' DO, pH and temperature are necessary (albeit not sufficient) conditions for healthy ecosystems. Auckland Council has planned a staged approach (i.e., '…*might simulate…periphyton, macrophytes in the future*…'). While this may be pragmatic, we believe that managers need to be alerted if important attributes and processes are not currently modelled satisfactorily and/or may not be amenable to modelling in the future (e.g., because of inadequate data, an incomplete understanding of the processes and the absence of suitable sub-models). Of concern are statements that "…*future FWMT iterations might simulate instream ecological outcomes (e.g., periphyton, macrophytes, macroinvertebrates, fish)*….". There should be a clear outline of whether specific variables or attributes (e.g., macroinvertebrates, fish) will be modelled explicitly – which is very challenging – or suitable proxies (e.g., temperature, DO) used to represent potential habitat for biota.

Recommendation 3: *The variables known to be simulated satisfactorily by the FWMT could usefully be linked to the goals of Auckland Council for managing its waterways, including relationships to the NPS-FM and instream ecological outcomes. This could also include a timeline for variables deemed important and feasible to include in further stages of the FWMT, based on experience from the recently completed modelling, as well as prioritising variables important in adaptive planning related to key investments ('grey' and 'green' infrastructure), policy development and community consultation. Attributes in the NPS-FM and NOF that are not amenable to modelling by the FWMT need to be identified and suitable alternative prediction strategies recommended. Auckland Council may need to manage some attributes without the support of the FWMT and could usefully predict what attributes these may be, how management of them can be undertaken and whether these attributes may be included in the FWMT in the future. This recommendation relates to the need to appreciate that the FWMT does not provide guidance on all the attributes currently in, or likely to be added to, the NPS-FM.*

Following on from discussion about model investment and prioritisation, we are concerned by the statement that "…*the FWMT/LPSC has the capability to model*…' and '…*future FWMT iterations might simulate*…". We warn against the reports 'over-selling' what the FWMT can currently model and might be able to model in the future (e.g., see Recommendation 3). Identifying the strengths of the FWMT (which are many) and the weaknesses (which there are clearly some) will increase the understanding and confidence managers and stakeholders will have in the FWMT. Clear communication among the modellers, Auckland Council, planners and managers, and the community is critical to build trust in model outputs, including scenarios, that can support the planning decision making process. We suggest the report identifies aspects of the FWMT where the modellers:

- are confident it can predict the baseline state of attributes;
- are confident it can quantify the impact of mitigation actions on attributes;
- are confident that advice provided by AC can be improved by model/data refinement;
- feel it can currently provide semi-quantitative information for baseline state and mitigations;
- feel it is only ever likely to indicate 'direction of change' for certain attributes and mitigations; and
- deem it unsuitable for addressing certain mitigations and attributes.

Once strengths and weaknesses of the FWMT have been clearly identified and the structure and processes in the model clearly communicated conceptually, we suggest:

- where the FWMT is perceived to be 'strong' – proceed with
    - refining calibration,
    - uncertainty analysis (including to input data) and
    - scenario modelling future state predictions and analysis.

This process would also involve considering the balance of data allocated between calibration and validation, noting that State of the Environment stations have been used extensively to date for calibration, to ensure region-wide performance, but not for validation purposes. A formal validation process should be considered in subsequent stages of the FWMT.

- where the FWMT is perceived to be 'weak' – decide:

- if it can be used to guide management even given high uncertainty (e.g., to indicate the direction of change) – this approach is suited to future state predictions with multiple dependencies rather than baseline state analysis.
- if the FWMT can be modified, including and assessment of time/effort, data availability, risks of failure, alternative approaches, and
- if not, seek alternative models or approaches that are complementary.

Recommendation 4: *To improve communication we suggest that a succinct summary of the model is included in the report, linking information that is currently dispersed through the text. A 'plain English' summary should also be included to synthesise the technical information and make it accessible to non-modellers (see Plain English Summary; this report). The strengths and weaknesses of the FWMT should be clearly identified to increase the understanding and confidence of managers and stakeholders in the FWMT. For example, the sediment model was generally well summarised and could be used as a basis for similar succinct summaries related to the models for nutrients, metals, and discharge. Relevant graphical or conceptual figures are recommended wherever possible.*

Recommendation 5: *Auckland Council could usefully map out an independent communication strategy document that helps with familiarisation of the FWMT and outlines expectations of how the model will be used to develop planning and policy to optimise investment and meet the goals of the NPS-FM; in effect the value of the model will be tightly linked to Auckland Council's ability to explain it to managers and stakeholders.*

The model summary (*Recommendation 4*) could usefully make greater use of the 'pressure-state-response' framework to link information about the 'pressure' (viz., the contaminant(s) of concern, their source(s) including how the FWMT uses the available information to estimate contaminant generation, their delivery processes to groundwaters, streams, lakes and estuaries including attenuation and lag times), what remedial measures are available and how they will be modelled, the 'state' (viz., how the FWMT predicts concentrations and loads and their impacts including relevant NOF standards), and 'response' including the range of parameters which are adjusted to simulate Council mitigation actions to reduce adverse impacts. This link is important not only for the review team but also for those who will further scrutinise and review model outputs, especially as they relate to those tasked with prioritisation of investments to address the NPS-FM and NOF.

Specific comments

A1: Process summary descriptions

Section 3.9.3 of the Baseline C&P report. These paragraphs provide an informative summary of how sediment is modelled. Summaries at the start of other FWMT sub-models to explain how other variables are modelled would be valuable in supporting good communication about the model generally.

A2: Factors affecting model fit

Section 3.11 of the Baseline C&P report. The reviewers would like to see acknowledgement in the C&P Report that the ability to calibrate the model is affected not only by monitoring data used in calibration but also by the quality/accuracy of the input data and the simplifications made in the model itself. Such a statement may help to clearly identify where further effort should be directed to address data gaps versus model inadequacies.

A3: FWMT staged improvement

The FWMT Stage 1 contributes to understanding aspects of contaminant generation and delivery, as well as identifying information gaps. It may be useful to outline the next steps in relation to the FWMT 10-year continuous (staged) improvement programme (see Recommendation 2 above). We believe it is important to align the long-term vision with the current identified strengths and deficiencies of the FWMT.

B. Validity of assumptions in the model

This section addresses the validity of assumptions on which the model was built, including data preparation, data wrangling, model configuration and parameterization.

All modelling requires assumptions about model structure, input data, calibration methods and parameter values. These assumptions contribute to uncertainty and risk, and the challenge for modellers is to determine how assumptions affect the ability of the model to inform management, and to communicate uncertainty and risk to managers and stakeholders.

In our opinion, the Stage 1 reports reviewed do not currently provide reviewers, managers and stakeholders with clarity around uncertainty and risk, including the effect of assumptions that have been made by the modellers. One source of uncertainty arises from simplifications and omissions of key processes in the model structure (other sources are discussed below). For example, prior to calibrating the FWMT simplifying assumptions were made for zinc and copper (e.g., constant Zn/SS and soluble/particulate ratios), for pasture (two stocking rates) and for HRUs (see section D). While such simplifications are pragmatic, the modellers need to investigate and report on the implications these assumptions might have on model calibration, and particularly on how uncertainty could propagate through to increased risk that mitigation actions (based on parameter settings) might not be successfully represented or that baseline state predictions (to complement sparse monitoring data) might be inaccurate.

Some of the risks can be mitigated to some extent by Auckland Council adopting a staged approach to the modelling and its communication. It is often necessary to cycle through some, or all, of the modelling steps that include data collation, parameter calibration and model scenario development and generation, before returning to these steps to refine the model. We note that Table 2-1 of the

FWMT Baseline C&P report outlines a sequential process but has downplayed the level of iteration that might be required, not only within the modelling team but also among the modellers and Auckland Council and its stakeholders.

As above, we believe the reports would benefit from a more considered summary of the strengths and weaknesses of FWMT Stage 1, particularly as concentration predictions by the FWMT are sometimes classified by the modellers as 'unsatisfactory' (see below). The reasons why continuous performance metrics of concentration do not often yield 'satisfactory' performance needs to be succinctly elucidated (e.g., inadequacies in model input data, model configuration and differences between discrete observed and continuous modelled data, and because continuous observed data are not available to compare with the highly temporally resolved model output) together with the implications for future modelling (e.g., additional monitoring, model refinement, calibration and sensitivity/risk analysis).

Recommendation 6: *Further analysis is required to set out the strengths and weaknesses of FWMT Stage 1. This work includes: (1) clear delineation of calibration and validation processes, possibly scaled to 1-2 sub catchments so that computational times do not become too cumbersome, (2) uncertainty analysis to understand the relative effects of input data, model conceptualization and parameter calibration, and (3) examining in detail the processes of contaminant generation, delivery and attenuation which will help to better inform the possible mitigation actions*.

The draft reports have very limited mention of risks and uncertainties, including no sensitivity analysis or any form of other uncertainty assessment or techniques (e.g., Monte Carlo runs). This omission represents a risk for Auckland Council because scenarios may not necessarily be trustworthy or able to be differentiated (effective or not). Uncertainty analysis is required before scenarios can be relied upon and risk analysis provides managers and stakeholders with greater confidence when assessing the effectiveness of different scenarios. The uncertainty analysis may help to differentiate different model outcomes (e.g., 'likely' and planned for, 'plausible' and should be accommodated for in planning considerations, 'possible' and may need further assessment as observed data and model runs accrue, and 'unlikely' but could be informative for extreme events or outcomes). Framing risk and uncertainty in a similar manner to the IPCC (2014)[1] and its Statement for Policymakers may be useful in this respect.

Another source of uncertainty arises from model calibration. Techniques are available to increase confidence in calibrated parameters (e.g., 'bootstrapping') through calibration of the model to one period (or set of sub-catchments) and comparing observations and predictions for another period (or

[1] IPCC, 2014: Climate Change 2014: Synthesis Report. *Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

set of sub-catchments) not used in initial calibration. The reviewers do not consider this process to be overly complex and recommend that some such procedure is adopted. Such a process might identify, and hence allow an assessment of, the impact of issues such as equifinality (where multiple different combinations of parameters yield similar model outputs). Equifinality can be problematic when the model is being used outside of the range of forcing data for which it was calibrated (e.g., in scenario testing or climate change assessments).

The reviewers accept that limited observations may impose some limitation on the extent to which the different confidence tests could be applied to parameters in the FWMT. The reviewers suspect that running the model across the entire catchment for the full suite of contaminants is computationally challenging. An attractive alternative is to focus on optimising the flow calibration, and then to focus on a single class of contaminants (e.g., Cu and Zn, or TSS depending on what is highest priority) in a subset of sub catchments where those contaminants are of concern and where there is high-quality input and monitoring data. This simplifying approach reduces the required computer run times (currently believed to be 72 hours) and facilitates multiple simulations required to undertake uncertainty/risk analysis and increase confidence in model calibration.

A third source of uncertainty arises from input data as discussed in detail in Section C.

Recommendation 7**:** *Sensitivity/uncertainty analysis of the input data (e.g., meteorology, groundwater contaminant concentrations, etc.), model structure and calibrated parameters should be undertaken. This process will provide confidence in the validity of assumptions made in the model and the reliability of model output.*

Specific comments

B1: Nitrogen model performance

The nitrogen model output does not match well with the high observed variability in concentration at low flow but performs reasonably well at high flows. This indicates difficulties with the modelled structure for baseflow (HRUs, inflows, groundwater and possibly plant/DO/pH effects on attenuation) revealed in the unsatisfactory results for common Council statistics for mean annual low flow. Automatic calibration and sensitivity analysis are unlikely to eliminate this discrepancy if the model structure omits key processes (e.g., GW-SW mass balance, sources of N, seasonal variations in plant uptake, DO/pH effects on N transformations). It could be useful in the reports to expand the discussion on the key processes (or omissions) in the nitrogen model (see also A1 above and the recommendation for good communication of the various sub-models). It will be important in ongoing work (through targeted monitoring and iterative modelling) to improve the N model.

B2: Benthic algae

The LSPC manual describes a model for benthic algae but the implementation of this model in the FWMT is not clear, including reference by the modellers to various sub models being 'activated'. We noted that DO-BOD and plankton subroutines relied on default values. We believe that it may be possible to go a step further, adapting the periphyton model as a surrogate for aquatic plants and improving the sediment oxygen demand sub-model, particularly in light of the importance of DO and periphyton in the NPS-FM and for a number of subsidiary variables. Work by Wilcock et al. (1999)[2], for example, shows the importance of benthic deoxygenation and denitrification in muddy-bottomed streams in the Auckland region, as well as the large diurnal variations in DO and pH in these macrophyte-dominated streams. Modelling by this group (Wilcock et al. 1998)[3] could be used to examine diurnal changes in DO in lowland streams and the major driving factors. A useful starting point would be to establish the impact of benthic algae on the physico-chemical properties of the streams, including the benthic oxygen demand, but the reviewers would also like to direct the modellers attention to the importance of benthic macrophytes in subsequent stages of the FWMT, and make the required linkages to DO and pH.

B3: Dissolved oxygen

The FWMT would benefit substantially through more work done on the alignment of variables with the needs of Auckland Council and the NPS-FM. The reviewers consider that a serious omission from the list of state variable assessments is DO. Dissolved oxygen is an important parameter in the NPS-FM and is a critical determinant of ecosystem health. In parts of the Auckland region issues with DO have been identified (e.g., large diurnal swings – symptomatic of high plant biomass, and low night-time concentrations that adversely affect sensitive organisms). Reports suggest that Auckland Council has had considerable experience with monitoring DO continuously in streams, dating back to 2003, and has even developed its own set of measurement protocols[4]. Diurnal variations in DO are often associated with swings in pH that affect ammonia toxicity and other aspects of stream health and biochemical reactions (e.g., nitrification/denitrification, phosphorus sorption/desorption and benthic releases). Considerations of DO and pH in model simulations should help to 'close the loop' in terms of some of the critical drivers that influence biogeochemical fluxes and concentrations of nutrients and metals in streams.

In Stage 1 the modellers assumed default parameters when modelling DO. We note that the modellers had limited access to continuous DO datasets and there were some problems with some of the data (drift, noise, location, etc.). Nevertheless, it may still be useful to re-examine the available data, 'correct' existing data, and/or seek alternative sources of information. The fact that DO is an important variable in the NOF, and in some AC streams (see reference to Wilcock et al. 1999 above), reinforces

[2] Wilcock, R. J., P. D. Champion, J. W. Nagels, and G. F. Croker (1999). The influence of aquatic macrophytes on the hydraulic and physico-chemical properties of a New Zealand lowland stream. Hydrobiologia 416: 203–214. doi:10.1023/A:1003837231848.

[3] Wilcock, R. J., J. W. Nagels, G. B. McBride , K. J. Collier , B. T. Wilson and B. A. Huser (1998). Characterisation of lowland streams using a single-station diurnal curve analysis model with continuous monitoring data for dissolved oxygen and temperature, New Zealand Journal of Marine and Freshwater Research, 32: 67-79. DOI: 10.1080/00288330.1998.9516806.

[4] Wilcock, B., Young, R., Gibbs, M., and McBride, G. 2011. Continuous measurement & interpretation of dissolved oxygen data in rivers. NIWA Client Report No: HAM2011-010 to Horizons Regional Council.

that it should be explicitly reported on in the FWMT. Although DO was included in Stage 1 modelling, results appear not to have been presented (e.g., indicative diurnal variations).

B4: Dissolved and particulate contaminants

As part of the prioritisation of measurements to support future iterations of the FWMT, it may be useful to consider in more detail both dissolved and particulate Cu and Zn, to better understand the variability of dissolved:particulate metal ratios under different flow regimes and from different sub-catchments. For sediment and phosphorus, consideration could be given to further differentiation of the particulate fraction by particle size; it is likely that much of the particulate phosphorus and *E. coli* is associated with the finest material which may be more difficult to manage compared with the eroded coarse material. There are existing studies of particulate settling and metal absorption[5] for the Auckland region which may help inform parameterising the FWMT and future mitigation options.

B5: Groundwater concentrations

The reviewers note that during Stage 1 *a priori* concentrations (e.g., groundwater nitrogen concentrations) and rates (e.g., sediment/nutrient release rates) are occasionally specified in order to improve model calibration. This approach may be useful to support Baseline State Analysis (provided the uncertainty in specified concentrations and rates is accounted for). However, it restricts model use for Future State Planning (FSP) where changes in the catchment may alter the specified concentrations (e.g., groundwater) or rates (e.g., sediment-nutrient exchange). The reports note that the FWMT could be linked to groundwater models in the future if required and we concur. More challenging is phosphorus-sediment release/adsorption for which no satisfactory catchment-scale process model exists, meaning Council is unlikely to be able to model the effects of P source management satisfactorily using the FWMT.

B6: Communication of model structure and function

Overall, we support the direction for process-oriented modelling but explanation is required of catchment processes to provide the reader with information about how these processes are represented in the FWMT parameterisation processes. A clear explanation of these connections will then transition to the operational basis for management levers to represent in the FSP. There is a risk that, if catchment processes represented in the FWMT have not been represented and calibrated accurately, then FSP will be compromised, especially if a single parameter value is taken to be representative of multiple complex biophysical process. In these circumstances a sensitivity analysis may not address the problem and provide only a range of predictions of the effectiveness of a management action. For example, from the FWMT summaries provided in the draft reports it cannot

[5] Auckland Regional Council (1992) "An Assessment of Stormwater Quality and the Implications for the Treatment of Stormwater in the Auckland Region" Technical Publication 5

be determined how the diurnal ecosystem respiration and primary production processes are represented. We have identified that processes related to DO and aquatic plants should be a particulate focus of attention. These key freshwater processes have a governing role in summer dissolved nutrient levels, which are also an area identified by the Peer Review Team for improvement in the FWMT.

B7: Model fitness for purpose (FWMT phase 2)

Sediment sources (TSS) were estimated using a well-established albeit simple erosion model (a modified University Soil Loss Equation) together with instream deposition/erosion models. Model parameters were calibrated to match available TSS observations at downstream monitoring sites. For permeable surface the sources of three contaminants that sorb to soils (total phosphorus, total zinc and total copper) were estimated from sediment losses using 'potency factors' (viz., TP:TSS, TCu:TSS, TZn:TSS ratios). For impermeable surfaces the sources of contaminants were estimated using build-up and wash-off parameters linked to drivers of contaminant generation (e.g., traffic density and roof materials) and calibrated to observed concentrations. While this is a standard and pragmatic approach, in the absence of sensitivity analysis (not undertaken in Stage 1) there is no way to assess the accuracy of TSS, TP, TCu or TZn source estimates independently of instream attenuation processes. This has the potential to bias estimates of the baseline state and the effects of remediation (e.g., if modelled generation and attenuation in the model are lower than actual values, Council may decide not to invest in source reduction when it could be effective). One way to avoid such bias is to investigate uncertainty in model calibration which may identify several combinations of parameters/coefficients that give equally good fit to observations. If so, remedial actions should be modelled using several combinations of coefficients and the probability assessed of achieving/not achieving the target attribute levels.

Recommendation 8: *Auckland Council can communicate that there is uncertainty in the model parameters, particularly those associated with contaminant source estimates, and demonstrate attempts to quantify such uncertainty through model sensitivity analysis, carrying findings into Future State Planning scenario runs to assess the risk of achieving/not achieving targets for different attributes.*

C. Suitability of model input data

This section addresses the suitability of input data for the configuration, calibration and validation of the model, and whether the 15-minute/1-hour model outputs can be (a) validated with measured data and (b) scaled to annual data suitable for (i) the NPS-FM assessment and (ii) the approximate time scales on which land use-contaminant relationships are typically developed.

An incomplete process understanding, deficiencies in input and monitoring data, and the need to make simplifying assumptions are common problems in catchment modelling. When this occurs, it is incumbent on the modellers to communicate clearly to managers and stakeholders the implications of these deficiencies. The reports stress that model structure, input and calibration data were constrained because the "…*best available data*…" had deficiencies. The report suggests that the accuracy of the FWMT can be improved by additional monitoring. However, monitoring alone may not significantly improve the model if there are problems with the model structure (e.g., processes not included or incorrectly modelled).

Specific comments

C1: Suitability of input data

The reports present GIS layers of data relevant to contaminant sources (e.g., roof materials, traffic density, soils, rainfall, land use/cover) from which estimates of contaminant sources had been made. The reviewers note that these layers were selected based primarily on what data layers were readily available rather than on considerations of what information determines contaminant generation. While the choices made during Stage 1 are pragmatic, Auckland Council could usefully review Stage 1 findings to determine whether or not the current input layers are 'fit for purpose' or whether additional/replacement layers need to be sought. For example, are the current land cover layers adequate for quantifying the intensity of rural land use (viz., stocking rates, animal types, nutrient and sediment losses)?

Recommendation 9: *The input data for the HRU typology could be improved and sourced from a consistent data set similar to other spatial landscape attribute studies like the hydro-chemically guided landscape classification system developed in the Physiographic Environments project*[6]*.*

A model may have deficiencies related to the quantity and quality of the input data and yet can still be useful for informing management actions. The sensitivity of the model result to the input data (i.e., as part of a more general sensitivity analysis that includes parameter/coefficient values) can help guide the weight of confidence attributed to model results when there is a paucity of data. The reviewers consider that valuable information could be gathered by further analysis of Stage 1 results that provide additional post-hoc analysis of model results include:

- analysis of the HRU typologies for inconsistency with industry data and other GIS typological data sets like physiographic data sets.
- analysis of frequency distributions across temporal steps (flow/concentration/loads) duration analysis, hour of day, day of year);

[6]Rissmann, C.W.F, L. K. Pearson, M. Beyer, M. A., Couldrey, J. L. Lindsay, A. P. Martin, W. T. Baisden, T. J. Clough, T. W. Horton, and J. G. Webster-Brown, 2019. A hydrochemically guided landscape classification system for modelling spatial variation in multiple water quality indices: Process-attribute mapping. Science of the Total Environment 672: 815-833.

- analysis of observed data covariance ratios (TSS/NO_3-N) with modelled data;
- sensitivity analysis (e.g., turning processes (RQUAL) on/off and adjusting model input data);
- mass balance analysis from investigating generation models like APSIM/SPASMO and soil moisture balance models (SMWBM) to provide data series for sub-catchments baseflows; and
- exploring possible reasons for discrepancy between model output and observations (e.g., using published information, 'expert opinion' and local knowledge to inform parameter values).

These tasks can lead to improvements in the model performance but perhaps more importantly, they may address the important question: how can model representation of the catchments provide robust assessments of the effectiveness of different management actions?

Recommendation 10: *The reviewers foresee opportunities for additional field (and possibly lab) studies to address some of the shortcomings identified in model input data. Before undertaking these studies, however, we recommend further analysis and discussion of Stage 1 findings because we consider such an analysis could yield important additional information. Aligned with this analysis is the need to identify priority attributes and link these attributes to pressures, management needs and possible responses. This analysis should lead to increased confidence that further data collection is focused on improving the model and its ability to inform management. For example, further reflection on Stage 1 stormwater modelling may identify the benefits of carefully planned targeted intensive studies (e.g., over one or two storms) to capture periods of elevated flow and loads and complementing grab sampling with autonomous sensor data for identified key variables (e.g., stage height, temperature, dissolved oxygen) with the objective of improving the estimation of key model parameters, especially those that may affect possible remedial actions.*

C2: Model calibration

To date, parameters in the FWMT have been calibrated manually. Manual calibration can be efficient when the modeller has a good understanding of the model, contaminant behavior in the catchment, and the input and calibration data, and is often useful early in a study. However, the manual calibration does not necessarily ensure that the parameter values are 'correct' (see earlier discussion on equifinality). Automated calibration methods are available[7,8] that make provision for errors in both the input and calibration data and identify the sensitivity of model predictions to processes and parameters. The review team does not recommend blindly using autocalibration to replace the

[7] Doherty, J., 2015. Calibration and Uncertainty Analysis for Complex Environmental Models. Watermark Numerical Computing, Brisbane, Australia. ISBN: 978-0-9943786-0-6.
[8] Shoajei, N. and Wells, S. A. 2014. Automatic calibration of water quality models for reservoirs and lakes. Proceedings World Environmental and Water Resources Congress, EWRI, ASCE, Portland, OR, pp. 1020-1029.

manual calibration in future stages of the FWMT, however, and the two methods should be closely linked to ensure parameters are adjusted within known ranges of published limits.

There is also scope within the model structure to look at a mix of bootstrapping (as above), calibration hierarchy (currently focused on sedimentation which is dominated by peak-flows) and using 'pilot values' to examine the numerical stability of some of the parameter/coefficient default values.

Recommendation 11: *We recommend that automatic calibration be included alongside sensitivity, uncertainty and risk analysis as an integral part of Stage 2. Until sensitivity, uncertainty and risk analysis is carried out, we recommend that the Stage 1 findings be treated as indicative of the baseline state. The autocalibration could complement the manual calibration and align with knowledge of parameter values from the modellers and the literature. This procedure could help to identify an envelope of plausible outcomes (i.e., to demonstrate how the deterministic model in the FWMT can generate statistical distributions of simulation output) and better express to Auckland Council and its stakeholders the level of confidence that can be attributed to model scenario runs.*

C2: Use of the Virtual Climate Station Network (VCSN)

The use of the VCSN from NIWA is common practice for environmental modelling in New Zealand as a second-tier source of gridded climate data, relative to the hydrological units in the model structure. We are unsure why the NIWA climate modelling was not used to complete the temporal downscaling and why there are gaps in the VCSN climate grid. The current approach in the FWMT appears to make the FSP climate set data more complicated and may reduce the reliability of predictions. It is also unclear why a longer hydrological data set (1972 to present) was not used for the catchment streamflow calibration as it would likely improve the flow statistics through including greater variation in climate and flow across nearly 50 years of data. It is possible that the modellers' use of the 15-minute model time step curtailed the opportunity to use the longer data set. A longer time series (without water quality attributes) could be used to parameterise the flow related coefficients in the model and improve the performance and mass balance in the water budget. The water quality functions could then be introduced following a satisfactory flow calibration using the shorter (15-minute) time step for the current duration of the model run or using a longer time step (daily) over the extended VSCN period.

C3: Balance of data for calibration and validation

We note that the flow and water quality stations used for calibration are disproportionate to the stations used for validation. This makes interpretation of the results difficult. For example, there are only 7 flow sites of high data value used for c. 95% of the model domain. Similarly, for water quality there are only 5 sites for forestry, pasture and horticulture land uses across the entire FWMT model, and for TSS specifically, there are only 6 sites and none of these is in horticulture dominated sub-catchments.

Recommendation 12: *We recommend initially using the majority of the data stations over the full period of data collection (including VCSN data) for calibration of the hydrological component of the model as well as use of all of the data in a refined sub-catchment within the model domain. A validation for the model predictions can be completed at a later stage in the modelling process once the baseline state predictions are considered acceptable.*

C4: Rural land use variability, representation and models

It is stated (p. 68) that the "*FWMT Stage 1 simulations do not directly…account for seasonal variation in land use and omit inter-annual variation altogether*...". While this may be a good starting point, it may be one reason the model is not able to be calibrated well at some sites, notably some of the rural sites. For example, one would expect seasonal and year-to-year variations in nutrient runoff from farmland according to management practices that tend to vary seasonally[9] (e.g., tilling, fertiliser application, stocking rates, etc.) and seasonal variations in evapotranspiration, grass growth and soil processes. Further information on temporal variability of these practices may be available from applications of other models (e.g., SPASMO, APSIM and OVERSEER) at farm scale or from indirect lines of evidence that suggest seasonality in nutrient runoff because of seasonal changes in soil nutrient conditions[10] from pasture increase with stocking rate. In general, the reviewers consider there is more information that could be extracted from existing knowledge bases and models that may improve upon the current allocation of only two impact classes (high and low) for stocking units in rural land uses. The modellers have indicated a willingness to work with farm sectoral bodies to assess farm types, contaminant yields to edge-of-field, mitigation choices (opportunity, cost and benefit) and inform HRU typology in the second stage of the FWMT. This extends to the rotational basis of land use across the rural catchments in Auckland with the majority of vegetable crops being on a five-year rotation with green catch crops grown in the intervening years. The areas in horticulture and forestry could be based on the rotation (horticulture) and harvest (forestry) in the model as a fraction of the total potential areas in Auckland Council Horticulture maps (see proposed Auckland Unitary Plan data and evidence of Horticulture New Zealand and Pukekohe Vegetable Growers Association.

This process is strongly supported by the modelling team, recognising that there is a wealth of knowledge and experience in this area that is being heavily utilised across New Zealand in response to the NPS-FM.

---

[9] Roche, J. R., Ledgard, S. F., Sprosen, M.S., Lindsey, S.B., Penno, J.W., Horan, B. and Macdonald, K. A. 2016. Increased stocking rate and associated strategic dry-off decision rules reduced the amount of nitrate-N leached under grazing, Journal of Dairy Science 99: 5916-5925.

[10]Menneer, J., Ledgard, S., McLay, C., and Silvester, W. 2005. The effects of treading by dairy cows during wet soil conditions on white clover productivity, growth and morphology in a white clover-perennial ryegrass pasture. Grass Forage Science 60: 46-58.

Recommendation 13: *The reviewers consider that it would be useful for the modellers to work with Auckland Council to examine whether there are land use management practices, particularly in rural catchments, that could be prescribed as being seasonally dependent and which could therefore be represented temporally in the FWMT. Seasonal effects can be relatively easily tested in the FWMT.*

C5: Nitrogen simulations

The reviewers noted that concentrations of nitrogen under low flow were not simulated accurately in the FWMT but were considerably better at higher flows. The reviewers suspect that seasonal changes in macrophyte biomass, involving greater nutrient uptake in spring-summer, may be responsible for this discrepancy, but that variability in other nitrogen-related processes may be involved, e.g., denitrification and other oxygen-dependent reactions.

Recommendation 14: *Some issues with simulating variations in nitrogen concentrations under different flow regimes suggest that there may be a knowledge and/or data gap in the dynamics of this nutrient. The reviewers consider that improvements may be made as better relationships are developed among other state variables, including macrophytes and/or periphyton that can have strong seasonal and flow dependencies, and in turn affect flow themselves (e.g., presence of dense macrophyte communities).*

D. Suitability of model configuration

This section addresses the suitability of the model configuration including the approach taken for contaminant source modelling via Hydrological Response Units (HRUs). It will include appropriateness of parameterization, calibration and validation steps at the HRU scale, choice of performance measures and bands therein, as well as narratives on model performance.

Specific comments

D1. Suitability of the HRU approach

The HRU functional components are focused on the development land areas, which are a small fraction of the model domain. The resulting reduced typology across the rural areas may be another constraint for model calibration for flow (baseflow, interflow and groundwater) and nutrients. There are challenges with applying regional hydrological responses inside HRUs but some of these issues may be compounded by drainage splitting involving separation of coastal margins which do not propagate flow but represent a significant proportion of the catchment. The reviewers suggest that there are reasonable grounds for considering higher resolution for some of the typological classes inside the HRUs. This would include examining the effect on calibration outcomes for flow and water quality attributes for HRUs covering headwaters, riparian, horticulture rotation types.

The generation typology ('impact classes') inside the HRUs is also considered to be narrow. Consideration should be given to increasing the function amplitude (e.g., from 2 or 3 to 5) to allow greater flexibility. A change would also improve the options for mitigation actions and sensitivity analysis for the FSP. We also note that a common class for the reach groups was used but did not appear to follow existing and validated stream reach typologies like REC2[11].

Recommendation 15: *We recommend that a review of the HRU typologies is undertaken to compare the FWMT delineation with existing data sets from geospatial analysis including physiographic datasets and REC2. There may be computational constraints with an increase in HRUs across a regional model so it may be necessary to undertake modelling at a smaller catchment scale or alter the time step of the model accordingly.*

The FWMT is well suited to applications of stormwater-driven issues (e.g., heavy metals and sediment) in urban catchments as it includes fine-scale discretization of urban areas and model contaminant generation that is rainfall-driven. At the same time, the use of standard equations (e.g., the Universal Soil Loss Equation) should also mean that it is well suited to other land uses where there is higher permeability.

D2: Suitability of model configuration

Pressure-state-response scenarios for different catchments may require different modelling approaches for different contaminants (e.g., Model 1: erosion (pressure) – generation and transport of coarse sediment (attribute) and remediation using detention ponds (response), versus Model 2: intensive grazing (pressure) – leaching of nitrogen (attribute) into shallow groundwater – use of riparian filter strips and/or critical source area controls (response)).

The reviewers understand that the FWMT is calibrated across the whole of the Auckland region for a suite of contaminants despite differences in how the pressures-states-response framework plays out in different sub-catchments. This approach makes computer run times long (thereby discouraging multiple runs required for sensitivity analysis and autocalibration) and assumes model parameters are uniform across the catchment. If there are (subtle) differences in contaminant generation, transport and attenuation processes between sub-catchments that are not factored into the model (e.g., differences in sediment properties that affect erodibility, particle size distributions and settling) then a 'common' set of parameters may not capture such differences. The reports could discuss whether the treatment of the whole region might bias calibrated model parameters, potentially affecting scenarios in which input data are altered. An alternative approach is suggested in Section B above, namely, to focus on modelling contaminants in sub-catchments that are problematic (i.e., have high contaminant concentrations) and where there is high quality input and monitoring data, enabling

[11] https://niwa.co.nz/freshwater-and-estuaries/management-tools/river-environment-classification-0

better calibration in those sub-catchments and thereby increasing the confidence Council can have in model predictions when addressing management options.

The reviewers note that small streams and ephemeral stream channels in headwater catchments are not modelled explicitly. Rather, contaminant yields are reduced by an 'attenuation factor' to account for removal during transport through the headwater HRU before discharging into a stream channel further downstream. We note that attenuation and transformation are generally high in small headwater streams and ephemeral channels. The report could comment further on how the use of regional 'attenuation factors' in headwater HRUs makes it difficult for Council to investigate the effects of remedial measures (e.g., riparian re-planting of headwater streams).

D3: Calibration and validation

As we understand it, during Stage 1 the FWMT was calibrated by fitting the 15-minute, 30-minute or 24-hour flow-weighted averages to monthly grab samples for both concentration and load using r^2, PBIAS and NSE as goodness of fit measures. The reports acknowledge that timing errors affect the goodness of fit values (r^2, PBIAS and NSE). For example, the timing of peak rainfall in the model may not be correct, resulting in the predicted contaminant concentrations either leading or lagging observations while successfully matching the peak concentration, peak load and storm load. While R^2, PBIAS and NSE are commonly used goodness of fit measures in model calibration, they do not make allowances for timing differences, which may be critically important in a model. The reviewers suggest that Stage 2 modelling investigate more robust goodness of fit measures.

Comparisons were made of attribute state grades for monitoring data and modelling results using r^2, PBIAS and NSE values, with the results appearing to be 'more satisfactory' than for concentration/loads. This outcome is not unexpected given the small number of attribute grades. It could be argued that accurately predicting the effects of remedial measures on grades (as defined in the NPS-FM) is more important to managers than achieving satisfactory r^2, PBIAS and NSE values for concentrations and loads. However, we do not recommend calibrating the model to observed grades because it is likely to identify a wide range of parameter values that give an equally good fit to grade, leading to a model that is unable to accurately predict the effectiveness of remedial measures. The reviewers consider that calibration and sensitivity analysis should focus on matching concentrations and loads similar to what has been done using r^2, PBIAS and NSE but using a broader set of comparisons such as peak concentration, storm load, frequency distributions, etc. It may also be instructive to present a comparison of annual loads (log-lot plot) for the observed and modelled data.

Recommendation 16: *Additional goodness of fit diagnostics should be presented for comparisons of observed and modelled data, including peak concentration, peak load, storm load and the frequency distributions of flow, concentration and load. These diagnostics are designed to demonstrate the advantages of a highly temporally resolved (15-minute interval) model.*

D4: Concentrations versus loads (mass flow)

The reports note that goodness-of-fit measures were almost invariably better for loads than concentration. One likely reason, acknowledged by the modellers, is that in the FWMT contaminant generation is '…driven by rainfall…'. This implies that the model places more emphasis on correctly modelling runoff processes, and hence is likely to make better predictions during rainfall events (viz., high flows) than during low flows.

Contaminant loads are a key driver of water quality and ecosystem health in lakes and estuaries – both are important in Auckland. Where the FWMT predicts loads tolerably well (including the effects of mitigation) it may be useful for providing boundary conditions (viz., input data) for lake and estuary models. Even large errors in concentration when there is low flow may have only a minor effect on median, average and 95%ile loads. We understand that Auckland Council is coupling the FWMT with hydrodynamic models in the Waitemata Harbour. We also understand that Council is having discussions with the Ministry for the Environment about possible guidelines for loads rather than concentrations, but at the time of writing this report the reviewers understand that no decisions have been made.

The NPS-FM sets limits for many contaminants in terms of concentration. High concentrations at low flow (e.g., anoxia or NH_4-N/NH_3-N concentrations that are toxic) can have a serious adverse effect on stream organisms even though they occur infrequently and for short periods of time. The FWMT needs to predict concentrations at low flow to grade streams using current NPS-FM guidelines (median and 95^{th} percentile concentrations). Where the FWMT needs to more accurately predict low-flow concentrations, then Council will need to (a) review model structure (e.g., are there missing processes like plant uptake of nutrients or point sources of contaminants) and/or (b) calibrate the model to low-flow concentrations rather than to all concentrations including those at high flows that may have lesser impact on ecosystem health.

D5: Assessing the ability of the model to inform management

The Stage 1 reports focus on how well model predictions match observations for concentrations and loads. This is important for baseline state modelling. While goodness of fit is an important measure of the model's usefulness, a more important question for managers is whether the model can predict the baseline state at unmonitored sites, and the effectiveness of proposed alternative remedial measures, in the face of uncertainty. In other words, is the model good enough to help management make decisions rather than solely to ask does it provide good r^2, PBIAS or NSE values? It is especially important for Future State Planning to quantify model uncertainty and undertake risk assessment (viz., what is the probability that a proposed remediation will/will not achieve the attribute target). Our discussions with Auckland Council indicate that they are aware that performance metrics (r^2, PBIAS or NSE) do not address whether a model is fit for purpose, but we feel this point should still be emphasised.

Recommendation 17*. Council modellers need to progress beyond the goodness of fit between observed and predicted contaminant concentrations and loads and consider the suitability of the FWMT to* *predict the baseline state at unmonitored sites and the effectiveness of remedial measures. This encompasses whether the model structure allows remedial measures to be simulated and how well the model predicts the key attributes (including uncertainty), thereby allowing risk assessments to be made (i.e., that a target attribute state is not met) in association with remedial measures. This is especially important given the modellers' own assessment that Stage 1 calibrations were often 'unsatisfactory' (viz., poor match between observations and predictions). It was beyond the scope of our review of Stage 1 reports to address Future State Planning but we draw Council's attention to the need to plan how the FWMT will be used to assess risk in the future analysis. Where the modelling indicates there is a reasonable expectation of being able to predict concentrations (or loads) relevant to the NPS-FM then Auckland Council can have some confidence to continue to use the FWMT. Where the modelling shows up major problems (e.g., consistently poor calibration, wide and unexplained variability, etc.) then a re-think may be needed on how to manage those contaminants either with, or without, the help of the FWMT.*

***Recommendation 18**. The C&P reports should more clearly and critically identify the strengths and weaknesses of the model (assumptions, formulations, input data quantity and quality, calibration data and parameter values, ability to simulate remediations) in order to refine Stage 1 findings and guide Stage 2 planning.*

D6: Freshwater Accounting System

The NPS-FM (2020) requires Auckland Council to operate and maintain an accounting system for freshwater quantity and quality management purposes. The accounting systems need to provide baseline information for setting target attribute states, environmental flows and levels, and limits. The accounting system will allow Auckland Council to assess the allocation status of a freshwater management unit. Importantly, it will provide time series data required to evaluate whether the freshwater resources are degrading or improving.

The freshwater accounting system will store and provide the data required to adopt the NOF and report on progress towards achieving targets, limits and values. The system will also account for all freshwater takes. For the NPS-FM implementation, expert advice, data collection and sub-model recommendations will be required for providing the baseline information. For the FSP and strategy development, a series of recommendations are made in the FWMT Stage 1 report to support investment in the future stages of the FWMT as they relate to stormwater consenting, mitigation optimisations, practice standards, and infrastructure planning. However, Stage 1 of the FWMT also has the potential to provide data for baseline state or setting baselines for freshwater accounting purposes.

***Recommendation 19**: It is recommended that the FWMT Stage 1 report details for each of the six priority contaminants the values affected (e.g., what aspects of the environment are affected by Cu and Zn), pressures (what are the sources of Cu and Zn), targets (currently there are no* *standards/guidelines for Cu and Zn but AC may propose its own targets) and an assessment of the suitability of the FWMT model to help quantify the baseline state and predict future state scenarios.*

E. Reporting of current state

This section addresses reporting of baseline state and associated modelled information (e.g., contaminant sources, events or conditions associated with water quality variation). Additionally, comment may be made on whether FWMT configuration and performance is suitable for scenario assessment to optimized management strategy development (e.g., reporting changes in water quality grading, sources, pathways and events or conditions).

Specific comments

E1. Reporting of baseline state

The Stage 1 assessments of Baseline State in the FWMT focus on nutrients, sediment, faecal matter and heavy metals in stormwater, wastewater and diffuse discharges, as well as the associated modelled information (e.g., contaminant sources, events or conditions associated with water quality variation). The contaminants are those deemed likely to pose the most serious risk to coastal and freshwater quality. The reviewers accept that these contaminants are important drivers of degradation but note that Stage 1 does not address some aspects of ecosystem health for which the NPS-FM sets (or is likely to set) limits, notably macroinvertebrates, periphyton, DO, pH, water temperature, fish and deposited sediment. We understand from discussions that the FWMT was used to predict DO, pH and water temperature but that none of these attributes underwent calibration and were not reported. Council would benefit from a Baseline State assessment of DO, pH and water temperature and, given the sparse monitoring data, the FWMT could assist. We endorse the commitment by Council to include DO, pH and temperature in Stage 2 modelling.

Recommendation 20: *The Stage 2 FWMT should give a high priority to modelling DO, pH and temperature to improve the Baseline State Assessment and improve the model's ability to predict other contaminants affected by DO, pH and temperature.* The reviewers note that the assessment of baseline state includes both the use of monitoring data and FWMT predictions. In theory, the latter enable the spatial distribution of baseline state to be described across the whole catchment rather than just at a limited number of monitoring sites. Because the model runs with a 15-minute time step it also has the potential to examine short-term variability. Predicting the spatial and temporal distribution of attributes has the potential to optimize contaminant management by identifying 'hot spots', 'hot reaches' and 'hot periods'. However, this potential is predicated on the reliability of the FWMT predictions and the way that model results are used to define baseline state. The reviewers note that the ability of the FMWT to match observations is classified as 'unsatisfactory' for several contaminants and monitoring sites. Until the reasons for unsatisfactory predictions are investigated, understood and reduced/eliminated, baseline state assessments using FWMT results should be regarded as tentative. In the meantime, the uncertainty in model predictions for baseline state predictions needs to be communicated in order to aid management in setting priorities for remediation.

Notwithstanding, the Baseline State Rivers report clearly identifies widespread exceedance of guidelines for some contaminants (notably *E. coli*) and localized exceedance for others (e.g., heavy metals) and such assessments are likely robust to uncertainty in FWMT predictions. Further improvement of the FWMT and refinement of the way results are used to grade streams within the NPS-FM framework would increase confidence in the Baseline State Assessment (with second-order benefits to Council) and Future State Planning (a priority need for Auckland Council).

Recommendation 21: *Given the draft stage of the FWMT reports and that the modellers' own assessment was some of the Stage 1 predictions of attribute baseline state are "unsatisfactory", we recommend further investment in the FWMT calibration and refinements to the way results are analysed and reported (see previous recommendations).*

F. Effect of model configuration and boundaries

This section addresses how changes in HRUs or boundary conditions (e.g., climate) might affect an optimized strategy development.

The FWMT has HRUs and reach groups that offer exceptional capability for scenario testing of changes in nationally and regionally relevant contaminants and allow for spatial integration throughout the Auckland region from headwater springs to coastal receiving waterways. Along the succession of HRUs, freshwater is modified by climate, geology and a range of human-related factors that are mostly represented within the FWMT, although it may be useful to further explore the influence of geology on model outputs (e.g., suspended sediments and nutrients).

Specific comments

F1: Headwater streams

The reviewers accept that in headwater sub-catchments, in-stream processes are not simulated (e.g., sediment erosion, resuspension and deposition and nutrient transformations). However, some analysis is required of the decision not to include modelled stream segments in headwater sub-catchments. Headwater streams often have high attenuation rates associated with relatively high benthic contact areas. A sensitivity analysis involving varying the inputs of headwater sub-catchments at the boundary may be useful, particularly as it relates to potential for remediation actions at these sites to propagate through the modelled domain. See also the Reviewers' comments about headwater streams in section D2.

F2: Rainfall boundary conditions

P49. Auckland Council is aware that the 15-minute rainfall and runoff data capture aspects of temporal variability but not the actual values at the time/place samples were collected, which could affect calibration to observed pollutant concentrations/loads. This is because rain systems move across the catchment so that disaggregating VCSN daily totals to 15-minute rainfall at specific locations is unlikely to quantify accurately both the spatial and temporal variability in rainfall which drives the FWMT. Consideration could be given to focusing on a few ‘first flush’ and/or major storm events that could help inform the calibration and uncertainty in model predictions. See Section C2 for additional comments on the VSCN.

F3: Hydrologic Response Units

HRU parameters are regionalised in the FWMT according to soil, vegetation cover and impact class, but variations in climate will generate differences among HRUs that could theoretically have identical prescriptions of soil, vegetation cover and impact class. The impact class could be clearly delineated, i.e., by crop type for horticulture or stocking rate for pastoral areas of land cover, to allow a wealth of agricultural data in New Zealand to help inform variations among and within impact classes, including variations within impact classes from different management practices (e.g., impacts from land disposal of dairy farm washwater).

F4: Groundwater assumptions

The FWMT will generate continuous outputs which are transformed to steady state outputs over a 5-year period for scenario generation, to assess the efficacy of remediation actions. Consequently, FWMT scenario outputs will not incorporate lag times of groundwater responses to altered climate or land use unless boundary conditions for groundwater are altered. In the long term, Auckland Council will want to address localised areas of groundwater contamination as well as reducing groundwater contaminant levels regionally. The current situation of groundwater contaminant concentrations being sampled regularly by Auckland Council but without clearly identifying causal factors of variability points to an area of work on groundwater that could help support the FWMT and lead to a process-based justification for changing contaminant concentrations of the FWMT groundwater boundaries. This is important to avoid changing groundwater contaminant concentrations in the FWMT to arbitrarily defined future values in response to remediation actions or potentially being overly conservative by holding concentrations at a fixed level despite remediation actions. See Section B5 for further discussion on the groundwater boundary conditions, which also cautions against calibrating groundwater contaminant concentrations to achieve improvements in the FWMT fit, and thereby compromising its value for Stage 2 assessments.

G. Approach taken to water quality state assessment

This section addresses the approach taken to grading or assessing water quality (contaminant) state, including implications of using simulations for grading against National (and Regional) Objective Framework numeric attribute states (i.e., potential for the latter not to have considered modelling applications that provide continuous time-resolved outputs and their effects on model reporting).

Specific comments

G1: The NOF framework

The NPS-FW (2020) sets the framework for management as:

a. Identify **values** (3.7 2b) and environmental outcomes (3.7 2c)
b. Identify attributes for each value
c. Define the baseline (usually the current) **state** of each attribute (3.7 2d)
d. Identify the **pressures** that affect the attributes (3.30 2e)
e. Set target attribute **states** (3.7 2e)
f. Identify **response** (viz., actions to reduce pressures and achieve targets)
g. Set rules and action plans (3.7 2f)

The main role of the FWMT is to predict the effects of **pressures** on **attribute state,** and hence **values**, in order to guide **response.** The FWMT is also useful to define the baseline (current) state [c] above. In our opinion the FWMT has no role in [a], [b] and [g].

The NPS-FM states that "*…the compulsory values listed in Appendix 1A apply to every FMU…a regional council may identify other values… the regional council must identify an environmental outcome for every value…must include the environmental outcomes as an objective…describe the environmental outcome sought in a way that enables an assessment of the effectiveness of the regional policy statement, plans and action plans in achieving the environmental outcome…*".

Further, the NPS-FM states on p. 31 that "*…a description of the environmental pressures on each FMU (such as water takes, sources of contaminants, or water body modification) as indicated by information from the freshwater accounting systems…Rules, Plans and Action Plans need to address these pressures in order to prevent or reverse degradation of waterways…*"

Currently the contaminants (attributes) modelled in Stage 1 are listed but these are not explicitly linked to pressures, values and ecological outcomes. Information about these links is critical in clarifying what the FWMT can/cannot predict and the array of possible management interventions.

Recommendation 22: *Stage 1 reports would be significantly improved if environmental pressures were clearly identified and linked to attributes and values, as well as the range of possible interventions listed. It seems likely that Auckland Council has good understanding about these links and discussing them would, in our opinion, (a) help modellers focus to generate the information required, and (b) increase managers'/stakeholders' understanding of, and confidence in, the FWMT.*

G2: Grading for contaminants modelled by the FWMT

The Baseline State (Rivers) report identifies the challenge of comparing grades based on sampling and modelling given the significant differences in temporal scales. Predictions are made at 15- or 60- minute time steps and aggregated to flow-weighted daily averages, while observations are typically monthly grab samples. Attribute state boundaries are derived from a range of information including laboratory studies and discrete sampling. Differences in temporal scales and methods to define attribute state boundaries gives rise to uncertainty in gradings.

The report "…*explores*…" three approaches to grading sites based on observed and modelled concentrations, in an attempt to quantify uncertainty in grading. While we acknowledge the modellers' efforts, we feel that none of the three approaches is entirely satisfactory.

Generally, Approach 1 compares the median or 95%ile of observed and predicted concentrations (e.g., if the median DRP concentration is 0.015 mg/L then the site is Grade C (for which the band is 0.010 < [DRP] ≤ 0.018 mg/L based on median of monthly monitoring results over 5 years). Two questions arise. Firstly, are gradings based on monitoring and modelling consistent? As the report notes, for the 5-year reporting period recommended in the NPS-FW, this comparison might calculate the median of 1826 daily average predictions and the median of 60 monthly grab samples. There is a risk that monthly sampling may 'miss' high and low observed concentrations (although the median is more robust than the average in coping with extremes). There is also a risk that modelling will include more high and low predicted daily average concentrations than are likely to be sampled (e.g., by not sampling storms). Secondly, there may be a risk that continuous modelling results will bias gradings, given that attribute state boundaries in the NPS-FM may not be based on exposure to time-varying concentrations. We acknowledge that Auckland Council is constrained to work with the NPS-FM attribute state boundaries but suggest they remain open to how best to use the FWMT modelling results. We understand that Council has asked MfE for advice but has not yet received clear guidance.

One approach (not considered in the report but meriting consideration) could be to randomly sample model predictions once per month over a 5-year period, yielding 60 predicted concentrations. By repeating the sampling (say 100 times, via. bootstrapping) a frequency distribution of predicted concentrations could be generated from which statements could be made like '…there is a very low likelihood (e.g., 5% probability) of the site being Grade B, high likelihood (e.g., 75% probability) of Grade C and moderate likelihood (e.g., 20% probability) of Grade D…' This approach is arguably more informative and robust than saying '…it is (most likely to be) Grade C…' Plotting the distribution of predictions (a 'violin plot') on top of the grading bands would be very informative.

Approaches 2 and 3 appear to be attempts to account for variability in predictions and we acknowledge these attempts, but also have misgivings about them. Approach 2 reports whether SoE stations are predicted at and within an additional grade of observed (e.g., modelled grade = observed, observed minus a grade and observed plus a grade). Thus, if the model predicts some days in Grades A, B and C, the site is graded B regardless of the number of days in each grade. In our opinion, if for the majority of days the values at a site are categorised as Grades B and C, with only occasional days of Grade A, it may not be sensible to assign Grade B and ignore the significant number of days at Grade C. In our opinion it would be preferable to estimate the probabilities of the site lying in the different grades. Approach 3 uses the width of the attribute bands (e.g., for range of DRP for the Grade C band extends over 0.008 mg/L) as a surrogate for variability in predicted and/or observed concentrations. In discussion, Auckland Council confirmed the methods are 'nonstandard' and that each approach suffers limitations (in Council view, Approach 1 is conservative, Approach 2 permissive and Approach 3 somewhere in between). Although none of the reviewers is an expert in the topic, we are aware of well-proven methods for estimating, from a given number of samples, whether there has been a breach of a standard (grade).

Recommendation 23: *The reviewers recommend that Auckland Council further investigates the approach taken to using modelling results for grading, in consultation with Ministry for the Environment (MfE), and extending the initial consultation between MfE, Auckland Council and the Peer Review Team of 25 May 2021. We do not recommend 'grade' be the main focus of model calibration – concentration and load should be the main focus albeit using a suite of measures and not just r^2, PBIAS and NSE as in the Stage 1 reports (see also Recommendation 16).*

H. Approach used in the lake component of the model

This section addresses the appropriateness of the approach used in the lake component of the model (e.g., identification of limitations and recommendations for addressing limitations).

The FWMT Stage 1 Lakes application included 16 dune lakes and one volcanic lake. The lake report should be considered highly preliminary as it is repeatedly acknowledged that there is a lack of suitable and representative data for modelling and assessment of lake health. For the FWMT application this meant application of a generalised New Zealand model to the Auckland lakes[12]. The model enabled external total phosphorus and nitrogen loading to be used to predict the 'steady state' of four measures of trophic state in the lakes: concentrations of total phosphorus, total nitrogen, chlorophyll *a* and Secchi depth. A critical part of being able to apply these types of loading-trophic state ('Vollenweider') models is to capture the water mass balance that allows for derivation of the water residence time in the lake; a key variable in the application of the Vollenweider model. In this regard there is clearly a major problem, with residence times (in years) ranging from 8.2 years (Lake Kereta) to 11531 years (Lake Pupuke). Logic would tell us that these values are wrong. For example, taking the deepest lake (Pupuke, 56 m) and a representative rainfall of 1 m/year, the residence time based on rainfall input alone (i.e., to the lake surface) would be 56 years, not the 11531 years given in Table 3-3. The catchment areas given in Table 2-1 also make the water residence times given in Table 3-3 non-sensical. The values given in Table 3-3 (sources from the Freshwater Ecosystems of New Zealand (FENZ) Geodatabase) are in fact acknowledged as being wrong and that "extensive modelling and verification exercises would be required to amend this issue" but this should not excuse the modellers from proceeding with incorrect values nor making the effort to at least partially resolve the situation (e.g., with some straightforward modelling, particularly given that digital elevation maps had been created for lake catchments, that evaporation rates could be modelled for the lake water surface and that the FWMT Stage 1 Lakes Module could produce discharge values used as input to the lakes). Obviously, this gross inaccuracy largely discredits the subsequent work, including attempts to carry out a sensitivity analysis on lake residence time, which unsurprisingly found that residence time was not particularly sensitive.

[12] Abell, J. and P. van Dam-Bates, 2018. Modelling Reference and Current Trophic Level Index for New Zealand Lakes. Consultant's report prepared for the University of Waikato by Ecofish Research Ltd, 6 November, 2018.

Recommendation 24: *The first stage of the lake modelling should have developed confidence in the hydrology, water balances and water residence times for the lakes. The lack of sensitivity of lake trophic state to residence time perturbations in the sensitivity analysis is a direct result of gross overestimates in the residence time. Trophic state modelling should not be undertaken until there is some attribution of confidence (including sensitivity analysis) in the lake residence times.*

The modellers acknowledge that the current lake model is steady state and therefore not suitable for capturing all aspects of dynamic variables like chlorophyll *a*, for which different prescriptive statistics are given in the NPS-FM (e.g., annual maximum value). Time variations also occur seasonally or more regularly as a result of stratification and from internal nutrient releases (e.g., dissolved nutrients released into bottom waters during anoxia). A dynamic model may be useful to introduce dissolved oxygen in bottom waters as it is an attribute in the NOF and, at least in seasonally stratified lakes, it may allow for long-term changes in trophic state of the lakes to be interpreted when there is not a regular sampling program.

On a positive note, the FWMT Stage 1 Lakes does provide a basis for seeking remediation measures as a result of the coupled catchment-lake model approach, which allows for identification of different land uses and sources (e.g., bank erosion) contributing to external nutrients (TP) to be identified and targeted. However, until the issues above are addressed, the FWMT Lakes does not allow for an understanding of the magnitude of internal TN and TP loading, including the way in which the residence time impacts recovery from eutrophication and the confidence that can be attributed to inter-annual variability.

Similar to the FWMT Baseline Rivers Baseline State Report, further lake modelling work should examine the issue of median (e.g., in the NOF) and mean (e.g., in the FWMT Stage 1) concentrations of chlorophyll *a*, although a dynamic modelling approach would be required to compare the annual maximum concentrations of chlorophyll *a* that are part of the NOF. A conversation with MfE may help to better clarify the guidance provided in the NPS-FM on modelling chlorophyll *a* concentration in lakes (at the same time as clarification is sought about river grading).

In general, the FWMT Stage 1 Lakes has identified some major problems with the data available for modelling lakes. These issues are addressed in Recommendation 20.

Recommendation 25: *Before embarking on further lake modelling (i.e., Stage 2 Lakes), it is critical to collate the available information on lakes to assess data quantity, quality, and suitability to support the modelling. The assessment should also include alignment with attributes and statistics in the NOF, data gaps (e.g., lake bathymetry, internal loading) and the frequency and coverage of current SOE reporting.*

I. Limitations of the FWMT Stage 1 for planning and operations

This section addresses limitations of the FWMT Stage 1 including issues that may arise for its use in strategic (regional planning) and operational (plan implementation) management aligned with the Future State Assessment.

Auckland Council's focus is initially on Stage 1, to develop the FWMT using the 'best available' data as input and calibration. It is signalled that Stage 2 will include scenario modelling. It would be useful in linking these stages to acknowledge that modelling is usually an iterative process. We suggest that it be made clear in these reports that the Stage 1 model may need to be re-visited once management options are explored through scenario modelling. It seems likely that the Stage 1 model will prove to be more useful in addressing some contaminants (e.g., sediment, Cu and Zn) than others (e.g., N, P and *E. coli*) as well as some attributes and management interventions (e.g., sediment using detention ponds) than others (e.g., nutrients using land use controls). In the latter case, Auckland Council should seek to improve the FWMT (further calibration-specific measurements, model refinement, re-calibration) or seek other tools (e.g., adopt BMPs for known 'hot spots' or critical source areas as well as adopting other models specific to the questions being asked).

As discussed at the earlier workshop (20 January 2021), extending the calibration period to capture additional data may indeed improve the calibration, but it does not necessarily improve validation nor the predictions and scenarios for other periods and it could introduce temporal bias. For example, the wet 2017 year could correspond to large loadings of SS and SS-associated contaminants were transported into harbours and coastal waters with low stream attenuation. The five-year calibration periods may partially circumvent this but long-term rainfall variations (e.g., related to El Niño–Southern Oscillation and the Southern Annular Model) can extend over multiple years[13].

J. Recommendations for improvements to FWMT Stage 1 and directions for FWMT Stage

This section summarises recommendations made in this review for improvements to the Baseline State Assessment in FWMT Stage 1 and directions for FWMT Stage 2, including potential expansion of capability. The PRT has prioritised the recommendations as high (**HP**), medium (**MP**) or for noting (**LP**)

Recommendation 1: *The ability to clearly communicate the baseline state conditions outside the 36 stream reaches should be supported by a common language, with consistent use of terms for uncertainty, risk, likelihood and consequence, as well as consistent descriptors for likelihood of occurrence (e.g., very likely, likely, about as likely as not, unlikely, very unlikely). The FWMT can therefore predict the likelihood of whether the six main attributes meet designated NPS-FM grades (or some target concentration for metals) at a site.* **HP**

[13] Ummenhofer, C. C. H. and M. H. England, 2007. Interannual extremes in New Zealand precipitation linked to modes of Southern Hemisphere climate variability. Journal of Climate 20: 5418–5440.

Recommendation 2: *Detailed information should be provided to show how the FWMT can support activities undertaken by Auckland Council to manage its waterways according to the goals of the NPS-FM. This may include adaptive planning related to key infrastructure investments (assessments of 'grey' and 'green' infrastructure), policy development and community consultation. We recommend making greater use of the pressure-state-response framework when discussing the TSS, TN, TP,* E. coli*, Zn and Cu sub-models, which would include more detail on the sources, generation/transport/attenuation/modelling of these processes, the available remedial actions and how these were implemented in the model, and how risk analysis will be undertaken to determine the likelihood that a proposed policy or remedial action will achieve NPS-FM targets.* **HP**

Recommendation 3: *The variables known to be simulated satisfactorily by the FWMT could usefully be linked to the goals of Auckland Council for managing its waterways, including relationships to the NPS-FM and instream ecological outcomes. This could also include a timeline for variables deemed important and feasible to include in further stages of the FWMT, based on experience from the recently completed modelling, as well as prioritising variables important in adaptive planning related to key investments ('grey' and 'green' infrastructure), policy development and community consultation. Attributes in the NPS-FM and NOF that are not amenable to modelling by the FWMT need to be identified and suitable alternative prediction strategies recommended. Auckland Council may need to manage some attributes without the support of the FWMT and could usefully predict what attributes these may be, how management of them can be undertaken and whether these attributes may be included in the FWMT in the future. This recommendation relates to the need to appreciate that the FWMT does not provide guidance on all the attributes currently in, or likely to be added to, the NPS-FM.* **MP**

Recommendation 4: *To improve communication we suggest that a succinct summary of the model is included in the report, linking information that is currently dispersed through the text. A 'plain English' summary should also be included to synthesise the technical information and make it accessible to non-modellers (see Plain English Summary; this report). The strengths and weaknesses of the FWMT should be clearly identified to increase the understanding and confidence of managers and stakeholders in the FWMT. For example, the sediment model was generally well summarised and could be used as a basis for similar succinct summaries related to the models for nutrients, metals, and discharge. Relevant graphical or conceptual figures are recommended wherever possible.* **HP**

Recommendation 5: *Auckland Council could usefully map out an independent communication strategy document that helps with familiarisation of the FWMT and outlines expectations of how the model will be used to develop planning and policy to optimise investment and meet the goals of the NPS-FM; in effect the value of the model will be tightly linked to Auckland Council's ability to explain it to managers and stakeholders.* **MP**

Recommendation 6**:** *Further analysis is required to set out the strengths and weaknesses of FWMT Stage 1. This work includes: (1) clear delineation of calibration and validation processes, possibly scaled to 1-2 sub catchments so that computational times do not become too cumbersome, (2) uncertainty analysis to understand the relative effects of input data, model conceptualization and parameter calibration, and (3) examining in detail the processes of contaminant generation, delivery and attenuation which will help to better inform the possible mitigation actions*. **HP**

Recommendation 7**:** *Sensitivity/uncertainty analysis of the input data (e.g., meteorology, groundwater contaminant concentrations, etc.), model structure and calibrated parameters should be undertaken. This process will provide confidence in the validity of assumptions made in the model and the reliability of model output.* **HP**

Recommendation 8: *Auckland Council can communicate that there is uncertainty in the model parameters, particularly those associated with contaminant source estimates, and demonstrate attempts to quantify such uncertainty through model sensitivity analysis, carrying findings into Future State Planning scenario runs to assess the risk of achieving/not achieving targets for different attributes.* **MP**

Recommendation 9: *The input data for the HRU typology could be improved and sourced from a consistent data set similar to other spatial landscape attribute studies like the hydro-chemically guided landscape classification system developed in the Physiographic Environments project.* **LP**

Recommendation 10: *The reviewers foresee opportunities for additional field (and possibly lab) studies to address some of the shortcomings identified in model input data. Before undertaking these studies, however, we recommend further analysis and discussion of Stage 1 findings because we consider such an analysis could yield important additional information. Aligned with this analysis is the need to identify priority attributes and link these attributes to pressures, management needs and possible responses. This analysis should lead to increased confidence that further data collection is focused on improving the model and its ability to inform management. For example, further reflection on Stage 1 stormwater modelling may identify the benefits of carefully planned targeted intensive studies (e.g., over one or two storms) to capture periods of elevated flow and loads and complementing grab sampling with autonomous sensor data for identified key variables (e.g., stage height, temperature, dissolved oxygen) with the objective of improving the estimation of key model parameters, especially those that may affect possible remedial actions.* **MP**

Recommendation 11: *We recommend that automatic calibration be included alongside sensitivity, uncertainty and risk analysis as an integral part of Stage 2. Until sensitivity, uncertainty and risk analysis is carried out, we recommend that the Stage 1 findings be treated as indicative of the baseline state. The autocalibration could complement the manual calibration and align with knowledge of parameter values from the modellers and the literature. This procedure could help to identify an envelope of plausible outcomes (i.e., to demonstrate how the deterministic model in the FWMT can generate statistical distributions of simulation output) and better express to Auckland Council and its stakeholders the level of confidence that can be attributed to model scenario runs.* **LP**

Recommendation 12: *We recommend initially using the majority of the data stations over the full period of data collection (including VCSN data) for calibration of the hydrological component of the model as well as use of all of the data in a refined sub-catchment within the model domain. A validation for the model predictions can be completed at a later stage in the modelling process once the baseline state predictions are considered acceptable.* **LP**

Recommendation 13: *The reviewers consider that it would be useful for the modellers to work with Auckland Council to examine whether there are land use management practices, particularly in rural* *catchments, that could be prescribed as being seasonally dependent and which could therefore be represented temporally in the FWMT. Seasonal effects can be relatively easily tested in the FWMT.* **MP**

Recommendation 14: *Some issues with simulating variations in nitrogen concentrations under different flow regimes suggest that there may be a knowledge and/or data gap in the dynamics of this nutrient. The reviewers consider that improvements may be made as better relationships are developed among other state variables, including macrophytes and/or periphyton that can have strong seasonal and flow dependencies, and in turn affect flow themselves (e.g., presence of dense macrophyte communities).* **HP**

Recommendation 15: *We recommend that a review of the HRU typologies is undertaken to compare the FWMT delineation with existing data sets from geospatial analysis including physiographic datasets and REC2. There may be computational constraints with an increase in HRUs across a regional model so it may be necessary to undertake modelling at a smaller catchment scale or alter the time step of the model accordingly.* **MP**

Recommendation 16: *Additional goodness of fit diagnostics should be presented for comparisons of observed and modelled data, including peak concentration, peak load, storm load and the frequency distributions of flow, concentration and load. These diagnostics are designed to demonstrate the advantages of a highly temporally resolved (15-minute interval) model.* **HP**

Recommendation 17*. Council modellers need to progress beyond the goodness of fit between observed and predicted contaminant concentrations and loads and consider the suitability of the FWMT to predict the baseline state at unmonitored sites and the effectiveness of remedial measures. This encompasses whether the model structure allows remedial measures to be simulated and how well the model predicts the key attributes (including uncertainty), thereby allowing risk assessments to be made (i.e., that a target attribute state is not met) in association with remedial measures. This is especially important given the modellers' own assessment that Stage 1 calibrations were often 'unsatisfactory' (viz., poor match between observations and predictions). It was beyond the scope of our review of Stage 1 reports to address Future State Planning but we draw Council's attention to the need to plan how the FWMT will be used to assess risk in the future analysis. Where the modelling indicates there is a reasonable expectation of being able to predict concentrations (or loads) relevant to the NPS-FM then Auckland Council can have some confidence to continue to use the FWMT. Where the modelling shows up major problems (e.g., consistently poor calibration, wide and unexplained variability, etc.) then a re-think may be needed on how to manage those contaminants either with, or without, the help of the FWMT.*

HP

Recommendation 18*. The C&P reports should more clearly and critically identify the strengths and weaknesses of the model (assumptions, formulations, input data quantity and quality, calibration data and parameter values, ability to simulate remediations) in order to refine Stage 1 findings and guide Stage 2 planning.* **HP**

Recommendation 19*: It is recommended that the FWMT Stage 1 report details for each of the six priority contaminants the values affected (e.g., what aspects of the environment are affected by Cu* *and Zn), pressures (what are the sources of Cu and Zn), targets (currently there are no standards/guidelines for Cu and Zn but AC may propose its own targets) and an assessment of the suitability of the FWMT model to help quantify the baseline state and predict future state scenarios.* **HP**

Recommendation 20: *The Stage 2 FWMT should give a high priority to modelling DO, pH and temperature to improve the Baseline State Assessment and improve the model's ability to predict other contaminants affected by DO, pH and/or temperature.* **HP**

Recommendation 21: *Given the draft stage of the FWMT reports and that the modellers' own assessment was some of the Stage 1 predictions of attribute baseline state are "unsatisfactory", we recommend further investment in the FWMT calibration and refinements to the way results are analysed and reported (see previous recommendations).* **HP**

Recommendation 22: *Stage 1 reports would be significantly improved if environmental pressures were clearly identified and linked to attributes and values, as well as the range of possible interventions listed. It seems likely that Auckland Council has good understanding about these links and discussing them would, in our opinion, (a) help modellers focus to generate the information required, and (b) increase managers'/stakeholders' understanding of, and confidence in, the FWMT.* **HP**

Recommendation 23: *The reviewers recommend that Auckland Council further investigates the approach taken to using modelling results for grading, in consultation with Ministry for the Environment (MfE), and extending the initial consultation between MfE, Auckland Council and the Peer Review Team of 25 May 2021. We do not recommend 'grade' be the main focus of model calibration – concentration and load should be the main focus albeit using a suite of measures and not just* r^2*, PBIAS and NSE as in the Stage 1 reports.* **MP**

Recommendation 24: *The first stage of the lake modelling should have developed confidence in the hydrology, water balances and water residence times for the lakes. The lack of sensitivity of lake trophic state to residence time perturbations in the sensitivity analysis is a direct result of gross overestimates in the residence time. Trophic state modelling should not be undertaken until there is some attribution of confidence (including sensitivity analysis) in the lake residence times.* **HP**

Recommendation 25: *Before embarking on further lake modelling (i.e., Stage 2 Lakes), it is critical to collate the available information on lakes to assess data quantity, quality and suitability to support the modelling. The assessment should also include alignment with attributes and statistics in the NOF, data gaps (e.g., lake bathymetry, internal loading) and the frequency and coverage of current SOE reporting.* **MP**

K. Additional comments on fitness for purpose of the model

Other factors that affect the ability of the Model to meet its purpose and objectives are considered in this section. These factors may be beyond those explicitly stated in the Terms of Reference. This section will be completed following the meeting with Auckland Council, 5th May 2021.

One aspect that was not fully resolved in communications of the PRT in a workshop with Auckland Council and the Ministry for the Environment (MfE) during the project was how to best compare the attribute values in the National Objectives Framework (NOF) with the attribute output from the model. While the idealised measurement frequency is specified in the NOF (e.g., a median of 60 readings at monthly interval over 5 years), there is considerable uncertainty whether measurements should be taken under baseflow conditions, randomly each month or weighted in some way to reflect the annual or five-year load. On the other hand the model can provide near continuous output of an attribute and, if well calibrated, generate understanding of the variations in concentration, relate concentrations to flow, and better understand attribute load variability. The PRT recommends further discussions with MfE to understand the potential for sampling bias according to antecedent and current flow, duration of monitoring and choice of statistic.

The scope of the PRT review did not include consideration of the need for Auckland Council to give effect to Te Mana o te Wai. The model output has the potential for assist and strengthen the responsibilities of Auckland Council to Te Mana o te Wai through informing the long-term vision and discussion with communities and tangata whenua. Further stages of the FWMT could give consideration of how the FWMT is best integrated with Te Mana o te Wai under the obligations of the NPS-FM 2020.

Find out more: fwmt@aucklandcouncil.govt.nz

Auckland Council
Te Kaunihera o Tāmaki Makaurau