Waikato Regional Council Technical Report 2018/62

Modelling E. coli in the Waikato and Waipa River Catchments

Development of a catchment-scale microbial model

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, Sandy Elliott, Sharleen Yalden
(NIWA

This report was commissioned by the Technical Leaders Group for the Healthy Rivers Wai Ora Project
Report No. HR/TLG/2015-2016/2.6

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

September 2015

Document #: 3428411

Peer Reviewed by:
Bryce Cooper
(NIWA) Date December 2015

Approved for release by:
Mike Scarsbrook Date December 2018

NIWA
Taihoro Nukurangi

Modelling E. coli in the Waikato and Waipa River Catchments

Development of a catchment-scale microbial model

Prepared for the Technical Leaders Group of the Healthy Rivers/Wai Ora Project

September 2015

Authors/Contributors:
Annette Semadeni-Davies
Sandy Elliott
Sharleen Yalden

For any information regarding this report please contact:

Sandy Elliott
Programme Leader
Freshwater Quality
+64-7-959 1839
Sandy.Elliott@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
PO Box 11115
Hamilton 3251

Phone +64 7 856 7026

NIWA Client Report No:	AKL2015-017
Report date:	September 2015
NIWA Project:	EVW1517

Contents

Tables

Figures

Reviewed by

A. Hughes

Approved for release by

D. Roper

Formatting checked by

Executive summary

The study reported here is one of a suite of technical studies commissioned through the Healthy Rivers / Wai Ora Technical Leaders Group (TLG). The report presents the development of three models developed to provide input data to a farm cost model to evaluate the costs associated with mitigating farm practices to reduce *E. coli* loads.

This report presents three steady-state catchment models to estimate *E. coli* loads and concentrations in the Waikato and Waipa River catchments from Lake Taupo to Port Waikato. The catchment was split into 74 subcatchments, 62 of which have *E,coli* monitoring data The models were developed to inform a farm cost model which will be used to assess the economic impact of mitigations to improve water quality in the catchment. It is planned that the models will allow calculation of future concentration from loads modelled under different land use and farm practice scenarios. The models are:

1. **The detailed load model (DLM).** This model operates at the REC reach unit scale and was calibrated against measured mean annual *E. coli* loads from water quality monitoring stations in the study area. The DLM was developed specifically to provide parameters to the coarse load model.

2. **The coarse load model (CLM):** This model operates at the subcatchment scale and has a different representation of spatial input data to the DLM. The model parameters are the same as those calibrated for the DLM. The CLM was developed to be compatible with the scale of the farm cost model.

3. **The concentration model:** Two regression models were developed to estimate site median annual and 95th percentile *E. coli* concentrations, respectively, for monitoring sites where *E. coli* is not currently sampled. The predictands for both models are upstream land use and drainage characteristics.

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, and sources of model uncertainty. Where concurrent flow records were available, the rating curve approach was the preferred method for estimating loads. For sites with no flow monitoring, *E. coli* concentrations were estimated as a function of the estimated mean annual flow for the reach within which the monitoring site is located.

The loads models estimate mean annual *E. coli* loads from diffuse sources on the basis of the REC unit land use, rainfall and soil drainage. Point sources are also added to the in-stream load. The loads are then subject to attenuation in reservoirs. While instream attenuation was included in the models, the time of travel was calibrated as zero leading to no instream attenuation being modelled.

The results of the load models are comparable. The RMSE between the log-transformed measured and modelled loads for the DLM is 0.55 and the adjusted R^2 is 0.92. The RMSE between the log-transformed measured and modelled loads for the CLM is 0.60 and the adjusted R^2 is 0.89. Differences between the two models are likely to be due to scaling issues and the different methods of representing spatial data in the two models.

The concentration regression model was better able to estimate median annual *E. coli* concentrations than 95th percentile concentrations. The R^2 for the former is 0.64 and the standard error in log-space is 0.53. This model performance is comparable to that in national-scale modelling studies (Unwin et al., 2010). The R^2 for the 95th percentiles is 0.53 and the standard error is 0.72.

Estimates of model error have been provided in the report for loads estimated using the rating curve method and the calibrated parameters for the DLM and the concentration regression models. There is substantial uncertainty in the parameters calibrated for the DLM which are transferred to the CLM. 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 estimated annual mean 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.

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

1 Introduction

1.1 Preamble

Waikato and Waipa River iwi and the Waikato Regional Council (WRC) are partners in a project "Healthy Rivers: Plan for Change / Wai Ora: He Rautaki Whakapaipai" (HR/WO). The plan change will seek to achieve reduction, over time, of sediment, bacteria and nutrients (nitrogen and phosphorus) entering water bodies in the Waikato and Waipa River catchments.

The *Escherichia coli (E. coli*) load and concentration modelling presented here is one component of a suite of technical studies that have been commissioned through the Healthy Rivers / Wai Ora Technical Leaders Group (TLG). These assessments include information on the current state of the streams and rivers, sources of contaminants, catchment modelling to determine how contaminants accumulate and move through the catchment, and economic catchment modelling to determine the cost of meeting water quality goals and targets.

1.2 *E. coli*

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 National Policy Statement for Freshwater Management (NPS-FM, Ministry for the Environment, 2014). Under the NOF, 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, overland flow from excess irrigation water and drainage via artificial drains (Collins et al., 2007; Muirhead, 2015). While this report focuses on *E. coli* modelling, methods for mitigating *E. coli* loads entering the stream network are overviewed in Appendix A at the request of WRC.

1.3 Scope

This report describes the development of three related spreadsheet models for predicting the median annual stream concentration and mean annual loading of *E. coli* in the Waikato River catchment. The modelled catchment area extends from Taupo Gates to Port Waikato inclusive of the Waipa River catchment. The catchment area was divided into 74 subcatchments; the boundaries of which were delineated (in consultation with WRC) on the basis of water quality monitoring sites. There are two catchment load models referred to as the *detailed load* and *coarse load models* (DLM and CLM) and a concentration model.

The load models build on previous work carried out for the Waikato River Economic Joint Venture study (EJV; Doole, 2013; Elliott et al., 2013; Journeaux, 2013; Semadeni-Davies and Elliot, 2013). The DLM was developed in order to provide input parameters for the CLM. The CLM will in turn provide input data to the Farm Cost Model (FCM) being developed for the HR/WO which will be used to investigate the economic impact of implementing mitigation actions in the study area. While operating at different spatial scales, both models estimate *E. coli* loads as a function of land use and key environmental drivers such as mean annual rainfall and soil drainage class. The DLM operates at the subcatchment scale (>10 km^2) with the smallest spatial unit being the contributing area of river reaches derived from the River Environments Classification (REC, Snelder et al., 2010), called REC units throughout this report. The detailed load model has the same spatial scale as the *E. coli* modelling undertaken using the Catchment Land Use for Environmental Sustainability (CLUES) model by Semadeni-Davies and Elliot (2013) and Elliott et al. (2013). The CLM operates at the subcatchment scale. The performance of the two models is compared to determine whether the CLM has sufficient spatial resolution to represent *E. coli* loadings in the river system.

The concentration model is a regression type model and has been developed to predict annual median and 95th percentile concentrations for water quality monitoring sites which currently do not have any *E. coli* sampling and to provide input data for future load modelling with a range of land use and mitigation scenarios.

Finally, the report discusses sources of model uncertainty and gives recommendations for further work.

2 Methodology

This section presents an overview of model input and calibration data followed by descriptions of the load models and the concentration model.

2.1 Input data

This section overviews the spatial input data that characterise the study area used in the *E. coli* models.

2.1.1 Drainage network and monitoring stations

The drainage network in the study area consists of approximately 22,200 REC river reaches. A river reach is defined as a section of river between upstream and downstream confluences and is typically between 500-1500m in length with a contributing catchment area, called an REC unit in this report, of around 40 ha.

The area was divided into 74 subcatchments for modelling purposes by aggregating River Environments Classification (Snelder et al., 2010) drainage units between selected sites located along the drainage network (Table 2-1 and Figure 2-1). Each subcatchment represents the contributing area draining to its corresponding site. *E. coli* concentrations are monitored at 63 of the monitoring sites as part of monthly State of the Environment (SoE) monitoring. There are, however, a number of sites where river *E. coli* is not currently sampled, for example Waikato @ Port Waikato and Waipa @ Waingaro Rd Bridge represent the Waikato River mouth and the confluence of the Waipa and Waikato Rivers @ Ngaruawahia, respectively. Concurrent flow data, required to calculate loads, are available at or near 20 of the sites. Estimated annual mean flows have therefore been taken for the other sites from Woods, Hendrikx, et al. (2006).

Table 2-1: List of modelled subcatchments and water quality monitoring sites in order of flow from upstream to downstream. Sites for which concurrent flow data are available for load calculation are shaded. Sites with no available *E. coli* sampling are marked with an asterisk.

Map ID		Subcatchment	Area (ha)	Monitoring site	
				Location code	NZ reach
Upper Waikato River Catchment	1	Pueto	20029	EW-0802-001	3042044
	2	Waikato @ Ohaaki	29009	EW-1131-105	3039804
	3	Waikato @ Ohakuri	53139	EW-1131-107	3035123
	4	Torepatutahi	21721	EW-1057-006	3038300
	5	Mangakara	2235	EW-0380-002	3037027
	6	Waiotapu @ Homestead	20478	EW-1186-004	3037105
	7	Kawaunui	2134	EW-0240-005	3034452
	8	Waiotapu @ Campbell	6079	EW-1186-002	3034280
	9	Otamakokore	4573	EW-0683-004	3031549
	10	Whirinaki	1080	EW-1323-001	3031392
	11	Waikato @ Whakamaru	44665	EW-1131-147	3035301
	12	Waipapa	10049	EW-1202-007	3035556
	13	Tahunaatara	20816	EW-0934-001	3032435
	14	Mangaharakeke	5415	EW-0359-001	3032678
	15	Waikato @ Waipapa	69392	EW-1131-143	3030247
	16	Mangakino	22186	EW-0388-001	3036710
	17	Mangamingi	5175	EW-0407-001	3027230
	18	Whakauru	5302	EW-1287-007	3027821
	19	Pokaiwhenua	32701	EW-0786-002	3023849
	20	Little Waipa	10649	EW-0335-001	3023862
	21	Waikato @ Karapiro*	53969	Karapiro Dam	3020656
	22	Karapiro	6741	EW-0230-005	3020352
	23	Waikato @ Narrows	12987	EW-1131-101	3018977
Lower Waikato River Catchment	24	Mangawhero	5347	EW-0488-001	3020102
	25	Waikato @ Bridge St Br (Hamilton Traffic Br)	5072	NAT-HM03	3017901
	26	Mangaonua**	8096	EW-0421-010	3017726
	27	Mangakotukutuku	2708	EW-0398-001	3018237
	28	Mangaone	6760	EW-0417-007	3018213
	29	Waikato @ Horotiu Br	5405	EW-1131-069	3015830
	30	Waitawhiriwhiri	2223	EW-1236-002	3017487
	31	Kirikiriroa	1233	EW-0253-004	3016924
	32	Waikato @ Huntly-Tainui Br	17322	EW-1131-077	3013160
	33	Komakorau	16399	EW-0258-004	3014466
	34	Mangawara	35884	EW-0481-007	3013137
	35	Waikato @ Rangiriri	6853	NAT-HM04	3010604

Map ID		Subcatchment	Area (ha)	Monitoring site	
	36	Awaroa (Rotowaro) @ Harris/ Te Ohaki Br*	4730	EW-1097_1	3012631
	37	Awaroa (Rotowaro) @ Sansons Br	4561	EW-0039-011	3013581
	38	Waikato @ Mercer Br	45168	EW-1131-091	3006806
	39	Whangape	31767	EW-1302-001	3010847
	40	Whangamarino @ Island Block Rd	14365	EW-1293-007	3007681
	41	Whangamarino @ Jefferies Rd Br	9701	EW-1293-009	3008369
	42	Waerenga	1959	EW-1098-001	3009556
	43	Matahuru	10637	EW-0516-005	3010952
	44	Waikare*	10418	EW-326_10	3010071
	45	Opuatia	7067	EW-0665-005	3008985
	46	Mangatangi	19452	EW-0453-006	3006132
	47	Waikato @ Tuakau Br	15178	EW-1131-133	3007421
	48	Ohaeroa	2033	EW-0612-009	3007733
	49	Mangatawhiri	6808	EW-0459-006	3005110
	50	Waikato @ Port Waikato	28148	Terminal Reach	3009006
	51	Whakapipi	4648	EW-1282-008	3006346
	52	Awaroa (Waiuku)	2506	EW-0041-009	3007434
Waipa River Catchment	100	Waipa @ Mangaokewa Rd	3221	EW-1191-005	3036214
	101	Waipa @ Otewa	28665	NAT-HM01	3029370
	102	Mangaokewa	17419	EW-0414-012	3031564
	103	Mangarapa*	5443	444_4	3028468
	104	Mangapu	16170	EW-0443-003	3027166
	105	Mangarama*	5528	EW-1391_1	3031371
	106	Waipa @ Otorohanga	13889	EW-1191-012	3027129
	107	Waipa @ Pirongia-Ngutunui Rd Br	43607	EW-1191-010	3022669
	108	Waitomo @ Tumutumu Rd	4318	EW-1253-007	3028966
	109	Waitomo @ SH31 Otorohanga	4393	EW-1253-005	3026779
	110	Moakurarua*	20630	EW-553_5	3023962
	111	Puniu @ Bartons Corner Rd Br	22785	EW-0818-002	3023180
	112	Puniu @ Wharepapa*	16853	EW-818_40	3025988
	113	Mangatutu	12269	EW-0476-007	3024473
	114	Mangapiko	28069	EW-0438-003	3022010
	115	Mangaohoi	431	EW-0411-009	3023476
	116	Waipa @ SH23 Br Whatawhata	31506	NAT-HM02	3017829
	117	Mangauika	978	EW-0477-010	3023179
	118	Kaniwhaniwha	10259	EW-0222-016	3019566
Waipa	119	Waipa @ Waingaro Rd Br*	15484	Waipa Waikato confluence	3015066

Map ID		Subcatchment	Area (ha)	Monitoring site	
	120	Ohote	4041	EW-0624-005	3017348
	121	Firewood*	3372	EW-0124_8	3015451

**E. coli* not currently monitored
**Flow for this site is taken from the Mangaonua @ Dreadnaught flow monitoring station 7 km downstream, for this reason, the flow was corrected by multiplying by the ratio of the Woods, Hendrikx, et al (2006) estimated flows for the water quality and flow monitoring sites.

Additionally, data from two neighbouring monitoring stations (*E. coli* concentrations from Waikato @ Taupo Gates, EW-1131-127; and flow from Waikato @ Reids Farm NAT-RO06) near the inflow to the study area within the Ohaaki subcatchment are used to provide input flow and *E. coli* load data from Lake Taupo.

Reservoir reaches are identified as such in the REC. Both the load models include reservoir attenuation or decay terms. In the DLM, the decay factor is calibrated as a separate parameter for larger reservoirs (Table 2-2), which are generally hydro-lakes subject to flow regulation or lower Waikato shallow lakes, and other, smaller lakes. In the CLM, only attenuation in the larger lakes is taken into account. :

Table 2-2: Large reservoirs in the study area.

Type	Subcatchment	Reservoir
Hydro lake	Waikato @ Ohaaki	Lake Aratiatia
	Waikato @ Ohakuri	Lake Ohakuri
	Waikato @ Whakamaru	Lake Whakamaru Lake Atiamuri
	Waikato @ Waipapa	Lake Maraetai Lake Waipapa
	Waikato @ Karapiro	Lake Karapiro Lake Arapuni
Shallow lake	Awaroa (Rotowaro) @ Harris/Te Ohaki Br	Lake Waahi
	Whangape	Lake Whangape
	Waikare	Lake Waikare

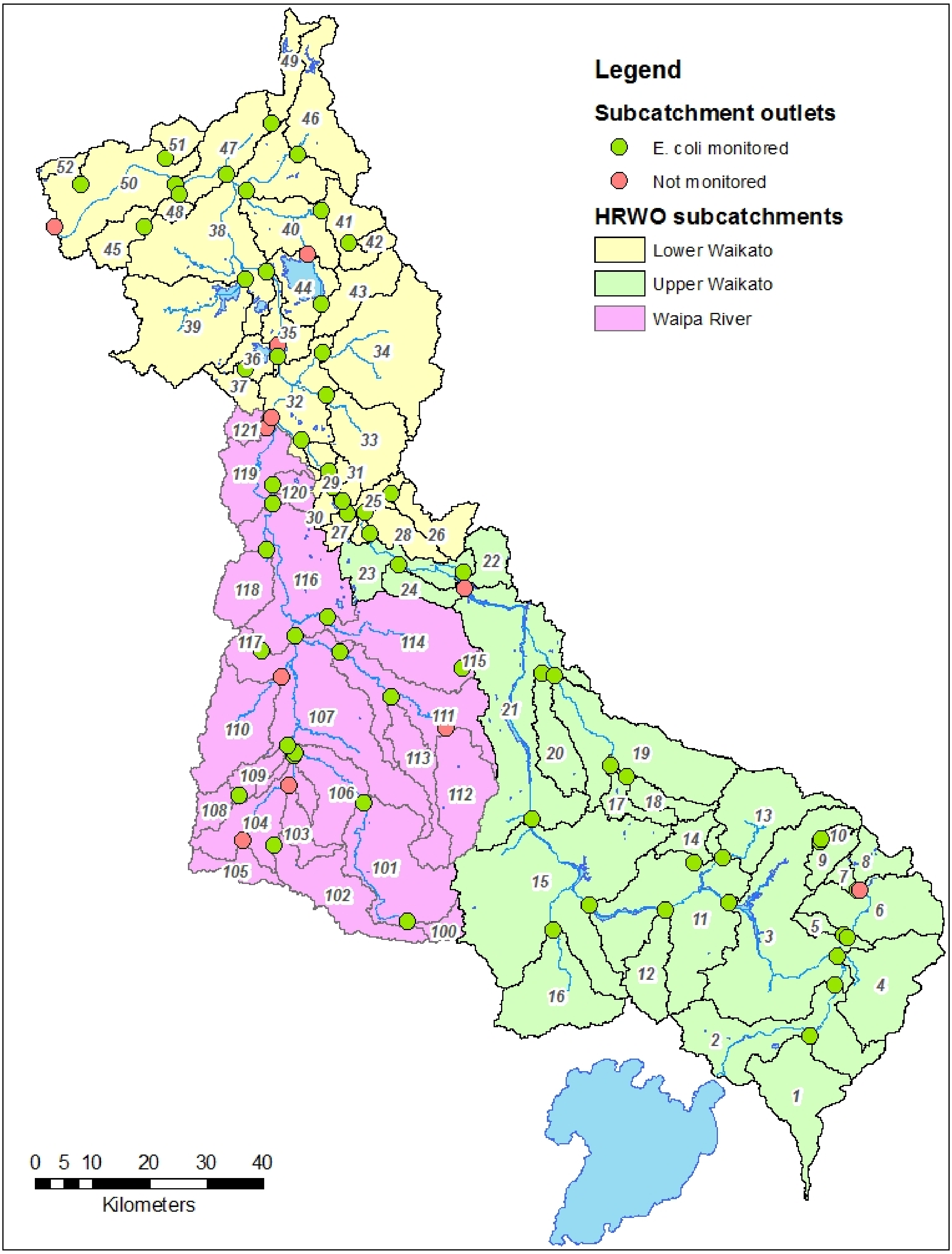


Figure 2-1: Subcatchments and their associated water quality monitoring sites. Catchment names are listed according to the map reference number in Table 2-3

2.1.2 Catchment characteristics

Soil drainage class and mean annual rainfall are input parameters for all the models. Soil drainage class is taken from the Land Resources Inventory Fundamental Soils Layer (FSL, Wilde et al., 2004; Newsome et al., 2008). The drainage class assigns scores from 1 (very poorly drained) to 5 (well drained) to soils. Drainage class used in the DLM is the areal weighted mean score for the reach unit and is taken from the CLUES model geospatial database(Woods, Elliott, et al., 2006). The aggregated drainage class is mapped in Figure 2-2.

REC reach aggregated mean annual rainfall used for the DLM has been also been taken from the CLUES model geospatial database.

Drainage class and mean annual rainfall are aggregated together with land use by subcatchment for use in the CLM. That is, each subcatchment is characterised by the percentage cover for unique combinations of drainage, rainfall and land use to account for the variable effects of drainage and rainfall on *E. coli* losses from different land uses. The spatial data was aggregated as part of the physical characterisation of the Waikato catchment also being carried out under the Healthy Rivers programme.

2.1.3 Land use

Regional land use data with the base-line year 2012 was supplied for this project by WRC as a polygon shape file with the same land use classes as those used in the CLUES model. These land uses were reclassified into 11 broader land use categories to be compatible with the FCM (Table 2-3, Figure 2-3). Dairy support, which is modelled in the FCM, was not assigned its own class for *E. coli* modelling.

The dominant land uses in the study area are dairy, sheep and beef, native forest and forestry. Most sheep and beef is classed as either intensive (lowland) followed by hill country. High country sheep and beef makes up less than 1% of the study area and is found predominantly in the hills east of Tokoroa and Taupo. Other animal farming also make up less than 1% of the area. Dairy dominates central Waikato in the Waipa and Lower Waikato catchments. Native and plantation forest is mainly located in the south-eastern Upper Waikato catchment area. Urban areas account for 3%, with Hamilton being the largest centre. All other land uses only account for 6% and are, for this reason, amalgamated in Figure 2-3.

The land use layer supplied by WRC was overlaid by the REC river reach drainage units layer to enable the proportional area of each land use within each unit to be determined. The land use areas were similarly aggregated to give the proportional area by subcatchment as summarised in Figure 2-4.

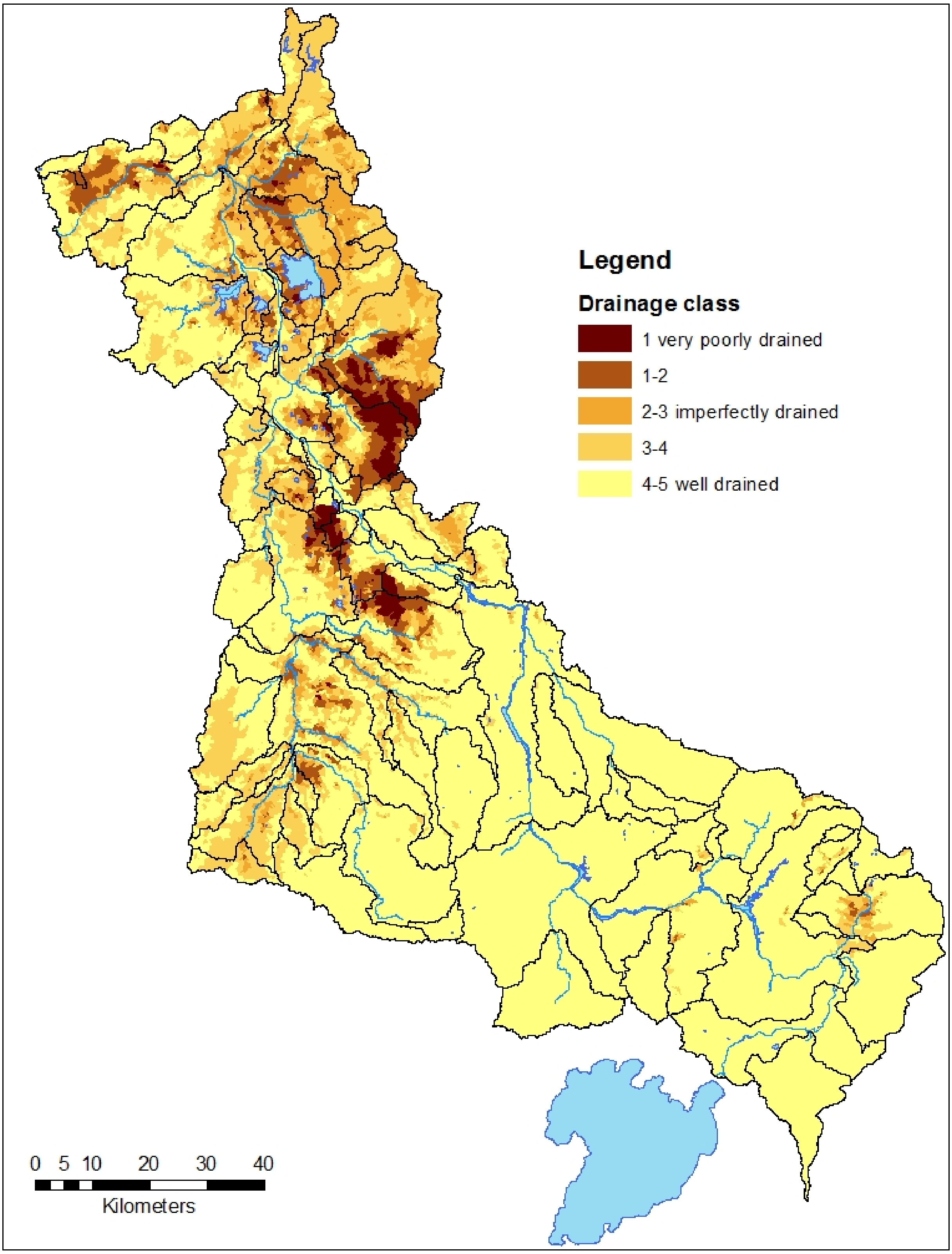


Figure 2-2: Area weighted mean average drainage class by REC unit. Derived from the LRI Fundamental Soil Layer.

Table 2-3: FCM land classes and corresponding CLUES classes; the coverage of each land use class is indicated for the entire study area.

CLUES land use class	Land use class	Percentage cover	Area (km^2)
DAIRY	Dairy	28	3053
SBINTEN	Intensive (lowland) sheep and beef	22	2426
SBHILL SBHIGH	Hill and high country sheep and beef	11	1249
DEER and OTHER_ANIM	Other stock	<1	17
PLANT_FOR	Forestry	15	1695
NAT_FOR and SCRUB	Native forest and scrub	16	1727
MAIZE	Maize	1	56
POTATOES and ONIONS	Horticulture	1	62
URBAN	Urban	3	350
APPLES, GRAPES, KIWIFRUIT, TUSSOCK*, UNGR_PASTURE, OTHER	Miscellaneous	4	388
Total area			11022

*not present in study area

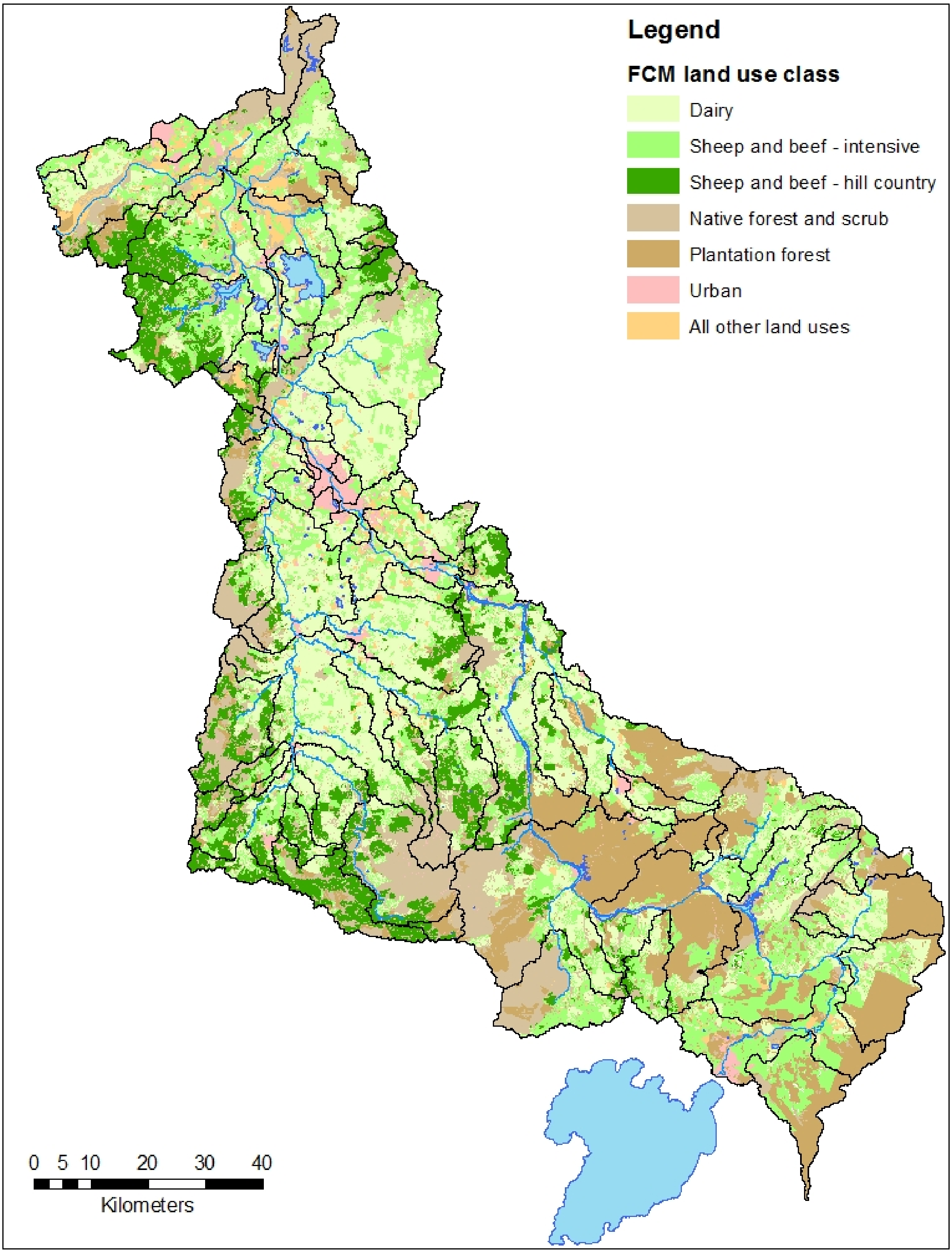
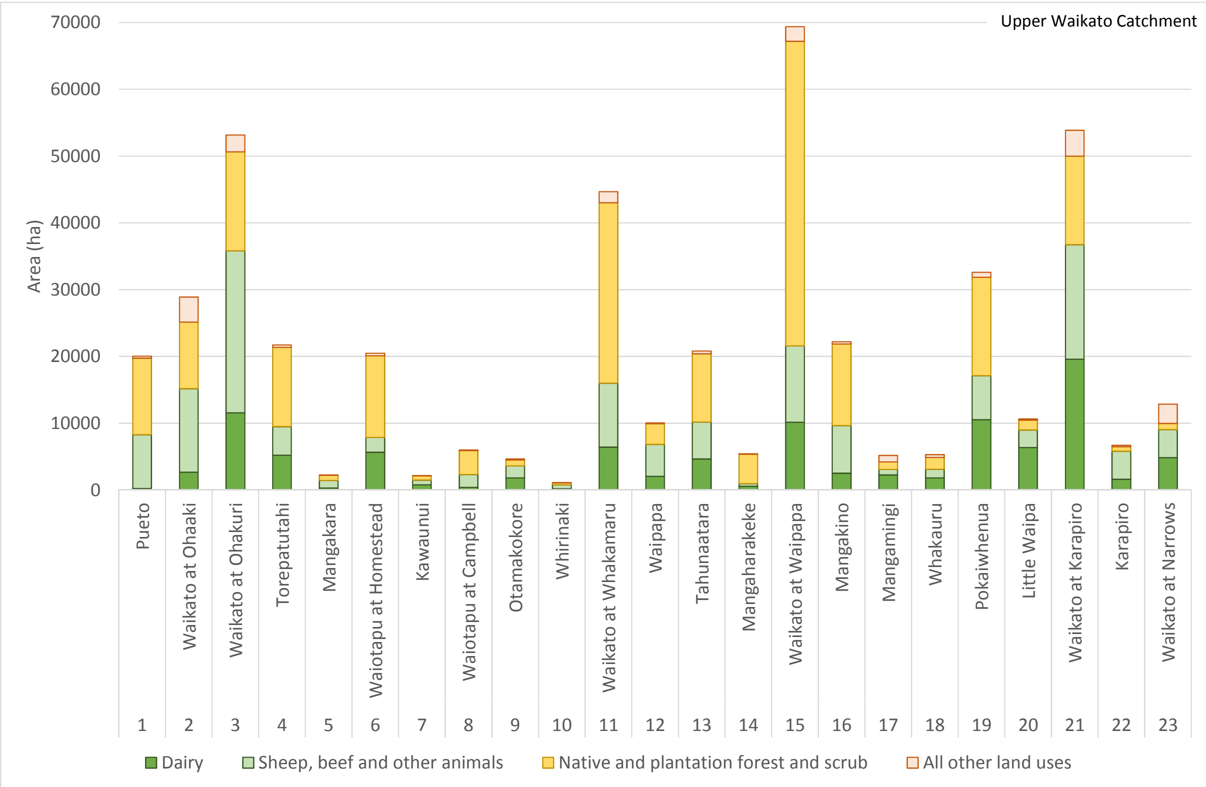


Figure 2-3: Land use from WRC classified into FCM classes. Maize, horticulture and miscellaneous land use classes have been amalgamated for display. Hill country sheep and beef also included a small area of high country sheep and beef.



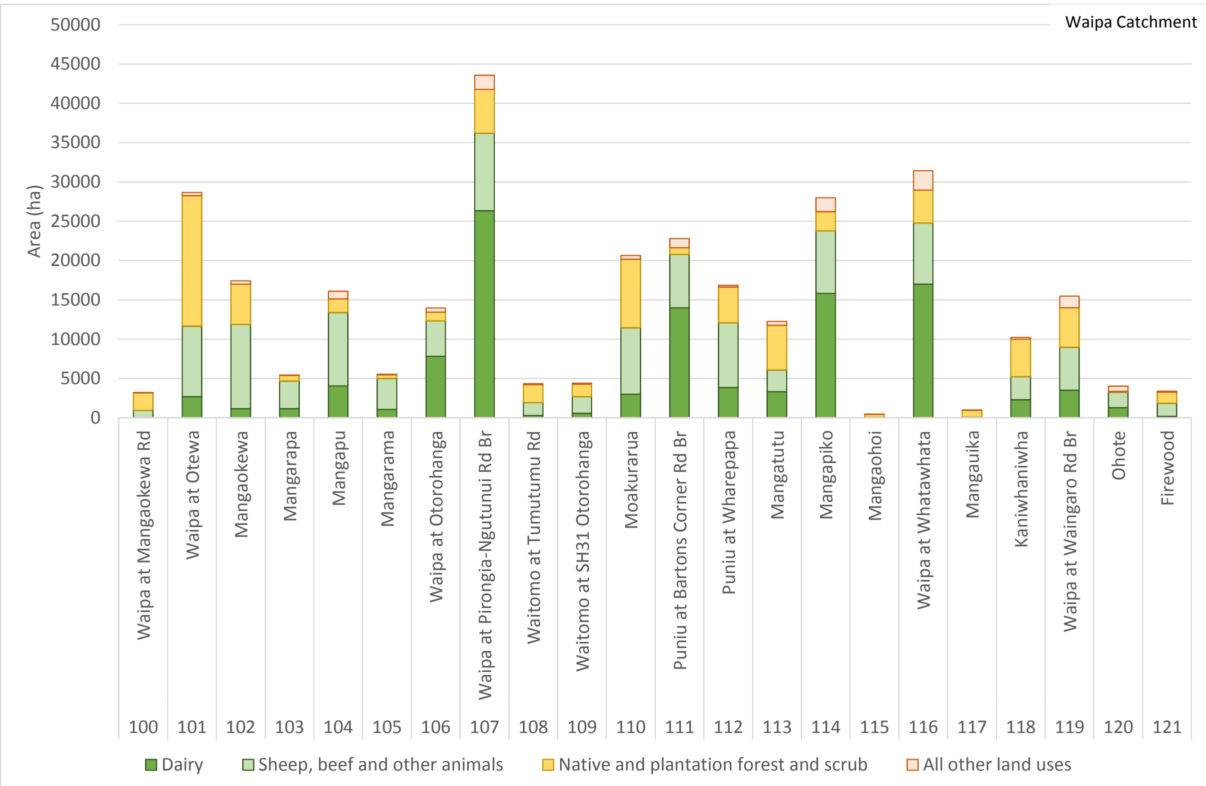


Figure 2-4: Land use by subcatchment. Dry stock, forest and all other land uses classes have been amalgamated for display. Top, Upper Waikato River catchment, Waipa River catchment; over page, Lower Waikato River catchment.

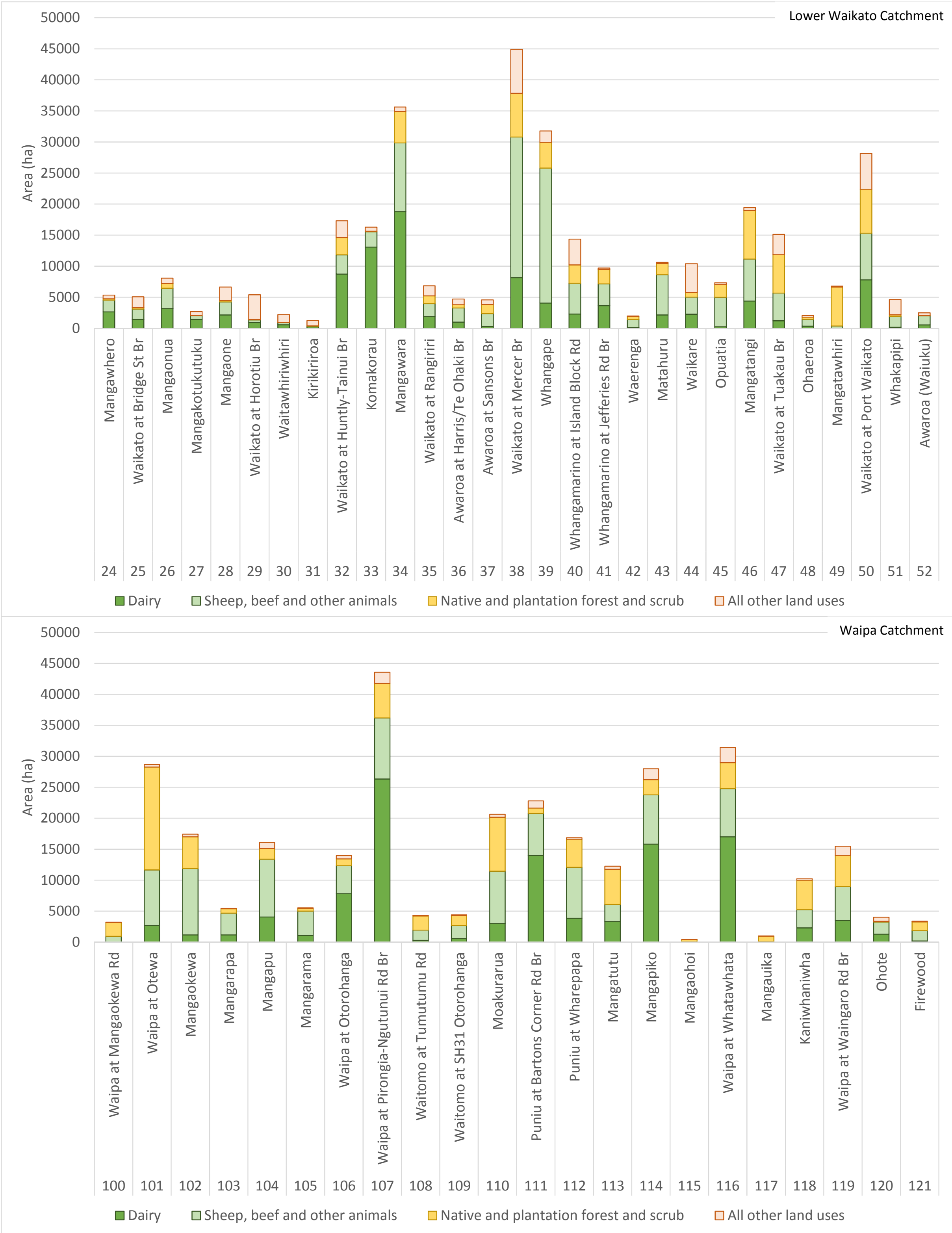


Figure 2-4 continued.

2.1.4 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) were added as model inputs for the REC units or subcatchments within which the sources are located. The loads associated with each source are given in Appendix B.

E. coli loads from point sources were obtained from the WRC (contact Bill Vant) and estimates calculated for the CLUES model (Woods, Hendrikx, et al., 2006). The WRC data were derived from consent monitoring between 2008-2013. Note:

- While the dairy factory @ Hautapu is located in the Mangaone subcatchment and most of the effluent is discharged via spray irrigation, some effluent from the factory is discharged to the Waikato River in the Waikato @ Narrows sub catchment. *E. coli* are monitored at the factory due to historic high loads in the early 2000's due to microbial growth in the pipeline. Recent *E. coli* loads from this source are minor compared to the other point sources.
- Since no *E. coli* monitoring data were available for the waste water treatment plant at Tokoroa, the annual load was estimated on the basis of population serviced by the plant and the nature of the treatment system (Hickey et al., 1989; Woods, Elliott, et al., 2006).
- The wastewater treatment plant at Te Kuiti had a major upgrade in 2013 (Bill Vant, personal comment, June 8 2015). The average discharge concentration before the upgrade was around 12400 cfu/100 ml. The post-upgrade average concentration is two orders of magnitude less at 230 cfu/100 ml. For this study, the average concentration for 2012-2013 (2600 cfu/100 ml) was used to calculate the average load from the treatment plant over the study period.
- The wood pulp mill at Kinleith has no associated *E. coli* monitoring data.
- *E. coli* data from the Cambridge waste water treatment plant was not included in the model as this data refers to inflow to rapid infiltration beds rather than the river.

The *E. coli* input loads from FDE 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 for each of the 97 sheds was provided by WRC. Most of the sheds are located in the Waipa and Lower Waikato river catchments. The annual load for each shed was calculated in the following manner:

i. The number of dairy cattle serviced by each shed 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). They found that the daily *E. coli* load (per cow) in pond effluent is in the order of 10^7-10^8 between September and December and in the order of 10^7 later in the milking season.

Like the loads estimated for the point sources, the estimated loads from dairy sheds were aggregated by REC reach and by subcatchment for input to the models.

2.2 Calibration data

This section presents the methodology used for determining the annual median *E. coli* concentrations and mean annual loads from monthly water quality data from the monitoring stations listed above that are used for model calibration. The data were obtained from the National Rivers Water Quality Network (NRWQN) database and WRC. Median concentrations were determined for each of the stations. Loads and concentrations were also calculated for the monitoring site at Taupo Gates in order to provide input data to the models. The calibration data are listed in Appendix C and Appendix D for concentrations and loads respectively.

2.2.1 Annual median concentrations

An investigation of water quality trends in 10 water quality monitoring sites between 1993 and 2012 (Vant, 2013) found few trends in *E. coli* concentrations in Waikato. The exceptions were a reduction in concentrations at Huntly and an increase at Whakamaru. Investigation of concentrations for each of the water quality monitoring sites for this project found trends in several other sites. Accordingly, it had been intended to calculate long-term (20-year) detrended median concentrations (similar to those used for nutrient modelling in Elliott et al., 2013) for sites with sufficiently long *E. coli* data records. However, there were concerns that the detrending method may be introducing error into the calculation. To avoid the effects of both trending data and detrending smoothing errors, five-year median concentrations, calculated for the period from January 2010 to December 2014, were used for this study. It should also be noted that a number of sites have limited data records available with *E. coli* monitoring commencing in early 2013.

For water quality monitoring sites with no monitored *E. coli* data, the concentration model described in Section 2.4 has been developed to provide estimated annual median and 95 percentile concentrations.

2.2.2 Mean annual loads

Two calculation methods; a rating curve method and a ratio method, were used to determine the measured mean annual *E. coli* loads. Where possible, the rating curve method was used in preference to determining loads from water quality monitoring sites with concurrent flow data from the nearest flow monitoring station. This method determines mean loads from loads calculated for each monthly water quality sampling event from the recorded *E. coli* and flow data. The ratio method was used to estimate loads for sites without suitable flow data from estimated mean annual flow rates (Woods, Hendrikx, et al., 2006) and the median annual concentration described above. These methods are described in more detail below.

Rating curve method

Measured mean annual *E. coli* loads were estimated using a rating-curve method at sites where there were sufficient concurrent flow data at the site or nearby. In this method, a rating curve is fitted to the natural log of measured monthly *E. coli* concentrations against the natural log of the flow rate using the following equation:

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

Where C is the *E. coli* concentration, s is a cubic spline smoothing function, Q is the flow at the time of the sample, 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. Equation (1) was applied to the hourly flow time-series over the period of the flow record to derive a time-series of concentrations, which was then multiplied by flow 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 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. This approach repeatedly took random samples of the original water quality data and estimated the mean annual load for each of these. On the basis of this assessment, the rating curve-derived loads for Pokaiwhenua and Waikato @ Whakamaru were flagged as unsuitable for calibration.

Ratio method

Measured *E. coli* loads for the sites without concurrent flow data, and from Pokaiwhenua and Waikato @ Whakamaru were estimated using the ratio method from the median annual concentration and estimated mean annual flow for the site taken from the model of Woods, Hendrikx, et al. (2006).

In this method, the median concentration is multiplied by a factor to convert to flow-weighted concentration, and this was then multiplied by the estimated mean annual flow to derive the load estimate. The conversion factor was determined from the results of the rating curve method (excluding Pokaiwhenua and Waikato @ Whakamaru) and is the average of the flow-weighted mean annual concentrations (i.e., the estimated mean annual load divided by the observed mean annual discharge) divided by the median concentrations for the sites. Two factors were determined, one for tributary monitoring sites and one for Waikato @Whakamaru which is a main-stem site. For the tributaries sites without flow data, the ratio was determined from the rating curve data for tributary sites with flow data. For Waikato at Whakamuru, the ratio was determined from the average from the two adjacent main-stem sites (Waikato @ Ohakuri Tailrace Br and Waikato @ Waipapa Tailrace).

Waiotapu @ Campbell is not included in the calibration since the load (0.0018 peta/year) was considered too low compared to other sites.

2.3 *E. coli* load models

2.3.1 Detailed load model

The DLM calculates input loads from each REC unit and routes these loads down the drainage network (described in Section 2.1.1) by adding loads from each reach entering the instream load and then accounting for instream attenuation or decay. The spatial data presented above are aggregated by REC unit.

Loads entering the drainage network from each reach unit are calculated as the sum of loads from point sources (see Section 2.1.4) draining to the reach and *E. coli* loads 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 2.1.2) multiplied by a corresponding source yield and then adjusted for surface losses. Five source yields were 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 surface decay between the summed diffuse source load and the drainage network is calculated using a first-order decay term as follows:

$$L_{att} = L_{\mathrm{int}} e^{(R_{reach.} k_{rain}) k_{drrain}} \tag{2}$$

Where, for each REC reach unit, L_{att} is the attenuated load from the diffuse sources entering the drainage network; L_{int} is the initial load summed for the diffuse sources; R is the mean annual rainfall anomaly; and k_{rain} and k_{drain} are the catchment wide rainfall delivery and drainage coefficients. The rainfall coefficient is calibrated. The drainage coefficient is set to 1 for imperfectly to well drained soils (drainage classes 3-5) and is assigned a calibrated value for poorly drained soil (drainage classes 1 and 2).

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

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

Where $R_{mean}$is the mean annual rainfall for the reach 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 geospatial database.

Once in the drainage network, the *E. coli* load is propagated downstream taking into account losses within the network by multiplying instream load by decay factors which relate to the proportion of the load remaining after attenuation. Separate losses are calculated for streams and reservoirs. Instream 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 instream 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 less the length of the reach in lakes divided by the flow rate.

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

$$Att_{res} = \frac{O_{res}}{(O_{res} + k_{res})} \tag{5}$$

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 (m/year) for the outlet reach taken from the SPARROW component of the CLUES model (Elliott et al., 2005); and k_{res} is the calibrated reservoir coefficient. Separate reservoir coefficients have been calibrated for the larger lakes listed in Table 2-2 and other lakes.

2.3.2 Calibration method

The model parameters defined above were calibrated 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 for all the water quality monitoring sites. The RMSE is used as a standard statistical metric to measure model performance in many fields, including meteorology, air quality, climate research and agriculture and assumes the errors are unbiased and follow a normal distribution (Chai and Draxler, 2014). The model calibration was constrained such that:

- Source yields ≥ 0.0001 peta (10^{15}) organisms/km^2/year
- k_{drain} ≥ 0.00001
- k_{time} and k_{res} ≥ 0

2.3.3 Coarse load model

The CLM calculates *E. coli* losses for each subcatchment from each unique combination of the drainage, rainfall and land use class listed in Table 2-4. The combinations were derived as part of a refined characterisation of the Waikato catchment for WRC under the Healthy Rivers programme (Wadhwa, 2015, under preparation). In all, there are 59 combinations present in the data set. The model has not been calibrated, instead the parameters calibrated for the DLM are used to inform the model.

In the CLM, *E. coli* losses for the different land use, rainfall and drainage class combinations are first attenuated for surface losses and then summed together with point source loads to give a subcatchment input load which is then added to the instream *E. coli* load and routed down the drainage network.

Table 2-4: Subcatchment aggregation classes. Each subcatchment is represented by the percentage area covered by each unique combination of land use, drainage and rainfall class.

Landuse class	Drainage class	Rainfall class (mm/y)
Dairy Intensive (low land) sheep and beef Hill and high country sheep and beef Other stock Forestry Native forest and scrub Horticulture Crop Miscellaneous	1-3 poor / moderate drainage 4-5 well drained	<1400 low rainfall 1400-1800 medium rainfall >1800 high rainfall

Urban		

For each land use, drainage and rainfall class combination in a specific subcatchment, the surface attenuated annual *E. coli* load reaching the drainage network from that subcatchment, $L_{att,com}$is calculated as:

$$L_{att,com} = L_{com}^{(R_{com}.k_{rain})k_{drrain}} \quad (6)$$

Where L_{com} is the initial load from the combination which is calculated as the product of the area within the subcatchment covered by the combination and the associated source yield calibrated in the DLM for the land use; R_{com} is the rainfall anomaly for the combination within the subcatchment; and k_{rain} and k_{drain} are the rainfall delivery and drainage coefficients determined for the DLM.

The rainfall anomaly for each combination within a subcatchment, R_{com} is calculated as:

$$R_{com} = R_{mean,com} - R_{catchment} \quad (7)$$

Where $R_{mean,com}$ is the area weighted mean annual rainfall for the section of the subcatchment covered by the combination; and $R_{catchment}$ is as defined above for the DLM. Note that while the mean annual rainfall has been grouped into rainfall classes for the aggregation, the mean annual rainfall (1981-2010) for each aggregated section of the subcatchment is used to derive $R_{mean,com}$.

Since the calibrated time of travel coefficient, k_{ime}, for the DLM is zero, instream losses are not calculated in the CLM. Reservoir attenuation is only applied to the subcatchments listed in Table 2-2. For subcatchments containing a single large reservoir, the entire subcatchment is assigned the same attenuation factor, Att_{res}, calculated for the reservoir outlet reach using Equation (5). For subcatchments with two reservoirs, the attenuation factor is the product of the factors calculated for the outlets of both reservoirs.

2.4 Concentration model

The concentration model was developed to provide estimates of annual median *E. coli* concentrations both for sites without current *E. coli* sampling and to allow future scenario modelling with a range of land use and mitigation scenarios.

2.4.1 Concentration estimated at unmonitored sites.

There are 11 sites, including the two virtual sites, in the study area where *E. coli* concentrations are not currently measured but where the concentrations are of interest. Estimates of the current and future (i.e., scenario) annual median and 95th percentile *E. coli* concentrations at these sites are required for the economic optimisation model.

To address this need, regression models were developed using the measured *E. coli* concentrations presented in Section 2.2.1 to predict concentrations at the sites where *E. coli* is sampled as a function of catchment characteristics. These models were then applied to predict the concentration at the sites for which there are no measurements, with the exception of the virtual sites which are discussed further below. Separate models were developed for the median and 95th percentile concentrations.

Waikato River main-stem sites were removed from the regressions because they are influenced by the hydro lakes and Lake Taupo outflow, which were not included in the regressions. The Pueto and Whirinaki were also removed from the calibration of the DLM as they are influential outliers. . Whangamarino @ Jefferies and Whangape were removed as they are also downstream of lakes.

The models used predictors that were also used in the load model and were initially of the form:

$$C = R^b\left(cF_{low} + F_{high}\right)\left(c_D F_D + c_{SBI} F_{SBI} + c_{SBH} F_{SBH} + c_{Urb} F_{Urb} + c_{Tree} F_{Tree} + c_{other} F_{other}\right) \quad (8)$$

Where C is the concentration, R is mean annual rainfall mean over the catchment upstream of the monitoring site; F_{low} and F_{high} are the fractions of the upstream catchment that have either poor to moderate or good to very good drainage respectively; F_D 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; and the other coefficients are calibration constants for drainage and the different land use classes listed above respectively.

The model was fit using non-linear least squares regression, with log-transformation to better condition the residuals. It was found that R^b was not statistically significant for concentration, so this term was removed from Equation (8). The dairy and intensive sheep and beef terms were lumped as these land uses were highly correlated such that the parameter for intensive sheep and beef was not credible.

A different approach was used for the Waipa confluence and Waikato @ Port Waikato virtual monitoring sites. The concentration for these sites was based on the measured concentrations from Waipa @ Whatawhata and Waikato @ Tuakau, (i.e., the closest *E. coli* monitoring sites respectively) with adjustments to take into account the modelled increase in load and flow between sites. That is, the concentration at the unknown site was determined by the concentration at the measured site, multiplied by the ratio of predicted loads, divided by the ratio of measured flows. The loads were taken from the load model and flow estimates described in Section 2.3.

2.4.2 Prediction of concentrations for future scenarios

The general approach for predicting concentrations is based on current (i.e. present-day) measured concentrations with adjustments for the modelled changes in loads. That is, the percentage difference between the loads calculated using a future scenario and the current scenario presented in this report would be applied to the current concentration to give a future concentration estimate. This approach assumes linearity between concentrations and loads, which is reasonable but has yet to be confirmed observationally. The approach is very similar to that undertaken by Semadeni-Davies and Elliott (2012) for WRC using the CLUES model.

3 Results

3.1 *E. coli* load models

3.1.1 Calibration

The DLM was calibrated against the measured mean loads using the method outlined in Section 2.3.2. The calibrated parameters were then applied to the CLM. This section presents the calibrated parameters and the estimated loads of both models.

The DLM calibrated parameters and their standard errors (se) are listed in Table 3-1. Parameter uncertainty was determined using a local linear approximation (Nikitas and Pappa-Louisi, 2000) to determine the se of each parameter. The errors in the parameters give, for example, an indication of whether different sources have different yield parameters, and whether a parameter is highly uncertain and could potentially be removed from the model.

Note that the Pueto and Whirinaki sites were both considered to be outliers and were therefore removed. For both sites, the modelled loads were high compared to the measured loads. The over-prediction could be due to unusual hydrological characteristics in these upper Waikato headwater subcatchments which are both at least partly spring-fed. In the case of Pueto, recent land-use changes may not yet be reflected in the stream concentrations.

No stream attenuation was calculated due to the time decay factor, k_{time}, having a calibrated value of zero. This result is unexpected given microbial die-off in streams. A possible reason for this calibrated value could be stable or naturalised *E. coli* growth in stream bed sediments (Pachepsky and Shelton, 2011; Perchec-Merien and Lewis, 2013). Forcing a positive term significantly reduced the performance of the model, therefore the zero decay term was retained.

Attenuation in hydro-reservoirs and large shallow lakes was calibrated separately from attenuation in other tributary lakes to reflect their differences in residence times. Although the factors have high associated errors, they were retained in the DLM to allows for decay in the drainage network.

The se values for the parameters show that there is substantial uncertainty in the model which reflects the difficulty in determining *E. coli* loads largely due to the high spatial and temporal variability of *E. coli* concentration measurements. While it is expected that the pastoral land uses would have the highest yields, it is surprising that hill and high country sheep and beef has the highest estimated yield. This may reflect the combined effects of lower stock exclusion and greater runoff associated with steep topography. The parameterised yields for dairy is roughly twice that for intensive sheep and beef. The 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 the DLM. Wilcock (2006) suggests annual yields of around 10^{11} organisms/ha/year (0.0001 10^{15} organisms / km^2/year) for hill country sheep and beef, which is an order of magnitude less than those parameterised for dairy and intensive sheep and beef and two orders less than for hill country sheep and beef. However, this rate does not include direct deposition of *E. coli* in streams, largely by dairy and beef cattle. At a stock rate of three cows per hectare, the yield could be in the order of 10^9 organisms /ha/year.

The relatively high yields for urban land use could point to unknown sources such as sewer or pumping station overflows that are not currently accounted for.

The reservoir coefficients are highly uncertain, removing reservoir attenuation from the models has little impact on the overall model fit bit but increased the estimated loads. The coefficients were kept in the models to reduce downstream loads.

Sources of error and uncertainty are discussed more fully in Section 4.

Table 3-1: Calibrated model parameters for the detailed load model with uncertainty estimates.

Parameters optimised	Unit	Parameters	se
Source yield: Dairy	10^{15} org/km^2/year	0.0072	0.0034
Source yield: Intensive sheep and beef	10^{15} org/km^2/year	0.0037	0.0028
Source yield: Hill and high country sheep and beef	10^{15} org/km^2/year	0.0183	0.0055
Source yield: other rural land uses - All other stock - Forestry - Native forest and scrub - Horticulture - Crops (maize) - Miscellaneous	10^{15} org/km^2/year	0.0011	0.0009
Source yield: urban	10^{15} org/km^2/year	0.0110	0.0063
Drainage coefficient k_{drain} (poorly drained soils)	dimensionless	4.3062	1.8247
Rainfall delivery coefficient k_{rain}	dimensionless	1.2605	0.6087
Reach time of travel k_{time}	/day	0.0000	0.2567
Reservoir attenuation coefficient k_{res}, Large lakes (see Table 2-2)	/year	3737.5055	1702.3051
Reservoir attenuation coefficient k_{res}, all other lakes	/year	15.2117	20.5341

3.1.2 Detailed load model

The measured and modelled mean annual *E. coli* loads for the water quality monitoring sites with *E. coli* sampling are shown in Table 3-2 along with the log-transformed loads and their residuals used for calibration.

The RMSE between the log-transformed measured and modelled loads is 0.55. The coefficient of determination (R^2) for the log-transformed loads, adjusted for the number of variables, is 0.91. Note that while Pueto and Whirinaki were removed from the calibration these sites are included in the calculation of the coefficient of determination. The relationship between the log-transformed loads is plotted in Figure 3-1.

Other than model over-prediction for Pueto and Whirinaki, key findings are as follows:

- Although the reservoir decay coefficient, k_{res}, is calibrated separately for main-stem hydro-lakes and tributary lakes, there is a considerable drop in measured and modelled loads at Waikato @ Ohakuri compared to the combined upstream loads delivered from Waikato @ Ohaaki, Torepatutahi, Mangakara, Waiotapu @ Homestead and Otamakokore, The drop in load between the two main-stem sites could be due to a long residence time in the Lake Ohakuri hydro-reservoir, which results in an estimated travel time of around five days for this stretch of the river (Brown, 2005).
- While the loads predicted for the predominantly urban tributary subcatchments of Kirikiriroa and Waitawhiriwhiri are reasonable, the calibrated yield for the urban land use class is higher than for the pastoral land uses. This could point to sources other than the sewage plant in Hamilton, such as water fowl or sewer overflows or pumping station overflows. Indeed, the increases seen in main-stem loads from Hamilton to Huntly (Waikato @ Bridge St Br, Waikato @ Horotiu Br and Waikato @ Huntly-Tainui Br) may be due to urban sources not represented in the model. Similarly, Whakaruru has under predicted loads. The higher measured loads from this monitoring station may be influenced by urban runoff from Tokoroa just upstream of the monitoring site.
- Conversely, if urban land cover has been overestimated in Kirikiriroa and Waitawhiriwhiri subcatchments, then the *E. coli* sources in these sub-catchments are more likely to be rural meaning that the calibrated urban yield is too high.
- The results for Mangaonua and Mangaone show the greatest model under-prediction of measured loads. Both subcatchments are immediately upstream of Hamilton city. The high measured loads in these subcatchments may be due to lifestyle blocks and septic tanks, but this is by no means certain and warrants further investigation.

Table 3-2: Measured mean annual *E. coli* loads (peta, 10^{15} organisms / year) against loads modelled using the detailed load model. Log-transformed loads and their residuals used in model calibration are also shown.

Map id	Monitoring site	Mean annual loads		Log-transformed loads		
		Measured	Modelled	Measured	Modelled	Residual
1	Pueto	0.100	0.450	-2.304	-0.798	
2	Waikato @ Ohaaki Br	0.992	1.380	-0.008	0.322	-0.331
3	Waikato @ Ohakuri Tailrace Br	0.300	0.771	-1.203	-0.260	-0.943
4	Torepatutahi	0.256	0.609	-1.361	-0.496	-0.864
5	Mangakara	0.065	0.060	-2.726	-2.807	0.081
6	Waiotapu @ Homestead	0.922	1.081	-0.082	0.078	-0.159
7	Kawaunui	0.126	0.077	-2.070	-2.558	0.488
9	Otamakokore	0.240	0.233	-1.426	-1.458	0.032
10	Whirinaki	0.005	0.066	-5.209	-2.725	
11	Waikato @ Whakamaru Tailrace	1.010	0.651	0.010	-0.429	0.439
12	Waipapa	0.242	0.539	-1.421	-0.618	-0.803
13	Tahunaatara	0.720	0.794	-0.329	-0.231	-0.097
14	Mangaharakeke	0.273	0.137	-1.299	-1.988	0.689
15	Waikato @ Waipapa Tailrace	1.548	1.183	0.437	0.168	0.269
16	Mangakino	0.343	0.781	-1.069	-0.248	-0.821
17	Mangamingi	1.463	0.756	0.381	-0.280	0.661
18	Whakauru	0.680	0.327	-0.385	-1.119	0.734
19	Pokaiwhenua	1.746	2.446	0.558	0.894	-0.337
20	Little Waipa	0.323	0.792	-1.131	-0.234	-0.897
22	Karapiro	0.380	0.664	-0.968	-0.410	-0.558
23	Waikato @ Narrows Br	14.364	10.377	2.665	2.340	0.325
24	Mangawhero	0.528	0.571	-0.639	-0.561	-0.078
25	Waikato @ Bridge St Br	17.241	12.214	2.847	2.503	0.345
26	Mangaonua	2.277	0.638	0.823	-0.449	1.272
27	Mangakotukutuku	0.244	0.499	-1.413	-0.695	-0.717
28	Mangaone	0.983	0.284	-0.017	-1.258	1.242
29	Waikato @ Horotiu Br	28.122	14.228	3.337	2.655	0.681
30	Waitawhiriwhiri	0.250	0.398	-1.386	-0.921	-0.465
31	Kirikiriroa	0.138	0.174	-1.983	-1.747	-0.236
32	Waikato @ Huntly-Tainui Br	69.821	58.673	4.246	4.072	0.174
33	Komakorau	3.379	2.706	1.218	0.996	0.222
34	Mangawara	8.608	4.367	2.153	1.474	0.679
35	Waikato @ Rangiriri	112.643	59.403	4.724	4.084	0.640
37	Awaroa @ Rotowaro	0.403	0.417	-0.910	-0.874	-0.036
38	Waikato @ Mercer Br	59.448	69.468	4.085	4.241	-0.156
39	Whangape	1.709	1.873	0.536	0.628	-0.091

Map id	Monitoring site	Mean annual loads		Log-transformed loads		
		Measured	Modelled	Measured	Modelled	Residual
40	Whangamarino @ Island Block Rd	1.851	2.657	0.615	0.977	-0.362
41	Whangamarino @ Jefferies Rd Br	1.719	1.433	0.542	0.359	0.182
42	Waerenga	0.251	0.222	-1.384	-1.506	0.123
43	Matahuru	1.630	1.020	0.489	0.020	0.469
45	Opuatia	0.753	0.809	-0.283	-0.212	-0.071
46	Mangatangi	1.747	1.115	0.558	0.109	0.449
47	Waikato @ Tuakau Br	66.223	70.446	4.193	4.255	-0.062
48	Ohaeroa	0.120	0.103	-2.119	-2.270	0.151
49	Mangatawhiri	0.345	0.132	-1.065	-2.028	0.963
51	Whakapipi	0.364	0.345	-1.011	-1.064	0.053
52	Awaroa @ Waiuku	0.120	0.185	-2.124	-1.687	-0.437
100	Waipa @ Mangaokewa Rd	0.352	0.283	-1.043	-1.262	0.219
101	Waipa @ Otewa	2.238	2.993	0.806	1.096	-0.291
102	Mangaokewa	3.751	2.737	1.322	1.007	0.315
104	Mangapu	7.556	7.077	2.022	1.957	0.065
106	Waipa @ Otorohanga	3.429	4.542	1.232	1.513	-0.281
107	Waipa @ Pirongia-Ngutunui Rd Br	26.479	26.112	3.276	3.262	0.014
108	Waitomo @ Tumutumu Rd	0.820	0.559	-0.199	-0.581	0.382
109	Waitomo @ SH31 Otorohanga	1.390	1.022	0.329	0.022	0.307
111	Puniu	2.494	4.499	0.914	1.504	-0.590
113	Mangatutu	0.229	0.774	-1.476	-0.256	-1.219
114	Mangapiko	1.961	3.240	0.673	1.175	-0.502
115	Mangaohoi	0.007	0.014	-4.962	-4.286	-0.676
116	Waipa @ SH23 Br Whatawhata	65.379	33.548	4.180	3.513	0.667
117	Mangauika	0.020	0.043	-3.935	-3.143	-0.791
118	Kaniwhaniwha	1.162	0.966	0.150	-0.035	0.185
120	Ohote	0.233	0.412	-1.457	-0.887	-0.570

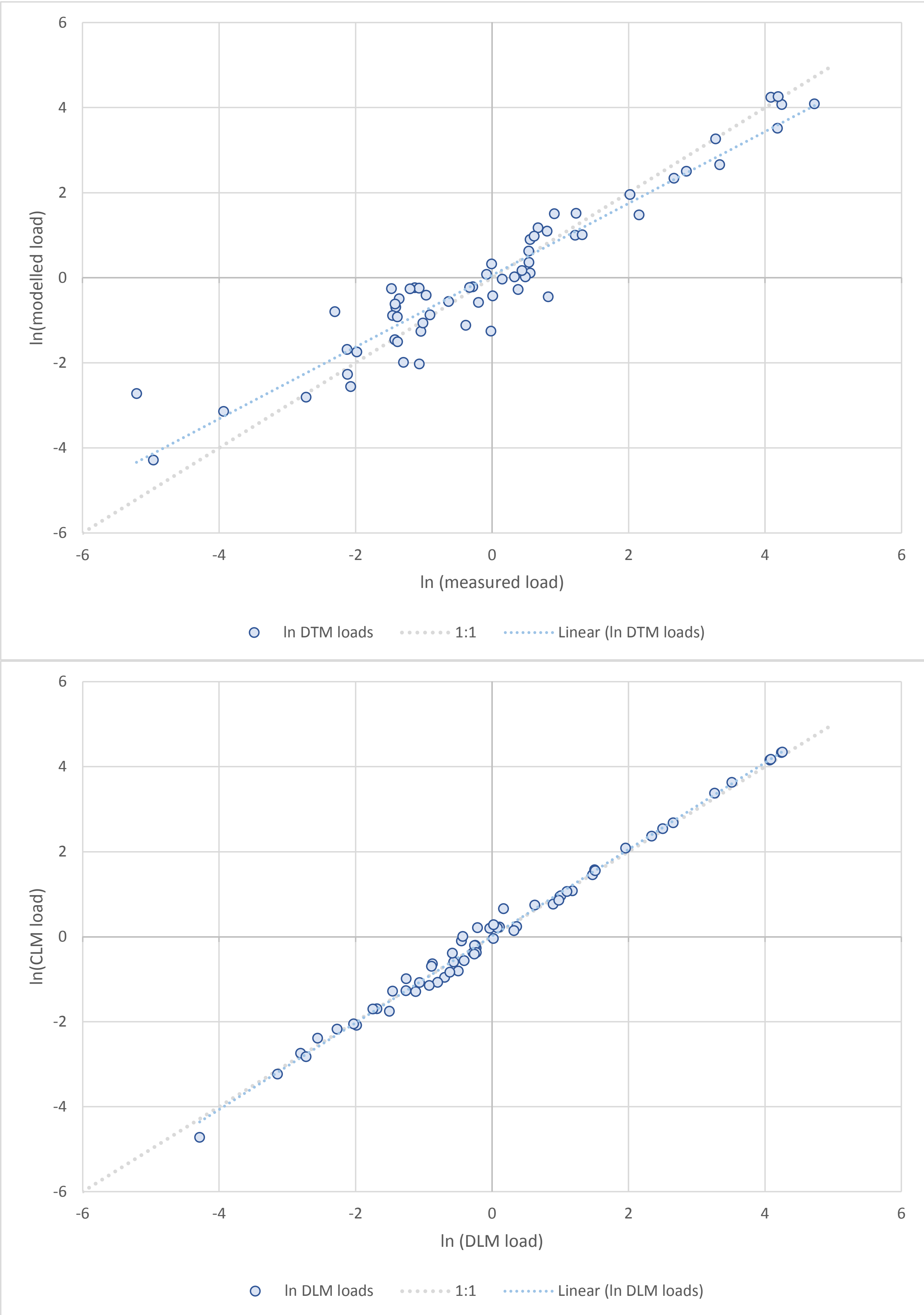


Figure 3-1: Log transformed loads estimated using the detailed load model against those measured.

- There is a discrepancy between the sum of the measured loads upstream of the Waipa / Waikato confluence (Waipa @ SH23 Br Whatawhata and Waikato @ Horotiu Bridge) compared to Waikato @ Huntly-Tainui Br downstream of the confluence. However, the load measured at Waikato @ Rangiriri downstream is much greater than the measured load at Waikato @ Huntly-Tainui Br and is more in keeping with the loads measured upstream of the confluence. This suggests that the measured load for Waikato @ Huntly-Tainui Br may be too low.
- Model over-prediction in the northern and eastern subcatchments of the upper Waikato catchment is partly explained by the pumice soils in the upper catchment. Despite having high drainage which would suggest low surface decay rates between the diffuse sources and the stream network, these soils have low microbial bypass vulnerability (Mcleod et al., 2008) and act as filters trapping *E. coli*. This offers a plausible explanation for why the measured loads in these sub catchments are lower than would be expected given the current land use.

3.1.3 Coarse load model

The relative pre-attenuation load estimated for each land use type in the models by subcatchment is given in Figure 3-2 along with the loads from point sources and dairy sheds. These data were used to derive the incremental increase in instream loads for each subcatchment. Pastoral land use is the dominant source of *E. coli* in the model. The proportion of the load contributed by either dairy or sheep and beef in each subcatchment is dependent on the relative coverage of these land uses in the catchment as well as the parameterised yields. Generally, the contribution from other land uses, dairy sheds and point sources are minor in comparision to the pastoral land uses. There are some exceptions, such as Waikato @ Horotiu Br, which has high loads from urban land use and point sources. The measured and modelled mean annual *E. coli* loads for the water quality monitoring sites with *E. coli* sampling are shown in Table 3-3 along with the log-transformed loads and residuals. The relationship between the log-transformed loads is plotted in Figure 3-3. The RMSE between the log-transformed measured and modelled loads is 0.60. The coefficient of determination (R^2) for the log-transformed loads, adjusted for the number of variables, is 0.89.

The results for the CLM are comparable to the results for DLM for most sites (Figure 3-4). Differences between the models are likely to be due to scaling issues (i.e. spatial smoothing between the REC reach scale and the subcatchment scales) and differences in the way in which land use and catchment characteristics are represented in the two models. Some of the differences are discussed below:

- The application of the reservoir attenuation factor calibrated to the outlet reach may not be applicable to the entire subcatchment. With the exceptions of Waikato @ Ohaaki, Waikato @ Ohakuri and Waikare[1] subcatchment, the loads calculated for the CLM are greater than for the DLM.

[1] Not monitored for *E. coli*

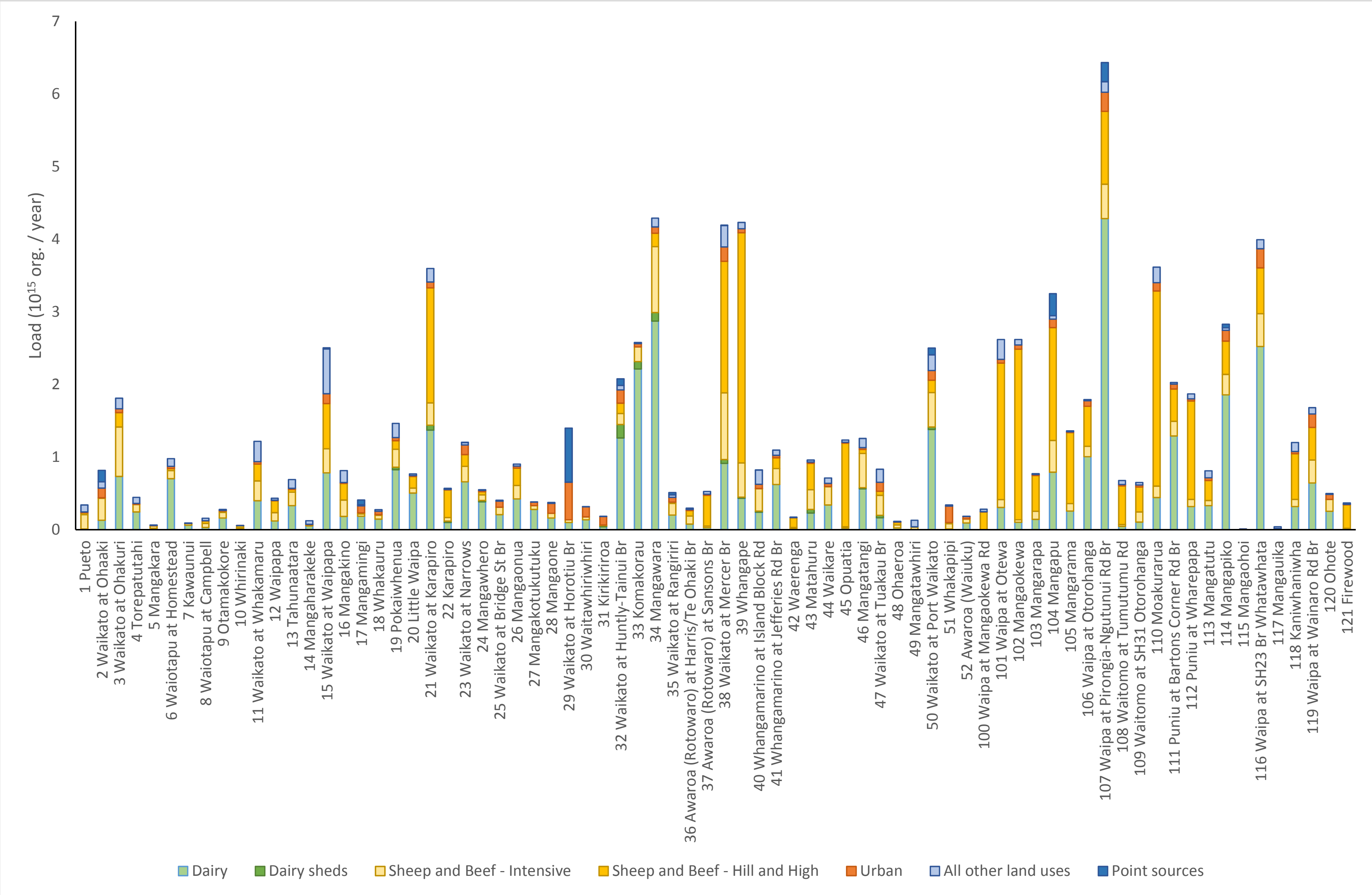


Figure 3-2: Pre-attenuation load contribution to the coarse model from different land use types and point sources. .

Table 3-3: Measured mean annual *E. coli* loads (peta, 10^{15} organisms / year) against loads modelled using the coarse load model. Log-transformed loads and their residuals are also shown.

Map id	Monitoring site	Mean annual loads		Log-transformed loads		
		Measured	Modelled	Measured	Modelled	Residual
1	Pueto	0.10	0.34	-2.30	-1.08	
2	Waikato @ Ohaaki Br	0.99	1.15	-0.01	0.14	-0.15
3	Waikato @ Ohakuri Tailrace Br	0.30	0.66	-1.20	-0.41	-0.79
4	Torepatutahi	0.26	0.45	-1.36	-0.81	-0.55
5	Mangakara	0.07	0.06	-2.73	-2.74	0.02
6	Waiotapu @ Homestead	0.92	1.23	-0.08	0.20	-0.29
7	Kawaunui	0.13	0.09	-2.07	-2.39	0.32
9	Otamakokore	0.24	0.28	-1.43	-1.28	-0.14
10	Whirinaki	0.01	0.06	-5.21	-2.83	
11	Waikato @ Whakamaru Tailrace	1.01	1.00	0.01	0.00	0.01
12	Waipapa	0.24	0.43	-1.42	-0.84	-0.58
13	Tahunaatara	0.72	0.69	-0.33	-0.37	0.04
14	Mangaharakeke	0.27	0.12	-1.30	-2.08	0.79
15	Waikato @ Waipapa Tailrace	1.55	1.93	0.44	0.66	-0.22
16	Mangakino	0.34	0.81	-1.07	-0.21	-0.86
17	Mangamingi	1.46	0.68	0.38	-0.38	0.76
18	Whakauru	0.68	0.27	-0.39	-1.29	0.91
19	Pokaiwhenua	1.75	2.15	0.56	0.76	-0.21
20	Little Waipa	0.32	0.77	-1.13	-0.27	-0.86
22	Karapiro	0.38	0.57	-0.97	-0.56	-0.40
23	Waikato @ Narrows Br	14.36	10.62	2.66	2.36	0.30
24	Mangawhero	0.53	0.55	-0.64	-0.60	-0.04
25	Waikato @ Bridge St Br	17.24	12.68	2.85	2.54	0.31
26	Mangaonua	2.28	0.90	0.82	-0.10	0.92
27	Mangakotukutuku	0.24	0.38	-1.41	-0.96	-0.46
28	Mangaone	0.98	0.37	-0.02	-0.99	0.97
29	Waikato @ Horotiu Br	28.12	14.58	3.34	2.68	0.66
30	Waitawhiriwhiri	0.25	0.32	-1.39	-1.15	-0.24
31	Kirikiriroa	0.14	0.18	-1.98	-1.70	-0.28
32	Waikato @ Huntly-Tainui Br	69.82	63.89	4.25	4.16	0.09
33	Komakorau	3.38	2.58	1.22	0.95	0.27
34	Mangawara	8.61	4.29	2.15	1.46	0.70
35	Waikato @ Rangiriri	112.64	64.76	4.72	4.17	0.55
37	Awaroa @ Rotowaro	0.40	0.53	-0.91	-0.64	-0.27
38	Waikato @ Mercer Br	59.45	75.91	4.09	4.33	-0.24
39	Whangape	1.71	2.10	0.54	0.74	-0.21

Map id	Monitoring site	Mean annual loads		Log-transformed loads		
		Measured	Modelled	Measured	Modelled	Residual
40	Whangamarino @ Island Block Rd	1.85	2.35	0.62	0.86	-0.24
41	Whangamarino @ Jefferies Rd Br	1.72	1.27	0.54	0.24	0.30
42	Waerenga	0.25	0.17	-1.38	-1.75	0.37
43	Matahuru	1.63	0.96	0.49	-0.04	0.53
45	Opuatia	0.75	1.24	-0.28	0.21	-0.49
46	Mangatangi	1.75	1.26	0.56	0.23	0.33
47	Waikato @ Tuakau Br	66.22	76.98	4.19	4.34	-0.15
48	Ohaeroa	0.12	0.11	-2.12	-2.18	0.06
49	Mangatawhiri	0.34	0.13	-1.06	-2.05	0.99
51	Whakapipi	0.36	0.34	-1.01	-1.08	0.07
52	Awaroa @ Waiuku	0.12	0.18	-2.12	-1.70	-0.43
100	Waipa @ Mangaokewa Rd	0.35	0.28	-1.04	-1.27	0.23
101	Waipa @ Otewa	2.24	2.90	0.81	1.06	-0.26
102	Mangaokewa	3.75	2.62	1.32	0.96	0.36
104	Mangapu	7.56	8.03	2.02	2.08	-0.06
106	Waipa @ Otorohanga	3.43	4.71	1.23	1.55	-0.32
107	Waipa @ Pirongia-Ngutunui Rd Br	26.48	29.25	3.28	3.38	-0.10
108	Waitomo @ Tumutumu Rd	0.82	0.68	-0.20	-0.39	0.19
109	Waitomo @ SH31 Otorohanga	1.39	1.32	0.33	0.28	0.05
111	Puniu	2.49	4.82	0.91	1.57	-0.66
113	Mangatutu	0.23	0.81	-1.48	-0.21	-1.26
114	Mangapiko	1.96	2.94	0.67	1.08	-0.40
115	Mangaohoi	0.01	0.01	-4.96	-4.72	-0.24
116	Waipa @ SH23 Br Whatawhata	65.38	37.64	4.18	3.63	0.55
117	Mangauika	0.02	0.04	-3.93	-3.24	-0.70
118	Kaniwhaniwha	1.16	1.22	0.15	0.20	-0.05
120	Ohote	0.23	0.50	-1.46	-0.70	-0.76

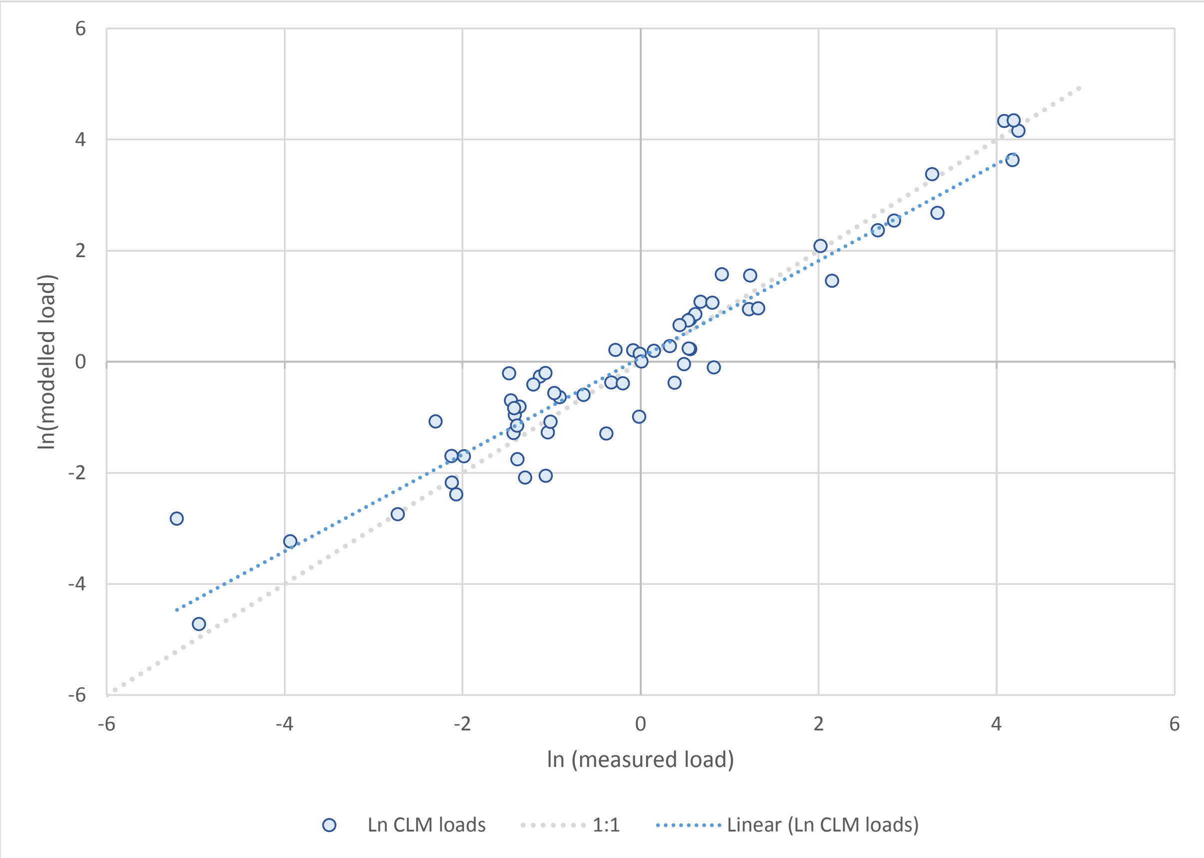


Figure 3-3: Log transformed loads estimated using the coarse load model against those measured.

- The CLM modelled loads for the two peri-urban subcatchments, Mangaone and Mangaonua, that are upstream of Hamilton, have higher estimated loads compared to the DLM, however, the loads are significantly less than those measured.
- Of the two urban subcatchments, Kirikiriroa has a slightly higher load estimated with the CLM compared to the DLM whereas the CLM estimate for Waitawhiriwhiri is slightly lower. The modelled loads for both models are greater than the measured loads.
- With the exception of the subcatchments listed above, the greatest differences between the detailed and coarse modelled loads are for Opuatia which has greater predicted loads for the CLM and Mangaohoi which has lower predicted loads. The former is located in the lower reaches of the Waikato and is dominated by sheep/beef farming and forestry, while the latter is largely covered in native forest (Maungatautari Sanctuary Mountain). The differences in loads for these catchments seems to be in the representation of soil drainage and rainfall classes.

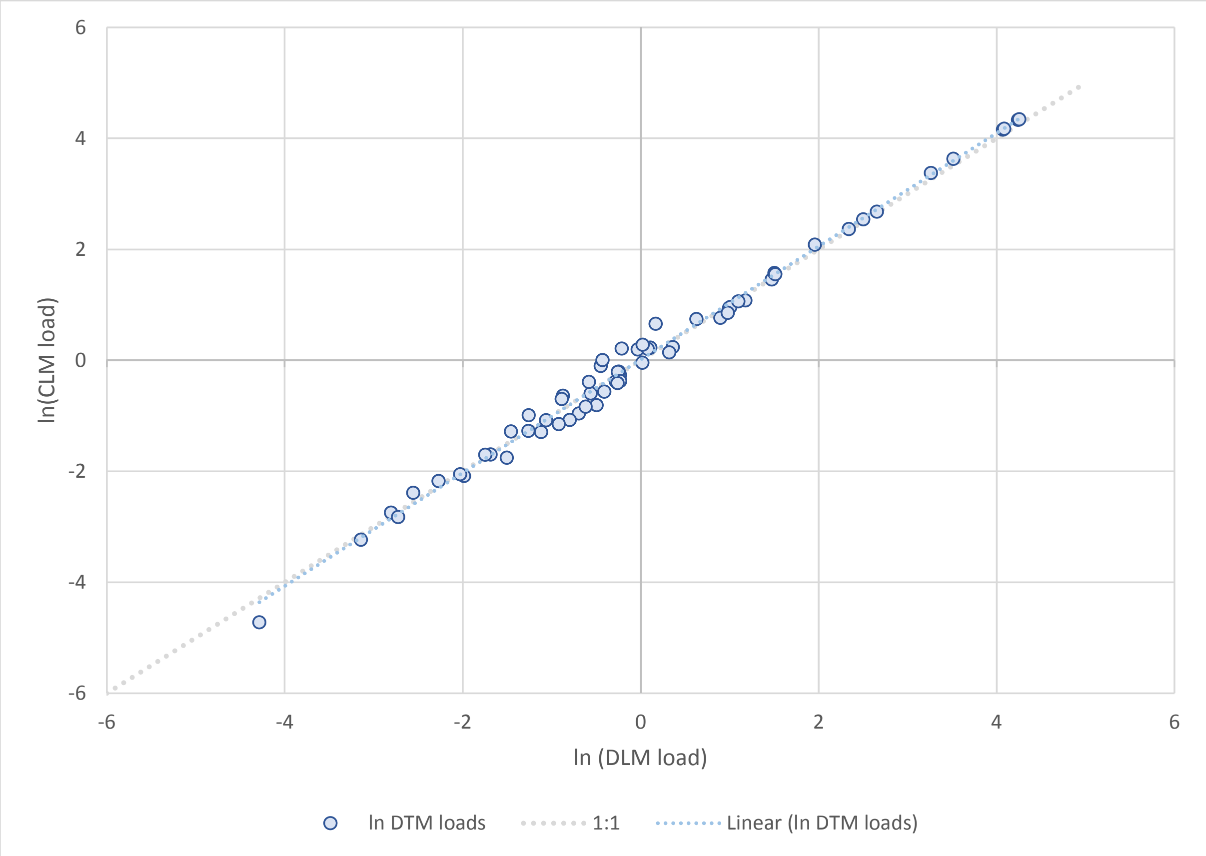


Figure 3-4: Comparison of log-transformed modelled loads simulated using the coarse and detailed models.

3.2 *E. coli* current concentration model

The regression models described in Section 2.4.1 were used to predict current concentrations for seven monitoring sites which currently do not have *E. coli* monitoring.

The model output for median concentrations, along with the standard errors for the model and the coefficients are presented in Table 3-4. The standard error is 0.52 in log space, or 67% for the predicted value, which should be borne in mind when using the predicted concentrations. This model performance is comparable to that in national-scale modelling studies (Unwin et al., 2010). The effect of poor drainage (c value of 4.46) indicates a strong increase in concentrations for poor drainage. The coefficient for dairy and intensive sheep and beef ($c_{D,SBI}$) was lower than that for hill and high country sheep and beef (c_{SBH}), although this will be offset to some degree by the predominance of the intensive land uses on poorly-drained land. The urban concentration was fairly low considering concentrations of about 700 per 100 mL that occur in some urban streams (such as Kirikiriroa and Waitawhiriwhiri), but those streams also have a component of intensive pasture in the catchment which increases the concentrations to close to the measured values. The coefficient for exotic and native forest and scrub (c_{Tree}) is uncertain, but is clearly lower than for the pastoral land uses.

Table 3-4: Regression model output for current median concentrations. Note that the coefficient for dairy and intensive sheep and beef has been combined.

R	R^2	Adjusted R^2	Std. error of estimate
0.80	0.64	0.59	0.52
Coefficient	**Value**	**Std. error of coefficient**	
c	4.76	1.40	
$c_{D,SBI}$	165.85	40.36	
c_{SBH}	425.91	97.16	
c_{Urb}	278.46	162.93	
c_{Tree}	52.60	21.94	
c_{other}	672.70	508.81	

The results for the model for 95-percentile concentrations are shown in Table 3-5. The performance of this model is worse than for the median concentrations. The concentration coefficients are generally an order of magnitude larger than for the median concentrations and the coefficient for hill and high country sheep and beef is particularly high, perhaps reflecting higher runoff variability in steeper areas. There is a stronger influence of soil drainage, but this is compensated by a lower coefficient for dairy/intensive sheep and beef land use which tends to occur more on poorly-drained soils.

Table 3-5: Regression model output for current 95-percentile concentrations.

R	R^2	Adjusted R^2	Std. error of estimate
0.73	0.53	0.47	0.72
Coefficient	**Value**	**Std. error of coefficient**	
c	8.02	3.08	
$c_{D,SBI}$	697.79	333.22	
c_{SBH}	4821.02	1419.55	
c_{Urb}	2288.44	1489.06	
c_{Tree}	776.24	300.28	
c_{other}	686.06	3463.81	

The predicted *E.coli* concentrations are shown in Table 3-6. As noted earlier, the concentrations for the virtual sites, Waipa @ Waingaro Road and Waikato @ Port Waikato in Table 3-6 have been estimated from adjacent sites rather than from the concentration model.

Table 3-6: Predicted current concentrations (organisms / 100 ml) for sites where concentrations are not measured.

Map ID	Site Name	Median Concentration	95th Percentile Concentration
21	Waikato @ Karapiro	30	214.0
36	Awaroa (Rotowaro) @ Harris/Te Ohaki Br	449	3520
44	Waikare	630	5278
50	Waikato @ Port Waikato	82	1644
103	Mangarapa	388	4636
105	Mangarama	481	6333
110	Moakurarua	222	2390
112	Puniu @ Wharepapa	228	2186
119	Waipa @ Waingaro Rd Br	387	4105
121	Firewood	255	2834

4 Model uncertainty

The standard errors reported for the calibrated DLM parameters (Section 3.1.1, Table 3-1) point to substantial uncertainty within the model. A general discussion on the sources of model uncertainty can be found in Walker et al. (2003). The potential sources of uncertainty with respect to *E. coli* modelling for the Waikato and Waipa River catchments are listed below:

- **Parameterisation**: The DLM was calibrated to minimise the RMSE between the modelled and measured loads for 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.

- ***E. coli* calibration data:** *E. coli* concentration data from 62 sampling sites was used to estimate mean annual loads for calibration. These data are subject to error in sampling and analysis. Moreover, *E. coli* concentrations (Muirhead, 2015) and yields (Wilcock, 2006) are highly variable over time making determination of average catchment concentrations and yields difficult.

  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 median annual load. As pointed out by Davies-Colley et al. (2011), this is not necessarily the case. They highlight the need for national protocols around the collection of water quality data in order to standardise monitoring and to provide data that is purpose collected for modelling.

  There is evidence in some sub-catchments that *E. coli* concentrations are increasing as a result of land use change. However, the concentration data were not trend adjusted due to concerns that detrending may introduce error. Some sites had fewer than five years of monitored data and some had upwards of 10% censored (below detection) data.

  The measured loads were determined using concurrent flow data where flow data were available. For other sites, estimated annual mean flows were used. Both methods are subject to error. In the case of the rating curve method, the confidence intervals for the measured loads are provided in Appendix D. The ratio method assumes that there is a relationship between the median and mean annual *E. coli* concentrations that is a function of the mean annual flow.

- **Point sources:** *E. coli* point source data used in the model include industrial and municipal wastewater discharges and treated effluent discharges from dairy farms. The point sources are variable over time introducing error into the assessment of mean annual loads. Some sources, like the Cambridge Wastewater Treatment Plant (which was removed from the model) 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, in turn estimated from the consented discharge. Since the consents are the maximum allowable discharge from each farm, which may not be reached, the estimated loads are conservative. The load from ponds is likely to be 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 ten land use types with diffuse loads from these sources represented in the load models by five calibrated source yields. The models do not include data on certain land uses, for example, irrigation or dairy support, which could affect *E. coli* generation. The land use type data come from WRC and were derived from a number of sources. The derivation and interpretation of the underlying land use data are subject to sampling precision errors and ground-truthing errors. The land use data is as at 2012, so recent land use changes, such as those that have occurred in the Pueto subcatchment are not represented in the model.

- **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 birds as well as unknown point sources such as such as sewer or pumping station overflows in urban areas. For example, water fowl can contribute significant loads of *E. coli* to freshwater bodies in Waikato (Wilcock, 2006) and other parts of the country (Moriarty et al., 2011). Indeed, water fowl living along the banks of the Waikato River in Hamilton have also been identified as an important source of *E. coli* in the city (Tonkin and Taylor Ltd., 2001). This could account for the high *E. coli* loads measured in and downstream of Hamilton.

- **Spatial resolution:** All three models are subject to spatial smoothing of heterogeneous input data (i.e., scaling effects). The smallest spatial unit of the DLM is the REC unit, and there are over 22,200 of these in the Waikato and Waipa River catchments. Spatial data within each unit are lumped and there are no linkages between the data types. 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 data.

  The CLM operates at the catchment scale, and the study area has been broken into 74 subcatchments, primarily based on the location of water quality monitoring sites rather than by grouping catchment characteristics. This means that there can be substantial spatial variability within a subcatchment. Quasi-spatial distribution is represented by splitting each catchment into proportional areas covered by unique combinations of land use, drainage and rainfall classes under the assumption that these catchment characteristics have the greatest bearing on *E. coli* loads reaching the stream network.

  While it is possible to similarly represent combinations within the REC units (to improve the spatial representation of the DLM) this would significantly increase the size and complexity of the model. Instead of this, it was assumed that the smaller unit size would adequately represent the interactions between land use and other catchment characteristics.

- **Temporal resolution:** The load models are steady-state models which predict mean annual loads. Given that the requirement is to predict at the average annual time-step (as per the *E.coli* attribute in the National Objectives Framework), such a model is 'fit for purpose'. Nevertheless, this means that seasonal changes in *E. coli* generation and transport are not separately captured by the models. The concentration regression model also does not specifically 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.

  Complex dynamic modelling is not possible at this point in time as the input data required to run such a model is not available and the detailed dynamics of E.coli generation and transport are not well understood.

- **Linearity**: The method for predicting future instream median annual *E. coli* concentrations from modelled mean annual loads assumes that there is a linear relationship between the two data sets. This assumption has not been tested.

5 Recommendations for further work

Section 4 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 monthly monitoring data, but have not been validated. 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 where possible to allow for better calculation of loads. See Davies-Colley et al. (2011) for a discussion of the need for national water quality monitoring to provide data for modelling.

  The models should be updated and recalibrated when sufficient data become available. Ideally, the model should be tested against data that have not been used for calibration.

- **Microbial source tracking:** Assessment of background *E. coli* from natural diffuses sources and point sources. This work would be required to determine the source of *E. coli* loads from forested catchments with unexpectedly high measured 95^{th} percentile concentrations. A related survey to determine the number of water fowl living along river banks, particularly in and around urban areas with high measured *E. coli* concentrations, should also be carried out.

- **Dynamic modelling:** As noted above, dynamic modelling is not possible at this time. The research and information requirements for developing a fit-for-purpose dynamic model *E. coli* should be scoped-out. Such a model would allow us to better understand the temporal processes in operation and a better representation of those processes.

6 Conclusions

This report presents three steady-state catchment models to estimate *E. coli* loads and concentrations in the Waikato and Waipa River catchments that have been developed to support a farm cost model. The models are:

1. **The detailed load model.** This model operates at the REC reach unit scale and was calibrated against measured mean annual *E. coli* loads from 62 monitoring stations in the study area. The DLM was developed specifically to provide parameters to the CLM.
2. **The coarse load model:** This model operates at the subcatchment scale and was developed to be compatible with the scale of the farm cost model.
3. **The concentration model:** Two regression models were developed to estimate site median annual and 95th percentile *E. coli* concentrations, respectively, for monitoring sites where *E. coli* is not currently sampled.

In addition to the development and calibration of the models, the report describes input and calibration data; the methodology used to determine median annual and 95th percentile concentrations and mean annual loads and sources of model uncertainty.

The load models estimate 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 instream load. The loads are then subject to attenuation in reservoirs. The RMSE between the log-transformed measured and modelled loads for the DLM is 0.55 and the adjusted R^2 is 0.92. The RMSE between the log-transformed measured and modelled loads for the CLM is 0.60 and the adjusted R^2 is 0.89. The concentration regression model is better able to estimate median annual *E. coli* concentrations than 95th percentile concentrations. The R^2 for the former is 0.64 and the standard error in log-space is 0.53. The R^2 for the 95th percentiles is as 0.53 and the standard error is 0.72.

It is recommended that current water and flow monitoring be continued or expanded to provide further data for water quality modelling. Additionally, the research and information needs required for dynamic modelling should be investigated.

Acknowledgements

We wish to acknowledge the support of the following:

- Mighty River Power and Genesis Energy for providing flow data.
- Waikato Regional Council for providing land use and water quality data.

The report was enhanced by constructive peer reviews by Andrew Hughes, David Roper, John Quinn, Bryce Cooper and Bill Vant.

References

Brown, E. (2005) Hydraulic travel times of major Waikato rivers, Enviroment Waikato, Technical Report 2005/04.

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

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.

Davies-Colley, R., Nagels, J., Smith, R., Young, R. and Phillips, C. (2004) Water quality impact of a dairy cow herd crossing a stream. *New Zealand Journal of Marine and Freshwater Research*, 38: 569-576.

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

Doole, G.J. (2013) Evaluation of policies for water quality improvement in the Upper Waikato catchment, Prepared by Waikato Management School, University of Waikato for Ministry for the Environment and Ministry for Primary Industries (in review)

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.

Elliott, S., Semadeni-Davies, A., Depree, C. and Harper, S. (2013) Catchment models for nutrients and microbial indicators: Modelling application to the upper Waikato River catchment, Prepared by NIWA for Ministry for the Environment, NIWA Client Report: HAM13-103 http://www.mfe.govt.nz/issues/water/freshwater/supporting-papers/niwa-catchment-models-jul14.pdf

Hickey, C.W., Quinn, J.M. and Davies-Colley, R.J. (1989) Effluent characteristics of domestic sewage oxidation ponds and their potential impacts on rivers. *New Zealand Journal of Marine and Freshwater Research*, 23(4): 585-600. 10.1080/00288330.1989.9516394

Journeaux, P. (2013) Modelling of E. Coli mitigation: Upper Waikato Catchment. Report Prepared by AgFirst for the Ministry for the Environment.,

Longhurst, B. (2012) Supporting data for CLUES project. Client report prepared for NIWA by AgResearch Ltd.

McDowell, R.W., Wilcock, B. and Hamilton, D.P. (2013) Assessment of Strategies to Mitigate the Impact or Loss of Contaminants from Agricultural Land to Fresh Waters, Prepared for MfE, AgResearch Client Report RE500/2013/066.

McKergow, L.A., Tanner, C.T., Monaghan, R.M. and Anderson, G. (2007) Stocktake of diffuse pollution attenuation tools for New Zealand pastoral farming systems, Prepared for Pastoral 21 Research Consortium under contract to AgResearch, NIWA client report HAM2007-161.

Mcleod, M., Aislabie, J., Ryburn, J. and McGill, A. (2008) Regionalizing Potential for Microbial Bypass Flow through New Zealand Soils. *Journal of Environmental Quality*, 37: 1959-1967.

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

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

Muirhead, R.W., Elliott, A.H. and Monaghan, R.M. (2011) A model framework to assess the effect of dairy farms and wild fowl on microbial water quality during base-flow conditions. *Water research*, 45: 2863-2874.

Nagels, J.W., James, T., Davies-Colley, R.J., Fenemor, A., Merrilees, R., Burton, A., Stuart, B. and Parshotam, A. (2012) The Sherry River - a success story. . *In: Advanced Nutrient Management: Gains from the Past - Goals for the Future. (Eds L.D. Currie and C L. Christensen). Occasional Report No. 25. Fertilizer and Lime Research Centre, Massey University, Palmerston North, New Zealand.*, http://flrc.massey.ac.nz/publications.html

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

Pachepsky, Y.A. and Shelton, D.R. (2011) Escherichia Coli and Fecal Coliforms in Freshwater and Estuarine Sediments. *Critical Reviews in Environmental Science and Technology*, 41(12): 1067-1110. 10.1080/10643380903392718

Perchec-Merien, A.-M. and Lewis, G.D. (2013) Naturalized Escherichia coli from New Zealand wetland and stream environments. *FEMS Microbiology Ecology*, 83(2): 494-503. 10.1111/1574-6941.12010

Quinn, J. (2012) Tools for improving freshwater quality in pastoral catchment. Presentation to NRC's Environmental Management Committee Workshop on Water Quality, May 1, 2012., Semadeni-Davies, A. and Elliot, S. (2013) Impact of stock exclusion on E. coli concentrations: Application of the CLUES model to the lower Waikato River catchment, Prepared for Minsitry for the Environment, NIWA Client Report No: HAM2013-108 (under client review).

Semadeni-Davies, A. and Elliott, S. (2012) Preliminary study of the potential for farm mitigation practices to improve river water quality: Application of CLUES to the Waikato Region. Prepared for Waikato Regional Council. NIWA Client Report: AKL2012-034.

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

Stott, R., Sukias, J., McKergow, L. and Tanner, C. (2014) Unexpected net export of E.coli from a constructed wetland intercepting draining water in an agricultural watershed. Implications for water quality and health. *21st Century Watershed Technology Conference, " Improving Water Quality and the Environment"*, University of Waikato, Hamilton, New Zealand, 1-6 november 2014.

Tonkin and Taylor Ltd. (2001) Microbial inputs to the Hamilton reach of the Waikato River, Prepared by Tonkin and Taylor Ltd. for Hamilton City Council, Reference number: 60322.

Unwin, M., Snelder, T., Booker, D., Ballantine, D. and Lessard, J. (2010) Predicting water quality in New Zealand rivers from catchment-scale physical, hydrological and land use descriptors using random forest models.

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

Walker, W.E., Harremoës, P., Rotmans, J., van der Sluijs, J.P., van Asselt, M.B.A., Janssen, P. and Krayer von Krauss, M.P. (2003) Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. *Integrated Assessment*, 4(1): 5-17. 10.1076/iaij.4.1.5.16466

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. .

Woods, R., Elliott, S., Shankar, U., Bidwell, V., Harris, S., Wheeler, D., Clothier, B., Green, S., Hewitt, A., Gibb, R. and Parfitt, R. (2006) The CLUES Project: Predicting the Effects of Land-use on Water Quality – Stage II. NIWA Client Report HAM2006-096.

Woods, R., Hendrikx, J., Henderson, R. and Tait, A. (2006) Estimating Mean flow of New Zealand rivers. *Journal of Hydrology (NZ)*, 45(2): 95-110.

Appendix A Mitigations for microbial load reduction

This appendix reviews the use and effectiveness of common farm mitigation practices with respect to the removal of faecal contaminants from agricultural land uses. There are few studies that have assessed the effectiveness of the mitigations listed. Muirhead (2015) suggests that mitigation is challenging because there is little information on *E. coli* losses at the farm scale (largely due to the difficulty of measuring these losses) and the data that is available has high variability.

Collins et al. (2007) evaluated the suitability of a range of mitigation practices for microbial removal, these are summarised in Table A-1. Each mitigation was discussed in relation to its contaminant sources. However, while potential removal processes were discussed, removal efficiencies were not presented. Effective mitigation methods include the exclusion of stock from waterways and riparian areas (Muirhead et al., 2011; Quinn, 2012), and the construction of bridges at stock access points (see Davies-Colley et al., 2004; Nagels et al., 2012). Other mitigations discussed for Farm Dairy Effluent (FDE) includes storage in ponds and resource reclamation via irrigation.

The efficiency ratings given in Table A-1 have been derived from literature reviews published in McKergow et al. (2007), Longhurst (2012) and McDowell et al. (2013). These reviews focused on New Zealand studies, however, international studies are also cited in some instances. For several of the mitigations listed from Collins et al. (2007) no efficiency ratings could be found. Longhurst (2012) was prepared for WRC under sub-contract to NIWA (Semadeni-Davies and Elliott, 2012) to provide background data in order to simulate the impacts of various mitigation options on water quality in Waikato. McKergow et al. (2007) and McDowell et al. (2013) are both stocktakes of common mitigation options in New Zealand, the former was prepared for AgResearch and Landcare Research under the Pastoral 21 Research Programme, while the latter was prepared for the Ministry for the Environment. The relative performance and cost ratings refer to percentile rankings against other mitigation options assessed by McDowell et al. (2013). Like Collins et al. (2007), McDowell et al. (2013) state that the cost effectiveness of a mitigation largely depends on it being applied at the right time and place. They give the example from Otago where a mitigation targeting total phosphorus was six to seven times more cost effective when applied to critical source areas compared to catchment wide applications.

Similar to McDowell et al. (2013), Muirhead (2015) ranked common FDE management practices in the following order of performance for the development of a risk-index for faecal contamination:

1. Two-pond (anaerobic and aerobic) treatment system discharged directly to stream;
2. Advanced pond system discharged directly to stream;
3. High application rate FDE irrigation system irrigating daily from a sump;
4. Low application rate FDE irrigation system including a 60-day storage system operating under a deferred irrigation strategy.

Table A-1: Mitigations suitable for reducing faecal contaminant loads. With reference to Collins et al. (2007). Removal and cost data from a. McKergow et al. (2007); b. Longhurst (2012); c. McDowell et al. (2013).

Mitigation class	Mitigation	Description	Enterprise	Relative *E. coli* removal	Relative removal of other contaminants	Relative cost with respect to *E. coli* removal	Key reference
Direct pathways	Herd crossings	Stream fords replaced by stock bridges or culverts to prevent stock access to streams.	Dairy	Medium	N (low), sediment (low)	Medium	c.
	Stock exclusion	Fencing or riparian planting to block stock access to water ways. Prevents direct deposition of excreta in streams and reduces treading damage and bed erosion.	All stock types	High, 20-35% Dependant on initial stock access to streams.	P (high), sediment (low)	High	a., b., c.
Treatment of surface runoff	Vegetated / grass buffer strips	Filter system to decrease sediment and particulate losses in surface runoff by a combination of filtration, deposition, and improving infiltration.	All stock types	Low	P (high), sediment (high)	V. high	a., c.
	Constructed wetlands	Water detention basins planted with aquatic vegetation designed to intercept surface and shallow subsurface runoff.	All stock types	around 80%, could be net exporter of *E. coli*	N (v. high), P (medium), sediment (high)		a., c.
Effluent management	Deferred FDE irrigation*	Pond storage of FDE prior to land application during periods of low leaching or runoff risk. Allows irrigation to be timed to maximise nutrient uptake.	Dairy	Medium	N (medium), P (medium)	Medium	b., c.
	Water irrigation / precision agriculture	Delaying irrigation by 7-10 days after grazing to allow microbial die-off to occur reducing leaching of microbes. Use of sensors to optimise water and nutrient application.	Irrigated land				
	Low application rate of FDE*	Pond storage of FDE prior to land application via a low rate system (<8mm depth).	Dairy	Medium, 1-10% depending on drainage	P (high), N (medium)	Medium	b., c.
	Advanced pond systems	Replacement of two-stage oxidation pond systems for FDE with four pond systems (anaerobic, high rate, algal setting and maturation). Can be combined with irrigation management.	Dairy and piggeries	V. high (combined with land disposal), >20%	N (v high), P (high)	V. high	b., c.
Grazing management	Restricted grazing of winter forage crops	Restrict winter grazing of forage crops when pastures are wet with stock rotation of 3-4	All stock with winter forage		P (high), sediment (medium)		

Mitigation class	Mitigation	Description	Enterprise	Relative *E. coli* removal	Relative removal of other contaminants	Relative cost with respect to *E. coli* removal	Key reference
		hours.	crops				
	Restricted grazing and herd housing and standoff pads	Restrict grazing away from paddocks near water ways when heavy rainfall is predicted by relocating stock from paddocks to feed or wintering pads of herd homes.	Dairy		N (high), P (medium), sediment (low)		
	Protection of groundwater supply	Stock grazing restricted to areas down gradient of wells for at least a week prior to or during border strip irrigation. This avoids infiltration of microbes through soil adjacent to wells.	All stock				

* most suited to soils with high potential for microbial attenuation

Missing from the table are two-pond systems which are occasionally used in Waikato for FDE pre-treatment prior to discharge to streams. An investigation of two conventional two-pond systems discharging to a stream in the Toenepi catchment found that this mitigation has poor microbial removal (Donnison et al., 2011). While pond treatment reduced *E. coli* loads in influent by around two orders of magnitude, the per cow effluent *E. coli* load was found to be in the order of 10^7-10^8 between September and December and in the order of 10^7 later in the milking season.

There are no specific data for *E. coli* removal from constructed wetlands cited in the reviews; McKergow et al. (2007) speculate that up to 80% of *E. coli* originating from stock could be removed. However, recent investigations, also in Toenepi catchment, suggests that there may be naturalised stream populations of *E. coli* and that wetlands may be net exporters of *E. coli* with water fowl being the most likely source (Perchec-Merien and Lewis, 2013; Stott et al., 2014).

Appendix B Point sources

Table B-1: Estimated annual *E. coli* loads from point sources in the study area. Data supplied by WRC.

Point source	Map ID	Subcatchment	NZREACH	Annual load peta/y
Hamilton sewage	29	Waikato @ Horotiu Br	3016614	0.7430
Hautapu dairy	23	Waikato @ Narrows	3020218	0.0006
Horotiu meatworks	32	Waikato @ Huntly-Tainui Br	3015715	0.0103
Huntly sewage	35	Waikato @ Rangiriri	3012533	0.0326
Meremere sewage	38	Waikato @ Mercer Br	3007635	0.0051
Ngaruawahia sewage	32	Waikato @ Huntly-Tainui Br	3014648	0.0785
Otorohanga sewage	107	Waipa @ Pirongia-Ngutunui Rd Br	3026372	0.2644
Te Awamutu dairy	114	Mangapiko	3022523	0.0002
Te Awamutu sewage	114	Mangapiko	3022524	0.0495
Te Kauwhata sewage	40	Waipa @ Pirongia-Ngutunui Rd Br	3026372	0.0008
Te Kuiti sewage	104	Mangapu	3030722	0.3023
Tokoroa sewage*	17	Mangamingi	3027689	0.0652
Tuakau rendering	50	Waikato @ Port Waikato	3006838	0.0081
Tuakau/Pukekohe sewage	50	Waikato @ Port Waikato	3006510	0.0840

*estimated on basis of population.

Table B-2: Estimated *E. coli* loads from dairy sheds with consented FDE discharges aggregated by subcatchment.

Map ID	Subcatchment	Estimated annual load (peta/y)	Number of consented discharges
19	Pokaiwhenua	0.0338	2
21	Waikato @ Karapiro	0.0675	4
22	Karapiro	0.0169	1
24	Mangawhero	0.0169	1
31	Kirikiriroa	0.0169	1
32	Waikato @ Huntly-Tainui Br	0.1858	11
33	Komakorau	0.1013	6
34	Mangawara	0.1182	7
38	Waikato @ Mercer Br	0.0507	3
39	Whangape	0.0169	1
40	Whangamarino @ Island Block Rd	0.0169	1
43	Matahuru	0.0507	3
46	Mangatangi	0.0169	1
47	Waikato @ Tuakau Br	0.0338	2
50	Waikato @ Port Waikato	0.0338	2
104	Mangapu	0.0169	1
105	Mangarama	0.0169	1
106	Waipa @ Otorohanga	0.0169	1
107	Waipa @ Pirongia-Ngutunui Rd Br	0.3040	18
110	Moakurarua	0.0169	1
111	Puniu @ Bartons Corner Rd Br	0.1182	7
114	Mangapiko	0.1013	6
116	Waipa @ SH23 Br Whatawhata	0.2026	12
118	Kaniwhaniwha	0.0169	1
119	Waipa @ Waingaro Rd Br	0.0507	3

Appendix C Calibration data: Concentration

Table C-1: Five-year measured median annual *E. coli* concentrations calculated from water quality monitoring sites in the study area. Sites with short *E. coli* data records (monitoring established in early 2013) are shaded .

Map ID	Monitoring site	Concentration (org / 100 ml)	
		Median	95th percentile
Input	Waikato @ Reid's Farm	2	20
1	Pueto	24	100
2	Waikato @ Ohaaki	14	82
3	Waikato @ Ohakuri	2	13
4	Torepatutahi	54	216
5	Mangakara	110	1675
6	Waiotapu @ Homestead	110	281
7	Kawaunui	200	2685
8	Waiotapu @ Campbell	1	21
9	Otamakokore	210	478
10	Whirinaki	16	98
11	Waikato @ Whakamaru	7	60
12	Waipapa	100	450
13	Tahunaatara	110	695
14	Mangaharakeke	160	680
15	Waikato @ Waipapa	8	144
16	Mangakino	40	251
17	Mangamingi	580	1975
18	Whakauru	495	2040
19	Pokaiwhenua	150	1375
20	Little Waipa	110	640
22	Karapiro	265	5980
23	Waikato @ Narrows	39	280
24	Mangawhero	530	3365
25	Waikato @ Bridge St Br	64	682
26	Mangaonua	1350	7060
27	Mangakotukutuku	475	14300
28	Mangaone	850	2260
29	Waikato @ Horotiu Br	90	680
30	Waitawhiriwhiri	585	7260
31	Kirikiriroa	570	3860
32	Waikato @ Huntly-Tainui Br	125	2020
33	Komakorau	1100	3575
34	Mangawara	1000	5445

Map ID	Monitoring site	Concentration (org / 100 ml)	
		Median	95th percentile
35	Waikato @ Rangiriri	112	2268
37	Awaroa (Rotowaro) @ Sansons Br	290	2100
38	Waikato @ Mercer Br	80	1560
39	Whangape	220	589
40	Whangamarino @ Island Block Rd	180	668
41	Whangamarino @ Jefferies Rd Br	600	5175
42	Waerenga	500	6225
43	Matahuru	600	6770
45	Opuatia	400	3180
46	Mangatangi	380	6125
47	Waikato @ Tuakau Br	80	1600
48	Ohaeroa	300	5125
49	Mangatawhiri	190	5615
51	Whakapipi	320	1910
52	Awaroa (Waiuku)	240	1070
100	Waipa @ Mangaokewa Rd	210	2625
101	Waipa @ Otewa	236	1986
102	Mangaokewa	490	6855
104	Mangapu	455	4800
106	Waipa @ Otorohanga	160	3775
107	Waipa @ Pirongia-Ngutunui Rd Br	300	4975
108	Waitomo @ Tumutumu Rd	190	2580
109	Waitomo @ SH31 Otorohanga	310	1575
111	Puniu @ Bartons Corner Rd Br	140	3040
113	Mangatutu	160	780
114	Mangapiko	325	7800
115	Mangaohoi	70	1038
116	Waipa @ SH23 Br Whatawhata	387	4108
117	Mangauika	38	1080
118	Kaniwhaniwha	250	2070
120	Ohote	275	2460

Appendix D Calibration data: Loads

Table D-1: Measured mean annual loads (peta, 10^{15} x organisms / year) calculated using the Rating Curve method for *E. coli* monitoring sites with concurrent flow data. Results of the boot-strapping assessment also provided.

Map ID	Monitoring site	Mean annual load	Boot-strapping results			
			Lower 90% confidence interval	Upper 90% confidence interval	Standard deviation	Mean of mean annual loads
Input	Waikato @ Reid's Farm	0.16	0.13	0.20	0.02	0.16
2	Waikato @ Ohaaki	0.99	0.82	1.21	0.12	1.01
3	Waikato @ Ohakuri	0.30	0.23	0.51	0.13	0.33
9	Otamakokore	0.24	0.10	3.39	8.82	1.42
11	Waikato @ Whakamaru	1.53	0.98	2265.95	754.16	222.33
13	Tahunaatara	0.72	0.41	8.72	15.06	3.60
15	Waikato @ Waipapa	1.55	1.11	5.36	4.12	2.51
19	Pokaiwhenua	1.31	0.73	16593.02	39438.96	6976.58
23	Waikato @ Narrows	14.36	10.63	23.44	3.91	15.52
25	Waikato @ Bridge St Br	17.24	11.86	28.12	5.56	18.95
26	Mangaonua*	2.27	1.16	6.18	4.12	2.99
29	Waikato @ Horotiu Br	28.12	20.31	41.26	6.67	29.16
32	Waikato @ Huntly-Tainui Br	69.82	50.28	103.98	18.08	73.41
35	Waikato @ Ranigiri	112.6	71.5	203.2	39.9	125.3
38	Waikato @ Mercer Br	59.45	40.76	88.12	17.87	64.76
47	Waikato @ Tuakau Br	66.22	42.31	96.15	16.70	69.16
101	Waipa @ Otewa	2.24	1.57	3.47	0.53	2.32
104	Mangapu	7.56	4.34	23.43	10.32	10.19
107	Waipa @ Pirongia-Ngutunui Rd Br	26.48	16.92	48.00	9.92	31.12
108	Waitomo @ Tumutumu Rd	0.82	0.36	2.95	1.09	1.09
113	Mangatutu	0.23	0.15	0.39	0.07	0.26
116	Waipa @ SH23 Br Whatawhata	65.38	44.19	100.97	17.38	68.88

*Flow adjusted by ratio of Wood's et al (2006) flow estimated for the Mangaonua @ Dreadnaught flow monitoring station and the water quality monitoring site.

Table D-2: Measured mean annual loads (peta, 10^{15} x organisms / year) calculated using the ratio method for sites without suitable flow data. *Loads from Pokaiwhenua and Waikato @ Whakamaru calculated using the rating curve method were deemed unsuitable for calibration.

Map ID	Monitoring site	Mean annual load
1	Pueto	0.10
4	Torepatutahi	0.26
5	Mangakara	0.07
6	Waiotapu @ Homestead	0.92
7	Kawaunui	0.13
10	Whirinaki	0.01
11	Waikato @ Whakamaru*	1.01
12	Waipapa	0.24
14	Mangaharakeke	0.27
16	Mangakino	0.34
17	Mangamingi	1.46
18	Whakauru	0.68
19	Pokaiwhenua*	1.75
20	Little Waipa	0.32
22	Karapiro	0.38
24	Mangawhero	0.53
27	Mangakotukutuku	0.24
28	Mangaone	0.98
30	Waitawhiriwhiri	0.25
31	Kirikiriroa	0.14
33	Komakorau	3.38
34	Mangawara	8.61
37	Awaroa (Rotowaro) @ Sansons Br	0.40
39	Whangape	1.71
40	Whangamarino @ Island Block Rd	1.85
41	Whangamarino @ Jefferies Rd Br	1.72
42	Waerenga	0.25
43	Matahuru	1.63
45	Opuatia	0.75
46	Mangatangi	1.75
48	Ohaeroa	0.12
49	Mangatawhiri	0.34
51	Whakapipi	0.36
52	Awaroa (Waiuku)	0.12
100	Waipa @ Mangaokewa Rd	0.35
102	Mangaokewa	3.75
106	Waipa @ Otorohanga	3.43
109	Waitomo @ SH31 Otorohanga	1.39
111	Puniu @ Bartons Corner Rd Br	2.49
114	Mangapiko	1.96
115	Mangaohoi	0.01
117	Mangauika	0.02
118	Kaniwhaniwha	1.16
120	Ohote	0.23