Waikato Regional Council Technical Report 2016/39

Modelling *E. coli* for the Hauraki Plains and Coromandel Peninsula

Development and calibration of catchment scale microbial models for load and concentration

www.waikatoregion.govt.nz
ISSN 2230-4355 (Print)
ISSN 2230-4363 (Online)

Waikato
REGIONAL COUNCIL
Te Kaunihera ā Rohe o Waikato

Prepared by:
Annette Semadeni-Davies, Michelle Greenwood and Sandy Elliott
(NIWA)

For:
Waikato Regional Council
Private Bag 3038
Waikato Mail Centre
HAMILTON 3240

August 2016

Document #: 9547483

Peer reviewed by:
Bevan Jenkins

Date January 2017

Approved for release by:
Mike Scarsbrook

Date December 2018

Disclaimer

Taihoro Nukurangi

Modelling *E. coli* for the Hauraki Plains and Coromandel Peninsula

Development and calibration of catchment-scale microbial models for load and concentration

Prepared for Waikato Regional Council

August 2016

Prepared by:
Annette Semadeni-Davies
Michelle Greenwood
Sandy Elliott

For any information regarding this report please contact:

Annette Semadeni-Davies
Urban Aquatic Scientist
Urban Aquatic Environments
+64-9-375 4532
annette.davies@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
Private Bag 99940
Viaduct Harbour
Auckland 1010

Phone +64 9 375 2050

NIWA CLIENT REPORT No: 2016060AK
Report date: August 2016
NIWA Project: EVW16102

Quality Assurance Statement	
Reviewed by:	J.P. Moores Jonathan Moores Group Manager, Urban Aquatic Environments
Approved for release by:	K.B. Ken Becker Regional Manager, NIWA Auckland

Contents

Tables

Figures

Executive summary

This report presents the development and calibration of models to estimate mean *E. coli* loads and median and 95th percentile concentrations in the Hauraki Plains and Coromandel Peninsula. The models are required by the Waikato Regional Council to aid catchment planning. The models follow from similar work undertaken for the Waikato and Waipa catchments under the Healthy Rivers / Wai Ora programme. The models were calibrated using water quality data from 24 monitoring sites located in the study area. In addition, this report also describes model input and calibration data; the methods used to determine median annual and 95th percentile concentrations and mean annual loads from monthly water quality sampling data, and sources of model uncertainty.

The *E. coli* loads model operates at the river reach scale and simulates first the load reaching the river network based on land use and catchment characteristics and then routes the load down the drainage network. The model was calibrated against mean annual *E. coli* loads estimated from measurements. It estimates mean annual *E. coli* loads reaching the stream network from each river reach sub-catchment on the basis of land use, rainfall and soil drainage. Point sources are also added to the in-stream load. The loads are then routed downstream and are subject to attenuation. The load model was calibrated against loads determined using the full flow record and with the 95th percentile flow record that removes the highest 5% of peak flows from the full flow record. The 95th percentile flow record represents the majority of flow conditions when people are most likely to have recreational contact with freshwater bodies and was used to remove potential bias in the model results towards high mean annual loads. The model performance for the model calibrated against loads calculated with the 95th percentile flow record is comparable to the model fit for the Waikato and Waipa catchments. The root mean square error (RMSE) between the log-transformed measured and modelled loads is 0.77 and 0.53 for the loads calculated with the full and 95th percentile flow records respectively. The model fit indicated by the Nash-Sutcliffe and adjusted R^2 values were 0.86 and 0.73 for the loads predicted using the full flow record and 0.90 and 0.85 for the loads predicted using the 95th percentile flow record.

The concentration regression model is better able to estimate the 95th percentile annual *E. coli* concentrations (RMSE 0.50, adjusted R^2, 0.55) than the median annual concentrations (RMSE 0.67, adjusted R^2, 0.31). Both regression models underestimate higher *E. coli* concentrations and overestimate lower concentrations.

The standard errors calculated for the calibrated load model parameters and the concentration model coefficients show that there is substantial uncertainty in the parameters calibrated for each of the models. Sources of this uncertainty in the models include:

- ***E. coli* calibration data:** *E. coli* concentration data from 22 sampling sites were used to estimate mean annual loads for calibration. These data are subject to error in sampling and analysis and are highly variable between sampling times and between sites. It is assumed that the SOE data are representative of the full range of *E. coli* concentrations.

  The measured loads were determined using concurrent flow data where flow data were available. For the other sites, TopNet modelled flows were used.

- **Point sources:** The point sources within the model are variable over time making it difficult to assess mean annual input loads. *E. coli* loads from consented dairy farm ponds were estimated using assumptions around the number of cattle serviced by the

ponds. These loads are likely to be conservative. There may be other point sources, such as urban sources, which have not been accounted for.

- **Diffuse sources:** The load from diffuse sources is calculated from land use and calibrated source yields. Land use is represented by a limited number of land cover types with diffuse loads from these sources represented in the load models by two calibrated source yields, i.e., pastoral and non-pastoral land uses. The derivation and interpretation of the underlying land use data are subject to imprecision (e.g. sampling precision errors and ground-truthing errors).

- **Spatial resolution:** The models are subject to spatial smoothing of heterogeneous spatial input data (i.e., scaling effects) due to the way in which these data are represented in the models as lumped or averaged values.

- **Temporal resolution:** Similarly, the load models are steady-state models which predict mean annual loads. This means that seasonal changes in *E. coli* generation and transport are not captured by the models. Dynamic modelling may be possible but would increase the input data needs and model complexity.

It is recommended that current water and flow monitoring be continued or expanded to provide further data for water quality modelling. Microbial tracking would help determine other sources of *E. coli* currently not included in the model. Point sources should be regularly re-evaluated to take into account changes in land use and contaminant management. Additionally, the feasibility of dynamic modelling should be investigated to capture seasonal changes in *E. coli* concentrations.

1 Introduction

Waikato Regional Council (WRC) has commissioned NIWA to develop spreadsheet models to estimate mean annual *E. coli* loads and median and 95th percentile annual *E. coli* concentrations for streams and rivers located in the eastern part of the Waikato Region covering the Hauraki Plains and Coromandel Peninsula. The models are required to help the WRC to set *E. coli* load limits in accordance with the National Policy Statement for Freshwater Management (NPS-FM, Ministry for the Environment, 2014).

E. coli is used as an indicator of freshwater faecal contamination as part of risk assessments of pathogen infection and is one of the attributes of the human health compulsory water quality objectives in the National Objectives Framework (NOF) under the NPS-FM. It is assumed that if *E. coli* are present in fresh water bodies, then other more pathogenic faecal micro-organisms are also likely to be present. The key source of faecal contamination in rural waterbodies is grazing livestock, although water fowl and other wild or feral animals can be additional sources. *E. coli* from stock enters the stream network via direct deposition of faecal matter into the stream or via indirect pathways including discharges of dairy effluent into streams, surface wash-off in areas of steep terrain, overland flow from excess irrigation water and drainage via artificial drains (Collins et al., 2007; Muirhead, 2015).

1.1 Scope

The modelling approaches applied in this study for both the load and concentration models are similar to those developed by NIWA for the Waikato and Waipa River catchment as part of the Healthy Rivers – Wai Ora (HRWO) programme (Semadeni-Davies et al., 2015).

The loads model operates at the sub-catchment scale with the smallest spatial unit being the contributing area of river reaches derived from the River Environments Classification version 2 (REC2, Snelder et al., 2010), these are called REC2 units throughout this report. The modelled loads are estimated as a function of land use and annual rainfall and soil drainage. Estimates of annual loads from point sources, for example sewage works and dairy factories, and from dairy shed effluent ponds are included in the model. The model is calibrated against load estimates derived from available measured data (i.e., loads calculated from recorded *E. coli* concentrations and concurrent flow data) using the same method as was used for the HRWO detailed model. Where paired measured flow data were not available, flow estimated using the TopNet model (Clark et al., 2008) was used to calculate load. The load model is calibrated against loads determined from the full flow record and the 95th percentile flow record. The 95th percentile flow record removes the top 5% of peak flows from the flow record, as the inclusion of these peak flows has been found to bias model calibration. The 95th percentile flow record thus represents the majority of flow conditions when people are most likely to have recreational contact with freshwater bodies.

The concentration model is a regression type model and has been developed to predict annual median and 95th percentile concentrations for locations where *E. coli* is currently not monitored. The model is developed using *E. coli* concentrations from 24 water quality monitoring stations in the study area and have the same range of predictors as the HRWO catchment concentration model (i.e., rainfall, soil drainage class and land use).

In addition to developing and running the models, this report discusses sources of model uncertainty and gives recommendations for further work.

2 Input data

This section describes input data required to develop, run and calibrate the loads and concentration models.

2.1 Drainage network

The modelled area covers 5762 square kilometres and includes all streams and rivers in the Waikato Region that flow into the Hauraki Gulf as well as the rivers and streams located on the eastern coast of the Coromandel Peninsula. This area includes the Piako, Waihou, Kauaeranga and Tairua Rivers and well as a number of minor rivers and coastal streams that have been aggregated into larger catchment areas for this study. The Piako and Waihou Rivers have been split into three smaller catchments each for reporting. The catchments are listed in Table 2-1 and are mapped in Figure 2-1 along with major river channels (i.e., Strahler stream order of four or more).

The drainage network in the modelled area consists of approximately 12,800 REC2 river reaches. A river reach is defined as a section of river between upstream and downstream confluences and is typically between 500-1500 m in length with a contributing catchment area, called an REC2 unit in this report, of around 40 ha.

There are a number of small lakes (area < 10 ha) in the study area; unlike the HRWO modelling, there are no large lakes or hydro-dams that need to be considered.

Table 2-1: Catchments in the modelled area. Map ID refers to Figure 2-1

Catchment	Map ID	Type	Area (km^2)
West Hauraki	1	Amalgamated	305.4
Piako River (Lower)	2	River	371.4
Piako River (Upper)	3	Tributary	568.7
Waitoa River (Piako)	4	Tributary	541.9
Waihou River (Lower)	5	River	428.4
Waihou River (Upper)	6	Tributary	1207.1
Ohinemuri River (Waihou)	7	Tributary	347.4
Kauaeranga River	8	River	129.1
East Hauraki / West Coromandel	9	Amalgamated	477.6
East Coromandel	10	Amalgamated	878.3
Tairua River	11	River	223.3
South Coromandel	12	Amalgamated	283

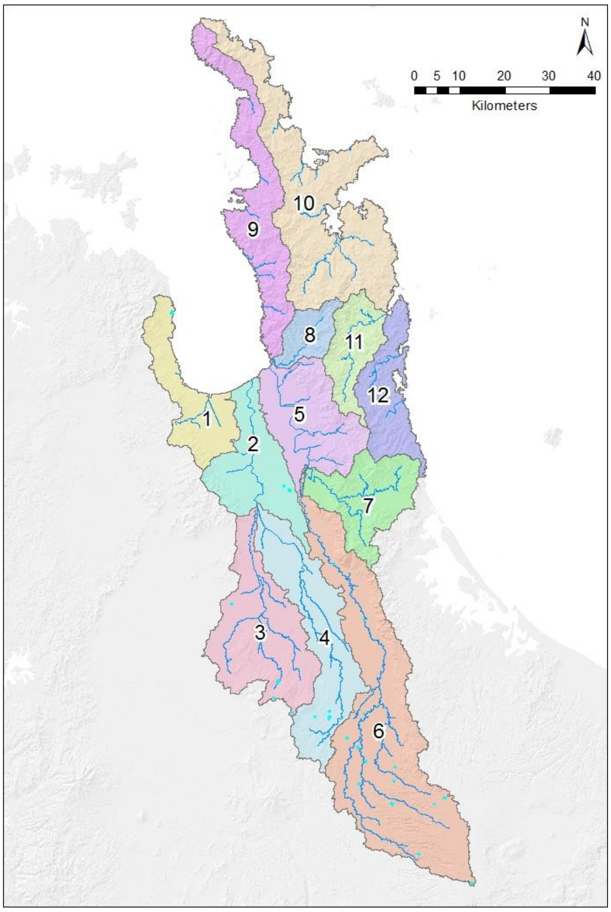


Figure 2-1: Study area showing catchments and main river channels. Sub-catchment names are given in Table 2-1.

2.2 Catchment characteristics

Soil drainage class and mean annual rainfall are input parameters for both models. The soil drainage class was derived from the Land Resources Inventory Fundamental Soils Layer (FSL, Wilde et al., 2004; Newsome et al., 2008) that assigns each soil type a score from 1 (very poorly drained) to 5 (well drained). Drainage class as used in the modelling is the areal weighted mean score for each REC2 unit and is taken from the Catchment Land Use for Environmental Sustainability (CLUES) model v10.3 geodatabase. REC2 reach aggregated mean annual rainfall has been also been taken from the CLUES model geospatial database and covers the period 1960 to 2006. This rainfall data was derived from the NIWA Virtual Climate Station Network (VCSN) which spatially distributes observed rainfall recorded at meteorological monitoring stations located around the country across a 5 km grid covering the entire country (Cichota et al., 2008). Drainage class and rainfall are shown by REC2 unit in Figure 2-2.

The area with the poorest drainage coincides with gley and peaty soils (Figure 2-3) found on the Hauraki Plains in the lower Piako and Waihou catchments. This area is largely artificially drained. The well drained areas are associated with allophanic, brown and pumice soils. Pumice soils, which were found to have a high influence on *E. coli* loads in the HRWO study, are restricted to the headwaters of the upper Waihou sub-catchment. The highest rainfalls are associated with the Coromandel and Kaimai Ranges, which run down the eastern and western Coromandel and upper Waihou sub-catchments, these areas are largely forested.

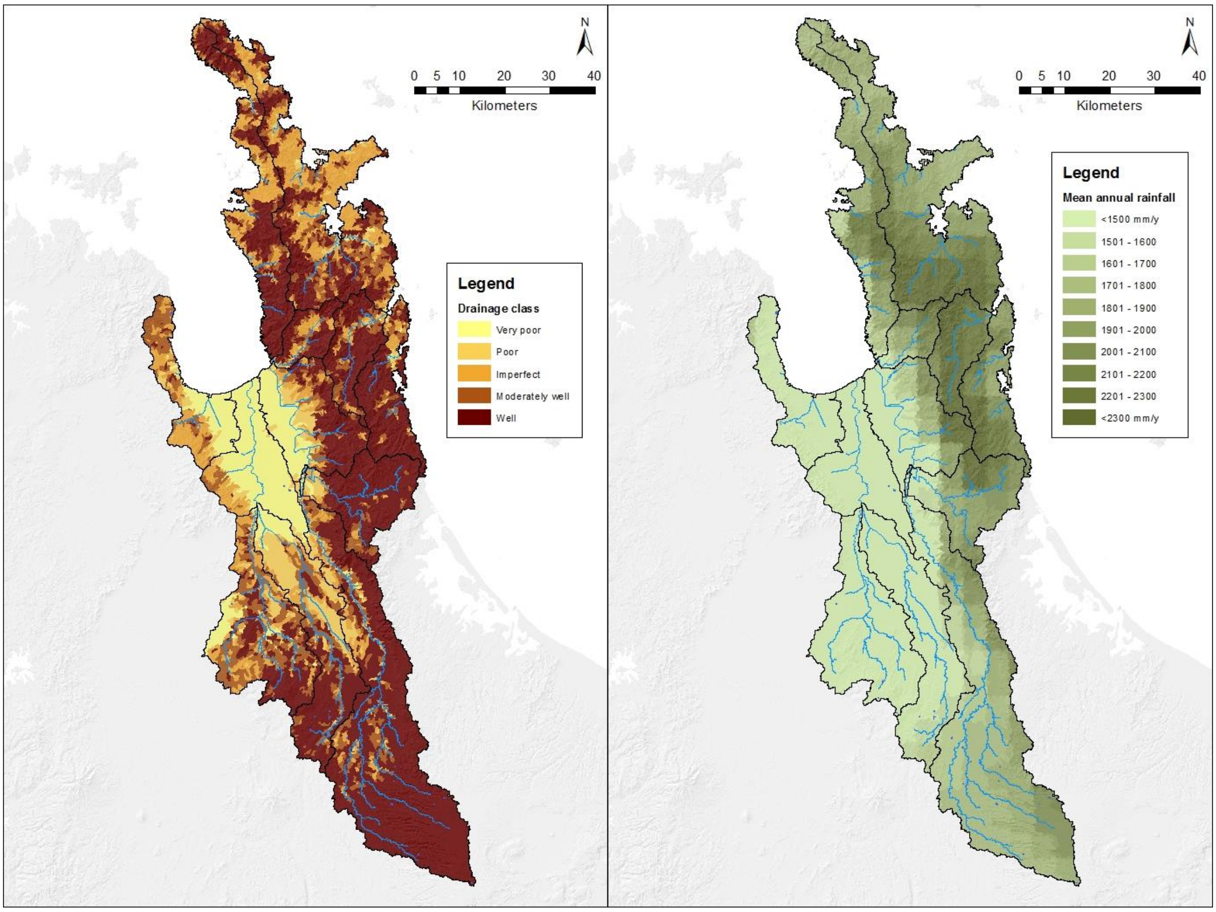


Figure 2-2: Area weighted mean average drainage class and mean annual rainfall by REC2 unit. Data sets taken from the CLUES 10.3 geodatabase.

The catchment rainfall and soil drainage data were aggregated for the upstream drainage area for each of the water quality monitoring sites listed in Section 3 both for use in the concentration model and to help interpret the load model outputs. These data are summarised in Appendix A.

2.3 Land use

Land use data was supplied for this project by WRC as a polygon shape file with the same land use classes as those used in the Catchment Land Use for Environmental Sustainability model (CLUES; Semadeni-Davies et al., 2016). These land uses were reclassified into 10 broad land use classes as listed in Table 2-2. The land use layer supplied by WRC was overlaid by the REC2 units layer to enable the proportional area of each land use within each unit to be determined.

The dominant land uses in the study area are dairy (33%), sheep and beef (18%), native forest (34%) and forestry (9%). The pastoral land uses are located largely on the Hauraki Plains. The forested areas are largely located on the Coromandel and Kaimai Ranges. Most sheep and beef is classed as intensive (lowland) farming, hill and high country sheep and beef makes up less than 1% of the study area. All other land uses only account for 6% and are, for this reason, amalgamated for display in Figure 2-4. As noted above, there are large peat lands located on the Hauraki Plains, these have been modelled as other land use, but are marked on the map for reference.

The landuse data were aggregated for the upstream drainage area for each of the water quality monitoring sites listed in Section 3, this aggregation is also summarise in Appendix A.

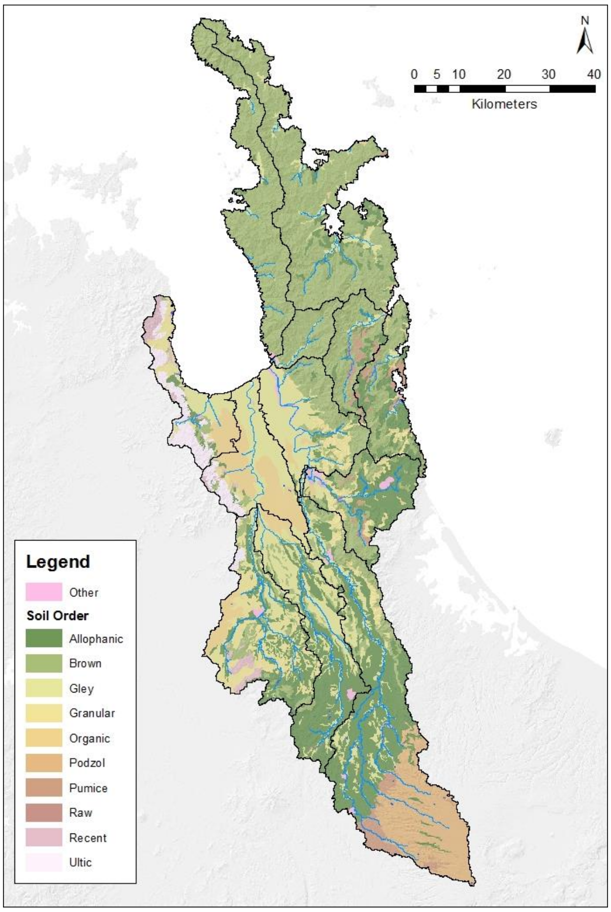


Figure 2-3: Soil order in the study area. Data derived from the Fundamental Soil Layer of the Land Resources Inventory.

Table 2-2: Modelled land use classes and corresponding CLUES classes. Data provided by WRC.

CLUES land use class	Land use class	Percentage cover	Area (km^2)
DAIRY	Dairy	33	1876
SBINTEN	Intensive (lowland) sheep and beef	17	998
SBHILL SBHIGH	Hill and high country sheep and beef	1	81
DEER and OTHER_ANIM	Other stock	1	36
PLANT_FOR	Forestry	9	517
NAT_FOR and SCRUB	Native forest and scrub	34	1959
MAIZE, POTATOES and ONIONS	Crops	1	33
APPLES, GRAPES*, KIWIFRUIT	Horticulture	0	6
URBAN	Urban	1	56
TUSSOCK*, UNGR_PASTURE, OTHER	Other	3	175

*not present in study area

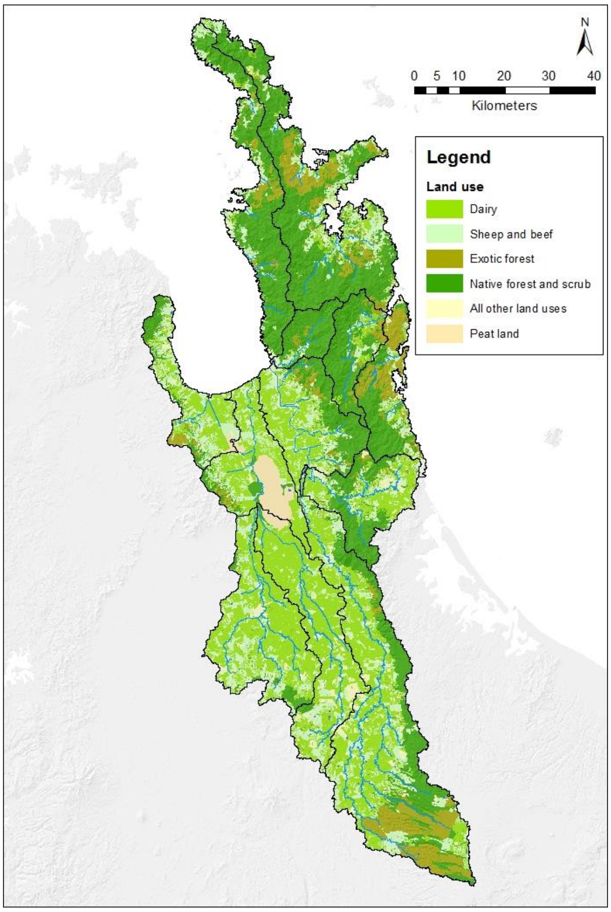


Figure 2-4: Broad land use classes for the study area. Data supplied by WRC.

2.4 Point sources and farm dairy effluent inputs

Estimated annual *E. coli* loads from point sources (i.e., waste water treatment plants, freezing works, and dairy factories) and farm dairy effluent (FDE) ponds were added as model inputs for the REC2 units within which the sources are located respectively. These loads are given in Appendix A.

Loads from point sources were estimated by the WRC (personal communication, Bill Vant) using consent monitoring data collected between 2006 and 2015. The *E. coli* input loads from FDE ponds were estimated on the basis of the consented discharge volumes for dairy sheds with two-pond treatment systems that discharge directly to the stream network. The location and discharge volume was provided by WRC. The annual load for each shed was calculated using the same method as for the HRWO modelling (Semadeni-Davies et al., 2015):

i. The number of dairy cattle serviced by each pond was estimated using the WRC rule-of-thumb of 20 cows per cubic metre consented discharge volume (personal communication, Amy Taylor, 23 February 2015).

ii. The mean average shedding of *E. coli* per cow was estimated to be 1.41×10^8 organisms per day on the basis of sampling of two two-ponds systems in the Toenepi catchment undertaken by Donnison *et al.* (2011).

Like the loads estimated for the point sources, the estimated loads from dairy sheds were aggregated by REC2 reach.

3 Calibration data

Water quality and flow data from monitoring sites located within the study area were used to calibrate both the loads and concentration models. The water quality and paired flow monitoring sites are mapped in Figure 3-1 and listed in Table 3-1. The water quality data were obtained from the National Rivers Water Quality Network (NRWQN; 2 sites) database administered by NIWA and from WRC state of environment (SOE; 22 sites) monitoring. The SOE sites have data available back to 1998 and were sampled quarterly up until 2012 from which time they have been sampled monthly. The NRWQN *E. coli* data are available monthly from 2005. Paired flow records are available for 12 of the 24 water quality monitoring sites. Note that the paired flow data for some sites comes from flow monitoring sites up or downstream of the water quality monitoring sites. For example, Wharekawa River at Adams Farm Br is 1 km upstream of the Wharekawa River at SH25 water quality monitoring site. In some cases, the flow data from neighbouring sites supplied by WRC were rejected from the study as they were considered to be unrepresentative of flow at the paired water monitoring sites due to either up- or downstream distance or location with respect to the drainage network. TopNet modelled hourly flow data was used for all but one of the other sites as discussed in Section 3.2. The method used to derive the mean annual loads is presented in Section 3.3.

3.1 *E. coli* concentrations

The measured *E. coli* concentration data are used to calculate mean annual loads that are in turn used to calibrate the loads model and to determine the annual median and 95th percentile concentrations used to develop and calibrate the concentration models.

Annual median and 95th percentile concentrations were determined using the Hazen method, recommended by the Ministry for the Environment, for the 18 years for which SOE data are available (March 1998-March 2016), the 11 years for which data is available from all the sites (March 2005-March 2016), and the most recent five year period (March 2011 to March 2016). The concentrations are given in Appendix C.

E. coli concentrations are sensitive to changes in land use, farm practices and wastewater management at sewage treatment plants and can therefore exhibit trends over time. A visual examination of the recorded *E. coli* concentrations showed that there are trends in the data for several sites (Figure 3-2), however, these trends are highly variable regionally. This finding is in keeping with Vant (2013) who investigated water quality trends between the years 1993 and 2012 in the Waikato Region. He found that most sites in the Hauraki Plains and Coromandel showed no significant trends for *E. coli*. The exceptions being Ohinemuri at Queens Head (619_19), Waitekauri River U/S Ohinemuri Confluence (1239_32), Tapu at Tapu-Coroglen Road (954_5) and Oraka at Lake Road (669_6) where there have been slight reductions in *E. coli* concentrations. Vant concluded that there was no evidence of a pattern in *E. coli* trends across the Waikato region as a whole. Like the HRWO modelling (Semadeni-Davies et al., 2015), the five-year median and 95th percentile concentrations are used to avoid the effects of both possible trending data and detrending smoothing errors.

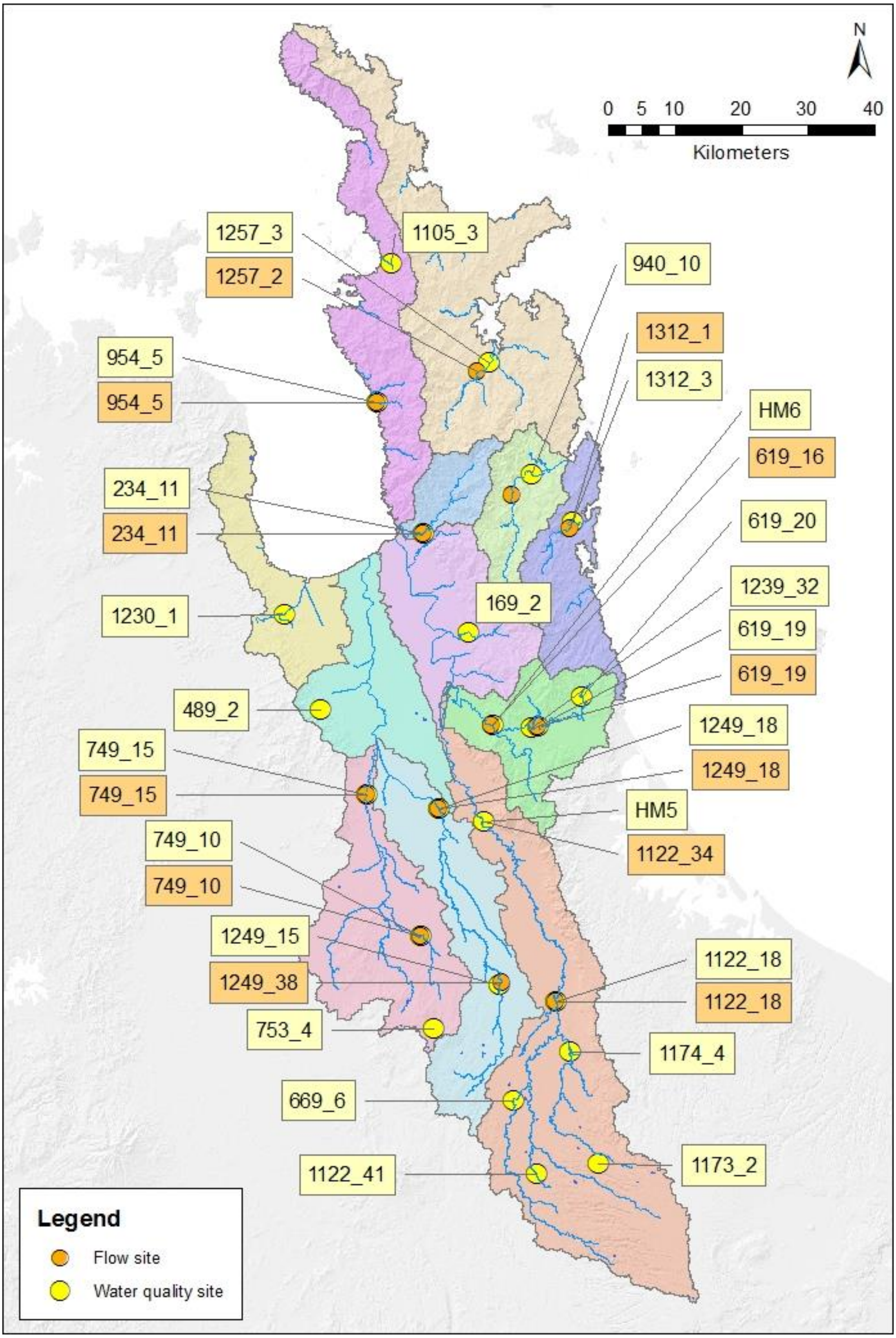


Figure 3-1: Location of water quality and paired flow monitoring sites in the study area. Yellow labels refer to water quality monitoring sites and orange to flow monitoring sites.

Table 3-1: Water quality monitoring sites and paired flow monitoring sites.

Reporting sub-catchment	Water quality site ID and name		REC2 segment	Flow monitoring site ID and name		REC2 segment
West Hauraki	1230_1	Waitakaruru River at Coxhead Rd Br	3047683	-	-	-
Piako River (Lower)	489_2	Mangawhero Stm at Mangawara Rd	3050877	-	-	-
Piako River (Upper)	749_10	Piako River at Kiwitahi	3059826	749_10	Piako River at Kiwitahi	3059826
Piako River (Upper)	749_15	Piako River at Paeroa-Tahuna Rd Br	3054261	749_15	Piako River at Paeroa-Tahuna Rd Br	3054261
Piako River (Upper)	753_4	Piakonui Stm at Piakonui Rd	3066020	-	-	-
Waitoa River (Piako)	1249_15	Waitoa River at Landsdowne Rd Br	3062720	1249_38	Waitoa River at Waharoa Control	3062720
Waitoa River (Piako)	1249_18	Waitoa River at Mellon Rd Recorder	3054693	1249_18	Waitoa River at Mellon Rd Recorder	3054693
Waihou River (Lower)	169_2	Hikutaia River at Old Maratoto Rd	3048594	-	-	-
Waihou River (Upper)	1122_18	Waihou River at Okauia	3064061	1122_18	Waihou River at Okauia	3064061
Waihou River (Upper)	1122_41	Waihou River at Whites Rd	3078605	-	-	-
Waihou River (Upper)	1173_2	Waiohotu Stm at Waiohotu Rd	3077848	-	-	-
Waihou River (Upper)	1174_4	Waiomou Stm at Matamata-Tauranga Rd	3067934	-	-	-
Waihou River (Upper)	669_6	Oraka Stm at Lake Rd				
Waihou River (Upper)	HM5*	Waihou River at Te Aroha	3055227	1122_34	Waihou River at Te Aroha	3055227
Ohinemuri River (Waihou)	1239_32	Waitekauri River U/S Ohinemuri Confluence	3051680	-	-	-
Ohinemuri River (Waihou)	619_19	Ohinemuri River at Queens Head	3051991	619_19	Ohinemuri River at Queens Head	3051991

Reporting sub-catchment	Water quality site ID and name		REC2 segment	Flow monitoring site ID and name		REC2 segment
Ohinemuri River (Waihou)	619_20	Ohinemuri River at SH25 Br	3050858	-	-	-
Ohinemuri River (Waihou)	HM6*	Ohinemuri River at Karangahake Gorge	3051925	619_16	Ohinemuri River at Karangahake	3051925
Kauaeranga River	234_11	Kauaeranga River at Smiths Cableway/Recorder	3044978	234_11	Kauaeranga River at Smiths Cableway/Recorder	3044978
East Hauraki / West Coromandel	1105_3	Waiau River at E309 Rd Ford	3037259	-	-	-
East Hauraki / West Coromandel	954_5	Tapu River at Tapu-Coroglen Rd	3040973	954_5	Tapu River at Tapu-Coroglen Rd	3040973
East Coromandel	1257_3	Waiwawa River at Sh25 Coroglen	3039645	1257_2	Waiwawa River at Rangihau Rd Ford	3039919
Tairua River	940_10	Tairua River at Morrisons Br Hikuai				
South Coromandel	1312_3	Wharekawa River at SH25	3044647	1312_1	Wharekawa River at Adams Farm Br	3044838

*NIWA NRWQN monitoring site

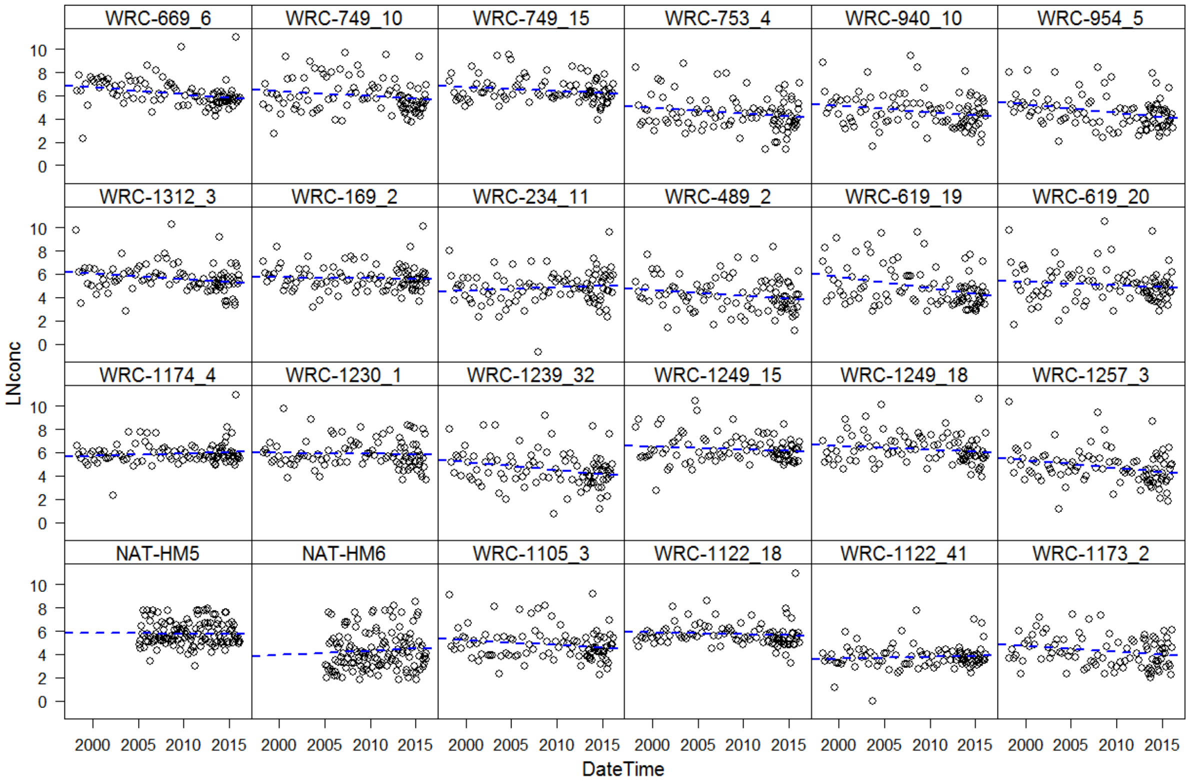


Figure 3-2: ***E. coli*** **concentrations over time for each of the water quality monitoring sites.** Concentration data are plotted on a log normal scale

3.2 Flow data

Paired flow data is available for half of the water quality sites. Flow data simulated by the TopNet model (Clark et al., 2008) is used to determine loads for all but one of the other sites, Tairua River at Morrisons Br Hikuai, which is located outside the calibrated area for TopNet which covered the Hauraki Plains and the west coast of Coromandel. TopNet was calibrated and validated as part of a project undertaken by NIWA for the Ministry of Business, Innovation and Employment to simulate sediment loads into the Hauraki Gulf (Christian Zammit, NIWA, personal communication 2016, unpublished data). Modelled flow records are available for each reach in the calibration area from 1972 onwards. Climate data, including rainfall, used as input to TopNet comes from the VCSN.

TopNet was calibrated against measured event flows from seven sites, four of which are paired to water quality monitoring sites used in this report (Waihou River at Te Aroha, HM5; Ohinemuri River at Karangahake Gorge, HM6; Piako River at Paeroa-Tahuna Rd Br, 749_15; and Tapu River at Tapu-Coroglen Rd, 954_5). The goodness-of-fit varied across the modelled area which has uncertainty implications for this study as discussed in Section 6. An example of the calibration time-series is given for the Ohinemuri River at Karangahake flow monitoring site in Figure 3-3.

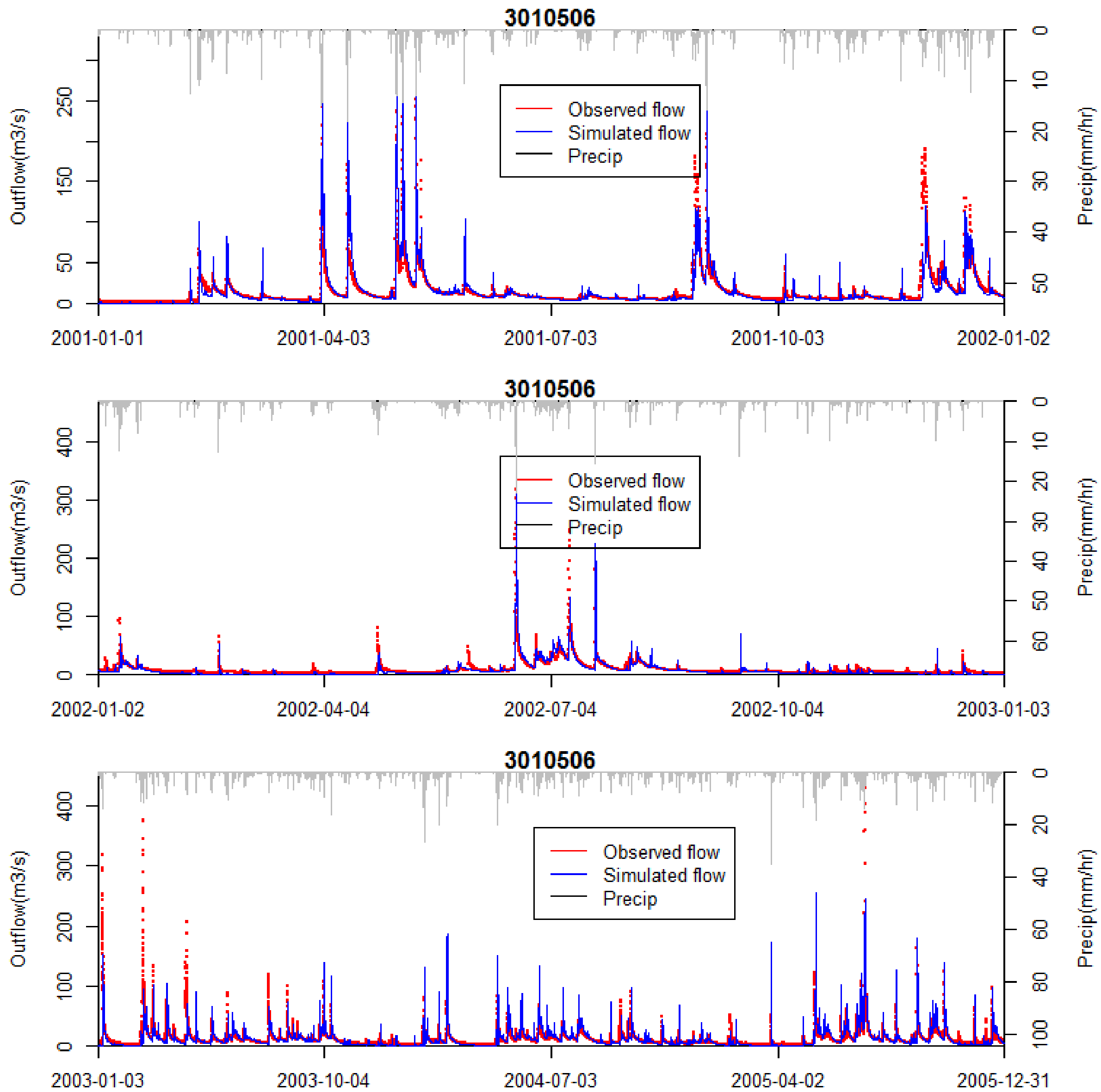


Figure 3-3: TopNet calibration results for Ohinemuri at Karangahake (HM6) flow monitoring. Source: Christian Zammit (pers. comm; unpublished data).

The measured and modelled mean annual flow rates are provided in Appendix D and are compared in Figure 3-4 for the water quality sites with paired flow records to give an indication of model fit across the study area. The modelled annual mean flow rates are within 20% of those recorded for all but two of the sites, Tapu River at Tapu-Coroglen Road (954_5) and Waitoa River at Landsdowne Rd Br (1249_15). The Tapu River site was one of the TopNet calibration sites and while peak flows are captured by the model, the model calibration and validation both had baseflows consistently less than those recorded. The TopNet baseflow was also underestimated for the Ohinemuri River at Karangahake Gorge and Waihou River at Te Aroha sites, but not to as great a degree. The overall overestimation of the annual flow rate at these sites is likely due to a consistent overestimation of peak flow rates. The model results for most of the other sites, with the exception of Waitoa at Landsdowne Rd Br which had flows underestimated by a factor of four, also suggest that peak flows are being overestimated. The reason for underestimation of flows at Waitoa at Landsdowne is not clear, TopNet was calibrated for Piako River at Paeroa-Tahuna Rd Br some 30 km downstream, and showed a fairly good fit for that site for both peak and low flow conditions. In any case, since measured flows are available for this site, the modelled flows were not used.

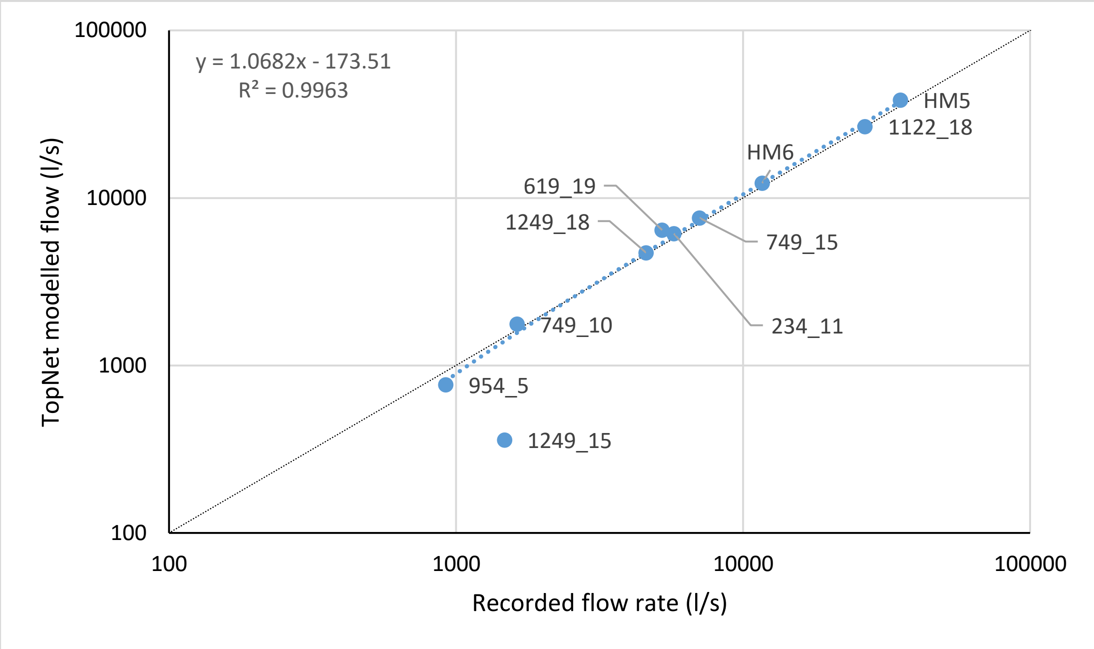


Figure 3-4: Comparison plot of mean annual flows for the paired water quality monitoring sites calculated from recorded and modelled flows.

In order to minimise uncertainty due to mismatches between the timing of hourly flows in the recorded and modelled flow records, loads were determined from daily flows using the rating curve method outlined below.

3.3 Mean annual loads

Mean annual *E. coli* loads were estimated using the same rating-curve method that was used for the HRWO modelling. Unlike that study, the ratio method (i.e. estimating mean loads from the catchment average mean load:median concentration ratio) was not used both because of the availability of TopNet modelled flows and as there was high variability in the ratios determined for the paired sites that precluded the use of an average ratio across the study area.

The method fits a rating curve to the natural log of measured monthly *E. coli* concentrations against the natural log of the flow rate using the following equation:

$$\ln(C) = s(t) + s(\ln(Q)) + a\sin(2\pi t) + b\cos(2\pi t) \tag{1}$$

Where C is the five-year median *E. coli* concentration, s is a cubic spline smoothing function, Q is the daily average flow rate for the date the sample was collected, t is time (in years), and a and b are coefficients. Cubic spline smoothing from the R statistical package was used, with a fixed effective degrees of freedom of two to restrict curvature. The fitted curves for each site are shown in Appendix E (Figure E-1. for sites with measured flow and Figure E-2 for sites with modelled flow).

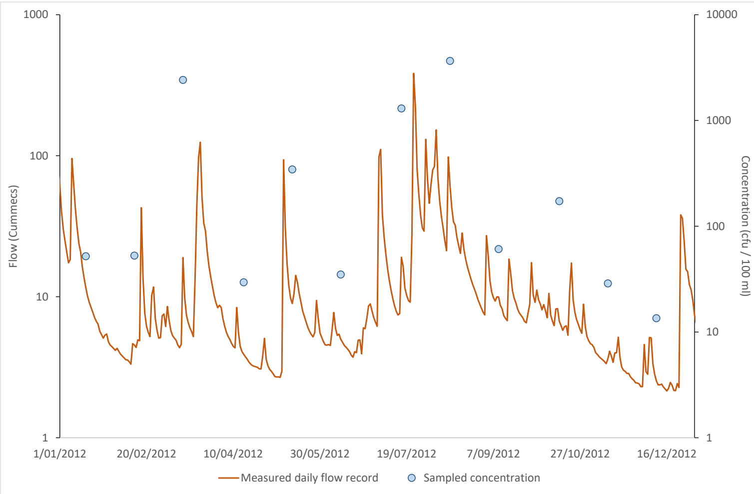
Equation (1) was applied to the daily flow time-series over the period of the flow record to derive a time-series of concentrations, which was then multiplied by volume (from flow x time over the flow monitoring time step) and summed to give the mean annual load. To account for retransformation bias the load was adjusted using the non-parametric smearing factor of Duan (1983).

The rating curve method is illustrated in Figure 3-5 for one site (HM6 - Ohinemuri River at Karangahake Gorge) over the course of one year. Figure 3-5 a. shows the measured daily flow record and the sampled concentration for 2012. These data, along with sampling and flow from the other load-estimation years, were used to develop the rating curve for the site shown in Figure 3-5 b. The rating curve relationship in Equation (1) was then used to calculate a load time-series which is shown in Figure 3-5 c, along with discrete loads calculated from the sampled concentrations. The series of figures gives a sense of the sampling frequency during the modelling period as well as the level of model uncertainty.

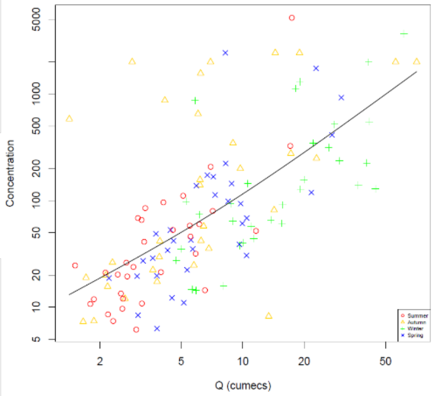
The suitability of the rating curve derived loads for model calibration were assessed by generating confidence intervals (90%) and standard deviations for the mean annual loads by repeating the rating curve procedure using a boot-strapping approach (see summary in Appendix E; Table E-1). This approach repeatedly took random samples of the original water quality data and estimated the mean annual load for each of these. Mean annual loads were calculated using both the full flow record and the 95^{th} percentile flow record (i.e., the top 5% of flow rates removed) and are given in along with the Root Mean Square Error (RMSE) and standard deviations of the log-transformed loads. A standard deviation of > 1 signals that the mean load calculated is likely to be unrepresentative of actual load for the site.

Only one site, Waihou River at Okauia, was rejected on the basis of the suitability assessment. The plot for this site (Appendix E; Figure E-1 - page 44) shows too few water samples were made during high flow periods (i.e., above 30 cumecs) causing a skew in the fitted rating curve so that the calculated mean annual load is likely to be too high. Note that no loads could be calculated for the Tairua River at Morrisons Br Hikuai (940_10) site calculated as there were neither measured nor modelled flow available.

a. Measured daily flows and sampled concentrations



b. Rating curve



c. Load calculated from measured concentration and estimated from daily flow data

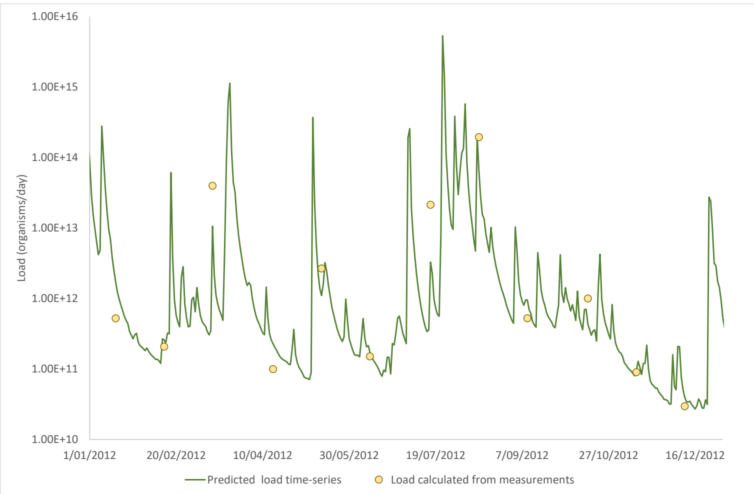


Figure 3-5: Rating curve method of determining daily load times series used to estimate annual mean and median loads, Ohinemuri River at Karangahake Gorge. The time-series plots show data for 2012, the rating curve has been derived for the 5-year estimation period.

4 Load modelling

This section describes the annual *E. coli* load model and its calibration and presents model results for the calibration.

4.1 Model description

The *E. coli* load model calculates sub-catchment loads discharged from each REC2 unit and routes these loads down the drainage network (described in Section 2.1) by adding loads from each reach entering the in-stream load and then accounting for in-stream attenuation or decay. Cumulative or in-stream yields were also calculated for each monitoring site (i.e., as the load for each monitoring site divided by the total catchment area upstream of the site) to normalise the loads for area. That is, all things being equal, a larger catchment will give a greater load than a smaller catchment with the same land use and catchment characteristics.

Loads entering the drainage network from each reach unit are calculated as the sum of loads from point sources (see Section 2.4) draining to the reach and *E. coli* losses from diffuse sources. The load from diffuse sources is calculated as the sum of the area of each diffuse source (i.e., land use type, see Section 2.3) multiplied by a corresponding source yield and then adjusted for surface losses. Five source yields were initially calibrated for the calculation of diffuse *E. coli* loads: dairy; sheep and beef intensive; sheep and beef hill and high country; all other rural land uses (including native and exotic forest); and urban. The loads are expressed as the number (peta) of organisms per year entering streams and are adjusted for rain and drainage according to:

$$L_a = L_{\text{int}} e^{k_{rain} R_{reach.} + k_{drain} D} \tag{2}$$

where, for each REC2 reach unit, L_a is the adjusted load from the diffuse sources entering the drainage network; L_{int} is the initial load summed for the diffuse sources; R_{reach} is the mean annual rainfall anomaly (m); k_{rain} and k_{drain} are calibrated catchment wide rainfall delivery and drainage coefficients, and D is set to 0 for subcatchments with good to well drained soils (average drainage class greater than 3.5) and 1 for poorly to imperfectly drained soil (average drainage class ≤3.5). Note that the drainage class does not take account of artificial drainage, but represents the natural soil profile from the LRI.

The rainfall anomaly, R_{reach} is calculated as:

$$R_{reach} = R_{mean} - R_{catchmen} \tag{3}$$

Where $R_{mean}$is the mean annual rainfall for the sub-catchment and $R_{catchment}$ is the mean annual rainfall in the study area. Here, the mean annual rainfall for each reach is taken from the CLUES model version 10.3 geospatial database.

Once in the drainage network, the *E. coli* load is propagated downstream taking into account losses within the network by multiplying in-stream load by decay factors which relate to the proportion of the load remaining after attenuation. Separate losses are calculated for streams and reservoirs. In-stream losses for each reach are modelled by a first-order decay term calculated as:

$$Att_{stream} = e^{-k_{time} T} \tag{4}$$

Where Att_{stream} is the in-stream attenuation factor for the reach; k_{time} is a calibrated time coefficient; and T is the time of travel in days for the reach and is a function of the reach length divided by the flow rate and is calculated as:

$$T = \frac{z\frac{1}{86400}}{0.36 * F^{0.241}} \tag{5}$$

Where F is the mean annual flow rate (m^3/s) taken from the CLUES model, z is the reach length (m) and the coefficients 0.36 and 0.241 were determined for New Zealand Rivers by Jowett (1998).

Losses for reservoirs are calculated for the outlet reach of each reservoir as

$$Att_{res} = \frac{O_{res}}{(O_{res} + k_{res})} \tag{6}$$

Where $Att_{res}$is the attenuation factor for the outlet reach of the lake, as identified in the REC; O_{res} is the reservoir overflow depth (m/year, calculated as the overflow volume divided by the lake area) for the outlet reach taken from the SPARROW component of the CLUES model (Elliott et al., 2005); and k_{res} is the reservoir coefficient. Since there are few lakes in the study area, and they tend to be rather small (<10 ha, median 1.5 ha), the reservoir coefficient was assigned the same value (15.2 m/y) as calibrated for the HRWO model calibration for small lakes.

4.2 Calibration and performance assessment methods

The model parameters defined above were calibrated by fitting modelled annual loads against estimated annual loads for each of the water quality monitoring sites. The calibration was performed automatically within Excel using the Solver function (GRC Nonlinear solving method) to minimise the Root Mean Square Error (RMSE) calculated for the residuals between the modelled and measured *E. coli* log-transformed loads. The RMSE is used as a standard statistical metric to measure model performance in many fields, including meteorology, air quality, climate research and agriculture (Chai and Draxler, 2014). RMSE represents the sample standard deviation of the residuals or difference between the predicted and observed values and has the same units as the parameter. The RMSE is calculated as:

$$RMSE = \sqrt{\frac{\sum(x-y)^2}{n}} \tag{7}$$

Where *x* and *y* are the observed and predicted values and *n* is the number of samples.

The annual loads used for calibration were calculated according to the methodology described in Section 3.3. The load model was calibrated separately against the measured mean loads calculated using the full flow record and the 95th percentile flow record. As was noted above, Waihou River at Okauia (1122_18) was excluded from the calibration on the basis of the assessment of its rating curve. The high variability for the site is due to a couple of concentrations sampled during high flow events that have skewed the rating curve leading to a bias towards high loads in the bootstrapping approach.

Parameter uncertainty was determined using a local linear approximation (Nikitas and Pappa-Louisi, 2000) to determine the standard error (SE) of each parameter. The SE indicates whether a parameter is highly uncertain and could potentially be removed from the model.

The calibration process was iterative and involved successively removing or combining different land use parameters in order to use as few parameters as possible without compromising model fit as indicated by the RMSE (i.e., model parsimony). Through this process, the model was optimised such that the initial five land use classes were regrouped into two aggregated land use classes (see explanation below); that is pastoral and non-pastoral land uses.

In addition to the RMSE, model performance is assessed using the Nash-Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970) and the adjusted coefficient of determination (adjusted R^2) between the modelled loads and those estimated from monitored data. The NSE is a measure of the scatter of model residuals around the 1:1 line. The value ranges from $-\infty$ to one. A NSE of one indicates that the modelled and measured values are the same whereas a NSE of zero indicates that the modelled values are only as accurate as the mean of the measured values. A value > 0.5 indicates that the model performance is satisfactory (Moriasi et al., 2007). A negative NSE means that the mean of the measured values is a better predictor than the model.

The NSE is calculated as:

$$NSE = 1 - \frac{\sum (y - f)^2}{\sum (y - \bar{y})^2} \tag{8}$$

Where y is the observed value, $\bar{y}$ is the mean of the observed values and f is the paired predicted value.

The adjusted R^2 is a modified version of the coefficient of determination that takes into account the number of predictors (parameters) in the model. The value increases if the new term improves the model fit more than would be expected by chance and decreases if the model improvement is less than would be expected by chance. The adjusted R^2 is calculated as:

$$R_{sq.adjust} = R_{sq} - (1 - R_{sq}) \frac{P}{n - P} \tag{9}$$

Where P is the number of parameters, n is the number of observations modelled and R_{sq} is the coefficient of determination calculated as:

$$R_{sq} = 1 - \frac{\sigma_{(y-f)}}{\sigma_y} \tag{10}$$

Where $\sigma_{(y-f)}$ and σ_y are the variances, calculated using the Excel var.s function, of the residuals and the observations respectively.

4.3 Load model results

4.3.1 Calibrated parameters

The calibrated parameters and their standard errors are listed in Table 4-1 for both the calibration against the full flow record and the 95th percentile flow record. The model fit for each calibration is summarised in Table 4-2 and discussed further in Section 4.3.2. The model fit is comparable to the fit obtained for the HRWO *E. coli* model.

Table 4-1: Model parameters and standard error (SE) calibrated using mean annual loads derived from the full flow record and the 95th percentile flow record

Parameters	Unit	Full flow record		95th percentile flow record	
		Parameters	SE	Parameters	SE
Source yield: Pastoral (dairy, all sheep and beef, other stock)	10^{15} org/km^2/year	0.0066	0.0037	0.0052	0.0024
Source yield: Non-pastoral (all other land uses)	10^{15} org/km^2/year	0.0002	0.0007	0.0006	0.0008
Drainage coefficient k_{drain}	dimensionless	17.54	10.94	5.94	2.99
Rainfall delivery coefficient k_{rain}	dimensionless	3.82	0.78	1.90	0.61
Reach time of travel k_{time}	/day	0.12	0.80	0.43	0.60

As noted above, it was found that grouping pastoral and non-pastoral land into two groups gave comparable model fit with lower SE values for both calibrations as having separate source yields for each land use class. This is because the pastoral land uses are dominated by dairy and intensive sheep and beef farming (other stock classes account for around 1% of the total study area), which had calibrated yield parameters in the same order of magnitude, whereas all the non-pastoral sources had calibrated *E. coli* yields that were an order of magnitude less than the pastoral yields. The SE values for both parameter sets show that there is substantial uncertainty in the model; that is, they are high relative to the parameter values.

The calibrated pastoral yields are comparable in order of magnitude to the yields estimated for different stock types by Wilcock (2006). He found a wide range of yields depending on stocking rates, farm grazing practice, and access to waterways, which underlines the difficulties in calibrating *E. coli* models.

The drainage parameter in each of the calibrations show that the model is sensitive to drainage class. The drainage parameter is largely being driven by several water quality monitoring sites located in areas with poorly to imperfectly drained gley and organic soils that have low to medium upstream pastoral land use fractions but higher than expected measured *E. coli* yields, such as Waiau River at E309 Rd Ford (1105_3) and Waitakaruru River at Coxhead Rd Br (1230_1). Similar high loads from forested catchments in poorly-drained areas were documented for the HRWO study. The higher yield for the poorly-drained class is reasonable given the greater surface runoff and likelihood of artificial drainage, both of which would be expected to increase losses of *E. coli*.

The stream decay coefficient is within the range of values expected for microbial decay, in the order of 1 day^{-1} (Hipsey et al., 2008).

4.3.2 Model performance

Modelled loads and yields are given in Appendix F along with loads estimated from monitoring data. Table 4-2 shows that the model fit for loads, is better for the calibration against the loads derived from the 95th percentile flow record than for the loads derived from the full flow record (which had a larger RMSE). The difference in fit is possibly due to skew in the rating curves resulting in high loads similar to that discussed above for Waihou River at Okauia (Section 4.2), albeit to a less noticeable extent.

Plots of the model fit for the calibration against loads calculated using the 95th percentile flow record are given in Figure 4-1 for loads and in Figure 4-2 for yields. These plots are discussed further below. None of the plots shows any difference in the goodness-of-fit for loads determined using recorded or TopNet modelled flows. Outliers are labelled on the plots.

Table 4-2: Goodness-of-fit for the log-transformed loads calibrated against the log-transfored mean annual loads derived from the full flow record and the 95th percentile flow record

Fit	Full flow record	95th percentile flow record
RMSE for ln(Loads) (calibration target)	0.77	0.53
Nash-Sutcliffe efficiency	0.86	0.90
Adjusted R^2	0.75	0.85

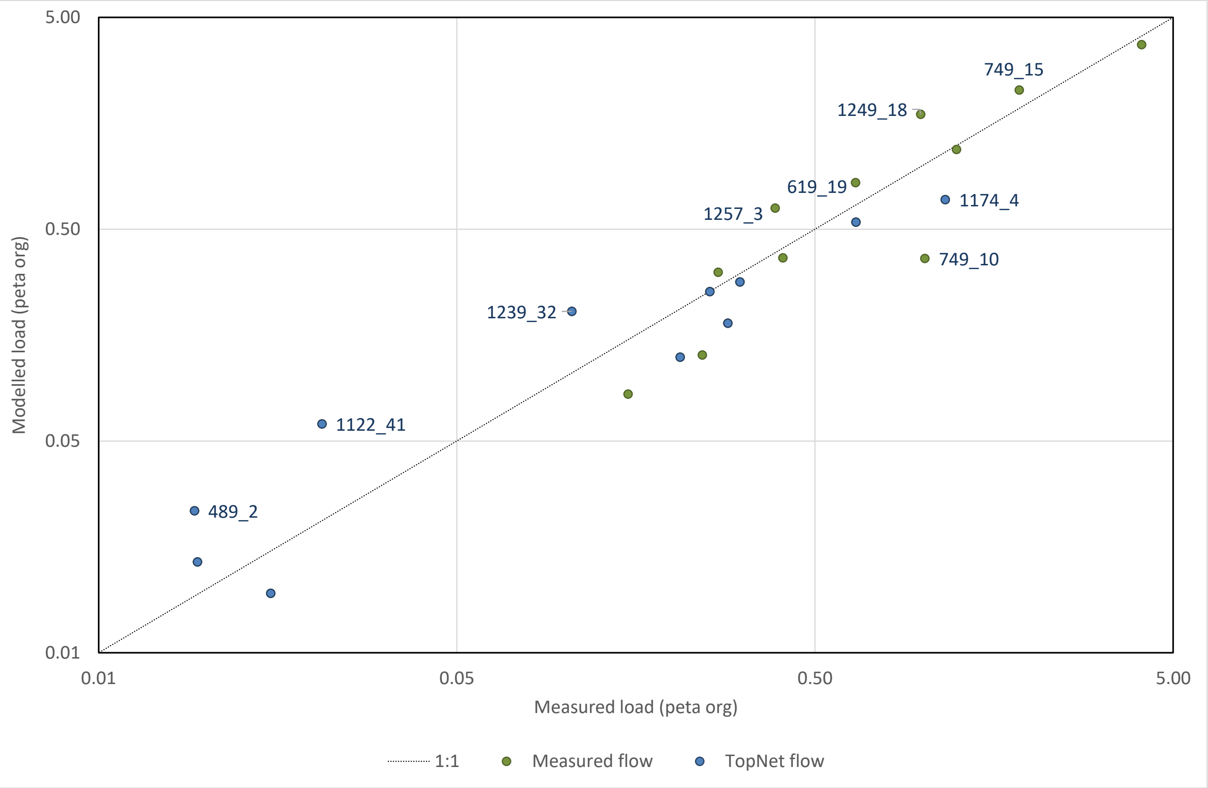


Figure 4-1: Comparison of modelled and measured load plotted on log scale Labels show load outliers discussed in the text.

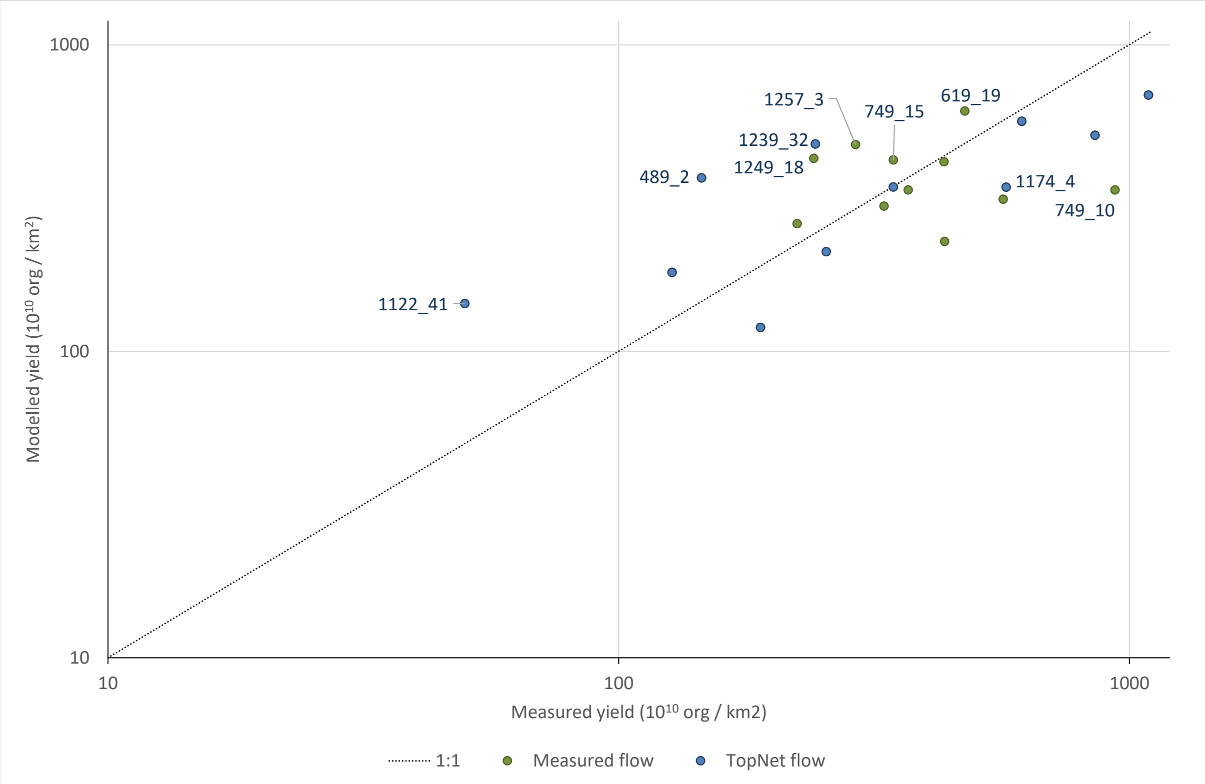


Figure 4-2: Yields determined for the measured and modelled loads plotted on log-scale. Labels refer to the outliers discussed in the text.

Five sites have loads that are overestimated by more than 50%, these are: Waitoa River at Mellon Rd (1249_18); Waiwawa River at Sh25 Coroglen (1257_3); Mangawhero Stm at Mangawara Rd (489_2) and Waitekauri River U/S Ohinemuri Confluence (1239_32); and Waihou River at Whites Rd (1122_41). The latter three sites have small upstream drainage areas and the absolute differences between the modelled and measured loads is small. The Whites Road site is spring-fed (this is the Blue Spring that is bottled as “Pump” water by Coca Cola) and the spring water is filtered through pumice sand and has a mean residence time of 50 years, which explains the over prediction by the model at this site. Of the five sites, three have a large fraction of their upstream area with imperfect soil drainage (between 20-50%; 489_2, 1249_18 and 1257_3) and, with the exception of the Waitoa site, which has 86% pastoral land use upstream, the sites with overestimated loads have largely forested upstream catchment areas. This finding supports the notion above that the drainage parameter is being driven by unexpectedly high *E. coli* yields in some catchments with poor to imperfect drainage which is causing overestimated loads at other sites with upstream poor to imperfect drainage.

In addition to these sites, the Piako River at Paeroa-Tahuna Rd Br (749_15) and Ohinemuri River at Queens Head (619_19) sites appear to be outliers in the load plot, the modelled loads for these sites are overestimated by 22% and 28% respectively. The catchment area upstream of the Piako site is dominated by pasture, however 37% of the upstream area has imperfect drainage which is likely to be behind the overestimation. The site is about 16 km downstream of two point sources (i.e., the sewage works and dairy factory at Morrinsville), but the *E. coli* discharges from these sources are negligible (<1%) compared to the total measured load for the site. The area upstream of the Ohinemuri at Queens Head site is similarly dominated by pasture but is well drained.

The only site that has a modelled load underestimated by more than 50% is Piako River at Kiwitahi (749_10); however the Waiomou Stm at Matamata-Tauranga Rd (1174_4) also appears to be an outlier. It is unclear why the loads have been underestimated for these sites. The Kiwitahi site is upstream of the Paeroa-Tahuna Rd site that had overestimated load. The upstream catchment area for this site is also dominated by pastoral land use and the soil is 88% well drained. There is a dairy shed effluent pond upstream of the catchment, however the estimated *E. coli* discharge from this point source accounts for only 3% of the total measured load. The Waiomou Stm at Matamata-Tauranga Rd site is located in the upper reaches of the Waihou River, has good drainage and around half of the upstream area is forested.

The reduction in model fit for cumulative yields shown in Figure 4-2 suggests that much of the model fit for loads can be explained by area. The Nash-Sutcliffe efficiencies for the yields calculated using the modelled loads against those estimated from monitored data are 0.31 and 0.32 for the full flow record and the 95th percentile flow record respectively, which indicates that the two calibrations result in the same goodness of fit for yields.

Overall, the model contains considerable uncertainty, but it reflects the expected influence of key drivers such as land use, rainfall, soil drainage, and stream decay in a reasonable fashion.

5 Concentration models

Two concentration non-linear multi-variate regression models were developed to provide estimates of median and 95th percentile *E. coli* concentrations for stream reaches in the study area. The models were developed using the measured *E. coli* concentrations presented in Section 3.1 and Appendix A. The models were initially of the form:

$$C = R^b\left(kF_{poor} + F_{good}\right)\left(c_{Dairy}F_{Dairy} + c_{SBI}F_{SBI} + c_{SBH}F_{SBH} + c_{Urb}F_{Urb} + c_{Tree}F_{Tree} + c_{Other}\right) \quad (11)$$

Where C is the concentration (cfu per 100ml), R is the annual rainfall averaged over the catchment upstream of the monitoring site (m/y); F_{poor} and F_{good} are the fractions of the upstream catchment that have either poor to imperfect (≤3.5) or good to very good drainage respectively; F_{Dairy} to F_{other} are the fractions of the upstream catchment in dairy land-use, intensive sheep and beef, hill or high country sheep and beef, urban, forest (native, scrub and exotic forest) or other land uses respectively; b (m^{-1}) and k (dimensionless) are calibrated coefficients for rainfall and drainage; and c_{Dairy}, c_{SBI}, c_{SBH}, c_{Urb}, c_{Tree} and c_{Other} are calibrated coefficients (per 100 ml) for land use.

The models were fitted using the SigmaPlot (Systat Software Inc) non-linear least squares regression tool, with log-transformation to better condition the residuals. Like the load model, the calibration process was iterative with parameters successively removed or amalgamated to optimise model fit. It was found that rainfall was not a statistically significant term for both the median and 95th percentile concentrations, so this term was removed from Equation (11). Likewise, pasture land uses were aggregated into a single land use class. The non-pasture land uses were similarly aggregated. The final model form is:

$$C = \left(kF_{poor} + F_{good}\right)\left(c_{Pasture}F_{Pasture} + c_{Non-pasture}F_{Non-pasture}\right) \quad (12)$$

5.1 Concentration model results

The model parameters and goodness-of-fit are given in Table 5-1 and Table 5-2 for the log-transformed median and 95th percentile concentrations respectively. The full SigmaPlot reports for each model are given in Appendix G. The land use coefficients for the 95th percentile model are generally an order of magnitude larger than for the median concentrations. Both models had similar soil drainage coefficients.

Modelled and measured median and 95th percentile concentrations are compared Figure 5-1 and Figure 5-2. This model performance is slightly poorer than for the models determined for the HRWO study. On average, the models underestimate higher *E. coli* concentrations (measured concentrations greater than 100 cfu/100 ml for median concentrations and 900 cfu/100ml for 95th percentile concentrations) and overestimate lower concentrations. The HRWO models showed a similar tendency.

Table 5-1: Goodness-of-fit and regression coefficients determined for the log-transformed median *E. coli* concentrations.

R	R^2	Adjusted R^2	RMSE
0.61	0.37	0.31	0.67
Coefficient	**Value**	**SE of coefficient**	
c	2.05	1.10	
$c_{Pasture}$	230.03	56.02	
$c_{Non-pasture}$	50.96	19.52	

Table 5-2: Goodness-of-fit and regression coefficients determined for the log-transformed 95th percentile *E. coli* concentrations.

R	R^2	Adjusted R^2	RMSE
0.77	0.59	0.55	0.50
Coefficient	**Value**	**SE of coefficient**	
c	2.38	0.92	
$c_{Pasture}$	2339.90	406.20	
$c_{Non-pasture}$	367.78	118.80	

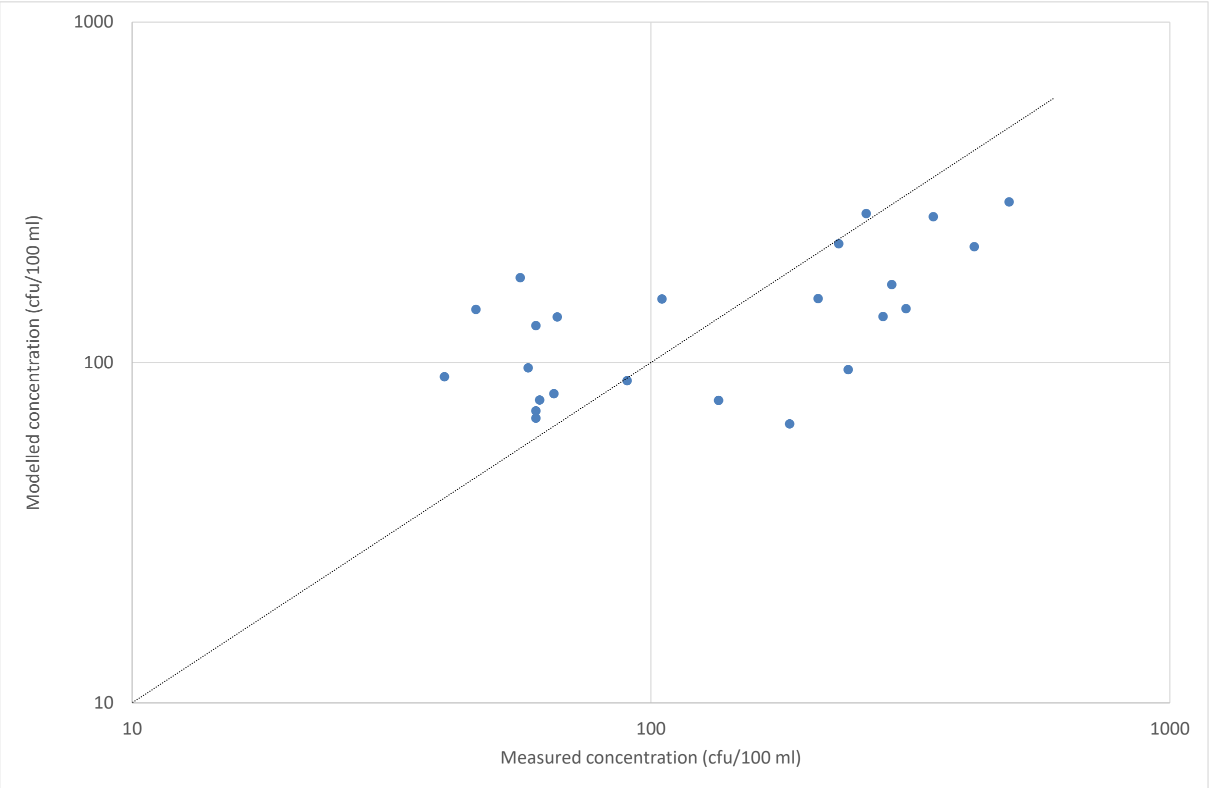


Figure 5-1: Comparison of measured and modelled annual median *E. coli* concentrations plotted with log-scales.

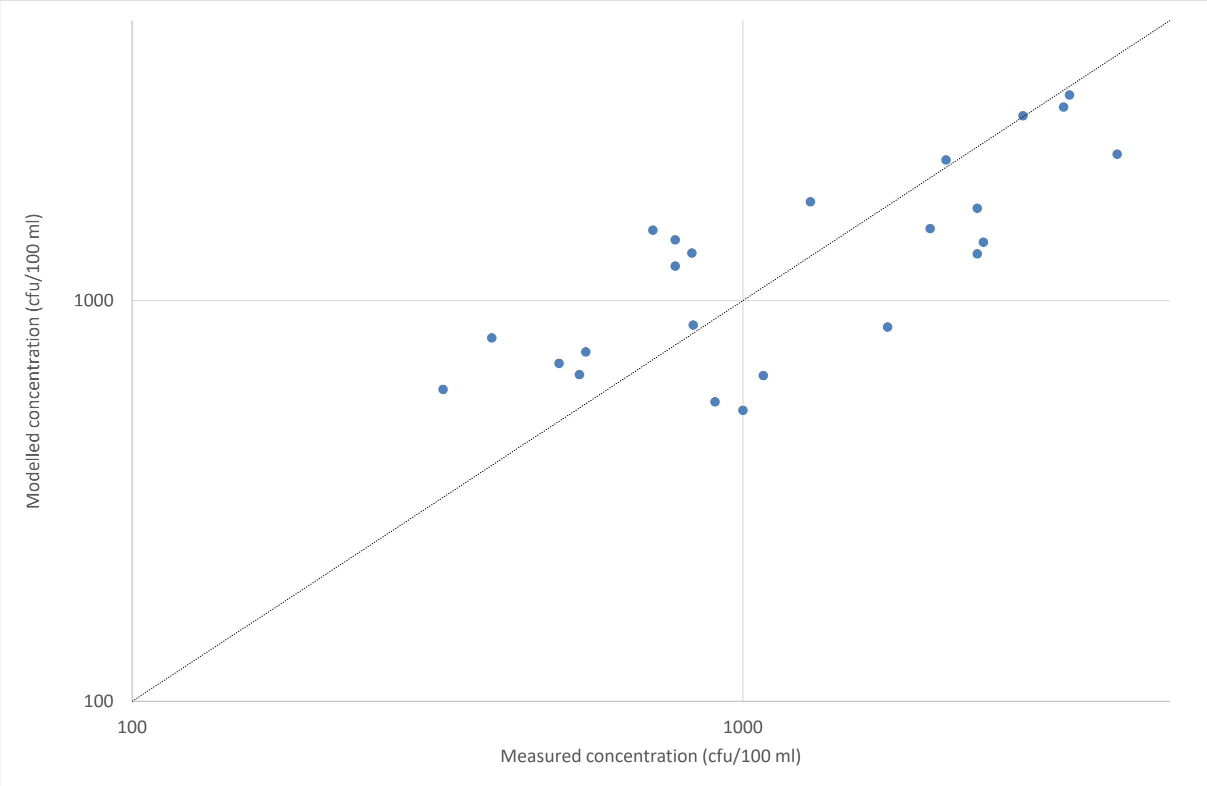


Figure 5-2: Comparison of measured and modelled annual 95th percentile *E. coli* concentrations plotted with log-scales.

6 Model uncertainty and error

In the modelling context, uncertainty refers to the limitations of the model due to, for instance, the choice and representation of model input and outputs; model structure and the simplification of complex physical, chemical and biological processes; and the choice and calibration of model parameters. Model error is separate from uncertainty and can refer to errors in the model code as well as errors in the input, calibration and validation data, due to, for example the accuracy and precision of data capture, data processing methods and storage. Errors and uncertainties within a model propagate at each step in the modelling process such that a small error in input data can snow-ball into a substantial error in model outputs. It is also possible that errors and uncertainties can compensate for each other making it much harder to detect and evaluate them.

The standard errors reported for the calibrated load model parameters and concentration regression coefficients point to substantial uncertainty and error within the models. This reflects the difficulty in modelling *E. coli*, as the yield of microbes from diffuse and point sources is highly variable in time and space (Wilcock, 2006; Muirhead, 2015) making determination of average annual catchment loads and concentrations difficult. This uncertainty is compounded by potential errors in the input data and assumptions made in processing that data for use in the model.

The potential sources of uncertainty and error with respect to *E. coli* modelling for the Hauraki Plains and Coromandel Peninsula are listed below:

- **Load model parameterisation**: The load model was calibrated to minimise the RMSE between the log-transformed modelled loads and loads estimated from measurements at the water quality monitoring sites. Each parameter is applied to the entire catchment area, however, it is feasible that the parameters could be spatially and temporally variable. The model was calibrated against the loads estimated from measurements using the full flow record and the 95th percentile flow record. The SE values for the latter were, with the exception of the non-pasture yield, lower which suggests there may be bias in the rating curves developed for the full flow record causing over-estimation of measured loads.

  There is a strong correlation in the model calibrations between the non-pasture yield and the drainage parameter. This suggests that the drainage parameter is being driven by sites that have a high proportion of upstream imperfectly drained soils and forest, but which also have a higher than expected measured yield.

  Unfortunately, there are no monitoring sites located in the peatlands of the lower Piako and Waihou rivers where the soils have poor drainage and are artificially drained. The river reaches in this are also subject to tidal influence that can extend some 20-30 km upstream of the terminal reaches of these rivers. This means that the calibrated parameters for drainage and stream attenuation may not be valid for this area. The stream attenuation calculated for a reach, for example, is a function of the reach travel time which may be increased on the ebb-tide.

- **Modelled and measured flow**: the loads estimated from measurements used for the load model calibration were determined using TopNet modelled flows for water quality sites without paired flow monitoring. TopNet has been calibrated for all of the study area except the East Coromandel sub-catchment. While the mean annual flow rates compared favourably with those measured, the model was calibrated for peak flow events and has underestimated low flow conditions for some sites. To avoid possible mismatches in instantaneous flow that could influence load calculations, the load calculations were made using daily rather than hourly flow data.

- ***E. coli* concentration data:** *E. coli* concentration data collected from 24 water quality monitoring sites was used to estimate both mean annual loads and median and 95th percentile concentrations. These data were collected as part of WRC SOE reporting for 22 of the sites and from the NRQWN for 2 NIWA maintained sites. The data were not purpose collected for modelling and are subject to error in sampling and analysis. It is assumed that monthly data are representative of the full range of *E. coli* concentrations and that the median *E. coli* load calculated is representative of the true median annual load. As pointed out by Davies-Colley et al. (2011), this is not necessarily the case, as was demonstrated for the Waihou River at Okauia (1122_18) site where there is bias in the load rating curve due a single sampling event coinciding with a high flow event that lead to possible overestimation of the mean annual load.

  Although there may be trends in the monitored data due to land use changes and changes in farm practices at some sites, the concentration data were not trend adjusted due to concerns that detrending may introduce further error. An inspection of the concentration data showed that there is little discernible regional trend in E. coli, which is supported by and evaluation of water quality trends by WRC (Vant, 2013). Instead, five-year median concentrations were used under the assumption that a short monitoring period would limit possible trends at any affected sites.

- **Point sources:** *E. coli* point source data used in the model include industrial and municipal waste and effluent from dairy farms. The point sources are variable over time, making it difficult to assess mean annual loads. Some sources may have new processes in place to reduce contaminant discharge that may not be reflected in the historical water quality record and cannot be accounted for in a steady-state model.

  E. coli loads from consented dairy farm ponds were estimated by using assumptions around the number of cattle serviced by the ponds. The consents are the maximum allowable discharge from each farm which may not be reached, thus the estimated loads are conservative. The load from ponds is likely to be highly variable over time depending on the size and maintenance of the ponds and the size of the dairy herd milked.

- **Diffuse sources:** The load from diffuse sources is calculated on the basis of the area covered by each land use and its calibrated source yield. Land use is represented by 11 land use types with diffuse loads from these sources represented in the load models by only two calibrated source yields (i.e., pasture and non-pasture). This was because the yields from dairy and intensive sheep and beef farming had similar calibrated yields. Likewise the non-pastoral sources were each an order of magnitude lower than the pastoral sources. The aggregation had a very small effect on model fit, but considerably reduced the SE values associated with the model parameters.

  The models do not include data on certain land uses, for example, irrigation or dairy support, which could affect *E. coli* generation.

The derivation and interpretation of the underlying land use data are subject to sampling precision errors and ground-truthing errors. Also, recent land use changes may not be represented in the model input data.

- **Unknown sources:** There may be other microbial sources that have not been accounted for in the models. These could include background *E. coli* from natural sources including wild pigs and waterfowl (e.g., Moriarty et al., 2011) as well as unknown point sources such as such as sewer or pumping station overflows in urban areas.

- **Spatial resolution:** The load and concentration models are subject to spatial smoothing of heterogeneous input data (i.e., scaling effects). The smallest spatial unit of the load model is the REC unit, and there are over 12000 of these in the study area. Spatial data within each unit are lumped and there are no linkages between the data types such as forestry on steeper, well drained land compared to dairy forming on low lands. Land use within each REC unit is split into proportional areas while the area weighted means for each reach unit are used for rainfall and soil drainage class.

  The concentration model operates at the catchment scale, and uses the lumped upstream fractions of land use and soil drainage as input data.

- **Temporal resolution:** The loads model is a steady-state model that predicts mean annual loads. This means that seasonal changes in *E. coli* generation and transport are not captured by the models. The concentration regression model also does not take seasonality into account.

  Adding seasonality would require more complexity in the load models and either extra regression terms or separate regression models for the concentration model. In either case, there are too few data at some monitoring sites to allow seasonal modelling.

  Dynamic modelling may also be possible but would increase the input data needs and model complexity.

7 Recommendations for further work

Section 6 above outlined a number of possible sources of uncertainty in the model, this section suggests further work that could be undertaken to improve the model fit and robustness.

- **Model calibration and validation.** The models have been calibrated using SOE monitoring data, but have not been validated against independent data from other sites or other time periods. Continuation and expansion of *E. coli* monitoring within the catchment will provide further test model and calibration data. Water quality monitoring should be concurrent with flow monitoring were possible to allow for better calculation of loads.

  No monitored water quality data was available for the poorly drained lower Piako and Waihou sub-catchments, this means that the model is likely to not be representative of concentrations or loads from these areas. Water quality monitoring sites set up in the lower reaches of the two rivers would provide a basis for recalibrating and testing the models for these areas.

- **Microbial source tracking:** Assessment of background *E. coli* from natural diffuse sources. This work would help to determine the source of *E. coli* loads from forested catchments with unexpectedly high measured loads. Microbial tracking would also be able to distinguish between *E. coli* loads from farming and from other sources such as water fowl.

- **Re-evaluation of point sources:** Point sources should be regularly evaluated and updated to take into account hitherto unidentified point sources and changes in loads associated with known sources due to, for example, land use change or changes in contaminant management and disposal.

- **Dynamic modelling:** Assessment of the feasibility of dynamic *E. coli* modelling in the catchment at the REC scale to gain a better understanding of the temporal processes in operation and a better representation of those processes.

8 Summary and conclusions

This report describes the development and evaluation of steady-state catchment models to estimate *E. coli* loads and concentrations in freshwater ways located on the Hauraki Plains and Coromandel Peninsula. The work follows a similar set of models developed for the Waikato and Waipa catchments under the HRWO programme (Semadeni-Davies et al., 2015). The models were calibrated using water quality data from 24 monitoring sites located in the study area. The load model operates at the REC reach scale and was calibrated against measured mean annual *E. coli* loads. Where paired flow data were not available to calculate loads, TopNet modelled flows were used. The load model was calibrated against loads determined using the full flow record and with the 95th percentile flow record that removes the highest 5% of peak flows from the full flow record. This was done to remove potential bias in the model results towards high mean annual loads. Two regression models were developed to estimate site median annual and 95th percentile *E. coli* concentrations, respectively. The independent variables for both models are upstream land use and drainage characteristics.

The load model estimates mean annual *E. coli* loads from diffuse sources on the basis of unit land use, rainfall and soil drainage. Point sources are also added to the in-stream load. The loads are then subject to attenuation. The RMSE between the log-transformed measured and modelled loads is 0.77 and 0.53 for the loads calculated with the full and 95th percentile flow records respectively. The model fit indictated by the NSE and adjusted R^2 values were 0.86 and 0.73 for the loads predicted using the full flow record and 0.90 and 0.85 for the loads predicted using the 95th percentile flow record.

The concentration regression model is better able to estimate the 95th percentile annual *E. coli* concentrations (RMSE 0.50, R^2 0.59, adjusted R^2, 0.55) than the median annual concentrations (RMSE 0.67, R^2 0.37, adjusted R^2, 0.31). Both regression models underestimate higher *E. coli* concentrations and overestimate lower concentrations.

There is substantial uncertainty in the parameters calibrated for each of the models which reflects the high spatial and temporal variability in *E. coli* concentrations in freshwater waterways. Sources of this uncertainty in the models include: possible errors in *E. coli* modelling; the assumption that monthly *E. coli* concentration sampling is representative of the range of *E. coli* concentrations in the stream network and over time; the use of TopNet modelled flows to determine mean annual loads for monitoring sites without concurrent flow sampling; spatial and temporal scaling issues; uncertain estimates of loads from point sources; and the possibility that there may be other point sources, such as urban sewer overflows, that have not been taken into account.

Finally, it is recommended that current water and flow monitoring be continued or expanded to provide further data for water quality modelling. Additionally, the feasibility of dynamic modelling should be investigated to capture seasonal changes in *E. coli* concentrations.

Acknowledgements

Thank you to fellow NIWA colleague Christian Zammit for the use of TopNet modelled flow records for water quality monitoring sites with no paired flow records. Water quality data, land use data, point source estimates and flow records were provided for this study by WRC. Thank you also to Bill Vant of WRC for his peer-review and insights.

References

Chai, T. and Draxler, R.R. (2014) Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.*, 7(3): 1247-1250. 10.5194/gmd-7-1247-2014

Cichota, R., Snow, V. and Tait, A. (2008) A functional evaluation of virtual climate station rainfall data. *New Zealand Journal of Agricultural Research*, 51(3): 317-329.

Clark, M.P., Rupp, D.E., Woods, R.A., Zheng, X., Ibbitt, R.P., Slater, A.G., Schmidt, J. and Uddstrom, M.J. (2008) Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources*, 31(10): 1309-1324. http://dx.doi.org/10.1016/j.advwatres.2008.06.005

Collins, R., McLeod, M., Hedley, M., Donnison, A., Close, M., Hanly, J., Horne, D., Ross, C., Davies-Colley, R., Bagshaw, C. and Matthews, L. (2007) Best management practices to mitigate faecal contamination by livestock of New Zealand waters. *New Zealand Journal of Agricultural Research*, 50(2): 267-278. 10.1080/00288230709510294

Davies-Colley, R., Larned, S., Unwin, M., Verburg, P., Hughes, A., Storey, R., McBride, G., Ballantine, D., Hudson, N., Daughney, C. and Hamill, K. (2011) Dependable monitoring of freshwaters for nation scale environmental reporting, Prepared by NIWA for Ministry for the Environment, NIWA Client Report Ham2011-055.

Donnison, A., Ross, C. and McGowan, A. (2011) Escherichia coli and Campylobacter in two conventional Waikato dairy farm effluent ponds. *New Zealand Journal of Agricultural Research*, 54(2): 97-104. 10.1080/00288233.2011.558905

Duan, N. (1983) Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association*, 78(383): 605-610.

Elliott, A.H., Alexander, R.B., Schwarz, G.E., Shankar, U., Sukias, J.P.S. and McBride, G.B. (2005) Estimation of Nutrient Sources and Transport for New Zealand using the Hybrid Mechanistic-Statistical Model SPARROW. *Journal of Hydrology (New Zealand)*, 44(1): 1-27.

Hipsey, M.R., Antenucci, J.P. and Brookes, J.D. (2008) A generic, process-based model of microbial pollution in aquatic systems. *Water Resources Research*, 44(7): n/a-n/a. 10.1029/2007WR006395

Jowett, I.G. (1998) Hydraulic geometry of New Zealand rivers and its use as a preliminary method of habitat assessment. *Regulated Rivers: Research & Management*, 14(5): 451-466. 10.1002/(SICI)1099-1646(1998090)14:5<451::AID-RRR512>3.0.CO;2-1

Ministry for the Environment (2014) *National Policy Statement for Freshwater Management 2014*. Issued by notice in gazette on 4 July 2014, New Zealand Government. http://www.mfe.govt.nz/rma/central/nps/freshwater-management.html

Moriarty, E.M., Karki, N., Mackenzie, M., Sinton, L.W., Wood, D.R. and Gilpin, B.J. (2011) Faecal indicators and pathogens in selected New Zealand waterfowl. *New Zealand Journal of Marine and Freshwater Research, 45(4): 679-688*, 45(4): 679-688. 10.1080/00288330.2011.578653

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Binger, R.L., Harmel, R.D. and Veith, T.L. (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the American Society of Agricultural and Biological Engineers*, 50(3): 885-900.

Muirhead, R. (2015) A Farm-Scale Risk-Index for Reducing Fecal Contamination of Surface Waters. *J. Environ. Qual.*, 44(1): 248-255. 10.2134/jeq2014.07.0311

Nash, J.E. and Sutcliffe, J.V. (1970) River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3): 282–290.

Newsome, P.F.J., Wilde, R.H. and Willoughby, E.J. (2008) Land Resource Information System spatial data layers: Data Dictionary, Landcare Research New Zealand Ltd.

Nikitas, P. and Pappa-Louisi, A. (2000) Non-linear least-squares fitting with microsoft excel solver and related routines in HPLC modelling of retention. *Chromatographia*, 52(7-8): 477-486. 10.1007/BF02535723

Semadeni-Davies, A., Elliot, S. and Shankar, U. (2016) CLUES - Catchment Land Use for Environmental Sustainability User Manual Fifth Edition: CLUES 10.3, NIWA internal report: AKL2016-017. Available for open download: ftp://ftp.niwa.co.nz/clues

Semadeni-Davies, A., Elliott, S. and Yalden, S. (2015) Modelling E. coli in the Waikato and Waipa River Catchments: Development of a catchment-scale microbial model, Prepared for Technical Leaders Group of the Healthy Rivers / Wai Ora Project. NIWA Client report: HAM2015-089.

Snelder, T., Biggs, B. and Weatherhead, M. (2010) New Zealand River Environment Classification User Guide. March 2004 (Updated June 2010), ME Number 499.

Vant, B. (2013) Trends in river water quality in the Waikato region, 1993-2012, Waikato Regional Council Technical Report 2013/20.

Wilcock, B. (2006) Assessing the Relative Importance of Faecal Pollution Sources in Rural Catchments Report prepared for Environment Waikato, NIWA client report: AHM2006-104.

Wilde, R.H., Willoughby, E.J. and A.E., H. (2004) Data Manual for the National Soils Database, Spatial Extension. Landcare Research New Zealand Ltd. .

Appendix A   Upstream input data

This appendix presents the model input data that has been aggregated for use in the concentration models for each water quality modelling site. The data were derived by summing the relevant areas for each REC reach sub-catchment upstream of each monitoring site to get a cumulative total.

Table A-1: Aggregated model input data for the area upstream of each water quality monitoring site.

Water quality site ID and name		Upstream area (km^2)	Land use fraction					Drainage class fraction		Areal weighted mean rainfall (mm/year)
			Dairy	Sheep and beef	Forest	Urban	Other	Poor to imperfect (1-3)	Good to well (4-5)	
1230_1	Waitakaruru River at Coxhead Rd Br	50.1	0.23	0.31	0.44	0.01	0.01	0.80	0.20	1395
489_2	Mangawhero Stm at Mangawara Rd	6.4	0.05	0.18	0.77	0.00	0.00	0.53	0.47	1439
749_10	Piako River at Kiwitahi	108.1	0.66	0.16	0.14	0.00	0.03	0.12	0.88	1435
749_15	Piako River at Paeroa-Tahuna Rd Br	538.9	0.65	0.25	0.06	0.01	0.02	0.37	0.63	1373
753_4	Piakonui Stm at Piakonui Rd	8.0	0.10	0.33	0.52	0.00	0.05	0.00	1.00	1420
1249_15	Waitoa River at Landsdowne Rd Br	122.9	0.54	0.33	0.03	0.00	0.11	0.06	0.94	1530
1249_18	Waitoa River at Mellon Rd Recorder	408.9	0.66	0.21	0.04	0.01	0.08	0.28	0.72	1463
169_2	Hikutaia River at Old Maratoto Rd	73.7	0.10	0.11	0.78	0.00	0.01	0.06	0.94	2177
1122_18	Waihou River at Okauia	806.4	0.34	0.20	0.42	0.01	0.03	0.04	0.96	1797
1122_41	Waihou River at Whites Rd	42.0	0.15	0.07	0.77	0.00	0.00	0.00	1.00	1820
1173_2	Waiohotu Stm at Waiohotu Rd	7.4	0.00	0.12	0.81	0.01	0.06	0.00	1.00	1918
1174_4	Waiomou Stm at Matamata-Tauranga Rd	201.1	0.26	0.24	0.47	0.00	0.04	0.03	0.97	1873
669_6	Oraka Stm at Lake Rd	255.3	0.25	0.19	0.51	0.01	0.01	0.02	0.98	1782
HM5	Waihou River at Te Aroha	1106.9	0.38	0.17	0.39	0.01	0.03	0.10	0.90	1796
1239_32	Waitekauri River U/S Ohinemuri Confluence	43.1	0.14	0.09	0.69	0.00	0.05	0.00	1.00	2290
619_19	Ohinemuri River at Queens Head	136.3	0.45	0.15	0.22	0.03	0.05	0.01	0.99	2099
619_20	Ohinemuri River at SH25 Br	26.2	0.36	0.14	0.40	0.00	0.02	0.00	1.00	2247
HM6	Ohinemuri River at Karangahake Gorge	286.2	0.29	0.11	0.49	0.02	0.04	0.02	0.98	2072
234_11	Kauaeranga River at Smiths Cableway/Recorder	119.9	0.01	0.05	0.93	0.00	0.01	0.22	0.78	2002

Water quality site ID and name		Upstream area (km^2)	Land use fraction					Drainage class fraction		Areal weighted mean rainfall (mm/year)
			Dairy	Sheep and beef	Forest	Urban	Other	Poor to imperfect (1-3)	Good to well (4-5)	
1105_3	Waiau River at E309 Rd Ford	24.5	0.00	0.01	0.97	0.00	0.00	0.56	0.44	2118
954_5	Tapu River at Tapu-Coroglen Rd	26.5	0.01	0.04	0.93	0.00	0.00	0.09	0.91	2013
1257_3	Waiwawa River at Sh25 Coroglen	132.9	0.04	0.03	0.92	0.00	0.01	0.21	0.79	2303
940_10	Tairua River at Morrisons Br Hikuai	152.6	0.05	0.06	0.88	0.00	0.00	0.11	0.89	2321
1312_3	Wharekawa River at SH25	55.7	0.01	0.06	0.92	0.00	0.00	0.02	0.98	2200

Appendix B Point sources and dairy farm effluent discharges

Table B-1: Estimated loads from point sources. Estimates provided by WRC for this project (contact Bill Vant). *Council data quality assessment.

Point source	REC2 reach	Sub-catchment	CFU x 10^9 100ml per day	CFU x 10^{15} 100ml per year	Estimate quality*
Waitakaruru sewage treatment	3046745	West Hauraki	0.99	0.0004	good
Kerepehi sewage treatment	3048741	Piako (Lower)	1.29	0.0005	good
Ngatea sewage treatment	3048264	Piako (Lower)	15.87	0.0058	good
Morrinsville dairy factory	3058232	Piako (Upper)	1.00	0.0004	expert judgement
Morrinsville sewage treatment	3058054	Piako (Upper)	5.99	0.0022	patchy
Tahuna sewage treatment	3053863	Piako River (Upper)	0.99	0.0004	limited
Tatua dairy factory	3057019	Waitoa River (Piako)	1.00	0.0004	expert judgement
Waharoa dairy factory	3062391	Waitoa (Piako)	0.04	0.0000	limited
Waihou sewage treatment	3056068	Waitoa (Piako)	90.31	0.0330	patchy
Waitoa dairy factory	3056329	Waitoa (Piako)	1.00	0.0004	expert judgement
Waitoa meatworks	3057133	Waitoa (Piako)	213.00	0.0777	good
Waitoa poultry processor	3056838	Waitoa (Piako)	0.09	0.0000	good
Thames sewage treatment	3044986	Waihou (Lower)	63.83	0.0233	good
Turua sewage treatment	3047100	Waihou (Lower)	3.00	0.0011	good
Matamata sewage treatment	3066095	Waihou (Upper)	31.03	0.0113	patchy
Putararu sewage treatment	3079825	Waihou (Upper)	2.91	0.0011	good
Te Aroha meatworks	3055718	Waihou (Upper)	339.96	0.1241	good

Te Aroha sewage treatment	3054417	Waihou (Upper)	123.14	0.0450	patchy
Tirau dairy factory	3075526	Waihou (Upper)	1.00	0.0004	expert judgement
Tirau sewage treatment	3075486	Waihou (Upper)	0.04	0.0000	good
Paeroa meatworks	3051334	Ohinemuri (Waihou)	1.28	0.0005	good
Paeroa sewage treatment	3050636	Ohinemuri (Waihou)	1.84	0.0007	good
Waihi sewage treatment	3051779	Ohinemuri (Waihou)	0.35	0.0001	good
Waihi gold mine	3051570	Ohinemuri (Waihou)	0.10	0.0000	expert judgement

Table B-2: Estimated loads from FDE ponds by sub-catchment. Derived from WRC consented pond discharge volumes.

Sub-catchment	CFU x 10^{15} 100ml per year	Number of consented FDE ponds
Ohinemuri River (Waihou)	0.0062	1
Piako River (Lower)	0.0154	1
Piako River (Upper)	0.0867	6
Waihou River (Lower)	0.0198	3
Waihou River (Upper)	0.1414	8
Waitoai River (Piako)	0.0359	3
West Hauraki	0.0218	2

Appendix C Measured *E. coli* concentrations

Figure C-1: Median and 95th percentile *E. coli* concentrations (organisms/100 ml) calculated (Hazen method) from the measured water quality data.

Water quality site ID and name		Five years (2011-2016)			11 years (2005-2016)			18 years (1998-2016)		
		Median	95^{th} percentile	Number of samples	Median	95^{th} percentile	Number of samples	Median	95^{th} percentile	Number of samples
1230_1	Waitakaruru River at Coxhead Rd Br	260	3350	45	320	3225	69	290	3525	97
489_2	Mangawhero Stm at Mangawara Rd	46	775	45	47	821	69	60	727	97
749_10	Piako River at Kiwitahi	230	4100	45	300	5680	69	300	6635	97
749_15	Piako River at Paeroa-Tahuna Rd Br	490	3425	45	490	3280	66	500	4740	94
753_4	Piakonui Stm at Piakonui Rd	60	775	45	70	1200	69	80	1530	97
1249_15	Waitoa River at Landsdowne Rd Br	420	2150	45	460	2315	69	470	6125	97
1249_18	Waitoa River at Mellon Rd Recorder	350	2875	45	450	2215	69	480	4305	97
169_2	Hikutaia River at Old Maratoto Rd	240	1725	45	250	1705	69	260	1765	97
1122_18	Waihou River at Okauia	210	2025	45	240	1845	69	250	1765	97
1122_41	Waihou River at Whites Rd	40	388	45	40	439	69	40	328	97
1173_2	Waiohotu Stm at Waiohotu Rd	60	323	45	50	463	69	61	561	97
1174_4	Waiomou Stm at Matamata-Tauranga Rd	310	2475	45	310	2210	69	310	2200	97
669_6	Oraka Stm at Lake Rd	280	825	45	300	3595	69	410	2265	97
HM5*	Waihou River at Te Aroha	291	2419	61	255	2419	134	-	-	-
1239_32	Waitekauri River U/S Ohinemuri Confluence	58	829	44	73	1900	68	80	2740	96
619_19	Ohinemuri River at Queens Head	56	1290	44	80	1740	68	85	4570	96
619_20	Ohinemuri River at SH25 Br	105	712	44	135	2400	68	150	2510	96
HM6*	Ohinemuri River at Karangahake Gorge	66	2419	61	57	1986	134	-	-	-
234_11	Kauaeranga River at Smiths Cableway/Recorder	135	1080	44	135	1200	68	120	1170	96

Water quality site ID and name		Five years (2011-2016)			11 years (2005-2016)			18 years (1998-2016)		
		Median	95th percentile	Number of samples	Median	95th percentile	Number of samples	Median	95th percentile	Number of samples
1105_3	Waiau River at E309 Rd Ford	90	553	45	100	2515	69	100	2695	97
954_5	Tapu River at Tapu-Coroglen Rd	60	900	45	70	1240	69	90	2520	97
1257_3	Waiwawa River at Sh25 Coroglen	61	540	44	95	1180	68	100	1560	96
940_10	Tairua River at Morrisons Br Hikuai	65	500	44	80	1050	68	90	2340	96
1312_3	Wharekawa River at SH25	185	1000	44	240	1680	68	250	2160	96

*Data not available before 2005

Appendix D Recorded and TopNet modelled mean annual flow rates

Table D-1: Mean annual flow rates determined for the water quality monitoring sites from the observed and modelled flow records .

Water quality monitoring site		Mean annual flow rates (l/s)		Ratio
		Recorded	TopNet	
1230_1	Waitakaruru River at Coxhead Rd Br	-	1073	-
489_2	Mangawhero Stm at Mangawara Rd	-	240	-
749_10	Piako River at Kiwitahi	1632	1768	0.92
749_15	Piako River at Paeroa-Tahuna Rd Br	7046	7611	0.93
753_4	Piakonui Stm at Piakonui Rd	-	153	-
1249_15	Waitoa River at Landsdowne Rd Br	1475	360	4.10
1249_18	Waitoa River at Mellon Rd Recorder	4595	4719	0.97
169_2	Hikutaia River at Old Maratoto Rd	-	3489	-
1122_18	Waihou River at Okauia	26563	26747	0.99
1122_41	Waihou River at Whites Rd	-	1238	-
1173_2	Waiohotu Stm at Waiohotu Rd	-	421	-
1174_4	Waiomou Stm at Matamata-Tauranga Rd	-	8403	-
669_6	Oraka Stm at Lake Rd	-	7014	-
HM5	Waihou River at Te Aroha	35289	38429	0.92
1239_32	Waitekauri River U/S Ohinemuri Confluence	-	2014	-
619_19	Ohinemuri River at Queens Head	5215	6426	0.81
619_20	Ohinemuri River at SH25 Br	-	1319	-
HM6	Ohinemuri River at Karangahake Gorge	11657	12252	0.95
234_11	Kauaeranga River at Smiths Cableway/Recorder	5738	6144	0.93
1105_3	Waiau River at E309 Rd Ford	-	2012	-
954_5	Tapu River at Tapu-Coroglen Rd	920	769	1.20
1257_3	Waiwawa River at Sh25 Coroglen	7474	-	-
940_10	Tairua River at Morrisons Br Hikuai	-	-	-
1312_3	Wharekawa River at SH25	2144	-	-

Appendix E Rating curves and mean annual loads estimated from measured water quality data

Sites with measured flow

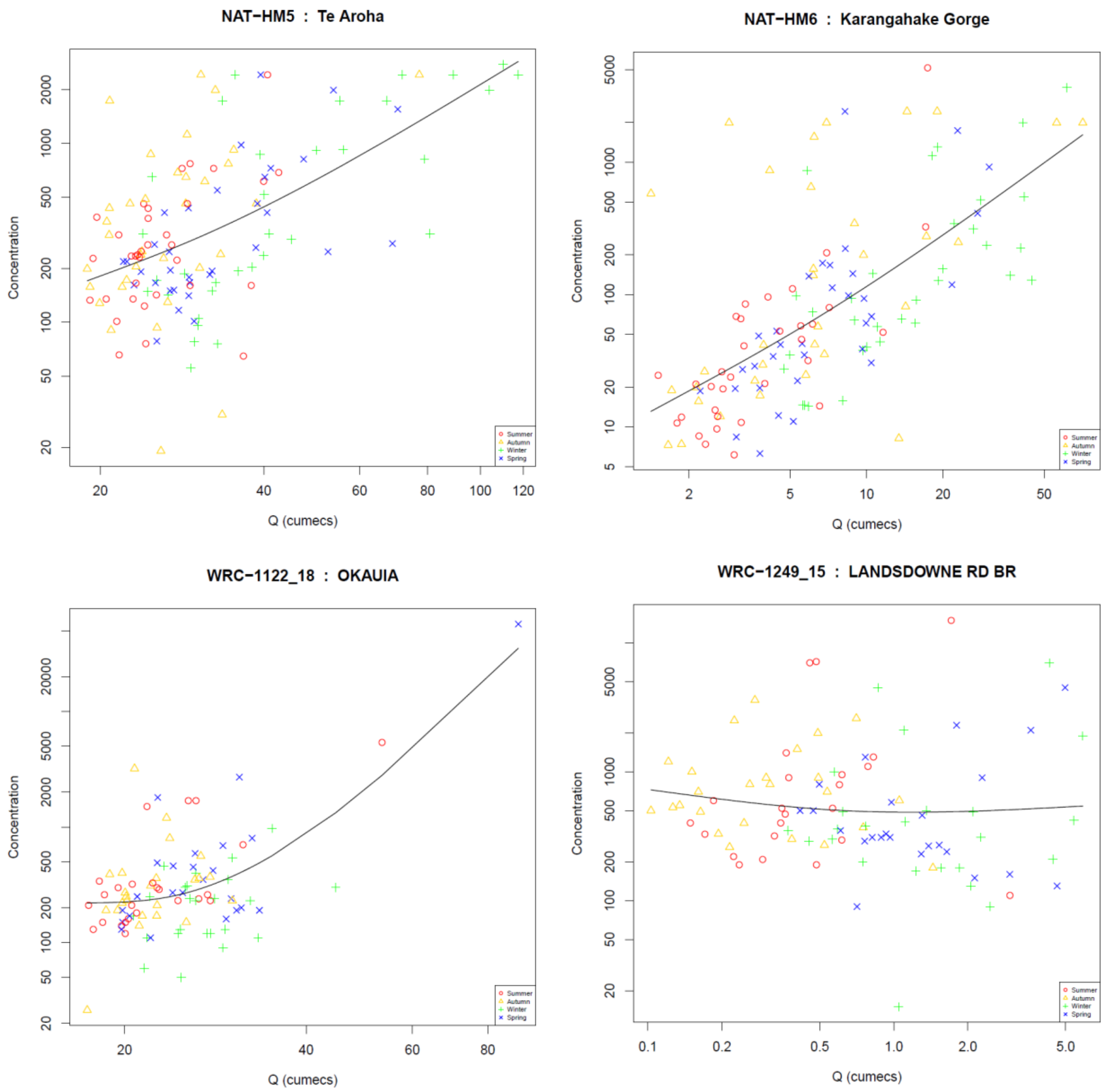


Figure E-1: Rating curves for the water quality monitoring sites with paired measured flow records.

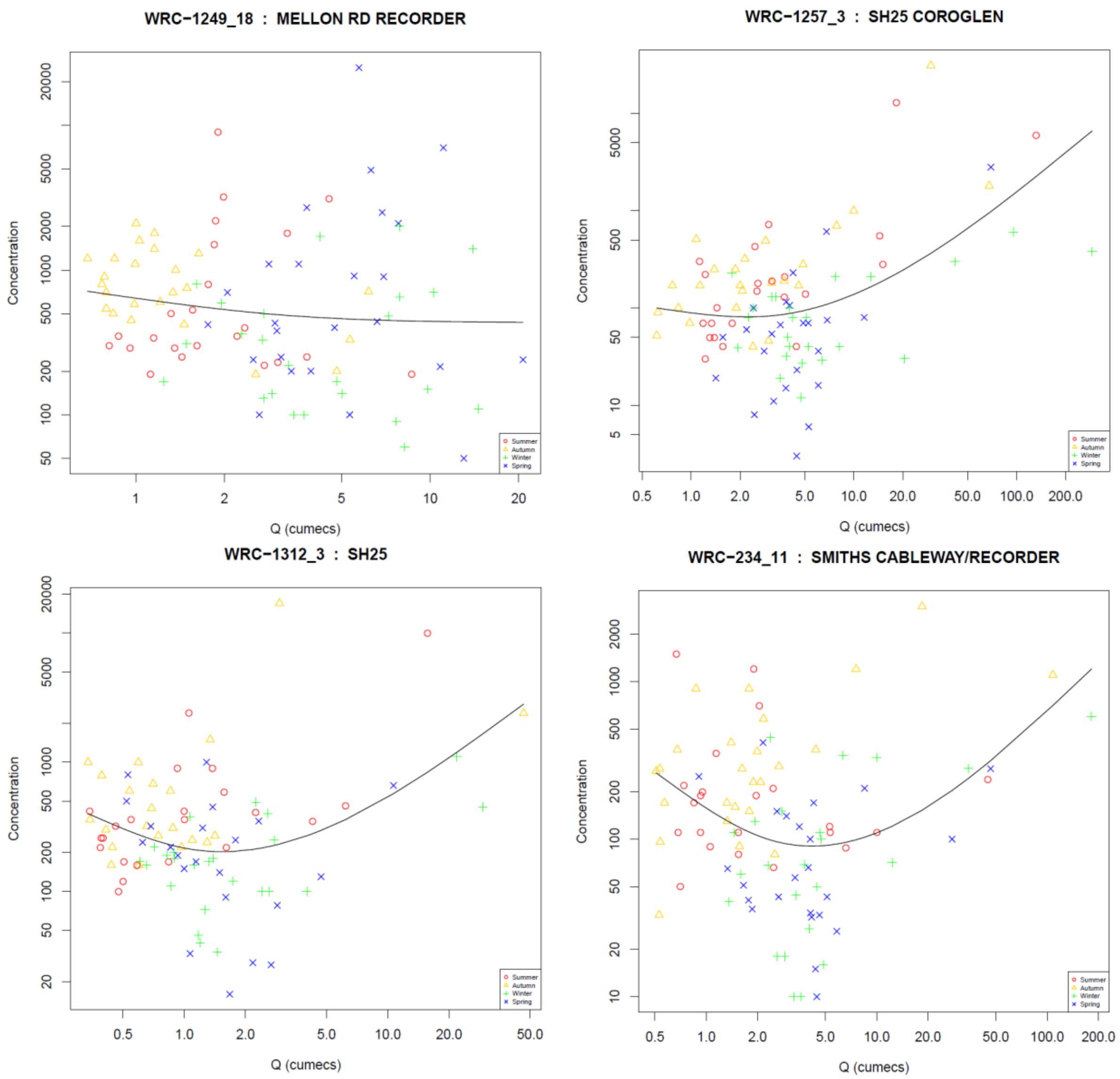


Figure E-1: Rating curves for the water quality monitoring sites with paired measured flow records. Continured

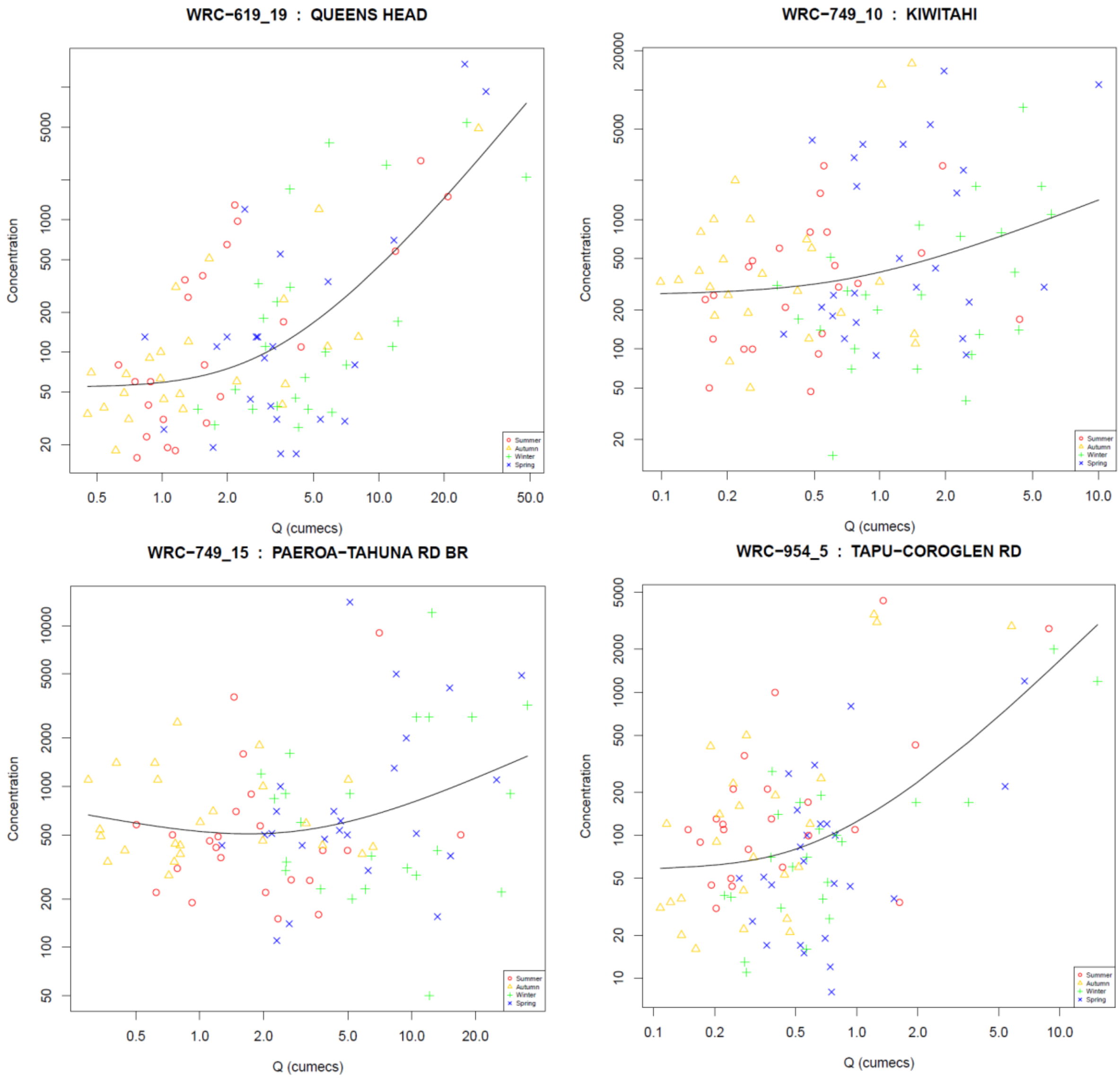


Figure E-1: Rating curves for the water quality monitoring sites with paired measured flow records. Continured

Sites with modelled (TopNet) flow

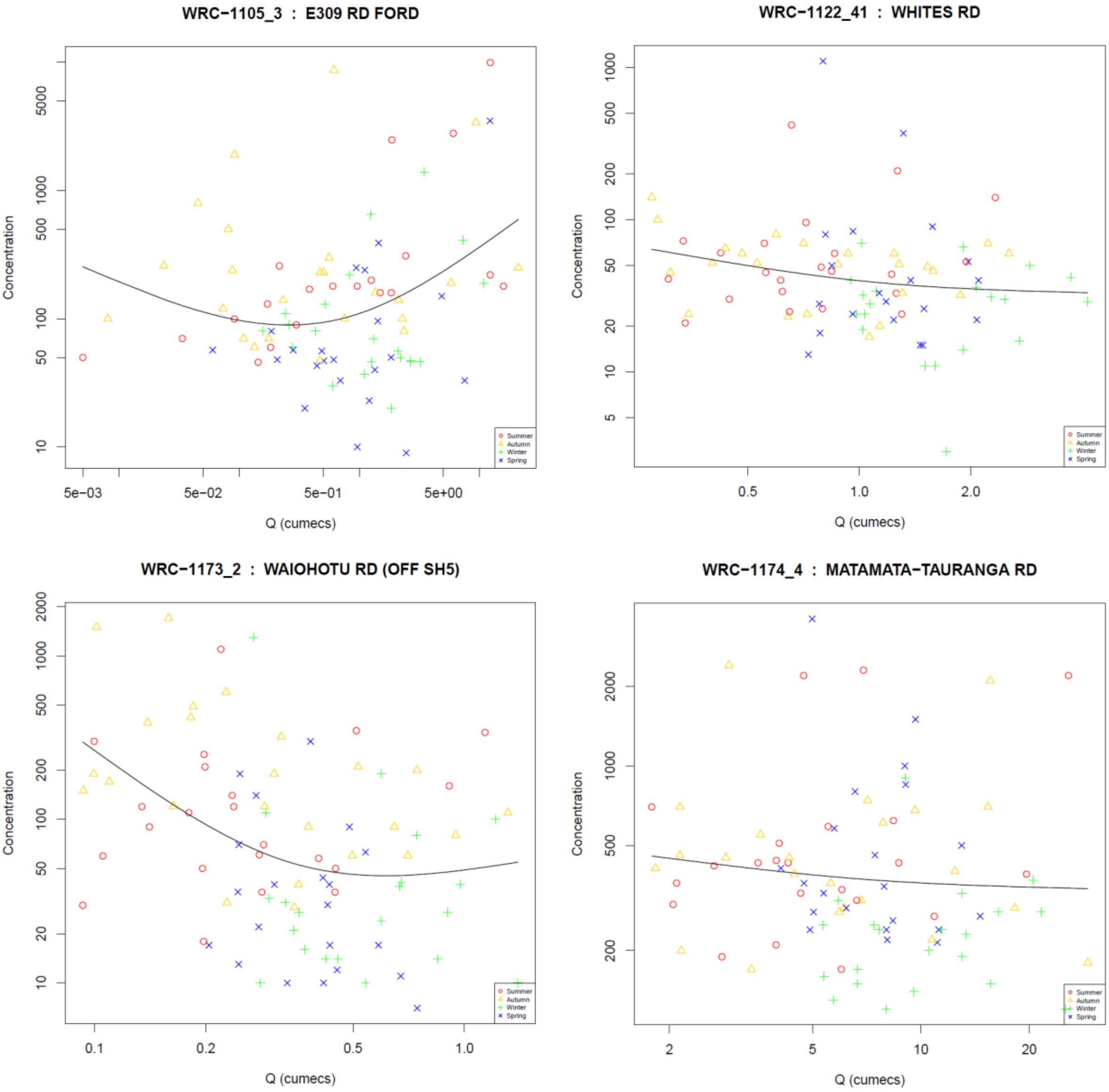


Figure E-2: Rating curves for the water quality monitoring sites with modelled (TopNet) flow records.

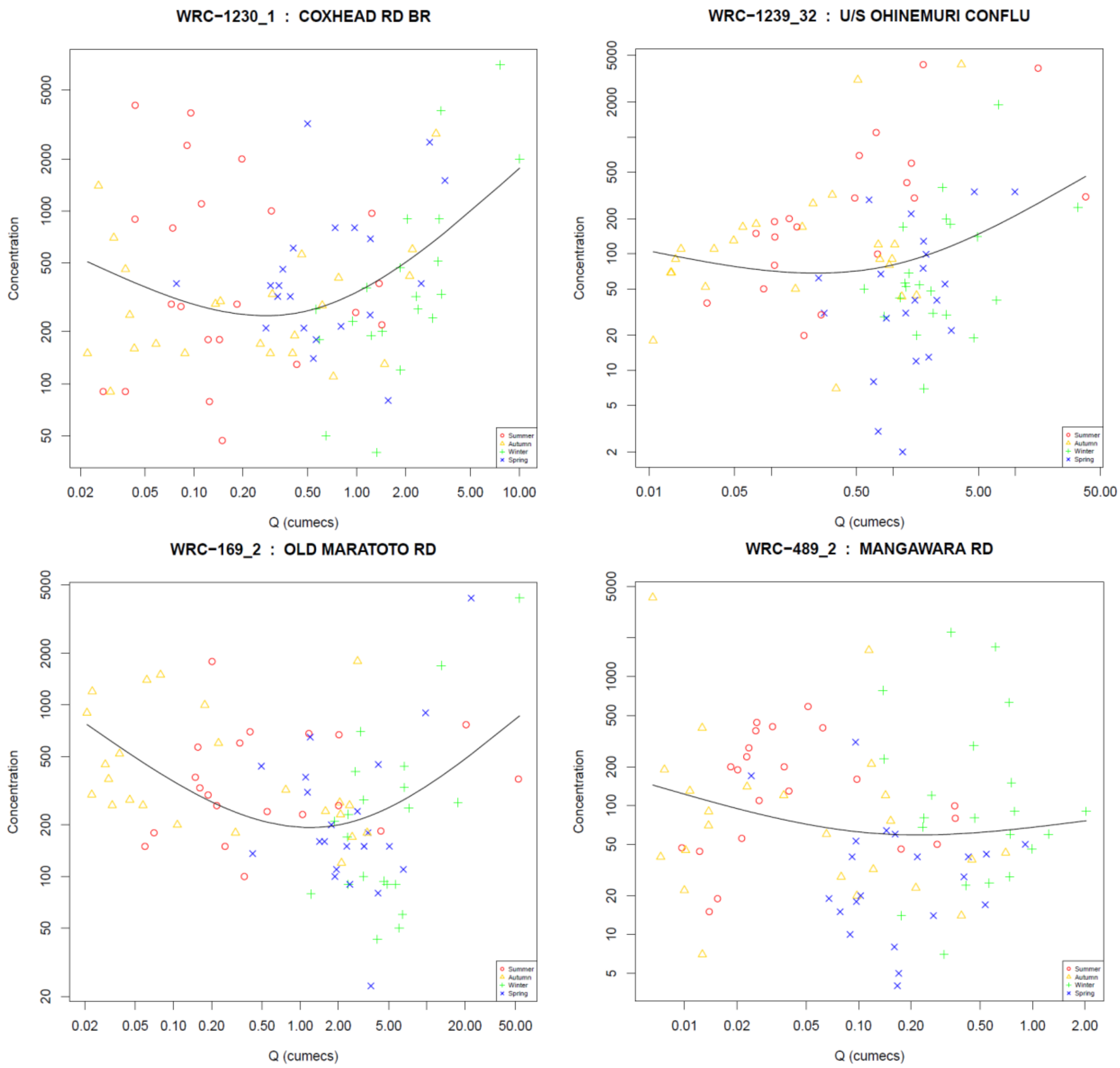


Figure E-2: Rating curves for the water quality monitoring sites with modelled (TopNet) flow records.- Continued

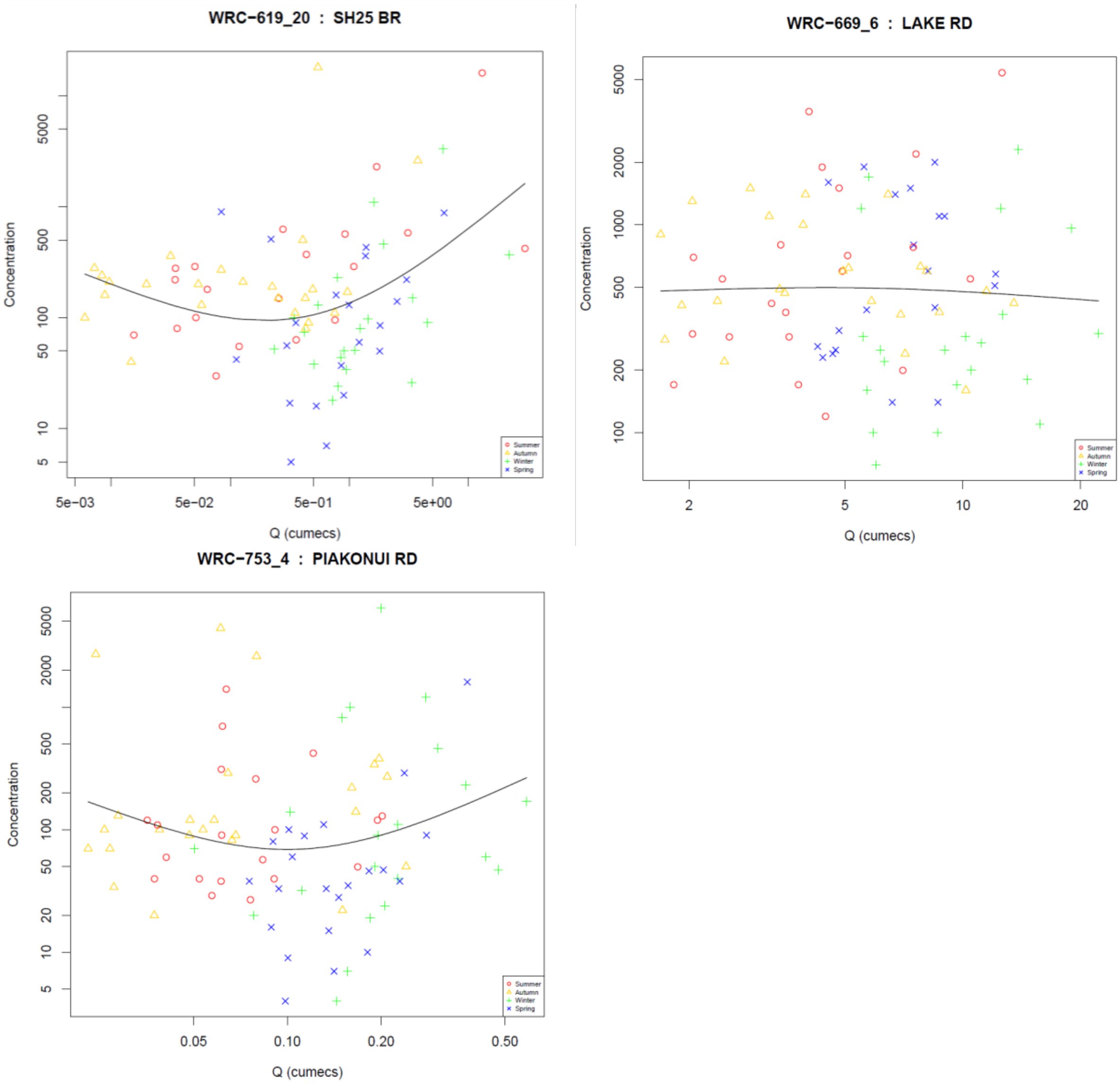


Figure E-2: Rating curves for the water quality monitoring sites with modelled (TopNet) flow records.- Continued

Assessment summary (overleaf)

Table E-1: Mean annual *E. coli* loads (10^{10} organisms) calculated from five year (2011-2016) median concentrations using daily flow rates from the full flow record and 95th percentile flow record.

Water quality monitoring site		Measured or modelled flow	Mean annual load Full flow record	Mean annual load 95th percentile flow record	Standard deviation of the natural log of loads	Rating curve RMSE
1230_1	Waitakaruru River at Coxhead Rd Br	TopNet	64831	30853	0.46	0.95
489_2	Mangawhero Stm at Mangawara Rd	TopNet	1000	925	0.23	1.20
749_10	Piako River at Kiwitahi	Measured	146326	101300	0.57	1.33
749_15	Piako River at Paeroa-Tahuna Rd Br	Measured	445193	185963	0.38	0.96
753_4	Piakonui Stm at Piakonui Rd	TopNet	1544	1512	0.52	1.34
1249_15	Waitoa River at Landsdowne Rd Br	Measured	49819	40666	0.46	1.05
1249_18	Waitoa River at Mellon Rd Recorder	Measured	129880	98566	0.39	1.06
169_2	Hikutaia River at Old Maratoto Rd	TopNet	49012	25419	0.39	0.83
1122_18	Waihou River at Okauia	Measured	9760215	363352	1.69	0.75
1122_41	Waihou River at Whites Rd	TopNet	2179	2101	0.14	0.76
1173_2	Waiohotu Stm at Waiohotu Rd	TopNet	954	944	0.15	0.95
1174_4	Waiomou Stm at Matamata-Tauranga Rd	TopNet	141249	115448	0.14	0.64
669_6	Oraka Stm at Lake Rd	TopNet	77688	65067	0.15	0.68
HM5	Waihou River at Te Aroha	Measured	1248601	408218	0.22	0.83
1239_32	Waitekauri River U/S Ohinemuri Confluence	TopNet	19141	10469	0.59	1.16
619_19	Ohinemuri River at Queens Head	Measured	711727	64891	0.61	1.13
619_20	Ohinemuri River at SH25 Br	TopNet	55264	28550	0.68	1.12
HM6	Ohinemuri River at Karangahake Gorge	Measured	1324077	124118	0.58	1.15
234_11	Kauaeranga River at Smiths Cableway/Recorder	Measured	88794	26813	0.31	0.93
1105_3	Waiau River at E309 Rd Ford	TopNet	44559	21002	0.53	1.12
954_5	Tapu River at Tapu-Coroglen Rd	Measured	32066	15026	0.33	0.98

Water quality monitoring site		Measured or modelled flow	Mean annual load Full flow record	Mean annual load 95th percentile flow record	Standard deviation of the natural log of loads	Rating curve RMSE
1257_3	Waiwawa River at Sh25 Coroglen	Measured	431578	38683	1.00	0.95
940_10	Tairua River at Morrisons Br Hikuai	TopNet	-	-	-	-
1312_3	Wharekawa River at SH25	Measured	77502	24227	0.60	0.85

Appendix F Comparison of measured and modelled loads and yields

Table F-1: Measured (March 2011-2016) mean annual loads and yields against those modelled using the full flow record.

Water quality monitoring site		Mean annual load (peta org / year)		Mean annual yield (10^{10} org / km^2 / year)	
		Measured	Modelled	Measured	Modelled
1230_1	Waitakaruru River at Coxhead Rd Br	0.65	0.47	1294	945
489_2	Mangawhero Stm at Mangawara Rd	0.01	0.04	157	587
749_10	Piako River at Kiwitahi	1.46	0.51	1354	471
749_15	Piako River at Paeroa-Tahuna Rd Br	4.45	4.12	826	765
753_4	Piakonui Stm at Piakonui Rd	0.02	0.01	194	72
1249_15	Waitoa River at Landsdowne Rd Br	0.50	0.53	405	429
1249_18	Waitoa River at Mellon Rd Recorder	1.30	3.64	318	890
169_2	Hikutaia River at Old Maratoto Rd	0.49	0.69	665	942
1122_18	Waihou River at Okauia		4.70		583
1122_41	Waihou River at Whites Rd	0.02	0.06	52	148
1173_2	Waiohotu Stm at Waiohotu Rd	0.01	0.02	129	334
1174_4	Waiomou Stm at Matamata-Tauranga Rd	1.41	1.48	702	735
669_6	Oraka Stm at Lake Rd	0.78	0.99	304	390
HM5	Waihou River at Te Aroha	12.49	11.48	1128	1037
1239_32	Waitekauri River U/S Ohinemuri Confluence	0.19	0.75	444	1744
619_19	Ohinemuri River at Queens Head	7.12	2.40	5221	1763
619_20	Ohinemuri River at SH25 Br	0.55	0.51	2109	1948
HM6	Ohinemuri River at Karangahake Gorge	13.24	3.48	4627	1216
234_11	Kauaeranga River at Smiths Cableway/Recorder	0.89	0.72	740	603
1105_3	Waiau River at E309 Rd Ford	0.45	0.39	1817	1604
954_5	Tapu River at Tapu-Coroglen Rd	0.32	0.39	1208	1483
1257_3	Waiwawa River at Sh25 Coroglen	4.32	2.89	3246	2175
940_10	Tairua River at Morrisons Br Hikuai				
1312_3	Wharekawa River at SH25	0.78	0.32	1391	570

Table F-2: Measured (March 2011-2016) mean annual loads and yields against those modelled using the 95th percentile flow record.

Water quality monitoring site		Mean annual load (peta org / year)		Mean annual yield (10^{10} org / km^2 / year)	
		Measured	Modelled	Measured	Modelled
1230_1	Waitakaruru River at Coxhead Rd Br	0.31	0.28	616	563
489_2	Mangawhero Stm at Mangawara Rd	0.01	0.02	145	368
749_10	Piako River at Kiwitahi	1.01	0.36	937	336
749_15	Piako River at Paeroa-Tahuna Rd Br	1.86	2.27	345	421
753_4	Piakonui Stm at Piakonui Rd	0.02	0.01	190	120
1249_15	Waitoa River at Landsdowne Rd Br	0.41	0.37	331	298
1249_18	Waitoa River at Mellon Rd Recorder	0.99	1.74	241	426
169_2	Hikutaia River at Old Maratoto Rd	0.25	0.25	345	344
1122_18	Waihou River at Okauia		2.17		269
1122_41	Waihou River at Whites Rd	0.02	0.06	50	143
1173_2	Waiohotu Stm at Waiohotu Rd	0.01	0.01	127	181
1174_4	Waiomou Stm at Matamata-Tauranga Rd	1.15	0.69	574	343
669_6	Oraka Stm at Lake Rd	0.65	0.54	255	211
HM5	Waihou River at Te Aroha	4.08	3.72	369	336
1239_32	Waitekauri River U/S Ohinemuri Confluence	0.10	0.20	243	475
619_19	Ohinemuri River at Queens Head	0.65	0.83	476	609
619_20	Ohinemuri River at SH25 Br	0.29	0.18	1089	686
HM6	Ohinemuri River at Karangahake Gorge	1.24	1.19	434	415
234_11	Kauaeranga River at Smiths Cableway/Recorder	0.27	0.31	224	261
1105_3	Waiau River at E309 Rd Ford	0.21	0.12	857	507
954_5	Tapu River at Tapu-Coroglen Rd	0.15	0.08	566	314
1257_3	Waiwawa River at Sh25 Coroglen	0.39	0.63	291	473
940_10	Tairua River at Morrisons Br Hikuai				
1312_3	Wharekawa River at SH25	0.24	0.13	435	228

Appendix G Concentration regression model reports

This appendix contains the regression reports generated by SigmaPlot for both the median and 95th percentile concentration models.

Median concentration model

Nonlinear Regression **Wednesday, August 24, 2016, 10:36:42 a.m.**

Data Source: Data 1 in Concentration_RegressionCMedian_Hauraki - Sandy.JNB
Equation: Section 1, FINAL model in Concentration_RegressionCMedian_Hauraki - Sandy.JNB
Frac_Past = Frac_Dairy+Frac_SBI+Frac_SBHillHi
Frac_NonPast = Frac_Urban + Frac_Trees + Frac_Other
f = ln(FracDrainHi + c * FracDrainLow) + ln(cPast*Frac_Past+ cNonPast * Frac_NonPast)

R	Rsqr	Adj Rsqr	Standard Error of Estimate
0.6080	0.3697	0.3097	0.6701

	Coefficient	Std. Error	t	P
c	2.0497	1.0995	1.8642	0.0763
cPast	230.0341	56.0189	4.1064	0.0005
cNonPast	50.9552	19.5164	2.6109	0.0163

Analysis of Variance:

	DF	SS	MS
Regression	3	572.0567	190.6856
Residual	21	9.4300	0.4490
Total	24	581.4867	24.2286

Corrected for the mean of the observations:

	DF	SS	MS	F	P
Regression	2	5.5317	2.7659	6.1594	0.0079
Residual	21	9.4300	0.4490		
Total	23	14.9618	0.6505		

Statistical Tests:

Normality Test (Shapiro-Wilk) Passed (P = 0.4984)

W Statistic= 0.9629 Significance Level = 0.0500

95^{th} percentile concentration

Nonlinear Regression **Wednesday, August 24, 2016, 10:52:27 a.m.**

Data Source: Data 1 in Concentration_RegressionC95_Hauraki.JNB
Equation: Section 1, FINAL model in Concentration_RegressionC95_Hauraki.JNB
Frac_Past = Frac_Dairy+Frac_SBI+Frac_SBHillHi
Frac_NonPast = Frac_Urban + Frac_Trees + Frac_Other
f = ln(FracDrainHi + c * FracDrainLow) + ln(cPast*Frac_Past+ cNonPast * Frac_NonPast)

R	Rsqr	Adj Rsqr	Standard Error of Estimate
0.7684	0.5904	0.5514	0.4985

	Coefficient	Std. Error	t	P
c	2.3774	0.9202	2.5834	0.0173
cPast	2339.9046	406.2404	5.7599	<0.0001
cNonPast	367.7764	118.7633	3.0967	0.0055

Analysis of Variance:

	DF	SS	MS
Regression	3	1217.3744	405.7915
Residual	21	5.2189	0.2485
Total	24	1222.5933	50.9414

Corrected for the mean of the observations:

	DF	SS	MS	F	P
Regression	2	7.5240	3.7620	15.1375	<0.0001
Residual	21	5.2189	0.2485		
Total	23	12.7429	0.5540		

Statistical Tests:

Normality Test (Shapiro-Wilk) Passed (P = 0.0923)

W Statistic= 0.9289 Significance Level = 0.0500