Addendum to Modelling E. coli to support implementation of the NPS-FM

Modification of CLUES E. coli model for Te Maunga

Prepared for Taranaki Regional Council

April 2024

Climate, Freshwater & Ocean Science

9 May 2024 5.53 pm

Prepared by:
Annette Semadeni-Davies
Sandy Elliott

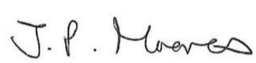
For any information regarding this report please contact:

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

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

Phone +64 9 375 2050

NIWA CLIENT REPORT No:	ELF24101
Report date:	April 2024
NIWA Project:	2024032AK

Quality Assurance Statement		
	Reviewed by and approved for release by:	Jonathan Moores Regional Manager, NIWA Auckland.

9 May 2024 5.53 pm

Contents

Executive summary

This investigation follows modelling undertaken by NIWA for Taranaki Regional Council (TRC) reported by Semadeni-Davies (2023) due to concern from TRC that the modelling had overestimated the *E. coli* annual loads and yields within Te Papakura o Taranaki (national park). The modelling was done using a version of the Catchment Land Use for Environmental Sustainability model (CLUES; Elliott *et al.* 2016; Semadeni-Davies *et al.* 2016) that had been re-calibrated with water quality data from Taranaki and Manawatū-Whanganui (Semadeni-Davies *et al.* 2023). The overestimation is due to the combination of the regional parametrisation of the CLUES rainfall delivery factor and extremely high orographic rainfall on the slopes of Taranaki Maunga.

The purpose of this work was to use water quality data collected by TRC from newly established monitoring sites within or in just outside the border of the national park to determine:

a. whether the yields are overestimated and by how much;

b. if simple adjustments to the model can improve model fit for Te Maunga;

c. what the impact of this improvement would be on the outcomes of the scenario modelling, and;

d. whether the adjusted model requires calibration based on the outcomes of b. and c.

TRC provided long-term (2015-2022) *E. coli* data from existing State of Environment (SOE) sites in the region and short-term (2022-23) data from the SOE sites and 11 newly established water quality sites located on the margins of the national park.

The workplan followed was to:

1. Compare measured concentrations taken at monitoring sites close to the national park, including the new sites and SOE sites, against the estimated annual median concentrations determined from Random Forest modelling (Whitehead *et al.* 2022).

2. Estimate mean annual yields[1] from water quality monitoring sites near the national park and compare the yield estimates against CLUES estimated loads for the sites. This has been done to get a sense of the scale of the discrepancies between the yields calculated using CLUES and "actual" yields discharged from the national park.

3. Make approximate adjustments to the model, to try to better represent the measured loads.

4. Run the adjusted model for current conditions to determine whether these adjustments have a significant impact on the total instream yields determined for the region and FMUs as well as for the SOE monitoring sites used for calibration.

5. Rerun the Stage 2 scenarios with the adjusted model to determine whether the adjustment impacts the water quality outputs of the scenario modelling. This was done to assess whether the yields reported in Stage 2 Report are reliable. Rerunning the cost and load target analyses was outside the scope of this work plan.

[1] We use yields in preference to loads since these are normalised for upstream area. Yields are calculated as the instream load divided by the upstream area.

Key findings

While the Random Forest modelling had good fit against long-term water quality data from the SOE sites, the *E. coli* concentrations at the border of the national park were overestimated due to a combination of scaling issues with respect to land use at the park border and the effect of elevation within the national park.

The comparison of the CLUES modelled yields against the long-term yields determined from the SOE data showed that the CLUES yields are lower than those measured. This is because CLUES was calibrated to yields determined using a bootstrapping method based on the 95th percentile flow record. However, yields determined for the full flow record using the same bootstrap method were comparable to the long-term yields.

For the new sites, we established that the CLUES yields could be overestimated by several orders of magnitude, but the uncertainty in the model outputs and the yields estimated from water quality data means we are unable to assess the models performance.

To adjust the model, we capped the maximum mean annual rainfall in the model at 3 m/year, which roughly coincides with rainfall at the boundary of the national park. The adjustment reduced the yields within the national park to values similar to those estimated for the Matemateaonga Range, which is also dominated by native forest. However, at the park border, the yields estimated for the new sites were still too high. For some sites (i.e., Mangaoraka, Pungaereere and Cold Stream), this is due to scaling effects where the model includes pastoral land outside the park's border in the yield calculation. For most of the other sites, the adjusted model yields are within the same order of magnitude.

The adjusted CLUES model reduced the downstream loads of the nearby SOE sites by variable amounts – for most of the sites, the modelled yields were between two and four times greater than the estimated measured yields (note that, there is considerable uncertainty in both the modelled and measured yields). However the estimated load for the more distant SOE sites used for calibration showed very little change due to the impact of loads from pastoral land. We found the adjustment had a minimal effect on yields regionally.

We reran the Stage 2 model scenarios and found that there were only minor differences in the water quality outputs compared to the outputs of the original model for both lower (<=3) and higher order (>= 4) streams. The similarities in results with and without the adjustment are because the same current state attributes were used to characterise baseline conditions and because the adjustment affects only a small part of the region and have little impact on the estimated changes in the NOF bands. Moreover, the downstream effect of the adjustment dissipates with distance and upstream area due to the cumulative effects of loads from the pastoral land that dominates land use on the ring plain. For streams that are not connected to the area affected by the adjustment, there is no difference in the estimated *E. coli* generated or instream yields.

From this work, we have confidence in the outputs of the scenario modelling reported previously (Semadeni-Davies 2023) without re-running the scenarios using the adjusted model. We also conclude that recalibration for the adjusted model is not warranted.

1 Background

This memorandum is an addendum to a modelling study undertaken by NIWA to determine the efficacy of two mitigation strategies to reduce *E. coli* loads from farms in the Taranaki region (Semadeni-Davies 2023). The original modelling was undertaken for Taranaki Regional Council (TRC) using a version of the Catchment Land Use for Environmental Sustainability model (CLUES; Elliott *et al.* 2016; Semadeni-Davies *et al.* 2016) that had been re-calibrated for water quality data from Taranaki and Manawatū-Whanganui (Semadeni-Davies *et al.* 2023). The work addresses TRC's concerns that the *E. coli* yields[2] estimated for subcatchments within Te Papakura o Taranaki (national park) are too high, particularly in comparison to other natural landscapes in the region such as the Matemateaonga Range to the east of the region (Figure 1-1).

The national park largely consists of low yield land covers (i.e., forest with some tussock and bare rock above the tree-line of Taranaki Maunga); the high estimated yields within the national park are an artefact of the model calibration of the model's rainfall and temperature delivery factors. These factors adjust the modelled yields from diffuse sources for climate. Of the two, rainfall has the greatest impact, this factor almost doubles the yield for every metre of rainfall. The elevation of Te Maunga means that the national park has the highest rainfall (i.e., orographic rise) and lowest temperature (adiabatic lapse rate) compared to the rest of the region. From the NIWA climate normal surface used in CLUES[3], the annual rainfall in the national park ranges from 3 metres just outside of the park boundary to 6.5 metres at the summit of Te Maunga. Lower temperatures at high elevation, on the other hand, lowers the estimated *E. coli* loads. The variability of rainfall and temperature on either side of Te Maunga led to noticeable differences between the eastern and western (rain shadow) slopes of Te Maunga. The rainfall on the eastern side of the summit is in the order of 0.5 to 1 m greater and the temperature around 1 to 2 °C warmer compared to the western lee side of Te Maunga.

Calibration of the original model was driven by water quality data from both Taranaki and Manawatū-Whanganui that are not representative of conditions on Te Maunga. While several water quality sites downstream of the national park were used for calibration, they are some distance away and their upstream catchment areas are largely dominated by pastoral land uses. These sites were not flagged as outliers in the calibration and the modelled instream loads at these sites had good agreement with loads calculated from monitored water quality and flow data.

The purpose of this work is to use new water quality data, provided by TRC, collected from 10 newly established monitoring sites within or just outside the border of the national park to determine:

- the downstream impact of the high yields estimated from Te Maunga on estimated downstream loads for the current state;
- the possible effects on the outcomes of the scenario modelling;

[2] In this study, subcatchment yield refers to the number of organisms generated by a unit area (peta organisms / km^2 / year) and is calculated for each subcatchment as the subcatchment generated load (peta organisms / year) delivered to the associated stream segment divided by the subcatchment area. The subcatchment generated load is sum of the loads generated by each land cover (or diffuse source) within the subcatchment. The generated load from each land cover is the product of the yield associated with the land cover and the land cover area modified by the catchment delivery factors. The instream or cumulative load refers to the total number of organisms found within the stream segment area (peta organisms / year) and is the sum of load from upstream sources and the subcatchment less instream losses in *E. coli*. Yields are preferred to loads for model evaluation as they normalise the model outputs for area, this is, large catchments will have a greater load than smaller ones with similar catchment conditions and land use because of contributions from a greater upstream area. Unless otherwise stated, yields used in this report refer to instream yields.

[3] Climate normal from 1991 to 2020.

- whether simple adjustments to the model can improve model fit for Te Maunga and what the impact of this improvement would be on the outcomes of the scenario modelling.

If a need is established on the outcomes of this work, further work that has been discussed with TRC including model recalibration and rerunning the scenario analyses reported in Semadeni-Davies (2023).

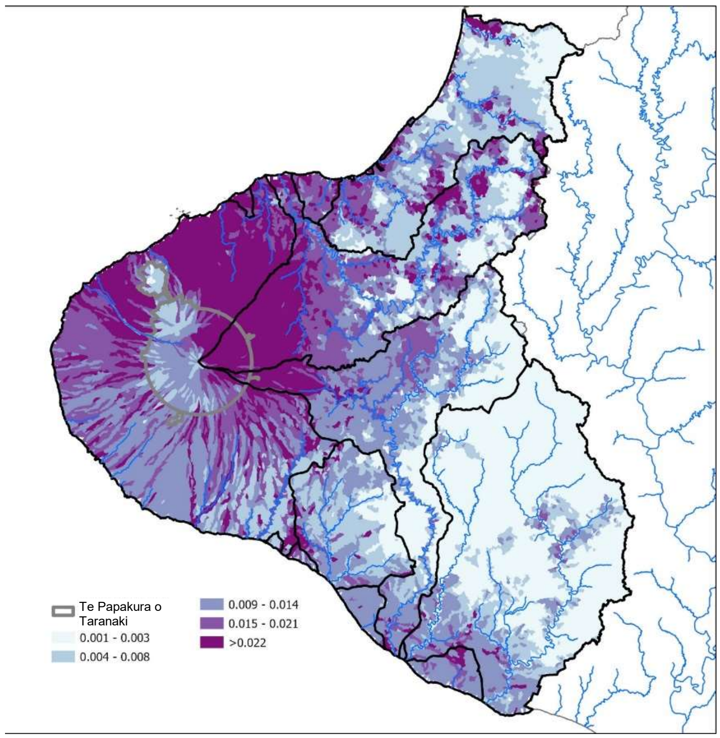


Figure 1-1: Estimated Baseline instream *E. coli* yield (peta/km²/y) mapped by REC2 subcatchment. Quantile distribution. The higher order streams (blue lines), FMU boundaries (bold black lines) and the location of the national park (grey lines) are marked for reference.

The workplan agreed with TRC is outlined below:

1. Comparison of the long term and short-term water quality data provided by TRC to determine if the short-term data collected over a single year is representative of the long term water quality.

2. Comparison of concentrations taken at monitoring sites close to the national park, including the newly established sites, against the estimated annual median concentrations determined from Random Forest modelling. The purpose is to check that the current state concentrations we use to represent the baseline conditions in the national park are in the right ball-park.

3. Estimation of mean annual yields from water quality monitoring sites near the national park and comparison of the load estimates against CLUES estimated loads for the sites. This has been done to get a sense of the scale of the discrepancies between the loads calculated using CLUES and "actual" loads discharged from the national park.

4. Make approximate adjustments to the model, as required, to try to better meet the measured loads.

5. Run the adjusted model for current conditions to determine whether these adjustments have a significant impact on the total instream yields determined for the region and FMUs as well as for the SOE monitoring sites used for calibration.

6. Run the model for the Stage 2 mitigation scenarios to indicate the impact of making adjustments in terms of the National Objective Framework attributes for *E. coli* (New Zealand Government 2023).
   Note that rerunning the cost[4] and load target analyses were outside the scope of this workplan.

[4] Note that the costs associated with the mitigations will not change.

2 Current concentrations and NOF attributes

2.1.1 Comparison of measured and modelled NOF attributes at SOE sites from past modelling

The four *E. coli* NOF attributes are the median and 95th percentile concentrations (C_{50}, C_{95}), and the proportion of exceedances of concentration thresholds of 260 and 540 *E. coli* 100mL^{-1} (G_{260}, G_{540}). The current state estimates of these attributes used in the scenario modelling were taken from random forest of modelling of stream water quality calibrated nationally (Whitehead *et al.* 2022). The modelling was done using State of Environment (SOE) data collected over the period 2016-2020. While regional random forest modelling has been undertaken for TRC data from Taranaki, Manawatu-Whanganui and Waikato (Fraser 2022), the national modelling had better fit for all the metrics for the 13 SOE sited used for calibration in Taranaki (Table 2-1). None of those sites are located within Te Papakura o Taranaki. For the nationally-calibrated *E. coli* models, the most important predictands for all of the *E. coli* metrics are the proportion of upstream intensive agriculture and stock density, mean catchment elevation and slope and the mean catchment coefficient of variation of annual rainfall. Of these, agriculture and stock density have a positive influence on predicted *E. coli* concentrations while the other predictands had a negative influence.

Table 2-1: Comparison of *E. coli* water quality metrics estimated for SOE sites (n = 13) in Taranaki with the metrics modelled using random forest models calibrated nationally and regionally. R^2 is the coefficient of determination and NSE is the Nash Sutcliffe Efficiency

Calibration	Median		C95		G260		G540	
	R^2	NSE	R^2	NSE	R^2	NSE	R^2	NSE
National	0.94	0.82	0.88	0.83	0.97	0.94	0.95	0.89
Regional	0.89	0.77	0.75	0.66	0.89	0.88	0.80	-0.08

Of the NOF *E. coli* attributes, we were only able to calculate median concentrations from the short-term data, for this reason, our analyses below looked only at median concentrations.

2.2 Description of new sites

To determine whether the estimated current state attributes are representative of water quality of streams flowing from the national park, we compared the random forest estimated median annual *E. coli* concentrations with data collected from 10 new sites located at or close to the border of the national park and five "nearby" SOE sites. The nearby SOE and new site locations are mapped in Figure 2-1 and are listed Table 2-2. We did not use data from an eleventh new site, Ōakura River SH45 (OKR000475); while fairly close to the national park as the crow flies, the flow distance is around 11 km and the upstream area of the site is dominated by pasture. For these reasons, we concluded that the site is not representative of the national park. The new sites were established by TRC in June 2022 to better understand the water quality from the national park. NIWA was provided with monthly *E. coli* grab sample data collected through to June 2023.

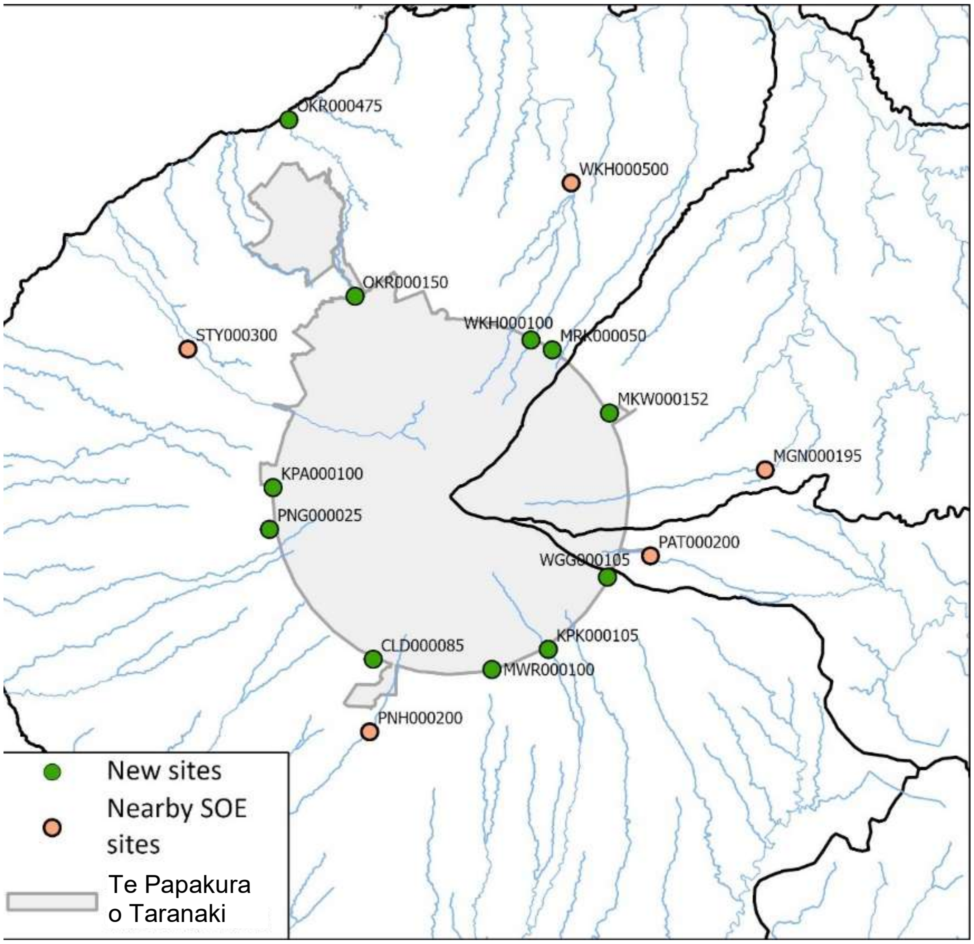


Figure 2-1: Location of newly-established water quality monitoring sites and 'nearby' SOE monitoring sites.

The nearby sites were selected as those within 10 km of the national park boundary. Site names are listed in Table 2.2. The nearby SOE sites were those located within 10 km distance of the boundary on rivers with headwaters flowing from the national park. We used data from the SOE sites to help us to assess whether the random forest estimated concentrations are indicative of the measured concentrations in the vicinity of the park, and whether the short-term record is representative of the long-term record. We considered that the more distant SOE sites downstream of the park would not provide a fair evaluation of the ability of random forest modelling to represent water quality within the park or just downstream of the park given the dominance of upstream pastoral land uses. We note that 10 km is still a fair distance from the national park, however we were limited in choice, and the key point was to identify SOE sites that may be representative of the national park. The closest site to the national park is Pātea River at Barclay Road Bridge which is about 3 km from the park border. The catchment area for this site is largely forest and the riparian margins upstream of the site are wooded. The other SOE sites are located within farmland and their catchment areas are dominated by agriculture.

Table 2-2: Water quality monitoring sites located within 10 km of Te Papakura o Taranaki boundary. The newly-established sites are shaded. Unshaded sites are 'nearby' SOE sites.

SiteName	Site Code
Cold Stream at Kopec Lodge	CLD000085
Kapoaiaia Stream at Bush line	KPA000100
Kaupokonui Stream at Bush line	KPK000105
Maketawa Stream at national park Boundary	MKW000152
Manganui River U/S of Railbridge (And Rumkeg Creek confluence)*	MGN000195 (NRWQN-00035)
Mangaoraka Stream at national park Boundary	MRK000050
Mangawhero Stream at national park Boundary	MWR000100
Ōakura River Carrington Rd.	OKR000150
Ōakura River SH45 (not used in the analysis)	OKR000475
Pātea River Barclay Road Bridge	PAT000200
Punehu Stream Wiremu Rd	PNH000200
Pungaereere Stream at Bush line	PNG000025
Stony River Mangatete Road	STY000300
Waingongoro River 30m Inside national park Boundary.	WGG000105
Waiwhakaiho River SH3	WKH000500
Waiwhakaiho River Waiwhakaiho Track, N.Pk.	WKH000100

*We did not use TRC long term data for this site since there are few samples in the dataset. Instead we substituted in data from the NIWA site, NRWQN_00035, where possible.

The full sampling record provided for the SOE sites covers the period January or April 2015 (depending on the site) to June 2023. The SOE data was split into two sets called the long-term data set containing data up to May 2022, and the short-term data set (or 2022-23) covering the last year of the monitored data for the comparison with data from the new sites. Note that one of the SOE sites (Waiwhakaiho River SH3) was used for CLUES calibration and another (Manganui River u/s of Railbridge) is in the same river reach as a NIWA-operated site (NRWQN-00035) that was also used for calibration (Semadeni-Davies *et al.* 2023). Since the TRC Manganui River u/s of Railbridge has only 24 samples; 12 made between April 2015 and May 2022 and 12 made over the short-term period. Hence, the long-term data from this site cannot be considered representative of the long term monitoring period. For this reason, where possible, we substituted data from the neighbouring NIWA site to represent the long-term.

Because the data from the new sites covers only a one-year period, we were unable to calculate the C95, G260 or G540 attributes for these sites. While we were able to calculate median concentrations, we surmised that these may not be indicative of the long-term *E. coli* concentrations at the sites.

2.3 Comparison between long-term and short-term concentrations

2.3.1 TRC analysis of SOE sites

TRC compared the *E. coli* concentrations collected over the 2022-23 period to the long-term concentrations and loads for all SOE monitoring sites in the region; this is reproduced in Appendix A. It was found that the annual concentrations collected over the 2022-23 year were generally higher than the long-term values across the region. Manganui River u/s of Railbridge had slightly lower *E. coli* concentrations over the 2022-23 year compared to the long-term data, but, as noted above, this site had few samples in the long-term dataset and should be excluded from consideration.

2.3.2 Additional analysis of SOE sites

In addition to the TRC comparison described above, we compared the medians for the nearby by SOE sites calculated over the full monitoring period and the 2022-23 period (Table 2-3). The long and short-term medians for the sites are plotted in Figure 2-2. Note that due to the low number of water quality samples in the TRC record for the Manganui River site, we substituted the long-term median calculated from the provided data with the median concentration calculated for the neighbouring NIWA site (NRWQN-00035_NIWA), this was calculated as part of calibration using water quality data collected for the period 2011-2020 (Semadeni-Davies *et al.* 2023)

Table 2-3: Comparison of median concentrations (cfu / 100 ml) calculated for the SOE sites from data up to June 2022 (long-term) and the 2022-23 year (short-term).

Name and site code	Median	
	Long-term	2022-23
Manganui River U/S of Railbridge (MGN000195)*	75	60
Pātea River at Barclay Road Bridge (PAT000200)	30	40
Punehu Stream Wiremu Rd (PNH000200)	90	200
Stony River at Mangatete Road (STY000300)	11	16
Waiwhakaiho River at SH3(WKH000500)	240	330

*The long-term median for the Manganui River site are from the neighbouring NIWA site (NRWQN-00035).

The long- and short-term medians are fairly close for the Manganui River site, Pātea River and Stony River sites. Generally, the medians are higher for the 2022-23 data period compared with the long-term medians. This is a similar finding to the TRC analysis of geometric mean concentrations across the region. The exception is the Manganui site where the long term median is higher than the short-term – which could reflect the different time period at the site.

From this analysis and the TRC analysis, it seems that the 2022-23 year had generally higher *E. coli* concentrations than the long term average across the region. We speculate that this could be due to the year being warmer and wetter than average[5]. This needs to be borne in mind when comparing modelled and measured values at these the new short-term sites.

[5] https://niwa.co.nz/climate/monthly

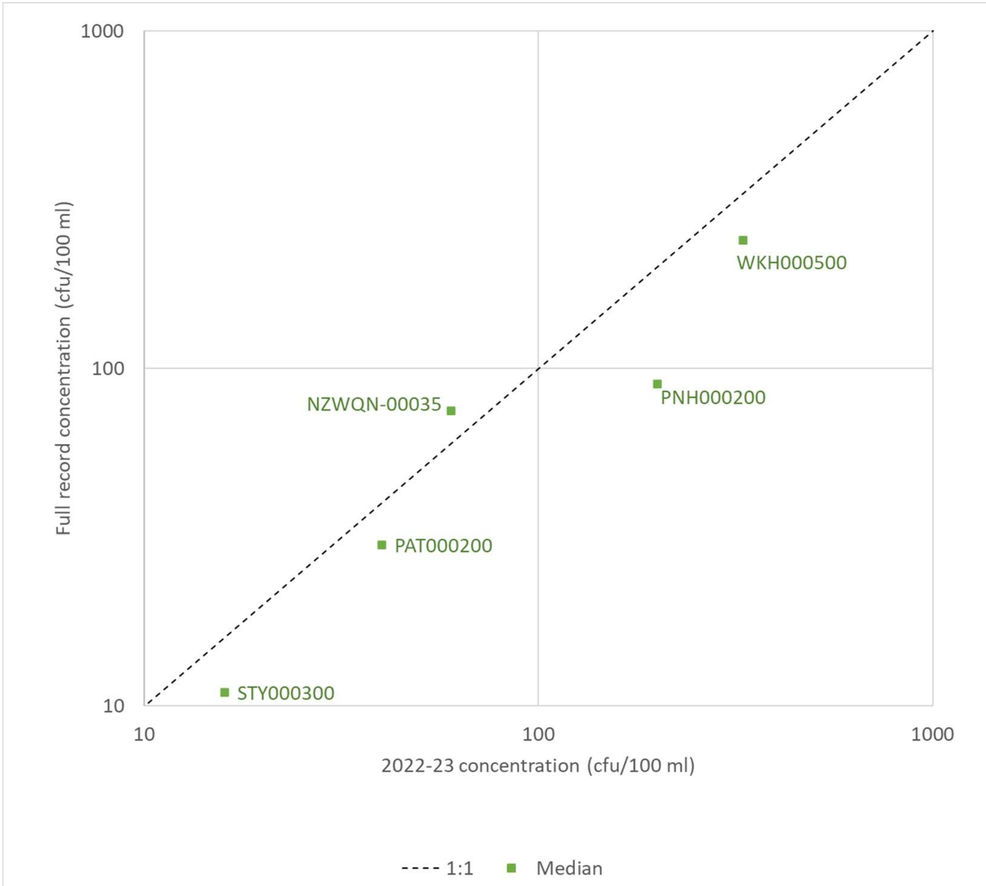


Figure 2-2: Comparison of the long term and short term (2022-23) median concentrations calculated for the nearby SOE sites.

2.4 Comparison of measured and modelled concentrations

2.4.1 SOE sites

There is good agreement between the long-term measured median concentrations for the nearby SOE sites and the Random Forest estimates (Figure 2-3)[6], and the long-term medians for all the sites are in the same median NOF band as the Random Forest estimates. In contrast, the NOF bands for the short-term medians are different for two sites, Punehu Stream at Wiremu Road and Waiwhakaio River at SH3. At both these sites, the measured median concentrations over the 2022-23 period are higher than the modelled concentrations.

[6] The median for the NIWA Manganui site NRWQN-00035 is used to represent the TRC Manganui River U/S of Railbridge site (MGN000195), which is located in the same river segment.

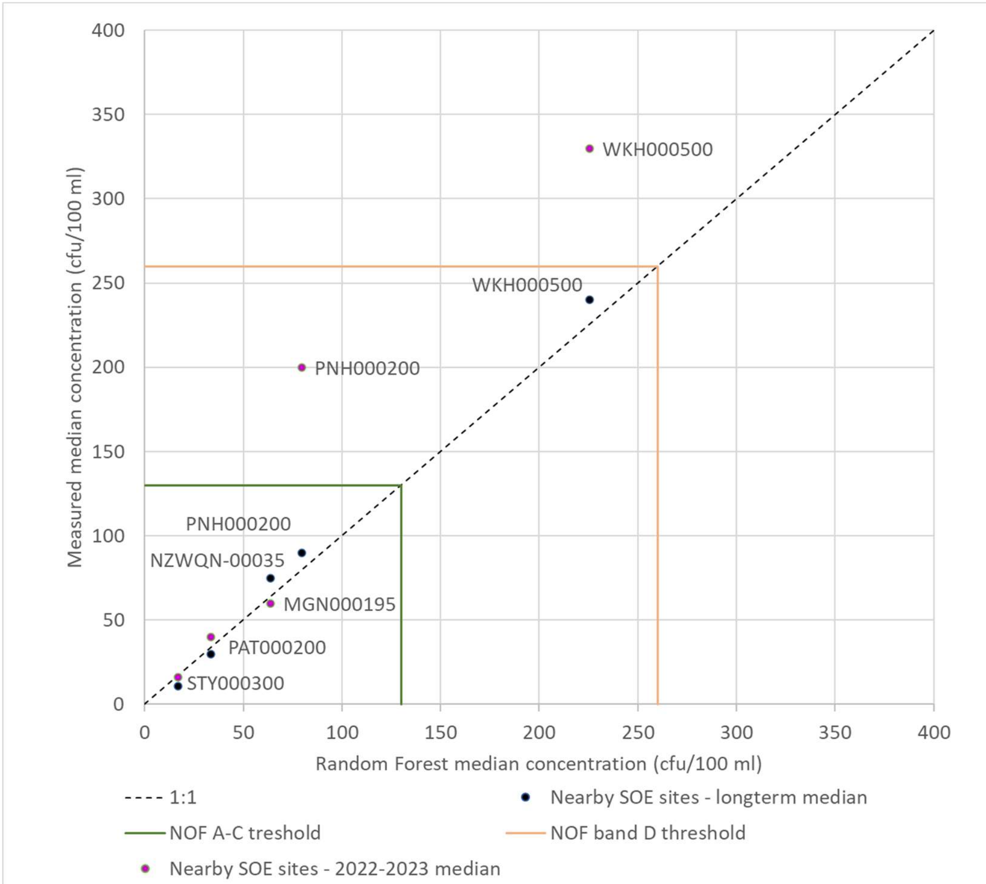


Figure 2-3: Long- and short-term median concentrations for the nearby SOE sites compared with the Random Forest current state median concentrations. The NOF thresholds for median *E. coli* concentrations are marked for reference.

2.4.2 New sites

For the new sites, the modelled Random Forest concentrations are higher than those measured (Figure 2-4), with the exception of one site (Mangawhero Stream at national park boundary). This over-prediction can be considerable (roughly a factor of 3-10). The new sites only have short-term data. Since we found that the long-term concentrations for the SOE sites were generally less than the short-term period (Section 2.3), the Random Forest model is likely to be over-predicting long-term median concentrations at the new sites by an even greater degree than for short-term-concentrations.

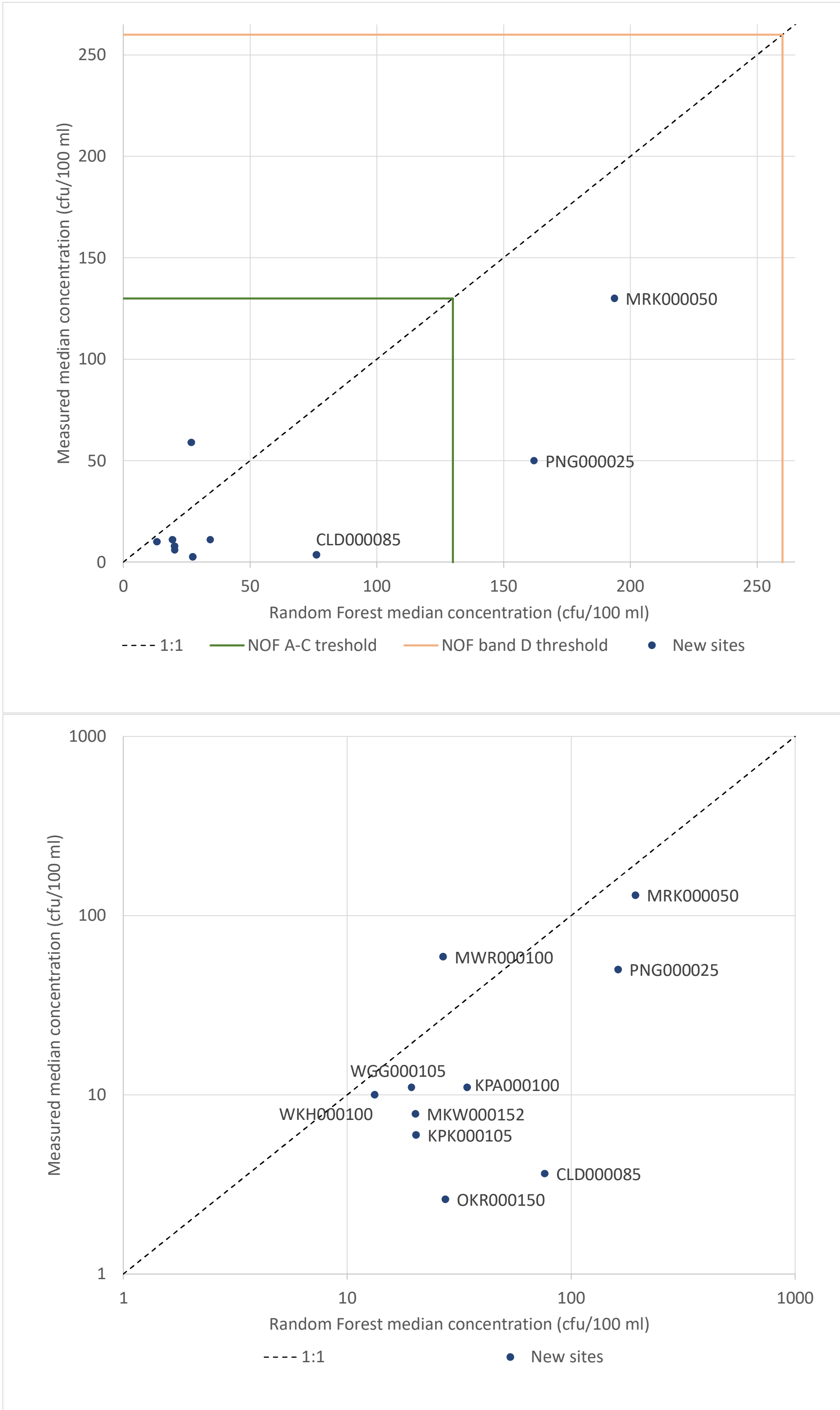


Figure 2-4: Comparison of median concentrations estimated by Random Forest modelling with median concentrations determined for the 2022-23 year for the new sites. Upper plot shows the concentrations in relation to the NOF bands for median concentration; the labelled sites are in mixed pastoral catchments and should be disregarded. The lower plot uses logarithimic axes, but shows the same data as in the first plot.

Part of the overestimation in median concentrations could be explained by the Random Forest model including the effects of pasture for some sites. The model calculations are at the spatial resolution of the REC 2.5 stream network, and it is not possible to estimate the concentration at the exact location of the sites. Instead the model estimates loads at the subcatchment discharge point. This means that while a monitoring site may be close to the bush line and have minimal impacts from pasture in actuality, the model outputs for the sub-catchment the site is located within may include the effects of local pasture (Table 2-4).

Table 2-4: Proportion of upstream catchment area with pastoral land cover. Catchment boundaries are defined by the REC 2.5 stream network. Land cover proportions are based on LCDB5.

Site		Catchment pastoral land cover
Cold Stream At Kopec Lodge	CLD000085	25%
Kapoaiaia Stream at bush line	KPA000100	8%
Kaupokonui Stream at bush line	KPK000105	5%
Maketawa Stream At national park boundary	MKW000152	2%
Mangaoraka Stream at national park boundary	MRK000050	47%
Mangawhero Stream At national park boundary	MWR000100	7%
Ōakura River Carrington Rd.	OKR000150	0%
Pungaereere Stream at bush line	PNG000025	37%
Waingongoro River 30m inside national park boundary.	WGG000105	4%
Waiwhakaiho River Waiwhakaiho Track, N.Pk.	WKH000100	1%

This effect is discernible for sites with even a low proportion of pastoral land, although there are other factors at play as discussed below. The sites with the highest proportions of pastoral land, Mangaoraka Stream (47%), Pungaereere Stream (37%) and Cold Stream (25%), labelled on Figure 2-4, upper plot, are in REC 2.5. headwater subcatchments that straddle the border of the national park. While the measured concentrations are usually low at these sites, there have been occasional high concentrations measured at each that could be linked to farm practices. For example, there is an artificial impoundment near the Mangaoraka Stream site that may be acting as a duck pond, which could be the source of the higher-than-expected measured *E. coli* concentrations (personal communications Thomas McElroy, December 2023). The Pungaereere Stream site has its measured median within NOF A-C band, but the Random Forest median is in band D. This site has high variability in sampled concentrations ranging from ≤10 to 700 cfu / 100 ml. Field staff from TRC (personal communication Josh Dowsing, 12 December 2023) found evidence of cattle at the site. Evidence of cattle was also found at the Cold Stream site, but the measured median concentration is fairly low, however the August 2022 sample had a concentration of 1200 cfu / 100 ml showing that high concentrations are possible at this site. It is conceivable that the other mixed catchments may also be accessible to stock.

While the presence of farmland in the model's catchment area will increase the modelled concentrations and explain part of the overestimation in some cases, the general overestimation of concentrations can also be partially due to the influence of elevation, which is one of the key predictors for *E. coli*, in the random forest model. A partial plot of elevation against modelled median concentration (Figure 4-4 in Whitehead *et al.* 2022, reproduced in Appendix B) shows a sharp change in median concentrations at an elevation of around 500 masl. Mapping the Random Forest median concentrations shows a commensurate jump in concentrations from <20 cfu/100 ml above 500 masl (which is in the same ballpark as the measured median concentrations for the new sites) to between 80-100 cfu/120 ml at the border of the national park (Figure 2-5), where the elevation ranges from 500 masl to the east and 400 masl to the west.

For the new sites, the measured and modelled NOF bands (based on median concentrations) are the same apart from Pungaereere Stream where the measured median is in the NOF A-C band and the modelled median is in the NOF D band. The measured and modelled median concentrations for Mangaoraka Stream are both within the NOF D band. As noted above, both of the Pungaereere Stream and Mangaoraka Stream sites are in mixed land use subcatchments. The measured and modelled median concentrations for the other sites are in the NOF A-C.

While the modelled median concentrations are higher than the measured values at the park border, the fact that: a) the NOF bands are the same for both the measured and modelled concentrations for most of the new sites; and b) there is good agreement between the modelled and measured median concentrations for the SOE sites across the region for all the NOF *E. coli* attributes, leads us to conclude that the outputs of the random forest modelling calibrated nationally can give an adequate representation of the current state of *E. coli* in Taranaki regionally. That is, downstream of Te Maunga, as upstream catchment size increases the fit between the measured and modelled median concentrations improves.

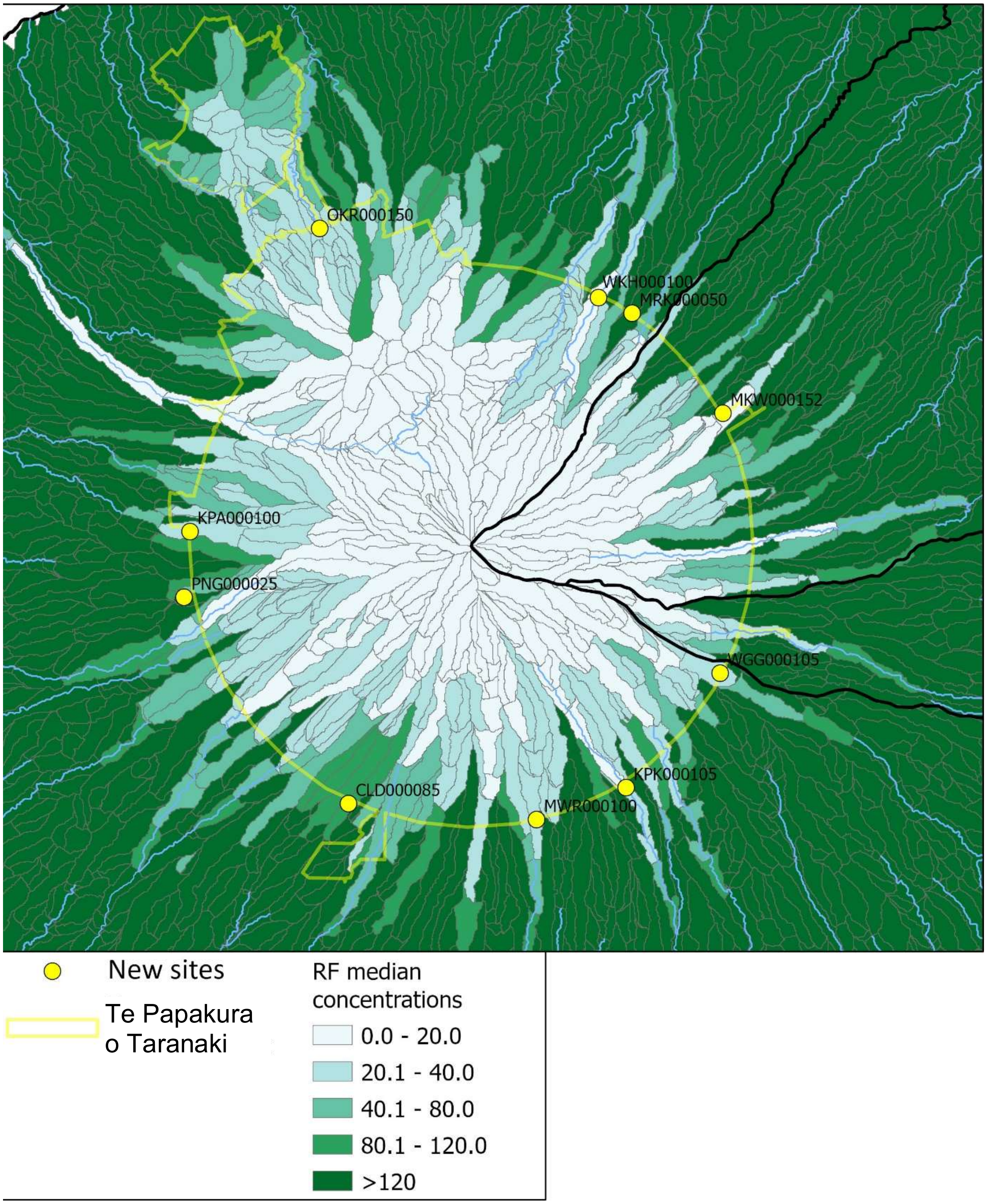


Figure 2-5: Random Forest estimates of median *E. coli* concentrations (cfu / 100 ml) within and at the border of Te Papakura o Taranaki. REC reaches are delineated for reference.

3 Loads and yields

In this section we compare loads from the CLUES model to the median loads and yields derived from data provided by TRC for the new sites and the nearby SOE sites. In all cases, the loads and yields refer to instream (cumulative) values rather than generated (within subcatchment values). We also compare the CLUES loads to the 'measured' loads determined for CLUES calibration for the two nearby SOE sites with continuous flow record (i.e., Waiwhakaiho River SH3 and Manganui River u/s of Railbridge). It is important to note that the loads derived from the TRC data were calculated using different methods to the loads used for calibration.

3.1 Mean annual loads and yields used for CLUES calibration

The CLUES model was calibrated against loads calculated using long-term *E. coli* concentrations from SOE sites in Taranaki and Manawatu-Whanganui that have concurrent continuous flow measurement. There were nine calibration sites from Taranaki including two that are less than 10 km from the national park boundary (Waiwhakaiho River SH3 and NRWQN-00035, which is next to the TRC Manganui River u/s of Railbridge site).

The method used is fully described in Semadeni-Davies *et al.* (2023). To summarise, we used a rating curve method to determine a relationship between *E. coli* concentration and flow rate followed by applying the relationship to the flow record to obtain estimates of the mean annual load, all with a bootstrapping (re-sampling) approach. We determined mean annual loads for the full flow record and to the 95th percentile flow record whereby the top 5% of flow rates were removed from the record.

The yields calculated with the 95th percentile flow record were ultimately used for calibration to: a) reflect that the NOF criteria are related to the median and 95th percentile concentrations; b) avoid large errors associated with estimating concentrations and loads at high flows for which there are little to no *E. coli* measurements; and c) reflect that load reductions in the lower flow range are more likely to influence median and 95th percentile concentrations, compared with removing storm loads. We used yields for calibration in preference to loads because they normalise for upstream area and avoid the inherent relation between catchment area and load.

The CLUES modelled loads and yields calculated with the 95th percentile flow record and with the full flow record for calibration sites in Taranaki are shown Table 3-1. The CLUES estimated loads and yields show good agreement with loads and yields determined for calibration.

3.2 Average loads derived from TRC data

Instantaneous loads were calculated by TRC for each of the *E. coli* concentration samples. Flow for the calculations was taken from continuous flow records for the Waiwhakaiho River SH3 and Manganui River u/s of Railbridge SOE sites and from site gaugings made at the time of sampling for all other sites. The TRC comparison of the long- and short-term loads for the SOE sites is reproduced in Appendix A).

Table 3-1: Taranaki measured and modelled mean annual loads and yields determined for the calibration monitoring sites. The shaded sites are within 10 km of the national park boundary. Modified from Appendix D of Semadeni-Davies *et al.* (2023). Measured loads and yields used for calibration were calculated with the 95 percentile flow record, loads and yields calculated with the full flow record are shown in parentheses for reference.

TRC site ID	Site name	LAWA ID used in calibration report	Load (peta organisms/y) Measured		Yield (peta organisms/km^2/y	
			Measured	Modelled	Measured	Modelled
MGN000195	WA2 Manganui at SH3	NRWQN-00035_NIWA	0.24 (0.62)	0.40	0.0166 (0.0418)	0.0272
WTR000800	WA1 Waitara at Bertrand Rd	NRWQN-00036_NIWA	25.99 (91.07)	18.76	0.0233 (0.082)	0.0168
MGH000950	Mangaehu at Raupuha Rd Bridge	TRC-00001	2.19 (5.64)	2.87	0.0053 (0.014)	0.0069
MRK000420	Mangaoraka at Corbett Rd	TRC-00003	1.1 (2.28)	1.58	0.0204 (0.042)	0.0294
PAT000360	Pātea at Skinner Rd	TRC-00005	3.01 (15.13)	2.39	0.0372 (0.187)	0.0295
WGG000500	Waingongoro at Eltham Rd Bridge	TRC-00009	0.85 (3.31)	0.85	0.0168 (0.066)	0.0169
WGG000900	Waingongoro at SH45	TRC-00010	1.8 (3.77)	2.82	0.008 (0.017)	0.0125
WKH000500	Waiwhakaiho at SH3	TRC-00011	3.12 (11.61)	2.52	0.0518 (0.193)	0.0418

3.2.1 Comparison of long and short term instantaneous yields

The mean yields calculated using the full record and the 2022-23 year are plotted in Figure 3-1. We chose the arithmetic mean values because these are the values modelled by CLUES[7]. The Manganui River at Railbridge site is not included in the analysis because we were unable to calculate a reliable value for the site from the long term data provided as noted above.

There are large differences in the mean yields calculated over the two time periods for most of the sites. The long-term mean for Stony River, for example, is more than 20 times larger than for the 2022-23 period. However, the high mean was driven by the high concentration event for this site cited above, which coincided with a high flow event. The long-term mean yield is around six times higher than that calculated for 2022-23 for Waiwhakaiho River at SH3, which can be attributed to three events with very high loads (e.g., 8 April 2015 and 8 November 2017). While the 2022-23 year had higher concentrations as evidenced above, the yields were not as high due to the absence of extreme events. This demonstrates the difficulty in calculating mean annual loads and yields from short-term data.

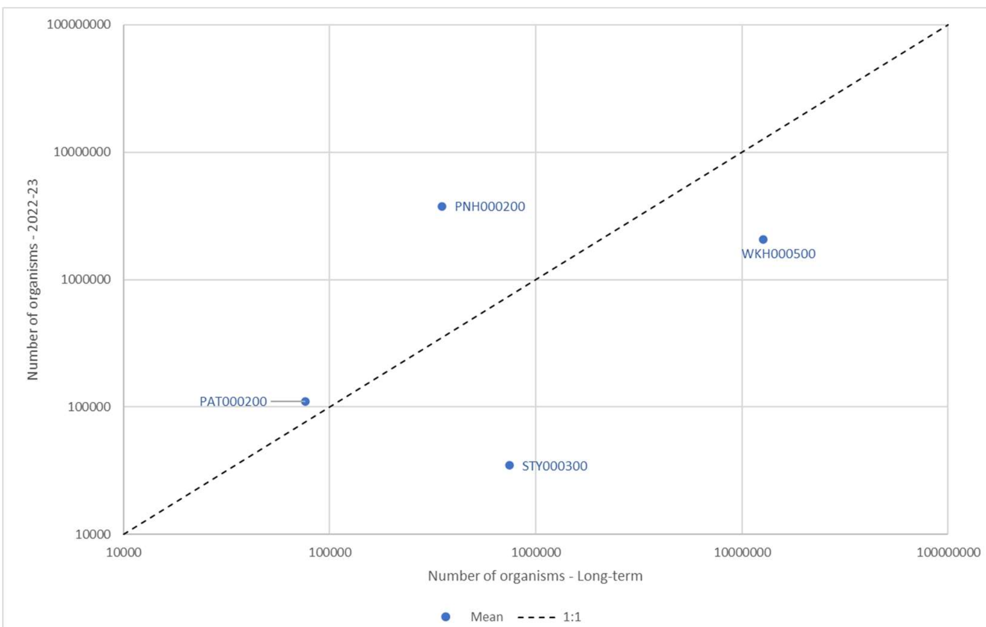


Figure 3-1: Comparison of mean yields calculated for the nearby SOE sites over the full and short time periods.

3.3 Additional methods for mean annual loads and yields

We used two methods of calculating annual loads from the *E. coli* concentration and instantaneous flow data provided by TRC:

1. Multiplying the mean of the instantaneous loads for each site by the number of seconds in a year;

[7] The arithmetic mean (rather than median or geometric mean) is the appropriate metric for annual average loads because these are essentially the average of the of daily loads over a long time series.

2. Dividing the sum of the instantaneous loads by the sum of the instantaneous flows for each site and multiplying by the estimated mean annual flow used by the CLUES model and by the number of seconds in a year;

Both methods gave comparable results as discussed below.

3.3.1 Long-term for SOE sites

Figure 3-2 shows the long-term yields determined using both methods against the yields predicted by CLUES. The bootstrapped yields determined for CLUES calibration are also given for reference for the Waiwhakaiho River SH3 and Manganui River u/s of Railbridge sites. The yields calculated with the long-term data were generally higher than estimated by CLUES, which is likely due to the fact that CLUES is calibrated to yields calculated from the 95th percentile flow record. That is, the data from TRC contains high flow events that are not represented in CLUES. For the two calibration sites (Manganui River and Waiwhakaiho River), the 95th percentile flows were exceeded several times by the instantaneous flows.

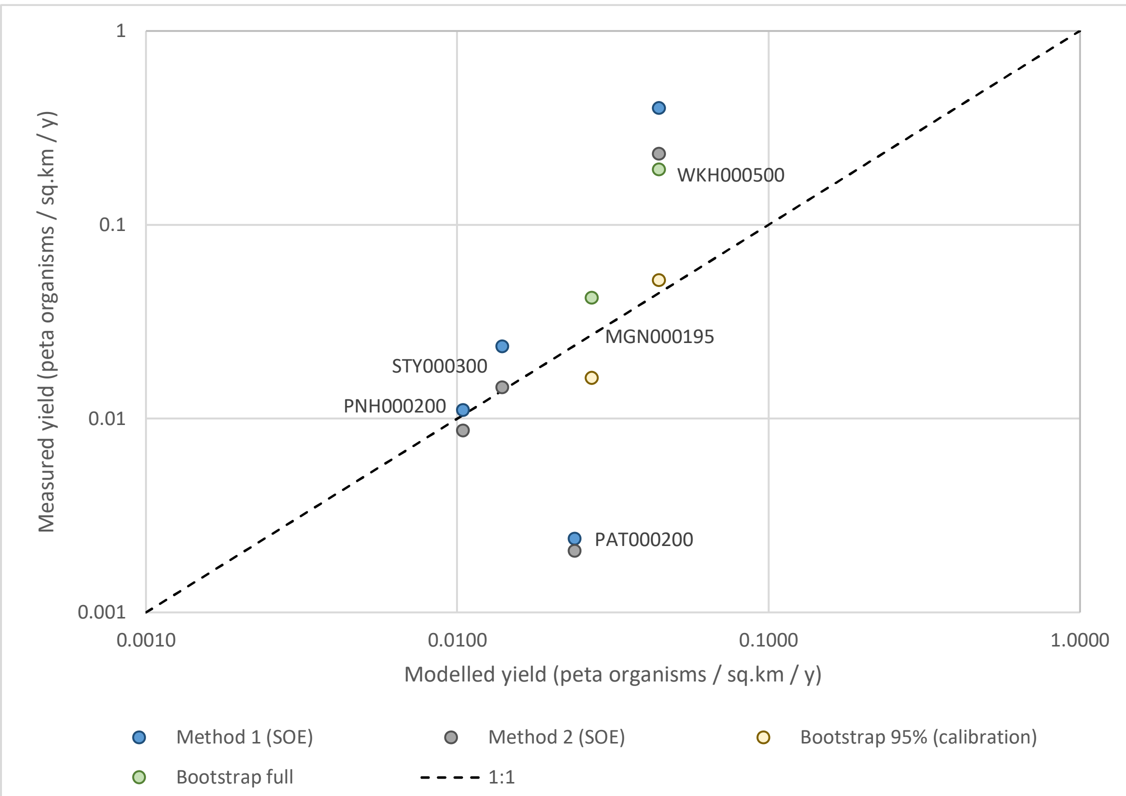


Figure 3-2: Comparison between the long-term annual yields calculated from water quality data and yields predicted by the CLUES model for SOE sites near the national park. The yields obtained from the bootstrapping procedure are shown for reference.

The exception was the Pātea River at Barclay Road Bridge site where the CLUES modelled yield is around 10 times higher than that estimated from the long-term data. This site is very close to the national park boundary and illustrates the suspected over-prediction of *E. coli* yields from the national park. In contrast, the Pātea River at Skinner Road site, which is located in pastoral land about 13 km downstream from the Barclay Road site, has a modelled yield that is lower than the measured yield (see Table 3-1). This suggests that the high generated yields estimated within the national park do not overly affect downstream model outputs.

To further demonstrate the differences between the yields estimated using the TRC long-term data and the CLUES yields, we also compared the yields from Methods 1 and 2 against the bootstrap yields obtained using the full flow record and the 95th percentile flow record for the other calibration sites in Taranaki (Table 3-2). The bootstrap yields calculated with the full flow record are closer to the yields estimated using Methods 1 and 2 than those calculated with the 95th percentile flow record, these are generally much lower (about 2 – 3 times).

Table 3-2: Comparison of yields determined from the long term data supplied by TRC and yields determined using the bootstrap method as part of CLUES calibration. CLUES modelled yields are also provided for reference.

SiteName	TRC and LAWA site codes	Method 1	Method 2	CLUES	Bootstrap 95th perc. flow record	Bootstrap Full flow record
Mangaehu River Raupuha Rd Bridge	MGH000950 (TRC-00001)	8.66	11.82	2.89	2.19	5.64
Manganui River U/S of Railbridge	MGN000195 (NRWQN-00035)	0.42	0.39	0.40	0.24	0.62
Mangaoraka Stream Corbett Rd	MRK000420 (TRC-00003)	5.93	4.40	1.62	1.10	2.28
Pātea River Skinner Road Bridge	PAT000360 (TRC-00005)	5.98	4.35	2.61	3.01	15.13
Waingongoro River Eltham Rd Bridge	WGG000500 (TRC-00009)	2.20	2.02	1.00	0.85	3.31
Waingongoro River SH45	WGG000900 (TRC-00010)	3.87	4.67	3.51	1.80	3.77
Waitara River Bertrand Road	WTR000800 (NRWQN-00036)	121.58	100.96	19.61	25.99	91.07
Waiwhakaiho River SH3	WKH000500 (TRC-00011)	24.06	14.00	2.68	3.12	11.61

3.3.2 New sites

Figure 3-3 shows the annual yields calculated using the two methods for the 2022-23 period against the CLUES predicted yields for both the nearby SOE sites and the new sites. The modelled and measured mean annual yields for the new sites are tabulated in Table 3-3. The key messages from Figure 3-3 are:

- While the yields estimated for the new sites using both methods are in the same range as the modelled yields from CLUES for the Matemateaonga Range to the east of the region (i.e., < 0.003 peta organisms / km^2 / year), the yields estimated using Method 1 are consistently higher than those estimated using Method 2.

Table 3-3: CLUES modelled yields and yields estimated from measured data for the new sites (peta organisms / km^2 / y). The shaded sites are those identified above as having a high proportion of mixed pastoral land cover.

Site		CLUES	Method 1	Method 2
Cold Stream At Kopec Lodge	CLD000085	0.0116	0.0050	0.0016
Kapoaiaia Stream at bush line	KPA000100	0.0109	0.0013	0.0008
Kaupokonui Stream at bush line	KPK000105	0.0148	0.0002	0.0003
Maketawa Stream At national park boundary	MKW000152	0.0387	0.0030	0.0009
Mangaoraka Stream at national park boundary	MRK000050	0.0669	0.0038	0.0151
Mangawhero Stream At national park boundary	MWR000100	0.0119	0.0018	0.0011
Ōakura River Carrington Rd.	OKR000150	0.0059	0.0006	0.0005
Pungaereere Stream at bush line	PNG000025	0.0163	0.0073	0.0038
Waingongoro River 30m inside national park boundary.	WGG000105	0.0256	0.0007	0.0012
Waiwhakaiho River Waiwhakaiho Track, N.Pk.	WKH000100	0.0404	0.0043	0.0014

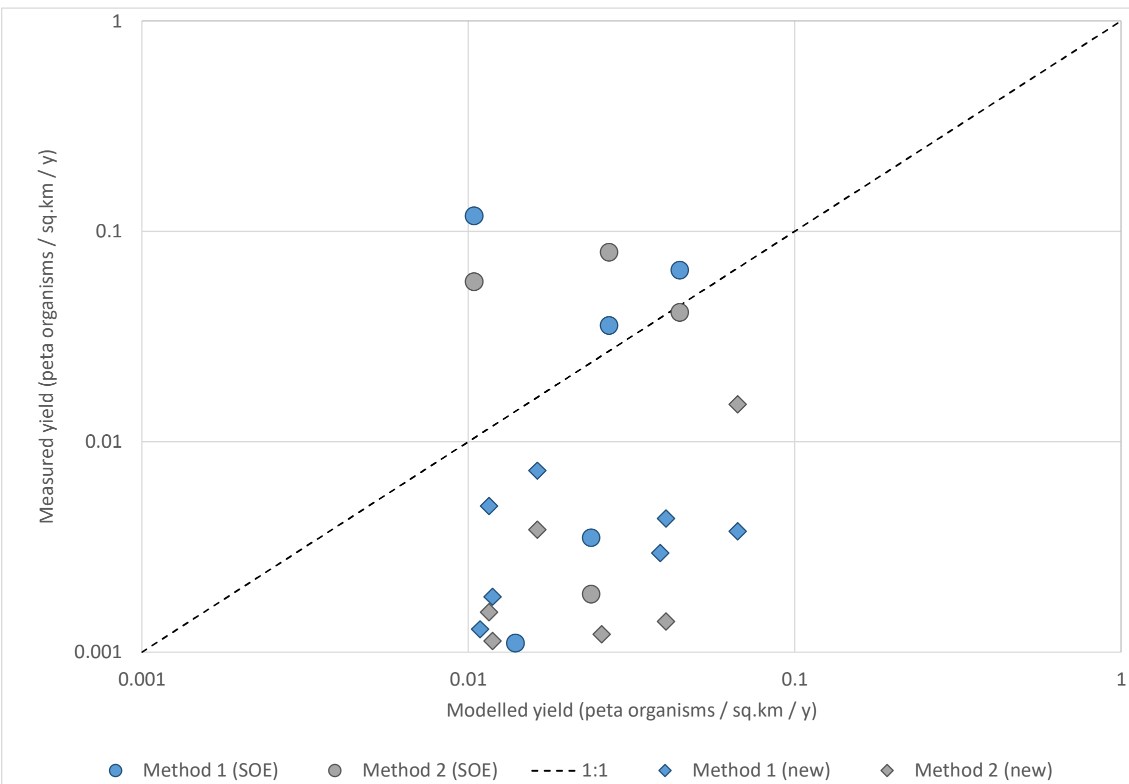


Figure 3-3: Comparison of 2022-23 annual yields calculated from water quality data and predicted by the CLUES model. Circles represent the nearby SOE monitoring sites while diamonds represent the new sites.

- There is variability in the yields determined using both methods between sites. For Method 1, the range is between 0.0002 and 0.0073 peta organisms / km^2, the range for Method 2 is between 0.0003 and 0.0151 peta organisms / km^2. The high yields determined for the Mangaoraka, Pungaereere and Cold Stream sites are likely a result of high concentrations linked to agriculture as discussed in Section 2.4.2. The lowest yields are for the Kaupokonui Stream and Ōakura Rivers sites.
- There is likewise variability in the CLUES yields which are due to varying land cover, estimated flow rates, rainfall and temperature.
- With the exception of the Pātea River at Barclay Road Bridge site, which is close to the national park, all of the yields estimated using Methods 1 and 2 for the SOE sites are higher than the CLUES predicted yields.
- The CLUES estimated yields are higher compared to the yields estimated using Methods 1 and 2 for the new sites and for the SOE Pātea River at Barclay Road site. The difference for most sites is an order of magnitude compared to both the Method 1 and 2 yields, however, the difference generally greater for Method 2. The greatest relative difference was for the Kaupokonui Stream site where the modelled load was 74 times higher than the Method 1 yield and 49 times higher than the Method 2 yield.
- The presence of pastoral land in the REC subcatchment some of the new sites are located within, especially the Mangaoraka, Pungaereere and Cold Stream sites, can partially explain the overestimation of yields by CLUES for those sites.

The analysis confirms the concern raised by TRC that the yields estimated by CLUES within the national park are too high. The overestimation is likely to be greater than shown in Figure 3-3 and Table 3-3 since the CLUES yields were calibrated to the bootstrap yields determined using the 95th percentile flow record. However, with only one year of data and variability in the yields calculated using Methods 1 and 2, it is not possible for us to determine the scale of overestimation reliably, but it could be in the range of two to three orders of magnitude. It was not possible for us to determine mean annual yields using the same bootstrap method for the ungauged SOE sites or the new sites. While we were provided with continuous flow data for some sites, these data were assessed from flow records taken at gauged sites some distance away.

4 Adjustment of the CLUES model

We made a simple adjustment to the CLUES model to reduce the effect of rainfall on the estimated yields from Te Papakura o Taranaki. The model was adjusted by capping the mean annual rainfall at 3 m per year, which is the approximate rainfall at the border of the park boundary within the CLUES model (Figure 4-1). This rainfall roughly coincides with the 400 masl elevation contour line and the national park boundary. The adjustment has the effect in CLUES of limiting the impact of the high orographic rainfalls. Since the rain-cap roughly corresponds with the park boundaries, the change will have little effect on the generated yields in the rest of the region with the exception of an area about 3-6 km wide to the east of Te Maunga abutting the park. We also tried a second adjustment that also capped temperature, but this resulted in outputs that were very close to capping just rainfall. We decided not to set a fixed yield for the national park because of the variability in the yields determined for the new sites. The outputs of the adjusted model are summarised in Table 4-1 and Figure 4-2.

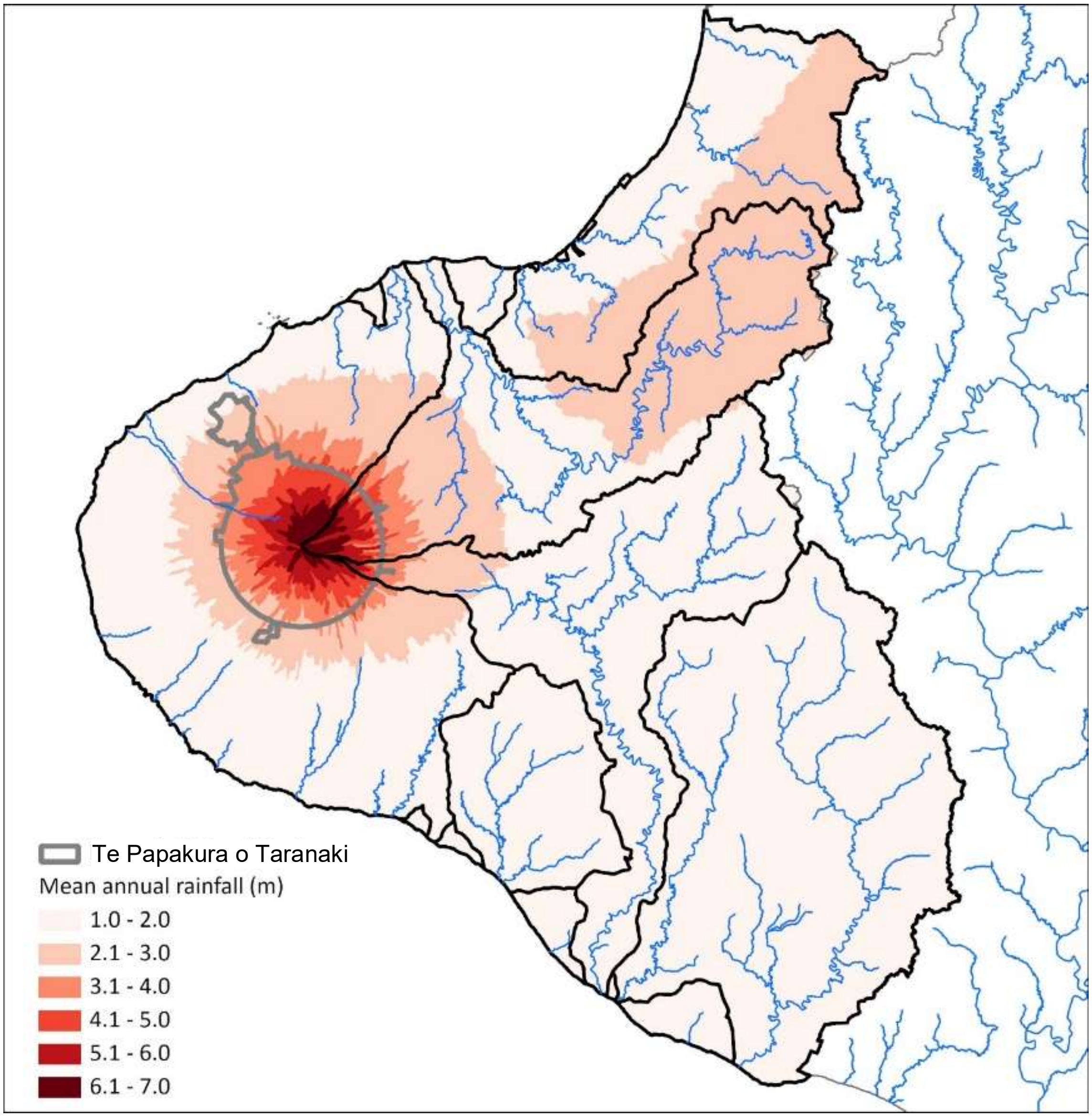


Figure 4-1: Mean annual rainfall (m) aggregated by subcatchment used in the CLUES model.

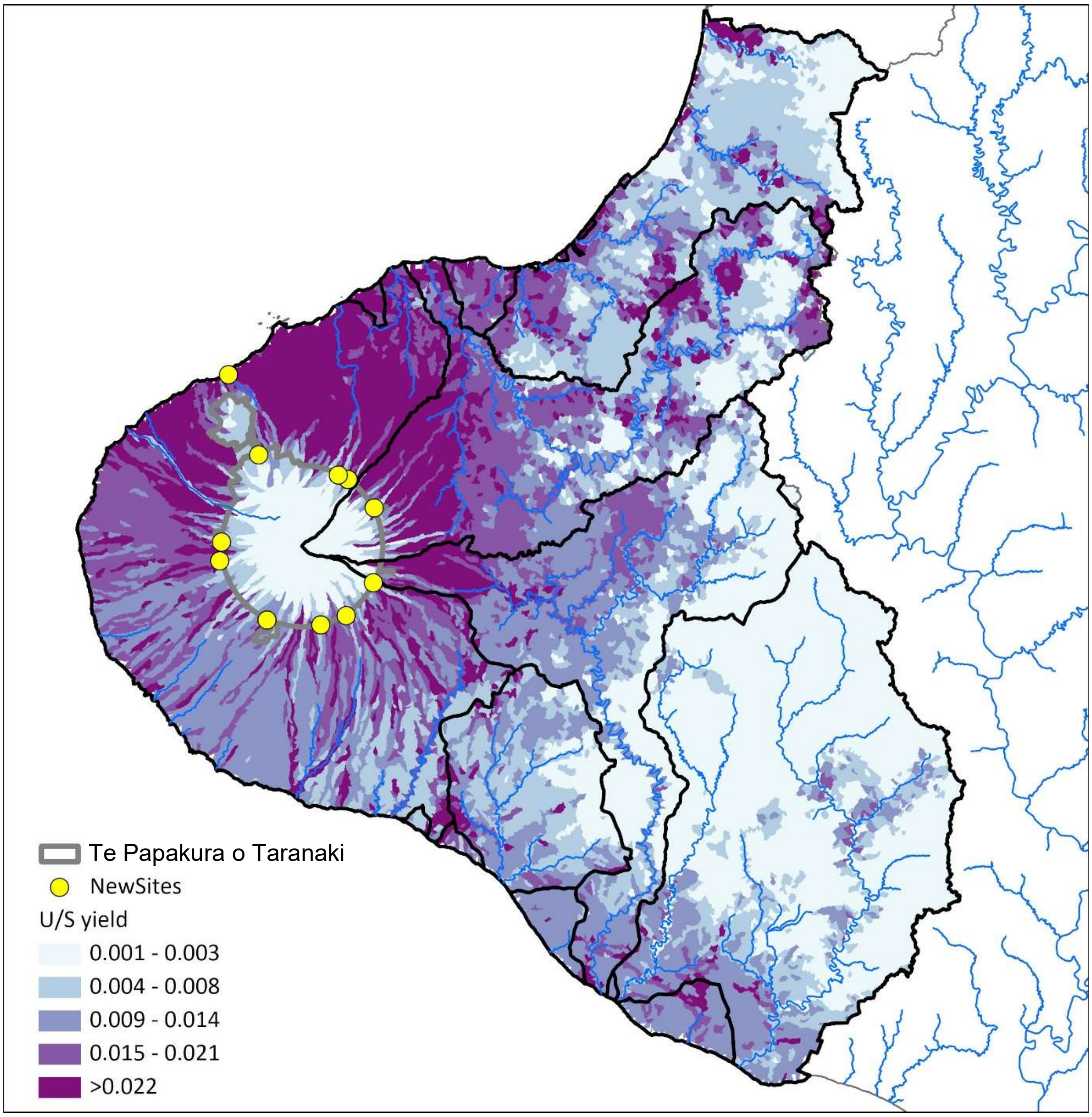


Figure 4-2: Baseline instream *E. coli* yields (peta organisms /km²/y) calculated with the adjusted model

Table 4-1: Original and adjusted CLUES modelled yields compared to yields estimated from measured data for the new sites (peta organisms / km^2 / y). Sites with high pastoral land cover are shaded.

Site		CLUES	CLUES adjusted	Method 1	Method 2
Cold Stream At Kopec Lodge	CLD000085	0.0116	0.0102	0.0050	0.0016
Kapoaiaia Stream at bush line	KPA000100	0.0109	0.0056	0.0013	0.0008
Kaupokonui Stream at bush line	KPK000105	0.0148	0.0035	0.0002	0.0003
Maketawa Stream At national park boundary	MKW000152	0.0387	0.0029	0.0030	0.0009
Mangaoraka Stream at national park boundary	MRK000050	0.0669	0.0171	0.0038	0.0151
Mangawhero Stream At national park boundary	MWR000100	0.0119	0.0042	0.0018	0.0011
Ōakura River Carrington Rd.	OKR000150	0.0059	0.0034	0.0006	0.0005
Pungaereere Stream at bush line	PNG000025	0.0163	0.0163	0.0073	0.0038
Waingongoro River 30m inside national park boundary.	WGG000105	0.0256	0.0029	0.0007	0.0012
Waiwhakaiho River Waiwhakaiho Track, N.Pk.	WKH000100	0.0404	0.0025	0.0043	0.0014

Given the uncertainty of the short-term yield estimates for the new sites, it is not possible to determine how well the adjusted model performs. It is likely that the model is still overpredicting the yields by around an order of magnitude compared to the yields determined using both Method 1 and 2, especially given that the 2022-23 year had higher concentrations than the long-term data across the region. This means that the discussion below should not be viewed as an evaluation of model fit but rather a generalised order of magnitude check.

The CLUES yields estimated with rainfall capped for subcatchments wholly within the national park ranges between 0.001 and 0.006 peta organisms / km^2/ year – this range is comparable to the yields modelled for the Matemateaonga Range. However at the margins of the national park where the new sites are located, the adjusted CLUES yields are between a factor of 2 to 4 times higher than those estimated using Methods 1 and 2 for most of the sites. For the Kaupokonui Stream site, which has very low estimated yields, the adjusted CLUES yields, while significantly reduced, are still much higher than those estimated using Methods 1 and 2 - 17 times for Method 1 and 12 times for Method 2. For the Waiwhakaiho River and Maketawa Stream sites, the adjusted CLUES yields are close to the yield estimated from one or other of the two calculation methods, but the fact that CLUES was calibrated to the yield determined from the 95th percentile means that the yields may still to be too high at these sites.

The adjustment caused a reduction in the estimated instream yields for all the nearby SOE sites, especially to the east where rainfalls >3 m / year extend beyond the park boundary. The adjusted yield for Pātea River at Barclay Road Bridge site is much closer to the yield determined from the measured data (around 0.002 peta organisms / km^2 / y for both Methods 1 and 2). This site is close to the park boundary and its catchment area is dominated by forested land cover (85%).

For the two calibration sites within 10 km of the national park, the estimated load at the Manganui River U/S of Railbridge, while underestimated, is closer to the calibration yield. The updated yield for the Waiwahaiho River at SH3 site is less than half the calibration yield (0.0210 vs 0.0518 peta organisms / km^2 / y).

The adjustment had very little effect on the instream yields at the more distant calibration sites downstream of the national park. For example, the yield modelled for the Pātea Road at Skinner site, which is about 13 km from the park border, remains at around 0.03 peta organisms / km^2 / year. The limited change in modelled yields at the calibration sites means that recalibration of the model is not warranted, especially since most of the calibration sites are in in Manawatū-Whanganui and are not affected by the adjustment.

While it is likely that the adjusted CLUES yields may still be too high by around an order of magnitude for the national park, we are unable to determine by how much, which precludes any further adjustments to the model at this stage.

Table 4-2: CLUES estimated *E. coli* yields (peta organisms / km2 / y) with and without the rainfall cap adjustment. The bootstrap load determined using the 95th percentile flow record for the two calibration sites are given for reference.

Site	CLUES Baseline	CLUES adjusted	Calibration (bootstrap)
Manganui River U/S of Railbridge (MGN000195)*	0.0272	0.0077	0.0162*
Pātea River Barclay Road Bridge (PAT000200)	0.0238	0.0056	
Punehu Stream Wiremu Rd (PNH000200)	0.0104	0.0071	
Stony River Mangatete Road (STY000300)	0.0139	0.0041	
Waiwhakaiho River SH3 (WKH000500)	0.0418	0.0210	0.0518

*Determined using data from the neighbouring NIWA Manganui at SH3 site

5 Scenario re-run

To determine the possible impact of adjusting the model, we reran the Stage 2 report (Semadeni-Davies 2023) scenarios and compared the scenario outputs the original scenario outputs. To do this, we first replaced the baseline yields used in the original model with the baseline yields determined with the adjusted model and then ran the model with the same settings as the original scenario runs. Output files from all the scenarios reruns have been supplied to TRC as supplementary files to this memo.

All the scenario reruns showed only minor differences compared to the original model runs for both lower (<=3) and higher order (>= 4) streams. To illustrate, here we report on the outputs of Scenario 1 (stock exclusion) run with the highest removal efficiencies. Both the original and adjusted versions of the model resulted in a 4% drop in loads compared to their relative baselines across the region. The scenario outputs for higher order streams (i.e., streams with an order or four or greater) are summarised in Table 5-1 through to Table 5-4. Table 5-1 and Table 5-3 have been copied from the Stage 2 report while Table 5-2 and Table 5-4 show the equivalent outputs of the adjusted model. The output tables with the 95th percentile concentration attribute (C_{95}) are reproduced in Table 5-1 and Table 5-2 and without C_{95} in Table 5-3 and Table 5-4. With the C_{95} attribute, there is a small increase in the length of higher order streams classed as Band D instead of Band E in the Waitara FMU. Without the C_{95} attribute, as well as differences in the Waitara FMU, there are minor differences estimated for the Volcanic Ringplain FMU. Similarly (not shown here), there are only minor differences in the NOF bands estimated for low order streams.

Table 5-1: Original model: Change in the length (km) of higher order streams compared to the current-state scenario in each NOF band determined for Scenario 1 (high removal efficiency) with C_{95} included in the band calculation. The percentage of the total stream length in each band is in parentheses.

FMU	A	B	C	D	E
Coastal Terraces (north)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Coastal Terraces (south)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Northern Hill Country	0 (0%)	0 (0%)	1 (0.5%)	-1 (-0.5%)	0 (0%)
Pātea	0 (0%)	0 (0%)	0 (0%)	5 (2%)	-5 (-2%)
Southern Hill Country (east)	0 (0%)	0 (0%)	0 (0%)	1 (0.3%)	-1 (-0.3%)
Southern Hill Country (west)	0 (0%)	0 (0%)	0 (0%)	0 (0.4%)	0 (-0.4%)
Volcanic Ringplain	0 (0%)	0 (0%)	0 (0%)	3 (1.3%)	-3 (-1.3%)
Waitara	0 (0%)	0 (0%)	0 (0%)	29 (9.4%)	-29 (-9.4%)
Taranaki Total	**0 (0%)**	**0 (0%)**	**1 (0.1%)**	**39 (2.7%)**	**-39 (-2.7%)**

Table 5-2: Adjusted model: Change in the length (km) of higher order streams compared to the current-state scenario in each NOF band determined for Scenario 1 (high removal efficiency) with C_{95} included in the band calculation. The percentage of the total stream length in each band is in parentheses.

FMU	A	B	C	D	E
Coastal Terraces (north)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Coastal Terraces (south)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Northern Hill Country	0 (0%)	0 (0%)	1 (0.5%)	-1 (-0.5%)	0 (0%)
Pātea	0 (0%)	0 (0%)	0 (0%)	5 (1.8%)	-5 (-1.8%)
Southern Hill Country (east)	0 (0%)	0 (0%)	0 (0%)	1 (0.3%)	-1 (-0.3%)
Southern Hill Country (west)	0 (0%)	0 (0%)	0 (0%)	0 (0.4%)	0 (-0.4%)
Volcanic Ringplain	0 (0%)	0 (0%)	0 (0%)	3 (1.3%)	-3 (-1.3%)
Waitara	0 (0%)	0 (0%)	0 (0%)	35 (11.1%)	-35 (-11.1%)
Taranaki Total	**0 (0%)**	**0 (0%)**	**1 (0.1%)**	**44 (3%)**	**-45 (-3.1%)**

Table 5-3: Original model: Change in the length (km) of higher order streams compared to the current-state scenario in each NOF band determined for Scenario 1 (high removal efficiency) with C_{95} not included in the band calculation. The percentage of the total stream length in each band is in parentheses.

FMU	A	B	C	D	E
Coastal Terraces (north)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Coastal Terraces (south)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Northern Hill Country	0 (0%)	2 (1.1%)	2 (1.1%)	-3 (-2.2%)	0 (0%)
Pātea	0 (0%)	0 (0%)	0 (0%)	5 (2%)	-5 (-2%)
Southern Hill Country (east)	0 (0%)	0 (0%)	1 (0.3%)	0 (0.1%)	-1 (-0.3%)
Southern Hill Country (west)	0 (0%)	0 (0%)	0 (0%)	0 (0.4%)	0 (-0.4%)
Volcanic Ringplain	0 (0.2%)	2 (1%)	0 (0.1%)	0 (0%)	-3 (-1.3%)
Waitara	0 (0%)	0 (0%)	1 (0.3%)	28 (9.1%)	-29 (-9.4%)
Taranaki Total	**0 (0%)**	**4 (0.3%)**	**4 (0.3%)**	**31 (2.1%)**	**-39 (-2.7%)**

Table 5-4: Adjusted model: Change in the length (km) of higher order streams compared to the current-state scenario in each NOF band determined for Scenario 1 (high removal efficiency) with C_{95} not included in the band calculation. The percentage of the total stream length in each band is in parentheses.

FMU	A	B	C	D	E
Coastal Terraces (north)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Coastal Terraces (south)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Northern Hill Country	0 (0)	2 (1.1)	2 (1.1)	-3 (-2.2)	0 (0)
Pātea	0 (0)	0 (0)	0 (0)	5 (1.8)	-5 (-1.8)
Southern Hill Country (east)	0 (0)	0 (0)	1 (0.3)	0 (0.1)	-1 (-0.3)
Southern Hill Country (west)	0 (0)	0 (0)	0 (0)	0 (0.4)	0 (-0.4)
Volcanic Ringplain	0 (0.2)	2 (1)	3 (1.1)	-2 (-1)	-3 (-1.3)
Waitara	0 (0)	0 (0)	1 (0.3)	34 (10.9)	-35 (-11.1)
Taranaki Total	**0 (0)**	**4 (0.3)**	**6 (0.4)**	**34 (2.3)**	**-45 (-3.1)**

The similarities in results with and without the adjustment are because:

- Although the baseline loads from the national park have changed, we use the same Random Forest current state attributes (i.e, C_{50} and C_{95}, G_{240} and G_{560}) to characterise baseline conditions within the park and across the region.
- The only changes in generated yields occur where rainfall > 3m, which is limited to the national park and a small band of pastoral land around 3-6 km wide east of the park. This means that for the majority of subcatchments, there is no difference in change in yields under the Scenarios calculated by the original model and the adjusted model. In more detail:

- For the subcatchments that are completely within the national park, there will be no difference in the yields between the adjusted baseline and the scenario model runs since the mitigations modelled are not applied within the national park.
- For lower order streams that are not within the affected area and higher order streams that are not connected to the area affected by the adjustment, there is no difference in the estimated *E. coli* generated or instream yields.
- For lower order streams in the affected area, there will be a change in both their generated and instream yields, while for higher order streams with their headwaters in the affected area, there will be no change in generated yields, but a change in instream yields. However, the change has little impact on the outcome of the scenario modelling with respect to NOF bands.
- For Scenario 1, which is represented by percentage changes in yields from pasture, the relative change in generated yields will have similar impacts on the future state as for the original model since we use the same current state attributes. Scenario 2 is represented by a reduction in loads from point sources. While the adjusted model does reduce the modelled instream loads downstream of the national park, the relative difference in instream loads, with respect to the point source loads, is not enough to change the assigned NOF bands.
- Finally, the effect of the adjustment dissipates with distance and upstream area due to the cumulative effects of the delivery to the stream network of loads from the pastoral land that dominates land use on the ring plain.

The similarity between the original and adjusted CLUES model outputs gives us confidence in the Stage 2 model outputs.

As noted, we did not re-run the cost and load target analyses as these were outside the scope of this work. The associated costs of each scenario will not change as the scenarios have not changed. Given the similarity in the load estimates, we would expect the differences in cost effectiveness to be minimal. The required reduction in load to meet concentration targets could potentially be influenced by the reduced loads from Te Maunga. However, we consider that the required load reduction (associated with a required concentration reduction, for example) will affected only to a minor degree by the reduced loads from Te Maunga. The reason is that required load reductions are generally controlled by conditions in the lower catchment. Those conditions dictate the proportion of manageable load that needs to be reduced. Also, the total load and manageable load at the downstream location will be dominated by inputs from developed land (pasture, urban) and point sources, rather than the load (if any) from Te Maunga. Hence, the required factor removal of local manageable load for subcatchments outside national park will not be affected by the load from Te Maunga. The local manageable load is also not influenced by Te Maunga loads. Hence the required reduction in local manageable load, and the reduction in local incremental load, will not be affected much by the load from Te Maunga. Exceptions might occur if the required load reduction is dominated by a site near Te Maunga. This would be an unusual and locally-restricted case. A further line of reasoning is that reduced reference yields in the Stage 2 modelling had only a small influence on the overall region load reduction. Hence, it seems unlikely the required reduction in load to meet concentration targets would be influenced to a significant degree by the reduced loads from Te

Maunga. To confirm this, the full load reduction analysis could be repeated with the reduced loads from Te Maunga.

6 Conclusion

This investigation has indicated that the CLUES model is overestimating the *E. coli* loads and yields within Te Papakura o Taranaki. This overestimation is due to the combination of the regional parametrisation of the CLUES rainfall delivery factor and extremely high orographic rainfall on the slopes of Te Maunga. To compensate, we made a simple adjustment to the model to cap the maximum mean annual rainfall in the model at 3m, which roughly coincides with the boundary of the national park.

The adjustment reduced the modelled yields within the national park, however, the yields may still be overestimated. Since the yields estimated using data from newly established water quality sites are highly uncertain and cannot be directly compared to the CLUES yields (which were calibrated to yields determined using the 95th percentile flow record), it is not possible to determine the model's performance for those sites. We estimate that the CLUES yields may be an order of magnitude higher than the actual yields at the sites.

The adjusted CLUES model reduced the downstream loads of the nearby SOE sites by variable amounts, however the yields for the more distant SOE sites showed very little change due to downstream cumulative effects.

To test whether the adjustment affects the outputs of the Stage 2 modelling, we reran the Stage 2 scenarios, the model outputs have been supplied to TRC as supplementary files. The adjustment had a minimal effect on the outputs of the scenario runs for both low and high order streams. This means that we have confidence in the outputs of the scenario modelling reported previously (Semadeni-Davies 2023). We also conclude that recalibration for the adjusted model is not warranted.

7 References

Elliott, A.H.; Semadeni-Davies, A.F.; Shankar, U.; Zeldis, J.R.; Wheeler, D.M.; Plew, D.R.; Rys, G.J. and Harris, S.R. (2016) A national-scale GIS-based system for modelling impacts of land use on water quality. *Environmental Modelling & Software*, 86: 131-144. http://dx.doi.org/10.1016/j.envsoft.2016.09.011

Fraser, C. (2022) Taranaki water quality state spatial modelling, Land and Water People, client report prepared for Taranaki Regional Council, LWP 2021-14.

New Zealand Government (2023) National Policy Statement for Freshwater Management 2020, as amended February 2023. https://environment.govt.nz/acts-and-regulations/national-policy-statements/national-policy-statement-freshwater-management/

Semadeni-Davies, A.; Elliot, S. and Shankar, U. (2016) CLUES - Catchment Land Use for Environmental Sustainability User Manual Fifth Edition: CLUES 10.3, NIWA internal report: AKL2016-017. Available for download from XXXXXX download site

Semadeni-Davies, A.; Elliott, S. and Yalden, S. (2023) Calibration of the CLUES *E. coli* model for the Taranaki and Manawatū-Whanganui Regions: Stage 1 Technical Report, Client report prepared for Tanaki and Horizons Regional Councils, NIWA report: 2023064AK.

Semadeni-Davies, A., Elliott, S., Matthews, Y. (2023) Modelling *E. coli* to support implementation of the NPS-FM
Stage 2 Technical Report, NIWA client report prepared for Taranaki Regional Council, TRC23101.

Whitehead, A.; Fraser, C. and Snelder, T. (2022) Spatial modelling of river water quality state: Incorporating monitoring data from 2016 to 2020, NIWA client report prepared for the Ministry for the Environment, 2021303CH. https://environment.govt.nz/assets/publications/spatial-modelling-river-quality.pdf

Appendix A TRC data comparison

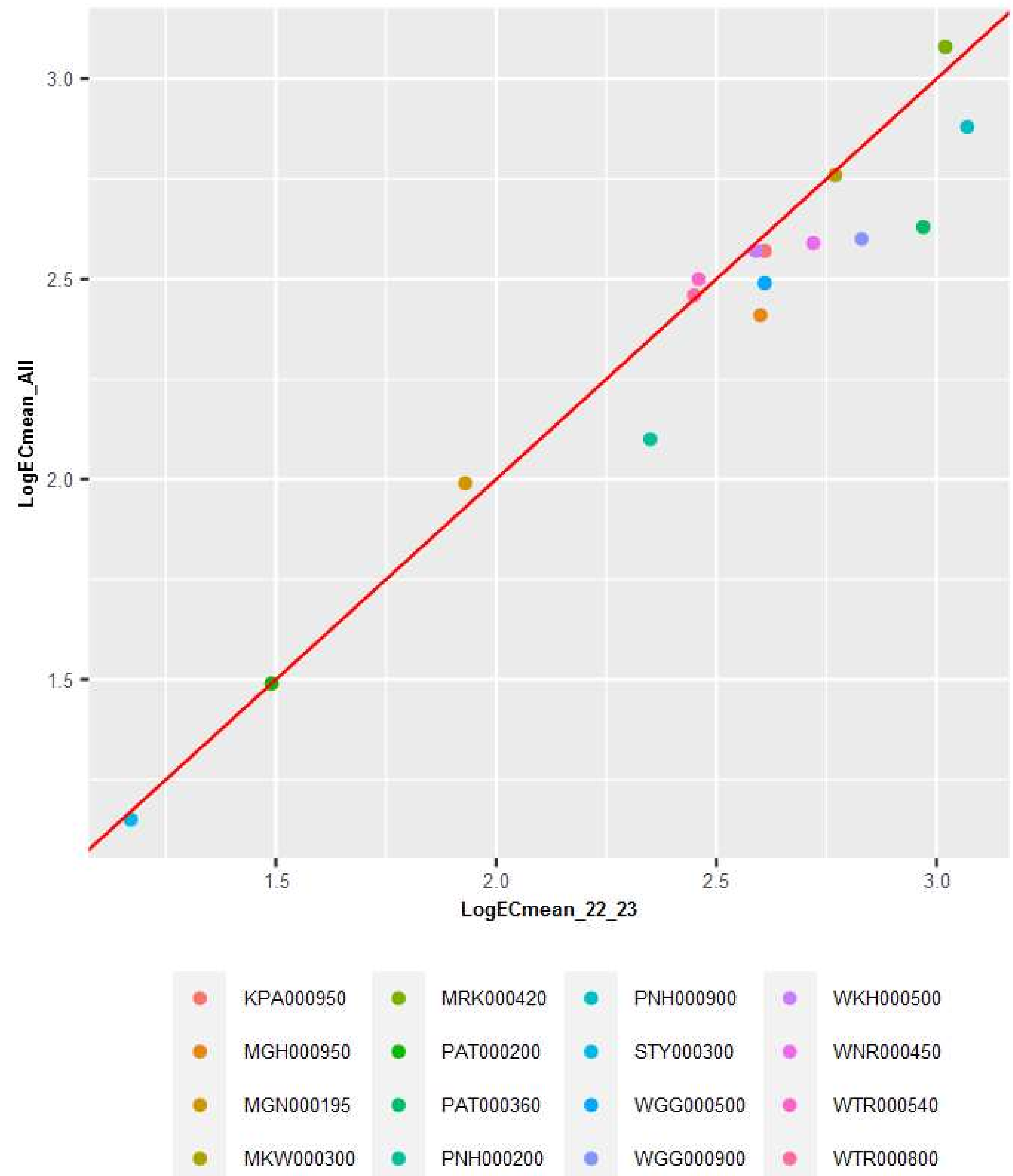
The following analysis was carried out by Jeremy Wilkinson at TRC and was provided to NIWA. The purpose of the analysis was to determine whether average *E. coli* concentration and flow data collected at SOE sites where flow is monitored over the same period as the new monitoring sites (June 2022 to June 2023) are similar to the average concentrations determined from the long-term water quality record. The rationale is that if the long and short term data are similar, we can be confident that the short term averages determined for the new sites can approximate the long-term averages. The following text was provided by TRC, with some minor modifications for clarity.

Tabulated statistics for all sites

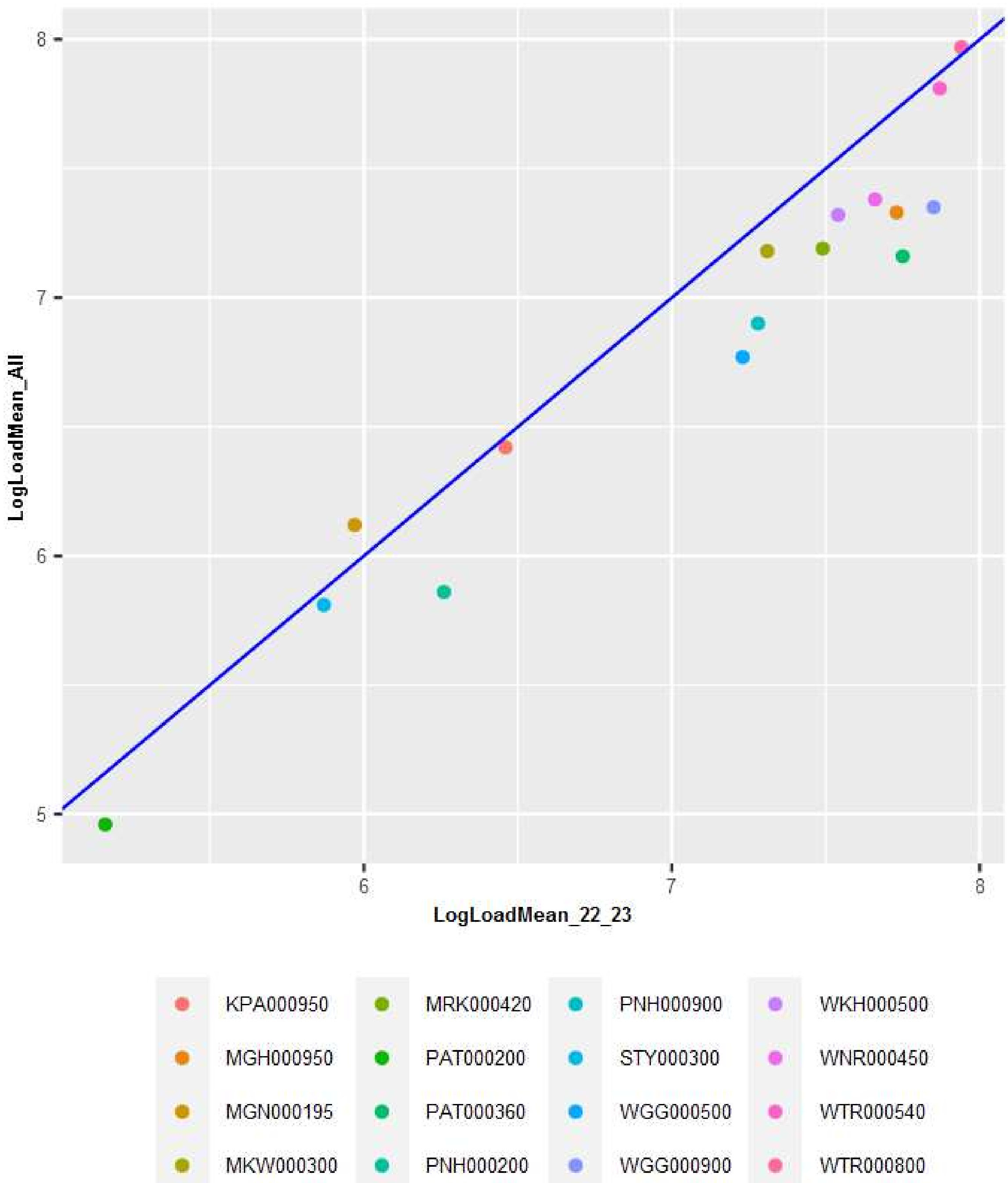
Here, we tabulate [mean of $\log_{10}$ of concentrations, LogECmean] for monitoring sites with extended records and those with only 1 year of record. We compare the statistics for the full period of data [All] and the last year only [22_23]. Data for the 2022/23 season are consistently higher than the overall record.

The data have been extended to include spot loads. The flows were taken from either spot flow gaugings, or gauged flow time-series. Missing spot flows were estimated from the correlation of existing spot gaugings with a flow time-series for a nearby site.

site_code	n_All	n_22_23	LogECmean_All	LogECmean_22_23	LogLoadMean_All	LogLoadMean_22_23	Qmean_All	Qmean_22_23
KPA000950	34	12	2.57	2.61	6.42	6.46	1.56	2.08
MGH000950	101	12	2.41	2.60	7.33	7.73	13.82	20.16
MGN000195	24	12	1.99	1.93	6.12	5.97	1.66	1.16
MKW000300	104	12	2.76	2.77	7.18	7.31	6.15	4.53
MRK000420	106	12	3.08	3.02	7.19	7.49	2.79	3.49
PAT000200	102	12	1.49	1.49	4.96	5.16	0.54	0.76
PAT000360	101	12	2.63	2.97	7.16	7.75	5.69	8.56
PNH000200	102	12	2.10	2.35	5.86	6.26	0.88	1.14
PNH000900	101	12	2.88	3.07	6.90	7.28	1.66	3.29
STY000300	101	12	1.15	1.17	5.81	5.87	6.69	5.55
WGG000500	112	15	2.49	2.61	6.77	7.23	3.18	5.59
WGG000900	101	12	2.60	2.83	7.35	7.85	8.27	12.87
WKH000500	101	12	2.57	2.59	7.32	7.54	10.94	12.17
WNR000450	97	12	2.59	2.72	7.38	7.66	9.51	12.18
WTR000540	95	12	2.50	2.46	7.81	7.87	36.14	36.66
WTR000800	71	22	2.46	2.45	7.97	7.94	56.62	43.25



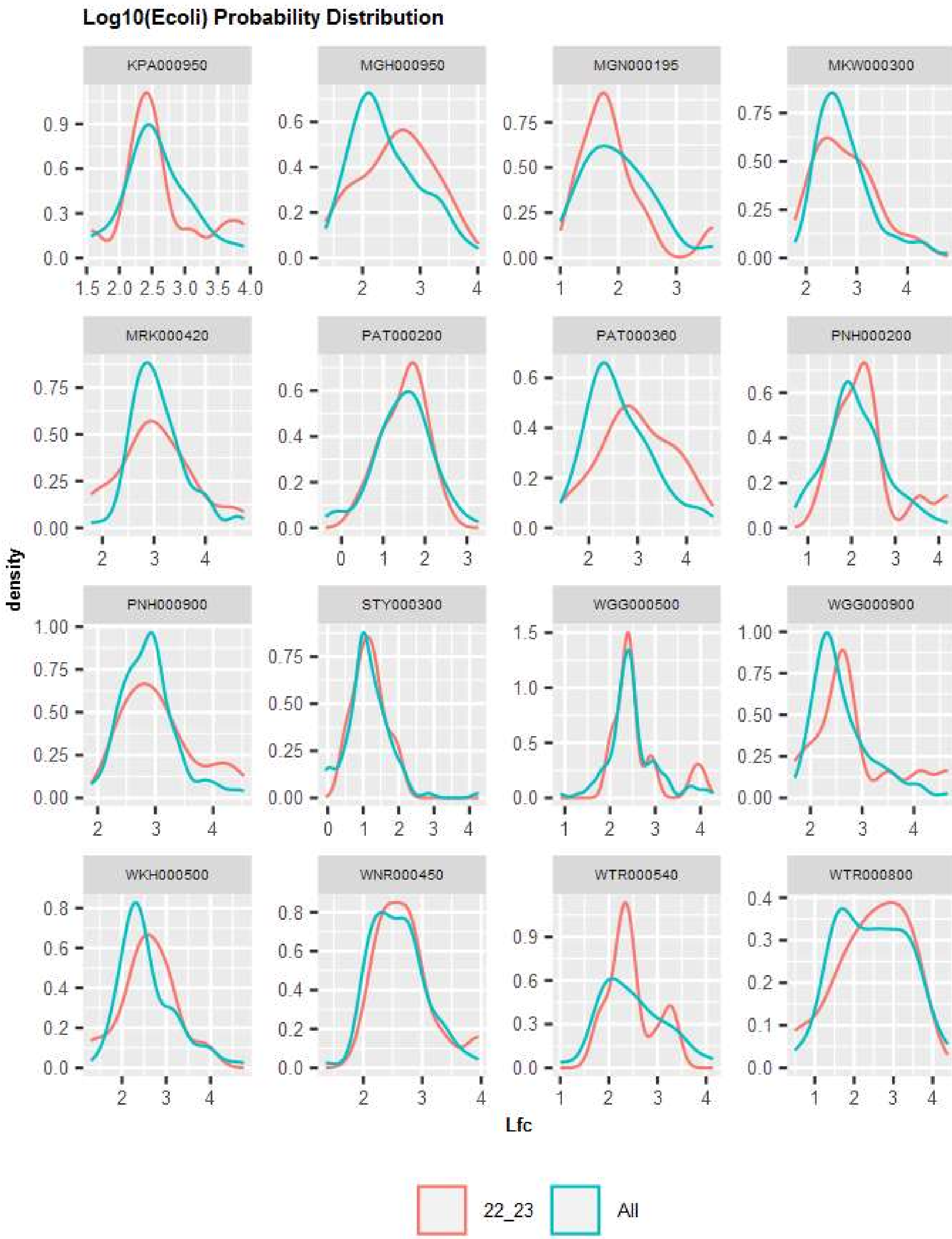
Site mean Log10 E. coli for 2022/23 season against full record



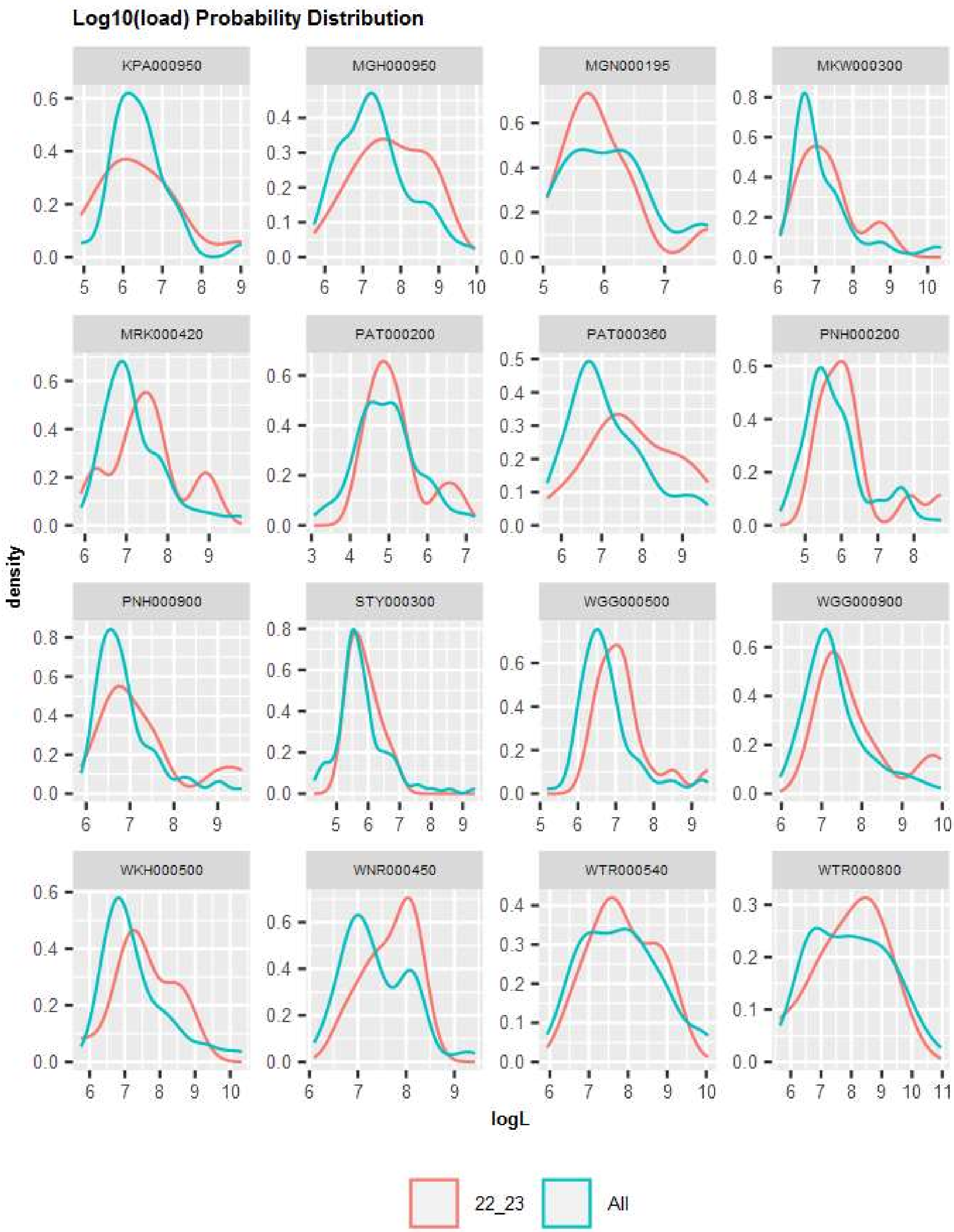
Site mean Log10 E. coli load for 2022/23 season against full record

Density functions for log10(*E. coli*) data

Here we compare data for sites with extended *E. coli* record and against the 2022/23 year for which new national park boundary data are available. The objective is to examine whether the year's data are likely to be adequately characteristic of the extended period.


Log10(Ecoli) Probability Distribution
KPA000950
MGH000950
MGN000195
MKW000300
MRK000420
PAT000200
PAT000360
PNH000200
PNH000900
STY000300
WGG000500
WGG000900
WKH000500
WNR000450
WTR000540
WTR000800
density
Lfc
22_23
All

Density functions for log10(*E. coli* - Load) data



Tabulated non-parametric Kolmogorov-Smirnov test

For log10 *E. coli* values

Here, the null hypothesis is that the data are drawn from the same continuous distribution. We are looking for low D and pVal approaching 1. The test compares the log10(*E. coli*) data, as per the distributions above.

site_code	D log10 Ec	p - log10Ec	D logLoad	p - logLoad
KPA000950	0.1136	0.9955	0.2121	0.7224
MGH000950	0.2904	0.2423	0.3300	0.1505
MGN000195	0.1667	0.9371	0.2083	0.8164
MKW000300	0.1378	0.9523	0.2276	0.5459
MRK000420	0.1478	0.9238	0.3278	0.1863
PAT000200	0.1345	0.9430	0.2294	0.5412
PAT000360	0.3680	0.0774	0.4274	0.0274
PNH000200	0.2108	0.5948	0.2647	0.3733
PNH000900	0.1815	0.7583	0.2649	0.3639
STY000300	0.0982	0.9930	0.1683	0.8669
WGG000500	0.1571	0.8009	0.3667	0.0417
WGG000900	0.3045	0.1973	0.3342	0.1431
WKH000500	0.1898	0.7260	0.3342	0.1431
WNR000450	0.1847	0.7460	0.3629	0.1781
WTR000540	0.2123	0.6176	0.1525	0.9302
WTR000800	0.0986	0.9790	0.0967	0.9870

Report date 19 September, 2023.

Appendix B Random Forest partial plots

This appendix reproduces the partial plots of the 12 most influential predictors for the random forest modelling reported by Whitehead *et al.* (2022). The most important predictors for median *E. coli* concentrations are upstream elevation (usElev), variation in upstream mean annual rainfall (usRainvar), the proportion of upstream bare surfaces (usBare), upstream stock density (usStockDensity) and intensive agriculture (usIntensiveAg) and upstream mean catchment slope (usSlope).

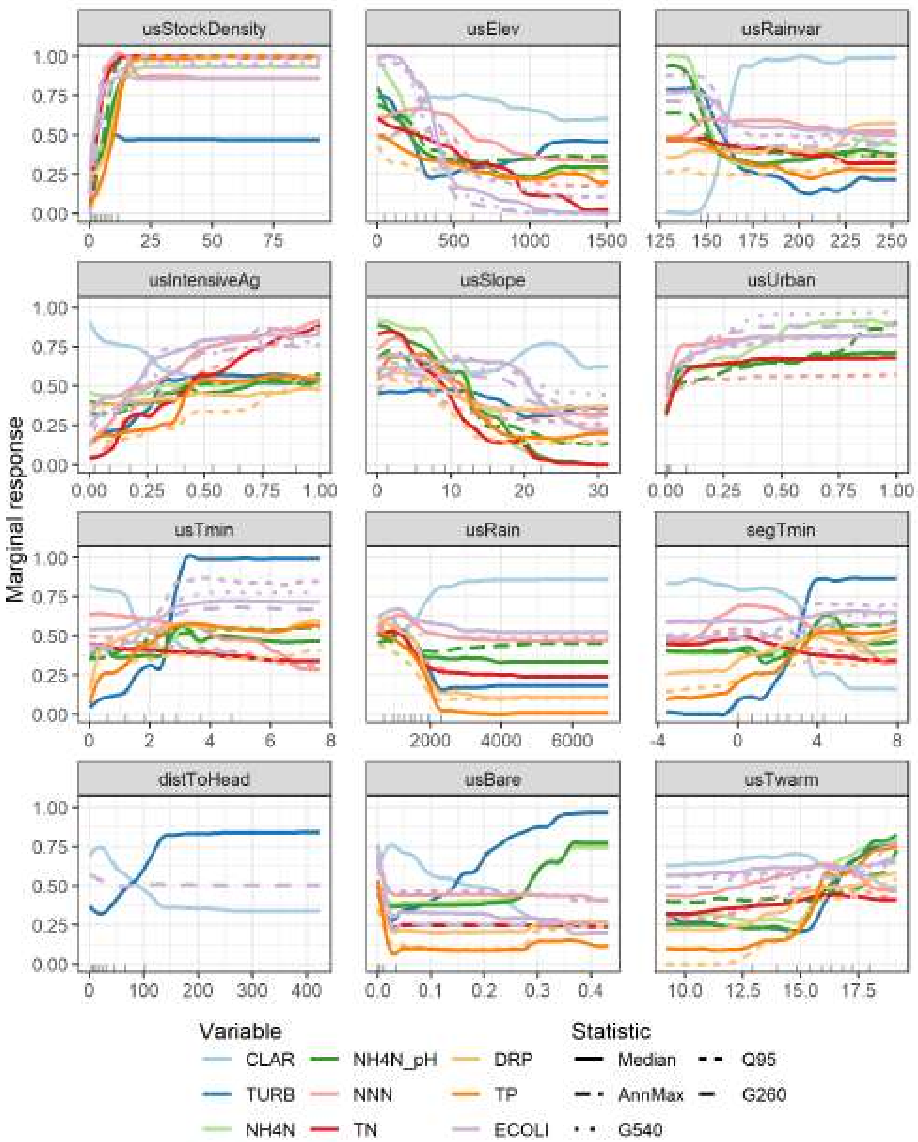


Figure 4-4: Partial plots for the 12 most important predictors in random forest models of current attribute state. Colours represent water quality variables, with the statistic indicated by line type (i.e., the combination of colour and line type represents an attribute). Each panel corresponds to one predictor, with predictors ordered by overall importance from most (top left) to least (bottom right) important. Y-axis scales represent marginal response standardised across all modelled attribute states. Plot amplitude (the range of the marginal response on the Y-axis) is directly related to a predictor's importance, with amplitude larger for predictors with higher importance. Units on X-axes are in Table 2-2.