WATER QUALITY IN AUCKLAND – FWMT CURRENT STATE & PROCESS MODELLING ADVANCES

Tom Stephens, Theo Kpodonu, Nick Brown (Auckland Council), Dustin Bambic (Paradigm Environmental), Caleb Clarke (Morphum Environmental).

ABSTRACT (500 WORDS MAXIMUM)

Auckland Council is responsible for the management of freshwater quality and quantity in the Auckland region. The National Policy Statement for Freshwater Management (NPS-FM) has driven an array of studies, programs, and policy considerations by Auckland Council, including development of a water quality accounting framework - the Freshwater Management Tool (FWMT).

The FWMT simulates contaminant concentrations and loading across nutrients (nitrogen, phosphorous), metals (copper, zinc), sediment (total suspended solids) and faecal indicator bacteria (*E.coli*) continuously (15-minute), for 2013-2017, across 5,465 sub-catchments and distributed into 3,085 km of permanent stream within the Auckland Region. Simulation performance has been assessed across flow and seasonal gradients, for several metrics (r^2, NSE, Bias). An interactive data-viewer is in development and peer-reviewed reports are in production.

From the country's first region-wide, continuous process model, Healthy Waters is uniquely placed to advance water quality accounting and scenario-planning in New Zealand. The resolution of process-driven flow and contaminant information throughout 489,000 Ha of rural and urban land distributed into 106 hydrologic response units, is a world first. The FWMT has even recently been applied to support the development of sound regulation by Local Government New Zealand.

However, development of the FWMT has raised a novel challenge for Auckland Council – how to utilize modelling and monitoring together for water quality decision-making, including: (1) to increase the availability and coverage of data; (2) grade discrete observational and continuous modelled data harmoniously; and (3) translate modelled flow and contaminant time-series into meaningful water quality outcomes (e.g., ecological and risk-based outcomes). For instance, the NPS-FM was developed primarily for discrete monitoring at lesser resolution, with such constraints included in recommended thresholds for water quality grading. Whereas, the FWMT produces 2.6×10^6 estimates of each contaminant (instream and to edge-of-stream) at 2,761 stream nodes. The FWMT can thereby better support decisions on acute and chronic contaminant effects than traditional monitoring alone, offering considerably greater coverage in space and time.

The presentation describes water quality state, causes thereof and accuracy in FWMT predictions for Auckland – producing the first definitive, whole-of-region, continuous assessment for the NPS-FM in New Zealand. The historic milestone has raised more questions than answers including, not simply the improvements to science needed but also the notion of whether New Zealand is planning for a future where modelling and monitoring combined, yield better outcomes and confidence in those outcomes being fair, efficient and effective.

KEYWORDS

Water quality, stormwater, contaminants, modelling, hydrology, national policy statement for freshwater management, modelling.

PRESENTER PROFILE

Tom Stephens is Principal-Integrated Catchment (Healthy Waters, Auckland Council) and responsible for developing the FWMT, New Zealand's first continuous, process-based, regional water quality model. Tom offers applied science, waterway management and peer-engagement expertise in the multi-discipline and agency group ensuring the FWMT delivers planning and operational value.

1 INTRODUCTION

Auckland Council (Healthy Waters) is developing a Freshwater Management Tool (FWMT) to inform plan change responses to National Policy Statement for Freshwater Management (NPS-FM), including Auckland Council's long-term infrastructure programme. The FWMT is a regional water quality accounting framework resolving daily contaminant concentration and loads in Auckland's stream network. The FWMT enables a more holistic assessment of water quality issues than monitoring permits, and for stakeholders to work together in identifying the best (optimal) options to improve water quality in Auckland waterways – enabling scenario testing and targeted catchment management, regionwide.

Early development of the FWMT has been shared previously at Stormwater 2019 (see Stephens et al., 2019). This paper revisits the requirements and purpose for the FWMT briefly, before delving into the "baseline" simulation capabilities (e.g., performance and outcomes for water quality grading, patterns and loading apportionment for 2013-2017). The paper explores how modelling advances in the FWMT challenge existing guidance for the NPS-FM, before sharing the first whole-of-region, continuously-simulated surface water quality assessment for any Auckland – the first of its kind for the NPS-FM and in New Zealand.

Notably, detailed analysis of uncertainty and outputs are not shared here but will be available in technical reports following external peer-review.

1.1 BACKGROUND TO THE NPS-FM

In 2014, the NPS-FM became operative, obliging all regional authorities in New Zealand including Auckland Council, to more robustly assess, manage and improve freshwater quality. The NPS-FM requires all regional authorities to notify compliant regional plans by 2025. To be compliant, water quality "values" and associated "attributes" require objectives within regional or unitary plans. No further degradation and where already heavily degraded, improvement in attributes to at-least "national bottom-lines" is required (i.e., achieving national bottom-lines does not assure regional planning compliance with the NPS-FM where communities request better water quality).

Several freshwater values are nationally mandated by the NPS-FM – ecosystem and human health, with cultural health also likely (e.g., Essential Freshwater – MfE, 2019). Ecosystem and human health value are underpinned by a national objectives framework (NOF), in which concentration-based statistics are graded for acute and chronic effects (e.g., median, 95^{th}%). By requiring quantitative attribute grades on contaminants and effects, an integrated (catchment) approach and with increasing consistency about objective and limit-setting, the NPS-FM is improving water quality management in New Zealand.

1.2 FWMT – BACKGROUND & PURPOSE(S)

The FWMT is the first region-wide, process-based and continuous hydrological and contaminant model for nitrogen (TON, NH4N, NO3N, TN), phosphorus (DRP, TP), heavy metals (Cu, Zn), *E.coli* and sediment (TSS), developed in New Zealand.

To enable NPS-FM decision-making for freshwater objectives, the FWMT simulates both baseline and scenario hydrological and contaminant processes (flow, loads, concentrations) arising from interventions for point and diffuse contaminants across 489,000 Ha of land (distributed into 5,465 sub-catchments) and 3,085 km of instream. FWMT capabilities include optimized life-cycle costing of such interventions, either of devices (grey and green infrastructure) or source-control (improved practices, altered land use).

The FWMT supports both regulatory decision-making, budgeting and management of private and public effects of discharge on waterways for the NPS-FM *and* operational implementation of the NPS-FM by Auckland Council (e.g., stormwater network discharge consent decision-making and reporting – see Figure 1). A key purpose for the FWMT is to inform decision-making about costs and interventions required for NPS-FM objectives. Initially, of contaminant effects and lifecycle management costs but over time of wider processes and ecological outcomes.

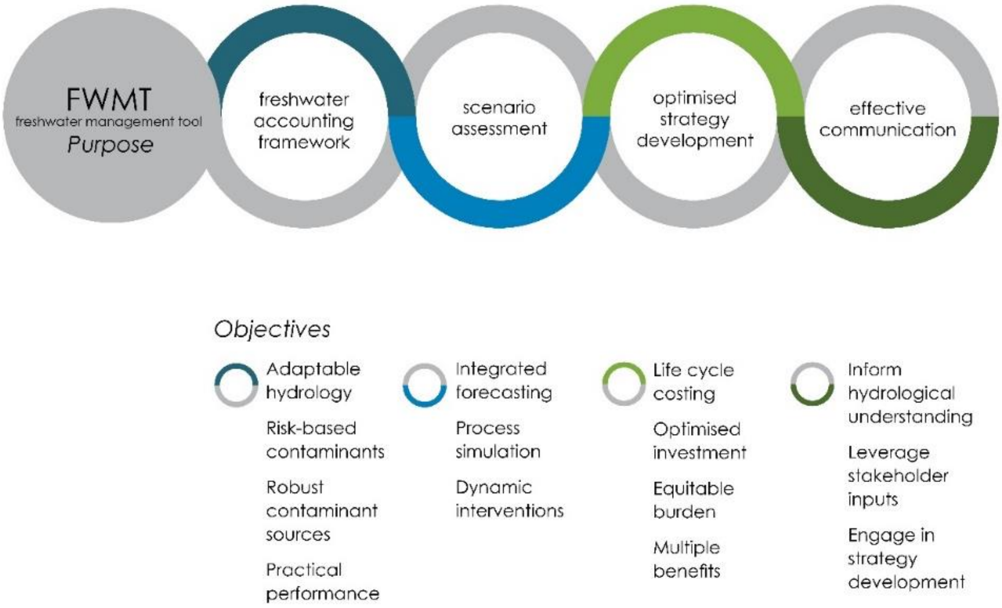

Figure 1. Purposes and objectives required of the FWMT presented as a value-chain for Auckland Council.

The FWMT development process builds-in continuous improvement. A 10-year programme will enable increasing complexity, reduced uncertainty and expanded capability to be developed adaptively, in response to planning and operational needs (Figure 2). A decadal development programme enables the FWMT to also respond to advances in water quality science (on grading and process-simulation), management (on opportunity, efficacy and cost of new or existing interventions), policy requirements (on water quality, urban development, indigenous biodiversity and greenhouse gas management) and performance (on results of targeted monitoring).

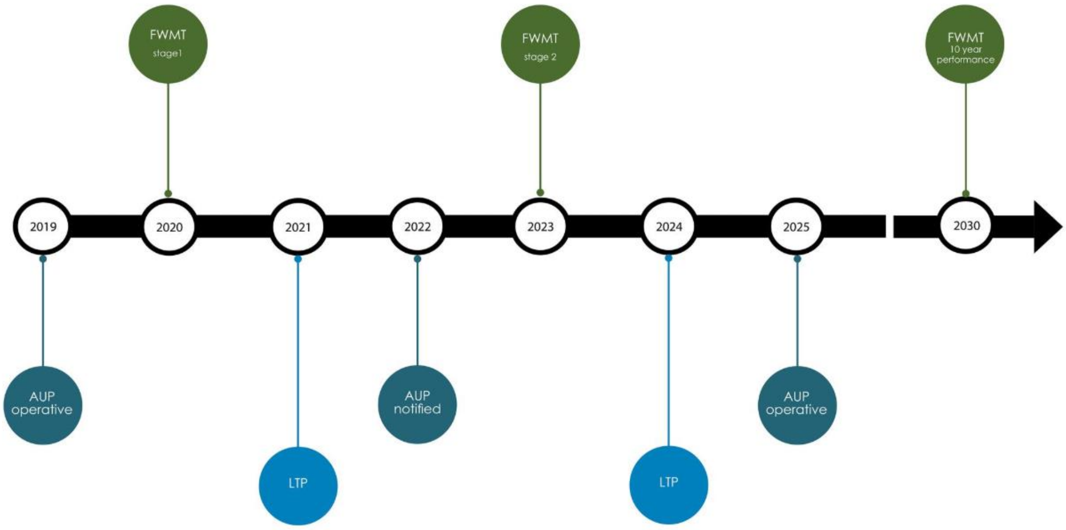

Figure 2. Iterative build programme underpinning the FWMT development by Healthy Waters.

For freshwater accounting purposes, the FWMT Stage 1 is intended primarily for grading-based decisions. Consequently, model performance is assessed by accuracy of predicted concentration and, specificity and sensitivity of grading. Precedents for similar continuous modelling exist in New Zealand (e.g., Greater Wellington – SOURCE in Porirua and Ruamanahanga whaitua [Blyth et al., 2018; Easton et al., 2019]; Bay of Plenty - SOURCE in Rangitikei water management zone [Loft et al., in prep.]). However, the FWMT is novel for its greater complexity of contaminant sources, scale, and operational purposes (e.g., investment, management and reporting of implementation actions).

2 BASELINE CONFIGURATION, CALIBRATION & VALIDATION

The FWMT is based on open-sourced US-EPA modelling software applied extensively in the United States (e.g., Loading Simulation in C++ for watershed modelling; System for Urban Stormwater Treatment and Analysis Integration for process-based interventions [Shen et al., 2004; Shoemaker et al., 2009]). LSPC simulates flow and contaminant build-up/wash-off and transformation (e.g., deposition, resuspension, scour, desorption, nitrification and denitrification).

2.1 BASELINE CONFIGURATION OF LSPC

FWMT Stage 1 baseline or "current state" conditions were simulated for the period 2013-2017. Hydrological and contaminant processes are regionally configured for each of 106 HRU's spanning natural, productive and developed land surfaces of varying topography, soil type and intensity of use, as well as into three model reach types (e.g., scour, nitrification and denitrification). Regionwide LiDAR combined with soil type (S-MAPS, NZLRI), land cover and use information (Agribase, LCDB4), defines HRU extent within each of 5,465 sub-catchments – each sub-catchment possesses a routing node for hydrology and contaminant mass either to stream or coastal receiving environment (2,165 sub-catchments lacked sufficiently large, permanent streams to warrant instream process simulation). FWMT Stage 1 baseline modelling applied a static configuration for HRUs with continuous variation in climatic boundary conditions.

Reticulated wastewater networks are simulated separately (using Watercare Ltd. MIKE URBAN or Infoworks ICM models for Rosedale, Warkworth, Army Bay, Mangere, Pukekohe and Waiuku) but incorporated via 15-minute continuous discharge time-series at 359 engineered overflow points. Only wastewater contributions from constructed overflows are accounted for by the FWMT Stage 1.

A mix of 40 gauged and 188 virtual climate stations drive regionalized hydrological and contaminant processes across HRU's and in stream. Climate is represented uniformly within but can vary between sub-catchments. VCSN data is disaggregated into 15-minute continuous time-series of precipitation and evapotranspiration for each sub-catchment.

LSPC simulates contaminants derived from HRU's largely through build-up/wash-off and subsequent transformation processes to edge-of-stream. For sediment, particle sizes of eroded material supplied by overland flow is divided into sand/silt/clay classes on basis of HRU hydrologic soil groups. Aside from build-up/wash-off, sediment is also lost via runoff-detachment processes from HRU's within the FWMT Stage 1 – only for pervious surface type HRU's. Runoff-detachment processes are driven by rainfall intensity, vegetation cover, and associated hydrologic soil group-defined detachment coefficients for HRU types.

Flow and contaminants are amalgamated at sub-catchment outlets, then routed through a regional stream network of 3,085 km comprised of 5,667 model reaches. Reach characteristics are uniform for the model reach length (<1-5km) from observational surveys and empirical predictions (e.g., of Manning's n, slope, width, bank height). The FWMT reach network downstream includes 15% of the 21,130 km of permanent and intermittent streams in the Auckland region, predominantly 3rd order but also all higher order reaches (Storey and Wadhwa, 2009).

FWMT reaches have been categorised into stream groups, to systematically configure nutrient (nitrification, denitrification, PO_4 adsorption) and erosion (scour) processes throughout the 3,085 km of simulated streams. FWMT reaches were assigned into their corresponding three groups for nutrient process parameterization from information on shade and upstream agricultural land cover – of 25 classes across breaks in both factors, a low, medium and high group was assigned using HRU land cover information, Watercourse Assessment Reports (WAR) and generalized rules from Freshwater Ecosystems of NZ (FENZ; Leathwick, 2010). Logically, low (less-shaded) nutrient groups were parameterized to receive severalfold more incident solar radiation than high (most-shaded) reaches. Land cover information was pertinent to classifying instream nutrient processes, through the associated and widespread effects of agriculture on instream and near-stream habitat (i.e., whether correlative or causative, altered instream function has been noted accompanying increasing proportions of agricultural land use – Larned et al., 2016; PMCSA, 2017). Nitrification and denitrification processes were regulated by dissolved oxygen and temperature, configured by stream group (e.g., parameterized into the three groups but able to vary markedly within group in response to oxygenation; denitrification occurring only where oxygenation <1 mg/L; nitrification simulated by first-order reaction linked to amount of NH4N in reach).

Reach erosion groups were defined by reach slope (averaged from up/downstream node elevation), bank material, bank cover and FWMT reach order (e.g., into low, medium and high as before). Both bank material and cover were estimated from WAR, FENZ, NZLRI and FWMT vegetation layer (LiDAR). Each of the three sediment groups were parameterized differently to represent varying sensitivity to scour (e.g., velocity-driven coefficient of scour from streambank soil matrix lesser for low [lined, hard-bed] than high sediment groups [soft-bed]; lesser critical shear stress for high than low groups). Scour processes were driven by estimated instream depth of flow, based on simulated discharge and channel profile.

Both HRU-delivered and stream-scoured sediment is processed separately as sand, silt and clay masses within each reach (assigning proportions based on hydrological soil group of contributing HRU). The FWMT Stage 1 includes reach suspension and deposition processes, adopting variable shear stress for both processes, driven by simulated depth of flow (e.g., deposition [suspension] at 5 Pa [14 Pa] and 1 Pa [9 Pa] for silt and clay, respectively; power function of velocity for sand). As before, depth of flow governed by cross-sectional profile and discharge which are configured uniformly within each reach.

An additional source of eroded sediment is accounted for by the FWMT Stage 1, gully erosion. The term describes the likely erosion of stream banks on non-modelled reaches. Gully erosion is represented by scour of pervious HRU surfaces via overland flow and is driven by surface runoff rate and a soil matrix factor (governed by hydrologic soil group). Gully erosion combined with erosion (scour) of modelled reaches, amounts together into "bank erosion" within the FWMT Stage 1.

Configuration of FWMT Stage 1 involved populating the model database with initial process parameters derived from several sources, including previous calibrated versions of the LSPC model (United States), modelling literature (NZ, international), prior regional modelling (Contaminant Load Model and CLUES) and contractor reports for Auckland Council (e.g., end-of-pipe, pervious and impervious surface stormwater data from URQIS, RIMU and HW).

2.2 CALIBRATION & VALIDATION APPROACH

The FWMT has been calibrated and validated for 46 continuous flow and 36 monthly water quality monitoring stations, distributed downstream of ~15% of the Auckland regional area, for the period five-year period 2012-2016[1].

Calibration proceeded sequentially, reviewing hydrological and then contaminant output. Monitoring stations downstream of more homogenous HRU mixes were prioritized for calibration Remaining monitored stations were utilized for validation of calibrated responses. All calibration stations were aligned to one of four dominant land cover classes, to inform associated HRU parameterization (e.g., extensive pasture, horticulture, forest and developed cover of upstream sub-catchment).

For hydrology, 16 monitoring stations were calibrated continuously by utilizing daily average flow time-series (~1,825 observations per station from 2012-2016). A further 30 were validated alike. All 46 hydrology stations tiered by dataset quality; record length and continuity, risk to free-flow and span of gauged stages were assessed to create five increasingly lower quality tiers from 1-5 (i.e., enabling investigation of performance for observational record quality) (Fordham, 2020)[2].

For water quality, 17 monitoring stations were calibrated discretely by utilizing monthly grab samples (observed) and daily average flow-weighted (modelled) estimates on the same day as field sampling (~60 observations per station from 2012-2016).

---

[1]Hydrology configuration was informed by longer simulations of 2002-2017.

[2]Tiers 1 stations possess >90% high quality data codes, not tidally influenced, not structurally or macrophyte impeded, and >75% of rating curve is verified with field measurements. Tier 2 pass latter except <75% of rating curve is verified by field measurement. Tier 3 include >75% high quality data codes, not tidally influenced and not structurally or macrophyte impeded. Tier 4 included stations with <75% high quality data codes and not tidally influenced, structurally or macrophyte impeded. Tier 5 are as per Tier 4 and either tidally influenced, structurally or macrophyte impeded.

2.3 GRADING STATE – CONCENTRATION

Continuously-simulated contaminant time-series for 5,667 model reaches[3] and discretely-observed contaminant time-series at 36 SoE stations were translated into water quality grades using a combination of national and regional objective guidance (Tables 1 & 2). Existing NOF guidance was utilized for grading nitrogen-toxicity (NO3N; NH4N) and faecal indicator bacteria (*E.coli*) (MfE, 2017). NO3N toxicity guidance was applied to simulated total oxidized nitrogen (TON) within the FWMT Stage 1, noting that analysis of monitored datasets suggested nitrate-nitrogen comprised the vast majority of TON (e.g., minimum >90%, typically >99% by mass at majority of monitored stations).

Table 1. Grading guidance for FWMT Stage 1 – for dissolved Zinc, modified from ANZ (2019). Bottom lines are proposed and not mandated by the NOF.

Attribute Grade	Dissolved Zinc[1]							
	Default Hardness = 30mg/L		Regional Hardness = 38.30 mg/L		Tamaki Hardness = 64.40 mg/L		Wairoa Hardness = 19.20 mg/L	
	Statistic	Value (ug/L)	Statistic	Value (ug/L)	Statistic	Value (ug/L)	Statistic	Value (ug/L)
A	Median	≤2.4	Median	≤ 2.9	Median	≤4.6	Median	≤1.7
	95th%	≤8	95th%	≤ 9.6	95th%	≤15.2	95th%	≤5.6
B	Median	>2.4 and ≤8	Annual Median	> 2.9 and ≤ 9.6	Median	>4.6 and ≤15.2	Median	>1.7 and ≤5.6
	95th%	>8 and ≤15	95th%	> 9.6 and ≤ 18.0	95th%	>15.2 and ≤28.5	95th%	>5.6 and ≤10.5
C	Median	>8 and ≤31	Median	> 9.6 and ≤ 37.2	Median	>15.2 and ≤58.9	Median	>5.6 and ≤21.7
	95th%	>15 and ≤42	95th%	> 18.0 and ≤ 50.4	95th%	>30.0 and ≤79.8	95th%	>10.5 and ≤29.4
Regional Bottom Line								
D	Median	>31	Median	> 37.2	Median	>58.9	Median	>21.7
	95th%	>42	95th%	> 50.4	95th%	>79.8	95th%	>29.4
Hardness multiplier ANZ (2019)	1.0		1.2		1.9		0.7	

[3]FWMT reaches are distributed in 3,300 of the 5,465 sub-catchments, with 2,165 sub-catchments configured to discharge to coast or other region (e.g., lacking a simulated reach with streams of 3rd order or greater). Yields across the full 5,465 sub-catchments are accounted for.

Table 2. Grading guidance for FWMT Stage 1 for *E.coli*, dissolved Cu, TON (graded as per NO3N), TAM (graded as per NH4N), DIN and DRP. Bottom lines for dissolved Cu are proposed and not mandated by the NOF.

Attribute Grade	**FRESHWATER CONTAMINANTS CALIBRATED & GRADED IN FWMT STAGE 1**											
	E. coli		Dissolved Copper[1] (DCu)		Total Oxidised Nitrogen (NO3N)		Total Ammoniacal Nitrogen[2] (NH4N)		Dissolved Inorganic Nitrogen (DIN)[3]		Dissolved Reactive Phosphorus (DRP)[3]	
	Statistic	MPN/100 mL	Statistic	µg/L	Statistic	mg/L	Statistic	mg/L	Statistic	mg/L	Statistic	mg/L
A	% over 540	< 5 %	Median	≤1	Median	≤1.0	Median	≤0.03	Median	≤0.24	Median	≤0.006
	% over 260	< 20 %										
	Median	≤130	95th%	≤1.4	95th%	≤1.5	Maximum	≤0.05	95th%	≤0.56	95th%	≤0.021
	95th%	≤540										
B	% over 540	5 - 10 %	Median	>1 and ≤1.4	Median	>1.0 and ≤2.4	Median	>0.03 and ≤0.24	Median	>0.24 and ≤0.50	Median	> 0.006 and ≤ 0.010
	% over 260	20 - 30 %										
	Median	≤130	95th%	>1.4 and ≤1.8	95th%	>1.5 and ≤3.5	Maximum	>0.05 and ≤0.40	95th%	>0.56 and ≤ 1.10	95th%	>0.021 and ≤0.030
	95th%	≤1000										
C	% over 540	10 - 20 %	Median	>1.4 and ≤2.5	Median	>2.4 and ≤6.9	Median	>0.24 and ≤1.30	Median	>0.5 and ≤ 1.0	Median	>0.010 and ≤ 0.018
	% over 260	20 - 34 %										
	Median	≤130	95th%	>1.8 and ≤4.3	95th%	>3.5 and ≤9.8	Maximum	>0.40 and ≤2.20	95th%	>1.10 and ≤ 2.05	95th%	>0.030 and ≤ 0.054
	95th%	≤1200										
National or Regional Bottom Line												
D	% over 540	20 - 30 %	Median	>2.5	Median	>6.9	Median	>1.30	Median	>1.0	Median	>0.018
	% over 260	>34 %										
	Median	>130	95th%	>4.3	95th%	>9.8	Maximum	>2.20	95th%	>2.05	95th%	>0.054
	95th%	>1200										
E	% over 540	> 30 %										
	% over 260	> 50 %										
	Median	>260										
	95th%	>1200										
Guidance	NPS-FM (2017)		ANZ (2019)		NPS-FM (2017)		NPS-FM (2017)		NPS-FM (2019)[3]		NPS-FM (2019)[3]	

[1]No Dissolved organic carbon adjustments have been made for this analysis.
[2]Attribute states shown are based on pH 8 and temperature of 20° C.
[3]Proposed guidance subject to ongoing review in Essential Freshwater: Healthy Water, fairly allocated revisions to the NPS-FM (MfE, 2019).

Proposed NOF guidance was utilized for grading instream eutrophication (DIN, DRP), conservatively applied to all reaches independent of substrate (MfE, 2019). Proposed regional objective guidance (aligned with ANZ, 2019) was used for heavy metal toxicity grading (Cu, Zn). Dissolved metal concentration was derived from simulated total concentration using regional conversion factors. Dissolved Zn was also hardness corrected.

All modelled grades were assigned from flow-weighted daily average concentrations, for the 2013-2017 interval (5-years[4]) (Table 2). All concentrations were estimated as Hazen percentiles (e.g., McBride, 2016). Modelled grades apply to the immediate downstream FWMT reach, to estimate length-weighted distributions of modelled instream water quality.

2.4 PERFORMANCE – CONTINUOUS & GRADING ACCURACY

Multiple performance metrics (NSE, %Bias, r^2) were estimated for various flow and seasonal conditions. Thresholds from Moriasi et al. (2015) were adapted for assessing *accuracy*. In line with Moriasi et al. (2015), more conservative thresholds were set applied to hydrology than contaminant concentration or loading prediction (Tables 3 and 4). Notably, performance is only one measure of uncertainty – reporting on the ability to represent observed conditions, not on the sensitivity of model predictions to parameterization of processes or choice of inputs. Sensitivity testing is ongoing for inclusion with baseline technical reports and guide Stage 2 development.

Table 3. Performance assessment metrics for hydrological accuracy – FWMT Stage 1.

Metric	Conditions	Hydrology Performance (modified from Moriasi et al., 2015)			
		Very good	Good	Satisfactory	Unsatisfactory
r^2	All flow	>0.85	0.75-0.85	0.60-0.75	≤0.60
	Seasonal flows	>0.75	0.60-0.75	0.60-0.50	≤0.50
	Highest 10% flows				
	Lowest 50% flows				
	Storm flows				
	Baseflows				
Nash-Sutcliffe Efficiency (NSE)	All flow	>0.80	0.70-0.80	0.50-0.70	≤0.50
	Seasonal flows	>0.70	0.50-0.70	0.40-0.50	≤0.40
	Highest 10% flows				
	Lowest 50% flows				
	Storm flows				
	Baseflows				
Percent bias (PBias)	All flow	+/-5	+/-5-10	+/-10-15	+/->15
	Seasonal flows	+/-10	+/-10-15	+/-15-25	+/->25
	Highest 10% flows				
	Lowest 50% flows				
	Storm flows				
	Baseflows				

[4] The 2017 water year was exceptionally wet and was included in grading output to ensure sediment outcomes during periods of exceptional stream erosion could be accounted for. Calibration/validation omits 2017 – similar high flows but of lower frequency were observed in 2012-2016.

Table 4. Performance assessment metrics for contaminant accuracy (concentration and loading) – FWMT Stage 1.

Metric	Conditions	Contaminant Performance (modified from Moriasi et al., 2015)			
		Very good	Good	Satisfactory	Unsatisfactory
r^2	All flow	>0.70	0.60-0.70	0.30-0.7	≤0.60
	Seasonal flows	>0.75	0.60-0.75	0.60-0.50	≤0.50
Nash-Sutcliffe Efficiency (NSE)	All flow	>0.85	0.75-0.85	0.60-0.75	≤0.60
	Seasonal flows	>0.75	0.60-0.75	0.60-0.50	≤0.50
Percent bias (PBias)	All flow	+/- <15	+/- 15-20	+/- 21-30	+/- >30
	Seasonal flows	+/- <20	+/- 20-30	+/- 31-40	+/- >40

Observations (grab samples) were compared to flow-weighted average *daily* concentration with equivalent performance thresholds for r^2, NSE and Pbias metrics as grades. SOURCE modelling in Greater Wellington reported model performance by comparing *monthly* average predicted concentrations to SoE sampling, for a single performance metric (PBias). SOURCE models for the Bay of Plenty apply both more and less conservative thresholds (e.g., NSE ≤0.05 or PBias ≥±0.15 is "unsatisfactory" for records of <100 observations).

Of the 36 SoE stations, 16 possessed both observed concentration and flow records. Loading estimates were derived as product of daily flow by grab-sampled concentration. Insodoing, potentially biasing observed loads at sites with marked daily flow variation.

Assessing performance at grading is a relatively novel test for continuous modelling in NZ – neither Blyth et al. (2018), Easton et al. (2019) nor Loft et al. (in prep.) assess grading-based performance. However, doing so aligns with the purpose of the FWMT as a decision-support tool for the NPS-FM.

Grading-based approaches are widely practiced in public health management, where modelling performance is linked to determining true positives (failing grades, precautionary reporting) and true negatives (passing grades, permissive reporting) (e.g., Nevers and Whitman, 2011; Thoe et al., 2014).

In Stage 1 FWMT, three alternative grading performance approaches were adopted for sensitivity (national bottom-line failing grades) and specificity (all grades): (1) reporting equivalent grades as those exactly alike (e.g., A=A through to E=E); (2) reporting equivalent grades as those within one band of observed (A=A,B, B=A,B,C through to E=D,E); and (3) reporting median or 95th% if modelled equivalents were within a fixed absolute range of observed – ranges being the average of upper and lower concentrations of A and B grades, or for C and D graded sites, the range in median or 95th% of C grade. Approach (3) is most complex but ensure SoE stations near a grading threshold to not be penalized for minor absolute difference in simulated concentration. Hence, approach (3) is likely a balanced grading approach, with greater performance than (2) but lesser than (1).

Table 5. Grading performance bands for approach (3), absolute concentration range.

Grading Measure		*E.coli*	DRP	DIN	NO3N	NH4N	Cu	Zn Region	Zn Tamaki	Zn Wairoa
Median	A or B	130	0.005	0.25	1.2	0.12	0.7	4.8	7.6	2.8
	C, D or E	130	0.008	0.5	4.5	1.06	1.1	27.6	43.7	16.1
95th%	A or B	500	0.015	0.55	1.75	0.2	0.9	9	14.25	5.25
	C, D or E	500	0.024	0.95	6.3	1.8	2.5	32.4	51.3	18.9

3 CURRENT STATE MODEL PERFORMANCE

3.1 REPORTING

To ensure transparency and rigour of FWMT development, four peer-reviewed reports are in preparation for baseline or "current state" FWMT output, including: baseline inputs; baseline configuration and performance; baseline outputs (rivers); and baseline outputs (lakes). Publication is expected in 2020. Only indicative summaries are presented here.

3.2 PERFORMANCE – CONTINUOUS ACCURACY

Tables 6 and 7 present calibrated and validated accuracy for hydrology and contaminants at stream SoE stations from 2012-2016 in Auckland. Despite stricter performance metric thresholds, the vast majority of continuously monitored stations were predicted with "satisfactory" or greater hydrological performance at "all flows".

Table 6. FWMT Stage 1 accuracy for continuous *flow* across 46 stations (2012-2016).

Tier	Hydrology Monitoring Locations	Drainage Area (km2)	Performance Metrics (Seasonal)															Performance Metrics (Flow Regime)														
			PBIAS					r-squared					Nash-Sutcliffe E					PBIAS					r-squared					Nash-Sutcliffe E				
			All	Winter	Spring	Summer	Fall	All	Winter	Spring	Summer	Fall	All	Winter	Spring	Summer	Fall	All	Top 10%	Storms	Low 50%	Baseflow	All	Top 10%	Storms	Low 50%	Baseflow	All	Top 10%	Storms	Low 50%	Baseflow
Tier 1	Hoteo River @ Gubbs	268	+	+	+	+	+											+	+	+	+	+										
	Wairoa River @ Tourist Road	161	-	-	+	-	-											-	-	-	+	-										
	Mahurangi @ College	46.8	+	+	+	-	+											+	+	+	+	-										
	Opanuku @ Vintage Reserve	26.5	+	+	+	+	+											+	+	+	-	-										
	Oratia @ Millbrook Road	22.9	-	-	-	-	-											-	-	-	-	-										
	Otara @ Hills Road Bridge	18.9	+	+	+	+	+											+	+	+	-	-										
	Wairau Creek @ Motorway	11.1	+	+	+	+	+											+	+	+	-	+										
Tier 2	Mahurangi Argonaut @ College	46.8	-	-	+	-	-											-	+	+	+	-										
	Swanson Stream @ Woodside Reserve	22.6	+	+	+	+	+											+	+	+	+	-										
	Opanuku Stream @Candia Road Bridge	15.9	+	+	+	+	+											+	+	+	-	-										
	Oteha River @ Days Bridge	12.2	+	-	+	+	+											+	+	+	-	-										
	Orewa @ Kowhai Ave	9.7	+	+	-	-	-											+	+	-	-	+										
	Motions Stream @ Western Springs.	7.5	-	-	+	+	-											-	-	-	+	+										
	Lucas @ Gills Road	6.3	+	+	+	+	+											+	+	+	-	-										
	Ngakoroa Stream @ Mill Rd	4.7	+	+	+	-	+											+	+	+	-	-										
	Mangemangeroa	4.6	-	+	+	-	-											-	+	-	-	-										
	Eskdale Stream at Lauderdale Reserve	3.9	-	-	+	+	+											-	+	+	+	-										
	Tamaki Trib at Bowden Road Crump Weir	3.1	-	-	-	-	-											-	-	-	+	+										
	Awaruku stream at Glenvar Road	1.7	-	-	-	-	-											-	-	-	-	-										
	Kaipatiki Stream at Kaipatiki road	1.5	-	-	+	+	-											-	-	-	+	+										
	Mairangi Bay Stream at Tennis Club	0.6	-	-	-	-	-											-	-	-	-	-										
	West Hoe @ Halls	0.5	+	-	+	+	+											+	+	+	+	+										
Tier 3	Kaipara River @ Waimauku	155.4	+	+	+	+	+											+	+	+	-	+										
	Rangitopuni River @ Walkers	81.5	+	+	+	-	-											+	+	+	-	-										
	Waiteitei River @ Sandersons	80.6	-	-	+	-	-											-	+	-	-	-										
	Ararimu River @ Old North Rd Bridge	66.8	-	-	+	-	-											-	-	-	-	+										
	Kaukapakapa @ Taylors	61.9	-	-	-	-	-											-	-	-	-	-										
	Mangawheau Stream @ Weir	30.4	+	-	+	+	-											+	-	-	+	+										
	Meola Creek at Motions Road Weir	14.7	-	-	-	-	-											-	-	-	-	-										
	Puhinui @ Drop Structure	11.6	-	-	-	-	-											-	-	-	+	-										
	Tamahunga River @ Quintals Falls	8	+	+	+	+	-											+	+	+	-	-										
	Oakley Creek at Richardson Road	6.1	+	+	+	+	+											+	+	+	+	+										
	Wairau Creek @ Chartwell Road	1.4	+	+	+	+	+											+	+	+	+	+										
Tier 4	Papakura @ Great South Road Bridge	51.6	-	+	+	-	-											-	+	-	-	+										
	Newmarket Stream @ AYR Street	5.5	-	-	-	-	-											-	-	-	-	-										
	Whau Stream at Crump Wier	4.7	-	-	-	-	-											-	+	+	-	-										
	Vaughn Stream @ Lower Weir	2.3	-	-	+	-	-											-	+	+	-	-										
	Taiaotea stream at Freyberg Park	2.2	+	+	+	+	-											+	+	+	-	-										
	Okura @ Weiti Forest	1.7	-	-	+	+	-											-	+	-	-	-										
Tier 5	Makarau at Coles	53.7	+	+	+	+	+											+	+	+	+	-										
	Kumeu @ Maddrens Weir	44.9	+	+	+	+	-											+	+	+	-	+										
	Waitangi @ S H Bridge	17.6	-	+	-	-	-											-	+	+	-	-										
	Oratia @ Parrs Cross Road	16.7	-	-	+	-	-											-	+	+	-	-										
	Waiwhiu Stream @ Dome Shadow	8.6	+	+	+	+	+											+	+	+	+	+										
	Okura Creek @ Awanohi Rd	5.8	-	-	-	-	-											-	+	-	-	-										
	Taiorahi Stream at Westbourne ave	1	+	+	+	+	+											+	+	+	-	+										

Very Good · Good · Satisfactory · Unsatisfactory
\+ Overpredicts · - Underpredicts

Across "all" observations at 46 continuous flow monitoring stations:

- 82-86% of Tier 1 and 2 stations achieved "satisfactory" or better accuracy, with 50-59% of "good" or better accuracy (i.e., range spans PBias, r^2 and NSE values);
- 76-80% of Tier 1-5 stations achieved "satisfactory" or better accuracy.

Amongst flow subsets, the FWMT Stage 1 has least consistent accuracy for lowest 50% of flows – for which 5-50% of Tier 1 and 2 stations achieved "satisfactory" performance across PBias, r^2 and NSE. However, baseflow predictions at 86-91% of Tier 1 and 2 stations were predicted with "satisfactory" or better accuracy. Table 6 also demonstrates increased hydrological modelling accuracy with increasing catchment area.

Table 7. Summary of calibrated and validated accuracy for r^2 under "all" conditions for *concentration* (and *loading* in brackets) at 36 SoE stream quality monitoring sites (2012-2016).

Accuracy	TSS	TN	TON	NH4N	TP	DRP	Cu	Zn	*E.coli*
Very good	0% (61%)	0% (69%)	0% (42%)	0% (62%)	0% (74%)	0% (69%)	0% (36%)	0% (48%)	0% (21%)
Good	6% (11%)	3% (14%)	3% (25%)	0% (33%)	0% (11%)	0% (14%)	0% (28%)	0% (16%)	3% (15%)
Satisfactory	17% (28%)	11% (14%)	11% (22%)	0% (5%)	14% (14%)	0% (17%)	8% (36%)	8% (36%)	6% (52%)
Unsatisfactory	78% (0%)	86% (3%)	86% (11%)	100% (0%)	86% (1%)	100% (0%)	92% (0%)	92% (0%)	91% (12%)

Station performance for contaminant concentration and loading has been assessed, for "all" as well as various subsets of flow and season. Combined, patterns for accuracy in TSS, TN, TON, NH4N, TP, DRP, Cu, Zn and *E.coli* are complex (i.e., varying markedly between contaminants by metric and no clear pattern of performance with sub-catchment size unlike hydrology).

The FWMT Stage 1 performance varies across all three accuracy measures, for both concentration and loading. "Satisfactory" or better performance for concentration was achieved at 12-50% (PBias), 0-22% (r^2) and 0-3% of stations (NSE). Loading performance was generally markedly better, being "satisfactory" or better at 12-32% (PBias), 88-100% (r2) and 12-60% of stations (NSE). r^2 performance was generally best, albeit no station predicted at "very good" accuracy for concentration but 21-74% of station loads predicted with "very good" accuracy. Weaker NSE contaminant performance is likely the consequence of greater simulated extremes, with NSE disproportionately weighted to extreme outliers (e.g., McCuen et al., 2006). Those greater simulated extremes might well be real but missed by monthly grab sampling, and possibly reduced using if instead flow-weighted monthly average were compared to observed (e.g., as per Bay of Plenty and Greater Wellington SOURCE models).

3.3 PERFORMANCE – GRADING ACCURACY

Table 8 presents grading performance. TSS and TP are not graded with regional and national instream guidance limited; ANZ can supply pass-fail "trigger value" thresholds but these ostensibly are not intended to report an "effect" (e.g., dissimilar from NOF attributes in the NPS-FM). Across seven attributes with national or regional attribute guidance, the FWMT Stage 1 performs well. For instance, 55-100% of "failing" stations were correctly graded at their equivalent grade, and 91-100% within an additional grade.

For a regionalized and continuous model, more likely to simulate short-term events and a reliance on 95^{th}% statistics for grading each attribute, there is remarkably little grading disagreement. By grading to a 6-year period (2013-2017), that agreement is reassuring (i.e., the longer observational records improves the likelihood that the observed 95^{th}% has not omitted short-term events better captured by the FWMT's continuous resolution).

Table 8. Summary of calibrated and validated grading performance across 36 stream monitoring sites (2013-2017). Upper values assessed as proportion of predicted grades exactly alike to observed grade. Lower values in brackets assessed as within ±1 grade of observed grade. All grades span "A" through to "E", failing grades refers only to "D" and "E". Buffered refers to within band-defined range of observed measure (e.g., median, 95th%).

Grading performance (approach in text)	Predicted grades as % of monitoring locations						
	Human health	Ecosystem health					
	E.coli	DRP	DIN	TON	NH4N	Cu	Zn
Failing grades (1)	100% (100%)	55% (91%)	71% (100%)	100% (100%)	100% (100%)	NA NA	75% (100%)
All grades (2)	86% (86%)	28% (94%)	56% (86%)	78% (100%)	28% (67%)	17% (71%)	50% (96%)
Buffered median (3) (Buffered 95th%)	56% (25%)	69% (31%)	78% (72%)	100% (100%)	97% (42%)	75% (50%)	92% (75%)

Grading performance reduced across "all" compared to just "failing" grades (e.g., A-E and D-E grades, respectively). For instance, 17-86% of stations were graded correctly across all grades and contaminants, whilst 67-100% were predicted within an additional grade. Equally, that medians and 95th% are frequently predicted within a grade's width of observed (56-100% and 25-100%, respectively). Grading performance also indicates lower contaminant concentrations are harder to predict. Likely causes include: greater sensitivity of less degraded sites to instream process errors; greater sensitivity at less degraded sites to non-hydrologic contaminant processes (unrepresented); increased error simulating minor or modest yields for equivalent hydrology (i.e., lesser concentrations for equivalent flow from low-yield HRUs); effects of varying ranges in medians or 95th% for grades; and, potential for configuration of FWMT Stage 1 not capturing recent changes in HRUs upstream of less degraded SoE stations.

Grading performance varied markedly between contaminants, being best for *E.coli* (e.g., 86% of all grades correctly predicted at their grade; 100% of fails correctly predicted – Table 8). Likely causes include the broader number of measures used to grade *E.coli* attribute state (e.g., 95th%, median, %>260 MPN/100 ml and %>540 MPN/100 ml). All other attributes are graded only by 95th% and median. Although a wide gradient of *E.coli* concentrations spanned by grades might be responsible, lengthy gradients of severalfold increases to median or 95th% exist in other attributes. So, a more likely additional cause is that observed and simulated "failing" *E.coli* concentrations were further above their national bottom-line thresholds (i.e., that *E.coli* concentrations observed were relatively higher than any "C/D" threshold in other contaminants).

Grading performance raise novel challenges for the FWMT and Auckland Council. Two are fundamental to ongoing FWMT development:

1. SoE stations might not represent regional variation in grading. The FWMT offers greater coverage, resolution and representativity of instream contaminant concentrations (however accurately). Unless observational sampling is distributed along gradients in regional water quality, any performance assessment is biased. Further examination of spread in regional HRU composition, stream network, and climatic conditions is needed to guide *where* to representatively sampling.
2. SoE sampling might not resolve numeric attribute state. A question remains of whether a strong association should be expected between the widely differing resolution of observed (monthly, grab-sampled) and FWMT modelled time-series (continuous, daily integrated). Further examination of diel variation in rainfall and instream flow, of process-responses and ultimately, of daily concentration by monthly SoE observations is needed to guide *how* to representatively sample.

Whilst ongoing analysis will determine and report on (1), McBride (2016) comments on (2), a challenge relevant to the FWMT but also all continuous water quality modelling in New Zealand, given the prominence of 95th% measures in NPS-FM attributes.

McBride (2016) notes the potential for "state switching" in reporting from infrequent SoE sampling (i.e., without change in state due to statistical sampling error). A recommendation of 40-60 samples should result in *sufficient* confidence about exceedances of 95th% thresholds being indicative of a real exceedance or sampling error – "sufficient" being subject to whatever misclassification error risk is acceptable, and whether a permissive, even-handed or precautionary reporting approach is needed. Regardless, less samples are required to confidently report on exceedance of median thresholds. McBride (2016) suggests a five-year period of monthly sampling assigning observed grades to SoE stations is likely sufficient provided no *bias* exists in sampling (e.g., FWMT Stage 1 reporting is for ~60 samples from 2012-2016). However, no analysis is available for stream SoE data to suggest sampling is not bias-free, represents full spread in regional contaminant concentrations or does not require flow-stratification.

Questions of SoE sampling location and frequency affecting grading are critical to catchment modelling for the NPS-FM. The reliance of NOF grading on "statistical tails" (95th%, maxima) raises the challenge further for continuous and process-based modelling. Empirical models can be developed that incorporate longer observed gradients (e.g., more stations from larger areas over longer periods). However, through their inherent lack of deterministic routines, statistical models are inherently of limited value for intervention or scenario modelling (e.g., cannot support FWMT purposes for operational programmes to implement the NPS-FM). Overall, a cautionary note remains that although the FWMT Stage 1 performs well at grading, whether that is an artefact of unrepresentative observational SoE datasets remains.

4 CURRENT STATE OUTPUTS

Only indicative, regional summaries of FWMT Stage 1 instream outputs follow. Full output is included in Baseline Outputs (Rivers) report (Bambic et al., in prep.). The focus here is on innovative means of presenting the wealth of information available from the FWMT Stage 1. All output produced below can be generated by watershed, catchment and sub-catchment for the 2,761 reach nodes and 3,085 km of moderate or larger streams distributed across the 3,300 sub-catchments. Yields from all 5,465 sub-catchments spanning the full region (489,000 Ha) are available to edge-of-stream and to instream or downstream FWMT reach nodes (e.g., attenuated for instream processing). Source-to-sink knowledge is provided by the process-based capability of the FWMT. Further investigation is needed to determine spread in yield and attenuation rates across various HRU's to inform confidence in findings by comparisons to other model builds.

4.1 WATER QUALITY GRADING

The first region-wide, comprehensive and integrated assessment of instream water quality for Auckland, over the period 2013-2017 and from the Stage 1 FWMT is presented in Figure 3. For comparative purposes an "integrated" grading outcome is presented, replacing modelled with observed grade, for all FWMT reaches of equivalent order to SoE station (approximately 9.6% or 297 km of FWMT reach network). Importantly and as above, limitations of SoE monitoring for informing modelling undoubtedly exist. Hence whilst broadly in agreement, modelled output now requires representative, targeted validation monitoring (e.g., of repeated events, HRU mixes, climate and stream types).

Regionwide		Attainment of Attribute State by Model Stream Length (km) or Number of Stations (#)				Percent of Stream Length or Stations Attaining Attribute State
		A	B	C	D or E	
Dissolved Inorganic Nitrogen	Predicted	1,276	667	664	478	41% 22% 22% 16%
	Observed	16	8	5	7	44% 22% 14% 19%
	Integrated	1,352	745	553	435	44% 24% 18% 14%
Dissolved Reactive Phosphorus	Predicted	283	351	636	1,814	9% 11% 21% 59%
	Observed	0	7	18	11	19% 50% 31%
	Integrated	278	362	799	1,647	9% 12% 26% 53%
Total Oxidised Nitrogen	Predicted	2,536	436	63	51	82% 14%
	Observed	29	5	1	1	81% 14%
	Integrated	2,620	350	65	51	85% 11%
Total Ammoniacal Nitrogen	Predicted	67	1,480	1,422	116	48% 46%
	Observed	19	10	6	1	53% 28% 17%
	Integrated	220	1,526	1,231	109	7% 49% 40%
Dissolved Copper	Predicted	1,538	399	888	261	50% 13% 29% 8%
	Observed	8	3	13	0	33% 13% 54%
	Integrated	1,576	401	887	220	51% 13% 29% 7%
Dissolved Zinc	Predicted	2,596	192	187	111	84% 6% 6%
	Observed	9	4	7	4	38% 17% 29% 17%
	Integrated	2,576	213	190	106	83% 7% 6%
E. coli*	Predicted	113	257	154	2,562	8% 83%
	Observed	1	5	0	30	14% 83%
	Integrated	124	264	149	2,548	9% 83%

Figure 3. Regionwide grading across 3,085 km of freshwater streams (2013-2017) combining FWMT Stage 1 (predicted) and SoE datasets (observed). Grades derived from worst of measures for attribute. Guidance for DRP and DIN is provisional and applied conservatively to all reaches. **E.coli* failing grades were predominantly "E" with <0.05% of FWMT reaches assigned "D" grades.

Uncertainty notwithstanding, 83% or 2,562 km of FWMT reaches fail national bottom-lines for *E.coli* suggesting a widespread challenge for NPS-FM implementation in Auckland. By contrast, failures in national bottom-lines for NO3N-toxicity appear localized, affecting just 2% or 51 km of FWMT reaches within the Manukau Harbour watershed (Figure 4). Whilst valuable, regional summaries can hide localized patterns. Fortunately, contaminant time-series and grading can be produced at regional through to sub-catchment scale. Such regional and sub-regional summaries will eventually be supplemented by further information on cost and spread-in-cost, for managing to varying instream objectives and whether simply to "C" grade or more improved objectives.

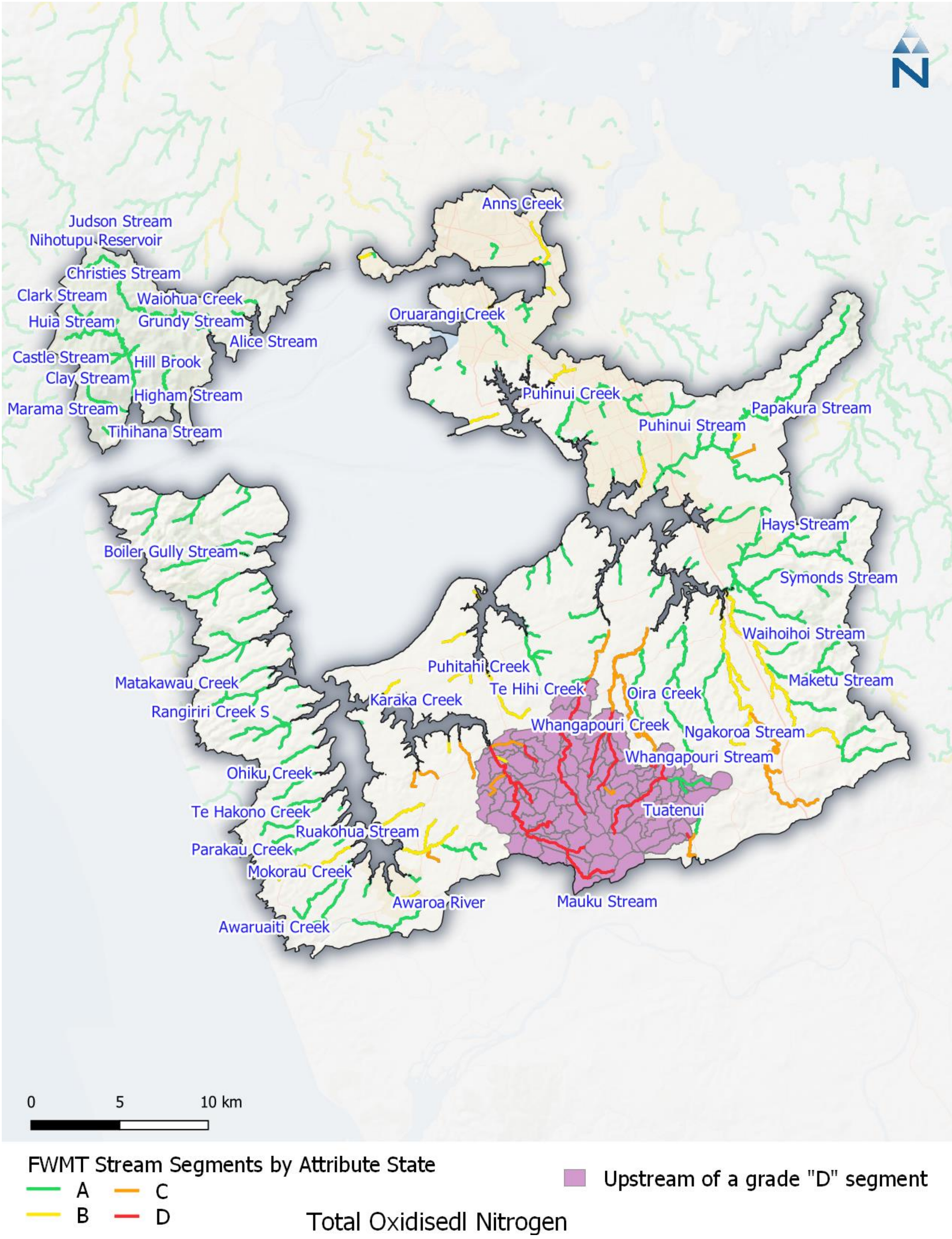

Figure 4. FWMT reaches within the Manukau Harbour watershed, graded by national objective for NO3N toxicity and sub-catchments upstream of reaching failing national bottom-lines highlighted.

Both Figures 4 and 5 help demonstrate the value of presenting grading outputs spatially. The continuous processing by LSPC enables all attribute measures to be reported, enabling prioritized management strategies for objectives. For instance, amongst *E.coli* measures, the 95^{th}% concentration was most frequently worst-graded and governed attribute grades for over three quarters of FWMT reaches. Consequently, 95^{th}% *E.coli* concentrations are likely to require greatest management for national objectives on improved human health for recreation. Corresponding pathways predominating during concentrations in excess of 1,200 MPN *E.coli*/100 ml can also be reported (i.e., during conditions governing 95^{th}% grading). Combined, the process-basis and continuous resolution of FWMT Stage 1 enables an understanding of conditions during which, pathways through which, and resultant attribute measures of greatest priority to achieving water quality objectives in Auckland streams – enabling targeted catchment management, regionwide.

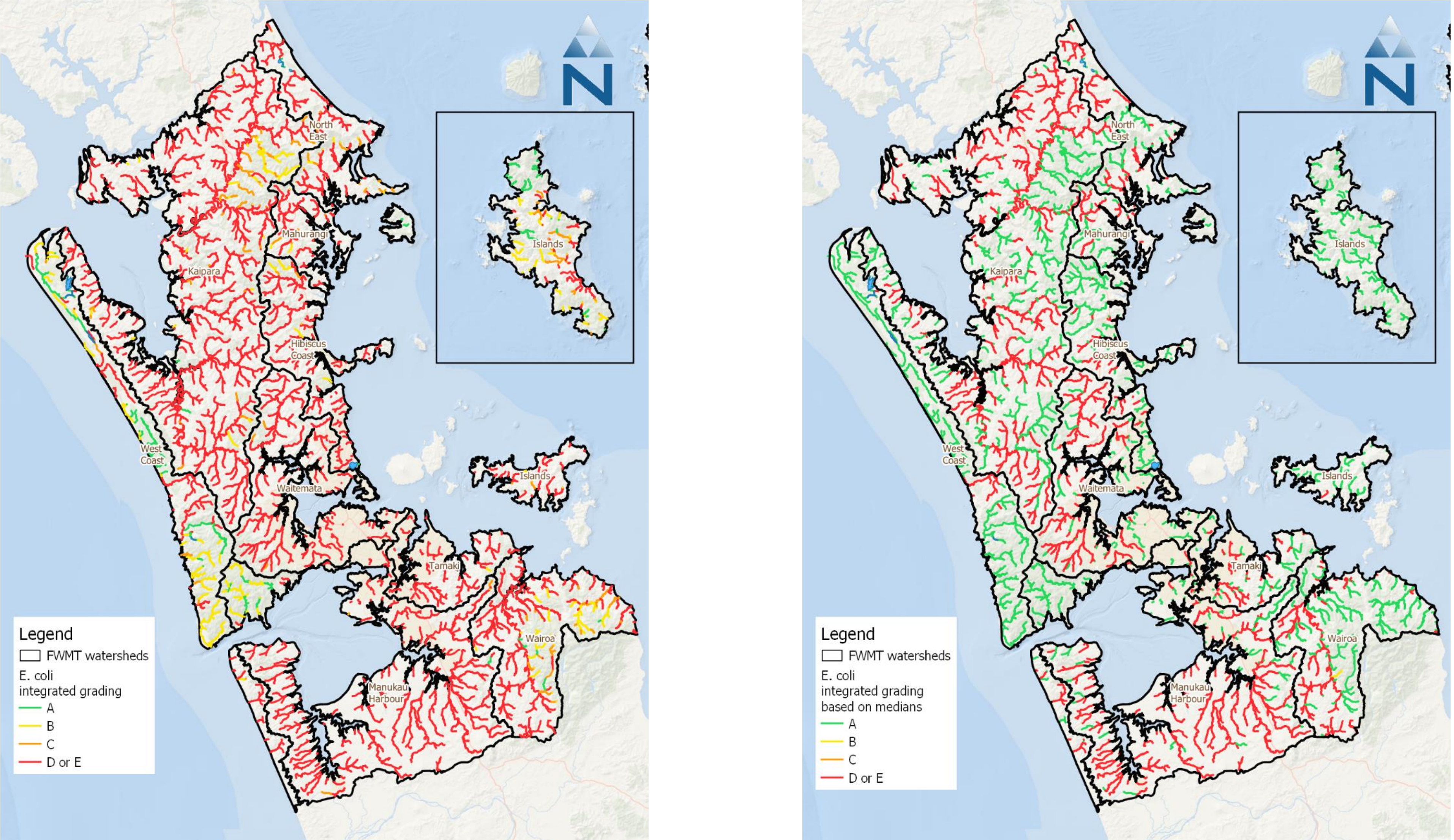

Figure 5. Regionwide integrated *E.coli* grading across 3.085 km of freshwater streams (2013-2017) combining FWMT Stage 1 and SoE datasets. Lefthand image of overall grade (worst of four measures: median, 95th%, %>260 and %>540 MPN/100ml). Righthand image is for median grade.

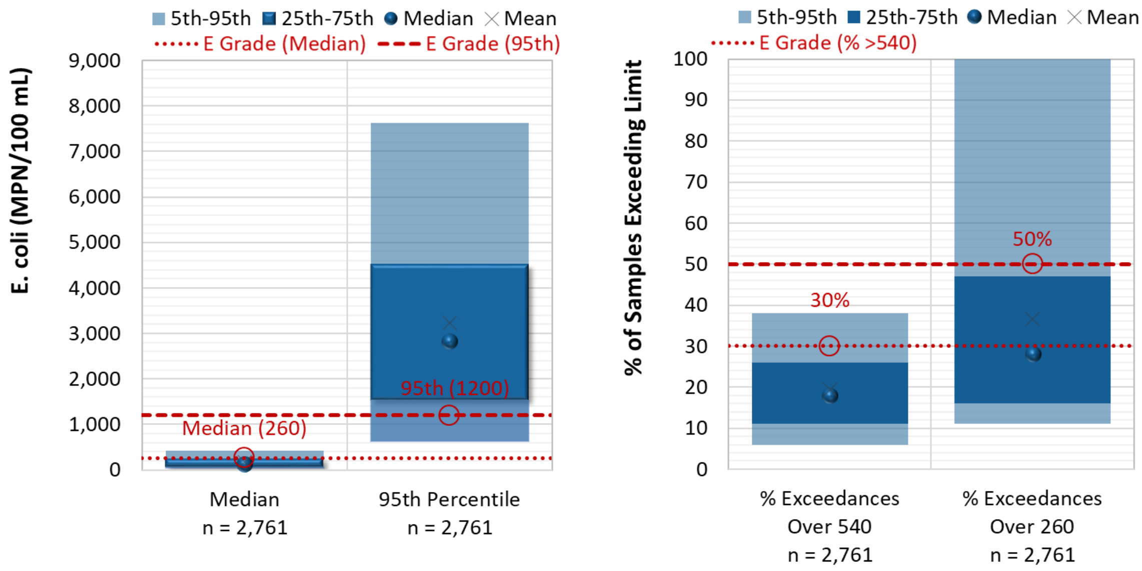

Figure 6. Regionwide integrated *E.coli* grading for 2,761 FWMT Stage 1 reach nodes (2013-2017) across all four measures governing overall attribute grade. Results demonstrate >75% of the 2,761 FWMT nodes failed the *E.coli* national bottom-line for 95^{th}% (e.g., 2,282 reach nodes exceeded a 95^{th}% of 1,200 MPN/100 ml).

4.2 LOADING & SOURCES

The process-simulation of FWMT Stage 1 enables *robust* source apportionment, with >88% of contaminant loads to SoE stations being predicted at "satisfactory", "good" or "very good" performance (2012-2016, across PBias, r^2 and NSE – Table 7).

Simulation of HRU and reach-group processes separately, enables loading to be reported to edge-of-stream (modelled FWMT reach) and for any FWMT node downstream or to coast. Bankside erosion and hydraulic scour are estimated for all modelled reaches. The FWMT Stage 1 can thereby account to stream, instream and to coast for a range of hydrological, hydraulic and contaminant processes.

HRU-specific yields can be estimated to guide risk-assessment or prioritisation of actions. Indeed, that is the basis of SUSTAIN decision-making. Prior to external peer review however, only regional average yields and loads to edge-of-stream are summarised in Table 9. More detailed summaries will be available for climatic variation represented by stormwater catchment and local board, following improved sensitivity analyses (i.e., once limitations of input and parameterization decisions are better understood).

Roofing and roading are large contributors to zinc losses, pasture and forest/open-space more generally across all contaminants, horticulture to TN loss and bank erosion to TSS losses (and associated TP, Zn, Cu). Regional summaries can mislead. For instance, roofing contributes 9.5% of regional dissolved Zn loads from only 5,896 Ha or 1.2% of the region, predominantly of urban sub-catchments. However, many regional findings are meaningful. For instance, that bankside erosion is a dominant sediment source in much of Auckland (57% of TSS load), supporting localized findings from more advanced modelling (B-STEM) and catchment investigative research across Auckland (e.g., Simon et al., 2015, 2016, 2017). A comparative assessment of FWMT Stage 1 TSS loading was made at 12 locations to empirical models available for the Auckland region, stressing remarkable agreement (see Figure 7 for Hoteo River).

Finally, FWMT Stage 1 outputs can be presented as heat-maps and/or apportionment pie-charts of yields, either to edge-of-stream or attenuated to locations instream or at coast.

Table 9. Regionalised contaminant yield by HRU types for 2013-2017, to edge-of-stream in FWMT Stage 1 (numbers in brackets are % of regional loading). **Note**: regionalized yields and total loading is not indicative of variation between watersheds. *Point source yields presented relative to combined paved urban, roof, roading and unpaved urban areas.

Source (HRUs combined by surface type)	TSS kg/Ha/yr (%)	TN kg/Ha/yr (%)	TP kg/Ha/yr (%)	TZn kg/Ha/yr (%)	TCu kg/Ha/yr (%)	*E. coli* (MPN billion/Ha/yr) (%)
Paved urban surfaces	203 (0.2%)	18.3 (1.2%)	1.1 (0.1%)	232 (2.5%)	54 (1.5%)	352 (3.5%)
Rooves	54 (0.1%)	7.0 (0.5%)	<0.0 (<0.1%)	788 (9.5%)	12 (0.4%)	NA (NA)
Roads and motorways	421 (0.6%)	8.6 (0.7%)	1.9 (0.2%)	604 (7.9%)	121 (4.0%)	128 (1.5%)
Unpaved urban surfaces	201 (0.4%)	2.7 (0.3%)	0.7 (0.1%)	46 (0.9%)	13 (0.7%)	16 (0.3%)
Septic Areas	310 (<0.1%)	9.4 (0.1%)	1.1 (<0.1%)	64 (0.1%)	20 (0.1%)	1171 (1.5%)
Horticulture	371 (0.9%)	115.3 (16.1%)	3.6 (0.8%)	46 (1.1%)	21 (1.2%)	5 (0.1%)
Pasture	357 (16.8%)	28.4 (76.3%)	18.0 (74.7%)	52 (24.1%)	24 (27.3%)	188 (78.3%)
Forest and Open Space	548 (23.9%)	1.9 (4.7%)	0.6 (2.3%)	82 (35.2%)	39 (41.7%)	34 (13.2%)
Bank Erosion (kg/m/yr)	89 (57.1%)	<0.0 (<0.1%)	0.4 (21.7%)	2.9 (18.4%)	1.5 (23.0%)	<0.1% (<0.1%)
Point Sources* (kg/Ha/yr)	2 (<0.1%)	0.5 (0.2%)	0.1 (0.1%)	3.0 (0.2%)	1.0 (0.1%)	33 (1.7%)

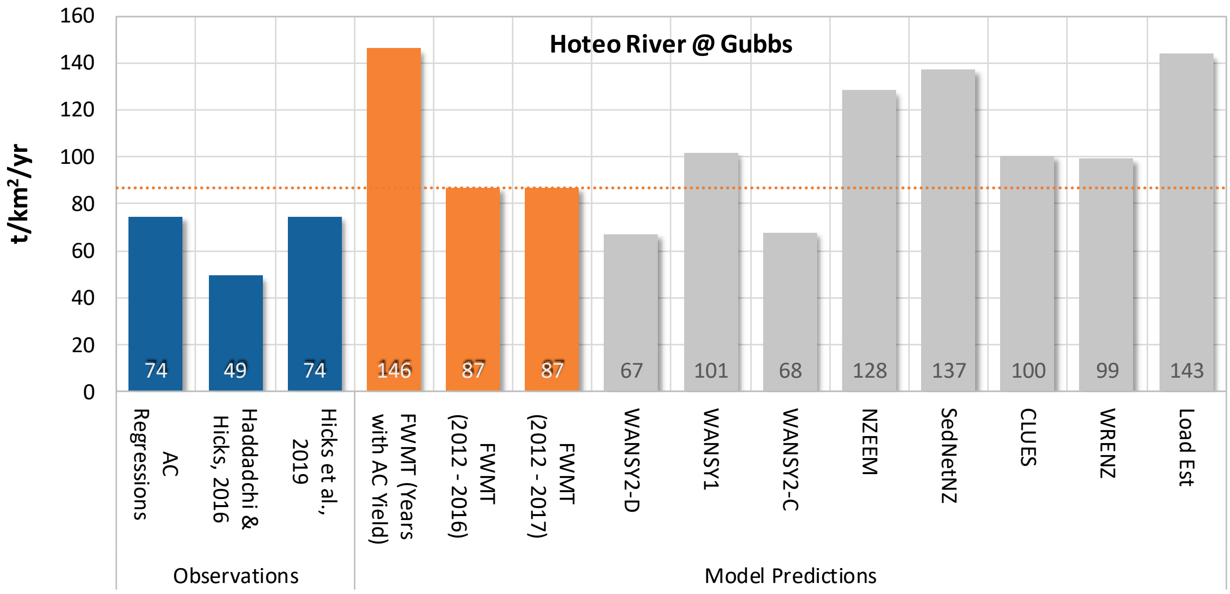

Figure 7. Sediment yield comparisons from the FWMT Stage 1 at the Hoteo Rivers @ Gubbs, over multiple years, to empirical erosional models available in Auckland (see Bambic et al., 2020 for references).

5 CONCLUSION – ADVANCES & CHALLENGES

The FWMT Stage 1 is the first whole-of-region, continuous and process-based water quality model in New Zealand. Completion of baseline (2013-2017) development marks a remarkable achievement and advance for water quality modelling in Auckland Council: for the diversity of sources (106 HRUs), overland and instream processes (build-up/wash-off, transport, suspension/deposition, nutrient speciation), and for remarkable hydrological and contaminant simulation performance.

A particularly stringent performance approach, including diverse metrics and thresholds to full and partial gradients have identified the FWMT Stage 1 is particularly well suited to simulating instream hydrology, loading and attribute grades – particularly of sites failing national or regional bottom-lines. Less so, absolute concentrations although comparison to other continuous model builds demonstrates comparable or greater accuracy across PBias, r^2 and NSE despite greater complexity and resolution of the FWMT Stage 1.

Challenges remain, not least with the lack of grading guidance for TSS and designing appropriate targeted validation monitoring to better represent regional conditions and extremes of contaminant concentration or loading. However, Auckland Council is now uniquely capable at implementing the NPS-FM through targeted and catchment-based planning with the FWMT. The FWMT offers evidence of which contaminant(s), measure(s) and causes for water quality degradation to manage.

ACKNOWLEDGEMENTS

The FWMT Stage 1 is undergoing continued development via the Water Quality Targeted Rate, funding a multi-agency approach with Paradigm Environmental Ltd., Morphum Environmental Ltd., Hydraulic Analysis Limited, Koru Environmental Limited and Perrin Ag Limited.

REFERENCES

Australia and New Zealand (ANZ) 2019. Guidelines for Fresh and Marine Water Quality. https://www.waterquality.gov.au/anz-guidelines

Blyth, J., Cetin, L., and Easton, S. 2018. Baseline SOURCE model build and calibration report: Ruamahanga Catchment. Jacobs client report for Greater Wellington Regional Council.

Easton, S., Shrestha, M., Cetin, L., Blyth, J., and Sands, M. 2019. Porirua Whaitua Baseline SOURCE Modelling Report. Jacobs client report for Greater Wellington Regional Council.

Larned, S., Snelder, T., Unwin, M., and McBride, G. 2016. Water quality in New Zealand rivers: current state and trends. *New Zealand Journal of Marine and Freshwater Research*, **50**: 389-417

Leathwick, J., West, D., Gerbeaux, P., Kelly, D., Robertson, H., Brown, D., Chadderton, W., and Ausseil, A 2010. Freshwater Ecosystems of New Zealand (FENZ) Geodatabase Version One – August 2010.

Loft, J., Mawer, J., Zhao, H., and Williamson, J. In prep. Kaituna and Rangitaiki Catchment Models. SOURCE catchment modelling analysis. WWLA client report for Bay of Plenty Regional Council.

McBride, G. 2016. National Objectives Framework Statistical considerations for design and assessment. NIWA client report for MfE publication umber: HAM16022.

McCuen, R., Knight, Z., and Gillian, C. 2006. Evaluation of the Nash-Sutcliffe efficiency index. *Journal of Hydrological Engineering*: **11**, 597–602

Ministry for Environment (MfE). 2017. National Policy Statement for Freshwater Management 2014 (amended 2017). Ministry for Environment publication number: ME 1234.

Ministry for Environment (MfE). 2019. Action for healthy waterways – A discussion document on national direction for our essential freshwater. Ministry for Environment publication number: ME 1427.

Moriasi, D., Gitau, M., Pai, N., and Daggupati, P. 2015. Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE*, **58**: 1763-1785.

Nevers, M., Byappanahalli, M., and Whitman, R. 2013. Choices in recreational water quality monitoring: New opportunities and health risk trade-offs. *Environmental Science and Technology*, **47**: 3073-3081.

Prime Minister's Chief Science Advisor (PMCSA) 2017. New Zealand's fresh water: Values, state, trends and human impacts. https://www.pmcsa.org.nz/wp-content/uploads/PMCSA-Freshwater-Report.pdf

Shen, J., A. Parker, J. Riverson, 2005. A new approach for a Windows-based watershed modelling system based on a database-supporting architecture. *Environmental Modelling and Software*, **20**:1127-1138

Shoemaker, L., Riverson, J., Alvi., K., Zhen, J.X., Paul, S., Rafi, T. 2009. SUSTAIN – A Framework for Placement of Best Management Practices in Urban Watersheds to Protect Water Quality. Office of Research and Development. U.S. EPA. Cincinnati, OH 45268.

Simon, A., Danis, N., Hammond, J. 2016. Channel and bank stability of the Hoteo River system, New Zealand: Loadings to Kaipara Harbour. Phase II Cardno client report for Auckland Council.

Simon, A., Bankhead, N., and Danis, N. 2015. Channel and bank stability of Lower Awaruku Stream discharging to the Long Bay – Okura Marine Reserve. Cardno client report for Auckland Council.

Simon, A., Hammond, J., and Schwar, H. 2017. Enhancing the Omaru Creek Corridor: Quantifying Stream Erosion, Mitigation and Conceptual Design. Cardno client report for Auckland Council.

Stephens, T., Brown, N., Bambic, D., and Clark, C. 2019. Current and future state of Auckland's watersheds: Stage 1 Freshwater Management Tool. Stormwater 2019 conference.

Storey, R. and Wadhwa, S. (2009) An Assessment of the Lengths of Permanent, Intermittent and Ephemeral Streams in the Auckland Region. Prepared by NIWA for Auckland Regional Council. Auckland Regional Council Technical Report 2009/028.

Thoe, W., Gold, M., Griesbach, A., Grimmer, M., Taggart, M., and Boehm, A. 2015. Sunny with a chance of gastroenteritis: Predicting swimmer risk at California beaches. *Environmental Science and Technology*, **49**: 423-431.