Development of an *E. coli* runoff risk matrix

Richard Muirhead, Sandy Elliott (NIWA), Ton Snelder (LWP), Annette Semadeni-Davies (NIWA)

June 2023

Report for Our Land & Water

Client Report Number RE450/2023/041

Contents

1 Executive Summary

Water quality in New Zealand needs to improve and this will require a reduction in contaminant losses from the land. Microbial water quality impacts are particularly difficult to understand as there is a dearth of fundamental data on *E. coli* losses from many land uses and landscapes. Nevertheless, decisions on future land use need to be made now. The NPS-FM 2020 requires Regional Councils to set limits to manage water quality and “must not delay making decisions solely because of uncertainty about the quality or quantity of the information available”.

To support these limit setting processes this project has developed an *E. coli* runoff risk matrix based on expert opinion. This expert opinion was supported by a review of the literature on known mechanisms of microbial transport and a review of all modelling studies published in NZ. We thus use multiple lines of evidence to support the matrix development.

Development of the risk matrix entailed several steps:

- Review of models to identify factors responsible for microbial loads and concentrations,
- Selection of the most important factors,
- Categorisation of each factor into discrete classes, and
- Developing a multi-dimensional risk table that included all reasonable combinations of the factor classes, assigning each combination to a risk ranking that ranged from 1 to 10, with 10 representing the greatest risk of stream contamination.

All of the catchment scale modelling studies that included *E. coli* were identified and summarized in an Excel database (Muirhead, 2022). Three classes of models were examined: mechanistic models, hybrid mechanistic/statistical load models, and random forest statistical models.

These modelling studies were investigated to identify landscape, hydrology, land use or other explanatory variables used to predict *E. coli* contamination. The development of the risk ranking matrix built on earlier development of typologies developed to explain landscape-scale variation in nitrogen and phosphorus losses to water.

The review identified 4 important factors influencing *E. coli* concentrations in streams: land use, soil drainage, soil wetness and elevation. The 4 factors were subdivided into various classes. Land use was subdivided into 5 classes: urban, pastoral, horticulture, arable and forestry/other. Soil drainage was subdivided into 3 classes: well drained, light soils and poorly drained. Wetness was subdivided into 3 classes: dry, irrigated and/or moist and wet. Elevation was subdivided into 2 classes: low and high. The resulting risk matrix is a ranking from 1 to 10 with 10 representing the highest risk. The risk matrix ranking can be used to indicate a direction of travel and does not represent a numerical risk factor. The *E. coli* risk ranking matrix would be best applied at the scale of a freshwater management unit.

The risk matrix is presented in Table i and a national scale map of this risk matrix and GIS layer is available through the Data Supermarket

https://landuseopportunities.nz/

Table i. The proposed *E. coli* runoff risk ranking matrix.

	Wetness →	Dry			Irrigated and/or Moist			Wet		
	Drainage →	Well	Light	Poor	Well	Light	Poor	Well	Light	Poor
LU ↓	Elevation ↓									
Urban	All	10	10	10	10	10	10	10	10	10
Pastoral	Low	5	6	7	6	7	8	7	8	9
	High	4	5	6	5	6	7	6	7	8
Horticulture	All	3	3	3	3	3	3	3	3	3
Arable	All	2	2	2	2	2	2	2	2	2
Forestry/other*	All	1	1	1	1	1	1	1	1	1

* includes other non-productive landuses

2 Introduction

Water quality in NZ needs to improve and this will require a reduction in contaminant losses from land. There is a need to understand how different land uses interact with the landscape and climate to impact on water quality. Microbial water quality impacts are particularly difficult to understand as there is a dearth of fundamental data on *E. coli* losses from many land uses and landscapes. A further complication of microbial water quality impacts is the difference between stormflow and baseflow conditions in rivers. During stormflow there is a large increase in total streamflow and the associated contaminant loads transported in the stream network, and the effect is greater for *E. coli* loads than for other contaminants (Davies-Colley et al. 2008; Ballantine & Davies-Colley, 2013). This means there is a much greater load of *E. coli* transported in the river during stormflow events. The total load of *E. coli* in a river has an impact on the waterbody (lake, estuary or ocean) that the river discharges into as the large pulses of storm water have to be diluted and dispersed over time. However, the river microbial water quality metrics are based on concentrations, not loads. Because a river spends more days per year in a baseflow state than stormflow state, *E. coli* concentrations during baseflow conditions have a large impact on the river water quality metrics – particularly the median concentration. Regardless of the relative size of the annual load of *E. coli* discharged to a river, sources that occur during baseflow conditions will have a disproportionately large impact on the microbial river water quality metrics. In a pasture-dominated catchment, these baseflow impacting sources are likely to be animal access to streams, farm dairy effluent (FDE) management and irrigation systems (Muirhead et al. 2011; Muirhead, 2019). Previous tool development in NZ has focused on managing *E. coli* impacts during these baseflow conditions (Muirhead 2015; Muirhead & Doole 2017). These have been incorporated into the farm support tool MitAgator (https://ballance.co.nz/MitAgator).

It is unknown whether runoff from the land during storm events will have an impact on *E. coli* concentrations under all flow conditions, including baseflow. As yet we do not yet have a good understanding of the extent to which *E. coli* that enter a stream during storm flows impact on the water quality guideline metrics i.e. do all the *E. coli* that enter a stream in runoff flow all the way to the river mouth and therefore only impact on storm flows? Do these stormflow conditions only impact on the 95th percentile values in the monitoring datasets? Or do some of the runoff *E. coli* get trapped in the stream sediments and subsequently bleed out during baseflow conditions, thus contributing to elevated stream median concentrations (Wilkinson et al. 2011; Davies-Colley et al. 2008; Drummond et al. 2022; Pachepsky et al. 2017)? Our understanding of the dynamics of *E. coli* concentrations in rivers is poor and this severely limits our ability to model and/or predict changes in microbial water quality that might occur in response to changes in land use or management (Oliver et al. 2016).

Nevertheless, decisions on future land use need to be made now. The NPS-FM 2020 requires Regional Councils to set limits to manage water quality and "must not delay making decisions solely because of uncertainty about the quality or quantity of the information available".

To support these limit setting processes, this project has developed an *E. coli* runoff risk matrix based on expert opinion. This expert opinion was supported by a review of the literature on known mechanisms of microbial transport and a review of all modelling studies published in NZ. We thus use multiple lines of evidence to support the matrix. The aim of this work was to develop a simple tool to help a regional council make decisions at a freshwater management unit (FMU) scale. The tool is designed to indicate whether a change of land use in an FMU is likely to increase or decrease *E. coli* concentrations in a river. It also provides an indication of where mitigation efforts could be focussed.

3 Method

3.1 Development of the *E. coli* runoff risk matrix

Development of the risk matrix entailed several steps:

- Review of models to identify factors responsible for microbial loads and concentrations,
- Selection of the most important factors,
- Categorisation of each factor into discrete classes, and
- Development of a multi-dimensional risk table that included all reasonable combinations of the factor classes, assigning each combination to a range from 1 to 10, with 10 representing the greatest risk of stream contamination.

For each factor in the *E. coli* risk matrix we reviewed the literature and modelling analyses to provide evidence for its selection and ranking. The focus of the risk matrix is on ranking risk between the parameters but does not quantify losses or associated stream concentrations. A ranking of 10 does not indicate 10 times the risk compared with a factor of 1. That is, if one land use, climate and soil type combination has a higher ranking number, then our expert opinion is that the *E. coli* losses from that combination will be higher, with implications for land use choice or prioritisation of mitigation measures. The ranking does not represent a numeric load and hence should not be used for modelling *E. coli* loads, concentrations or a weighted average risk in a catchment.

The factors and classes that are represented in the risk matrix were aligned with an existing land use typology framework developed to explain landscape scale variation in nitrogen and phosphorus losses to water (Srinivasan et al., 2021). The typologies of Srinivasan et al. (2021) were modified to include additional land uses and some factors were collapsed or changed where it made sense for an estimation of *E. coli* risk.

3.2 Review of NZ-based *E. coli* modelling studies

All of the catchment scale modelling studies that included *E. coli* were identified and summarized in an Excel database (Muirhead, 2022). These modelling studies were investigated to identify landscape, hydrology, land use or other explanatory variables used to predict *E. coli* impacts. Where a specific model explanatory variable is used in multiple modelling approaches, this was taken as an indication of the relative importance of the explanatory variable and hence it's potential importance as a factor in the *E. coli* risk matrix. Also, the importance of a model explanatory variable used in both national and regional scale model applications was interpreted as the scale at which that parameter was important. That is, if an explanatory variable was statistically significant at a national scale application but not at a regional or catchment scale, then this was interpreted as a potential risk factor for a national scale application, but not at a FMU scale. Three classes of models were examined: mechanistic models, hybrid mechanistic/statistical load models, and random forest statistical models. In this report we consider all these models as complementary lines of evidence. Additional emphasis in this review was placed on the random forest models as the outputs of this relatively new modelling approach have not been summarised before.

3.2.1 Review of the random forest models

There have been 5 published studies using the random forest (RF) model technique for *E. coli* in NZ - two national scale studies and 3 regional level models for Southland, Taranaki and Otago. Details of this technique can be found in the references but, in brief, the model is provided with *E. coli* data collected at river water quality monitoring stations and a range of associated data describing catchment characteristics. The characteristics include land use, climate, topography data from the River Environment Classification (REC) and Freshwater environments etc. RF regression models then use a non-linear model approach to identify which combination of predictor variables best predicts the pattern of *E. coli* concentrations in the monitoring data. Because the microbiological water quality guidelines include four water quality statistics (NPS-FM 2020), four different RF models were developed for each study i.e. one for each *E. coli* statistic. Except for the first RF modelling study by Whitehead et al. (2018) that only modelled the median *E. coli* concentration. These results are further complicated by the fact that for each of the different RF modelling studies different sets of predictor variables were provided to the model. The output from the model is a prediction of the *E. coli* statistic for each combination of predictor variables. The model outputs can also include a list of the relative importance of each predictor variable used in the RF model prediction. For the most important predictor variables (up to 8), we were also provided with the “marginal response curves”, also known as “partial plots”, for these predictor variables. While these RF marginal response curves are not necessarily linear, each of the responses were classified into a generally positive or negative response depending on whether the *E. coli* concentrations increased or decreased as the predictor variable increased or decreased.

To standardise the results from the 5 different RF studies we selected only the top 10 important predictor variables for each study/*E. coli* metric combination. This generated a reduced database of the top 10 parameters for 17 individual RF models. We then summarized this data to determine which predictor variables were most consistently identified as an important predictor of *E. coli* concentrations in the RF models. This summary is presented in Table 1.

Table 1. Summary of the frequency of occurrence of a predictor variable in the top 10 most sensitive inputs in a RF model. The row labels are the short code used to identify the individual input parameters. A full description of the short codes can be found in the individual RF modelling reports; the top 10 are presented in Table 2. The count is the number of times that the predictor variable appeared as one of the top ten most important predictors for the individual RF models.

Predictor variables	**Count**
uselev	17
usslope	15
ustwarm	14
usrnvar	13
usIntensiveAg	13
SUDensityTotal2017	12
ustmin	11
usPastoralLight	10
usBare	10
ushard	9
usNativeForest	8
uspsize	6
PropDairy2017	5
segElev	5
usPastoral	4
usrain	2
usUrban	2
PropSheep2017	2
usIndigForest	2
usLake	2
usWetland	1
usPhos	1
usExoticForest	1
usrd20	1
FRE3.Count.StandardisedByMeanFlow	1
usParticleSize	1
USCalcium	1
lcv.StandardisedByCatchArea	1
Grand Total	**170**

Table 2. Full description of the top 10 short codes contained in Table 1.

Short code	Full description
uselev	Mean elevation of the upstream catchment
usslope	Mean slope of the upstream catchment
ustwarm	Mean catchment January air temperature
usrnvar	Mean catchment coefficient of variation of annual rainfall
usIntensiveAg	Proportion of catchment occupied by combination of high producing exotic grassland, short-rotation cropland, orchard, vineyard and other perennial crops (LCDB4 classes 40, 30, 33)
SUDensityTotal2017	Stocking rate of all animals in a catchment using relative stock units (LCDB4)
ustmin	Mean catchment June air temperature
usPastoralLight	Proportion of catchment in low producing grassland (LCDB4 class 41)
usBare	Proportion of catchment occupied in bare or lightly-vegetated cover (LCDB4 classes 10, 12, 14, 15, 16)
ushard	Mean catchment induration (hardness) of regolith

4 Results and Discussion

4.1 Results of the modelling studies

4.1.1 Landscape features

From the RF model studies, the dominant landscape features were upstream elevation and upstream slope. Upstream elevation was a top 10 predictor variable in all 17 RF model studies and upstream slope was a top 10 parameter in 15 of the 17 studies. Both of these parameters were negatively correlated with *E. coli* concentrations i.e. *E. coli* concentrations were lower in catchments with higher mean elevations and higher mean slopes. The elevation parameter may be related to land use intensity as in NZ the most intensive agricultural land use generally occurs on the flat land at low elevation. The reference site analysis by McDowell et al (2013) also identified a strong correlation between catchment elevation and *E. coli* concentrations at water quality monitoring stations.

The potential relationship between stream *E. coli* concentrations and average catchment slope is conflicting. The RF models identified a consistent negative relationship with slope. The hybrid mechanistic/statistical model CLUES investigated the influence of slope as a potential explanatory variable but found it did not have a statistically significant effect and it was thus not used in the models. Slope is further complicated as it is negatively correlated with land use intensity (Srinivasan et al. 2021). While runoff generation and associated transport of *E. coli* may increase with increased slope, the source of *E. coli* from animal faeces may be decreased and hence cancel out other effects to leave a minimal overall effect of slope. Another complication with slope may be related to connectivity effects whereby overland flow is generated on the steeper slopes some distance from the stream, but the flow does not reach the stream due to capture in sinks or infiltration on lower slopes closer to the stream (Thomas et al. 2016).

From the RF models, upstream temperature was frequently identified in the top 10 parameters with usTmin identified in 11 of the 17 models and usTwarm identified in 14 of the 17 models. Temperature was positively associated with *E. coli* concentrations. This positive relationship was also identified in the reference site study (McDowell et al. 2013). The national scale CLUES model identified temperature as a significant parameter (positive effect) but temperature was not significant when CLUES was applied at the smaller regional or catchment scales. This could imply that temperature is important when there is a wide range of annual temperatures as seen across the whole country rather than at the smaller regional scale of annual temperature differences. Temperature may be related to productivity of the landscape with higher plant growth rates supporting higher plant and animal biomass on productive land and in forests.

The three factors of elevation, slope and temperature are all correlated. Therefore, the single factor of elevation, that had the strongest influence as a predictor variable, could be used in the risk matrix with the assumption that this factor will also capture some impact of slope and temperature on the *E. coli* risk.

A key explanatory variable of *E. coli* losses in all the process-based models and in the hybrid CLUES model is soil drainage class. The logic behind this is that the soil drainage class has a large impact on the amount of surface runoff generated from the landscape, and poor drainage is likely to lead to surface or near-surface runoff and artificial drain flows that are likely to have high concentrations. The RF models did not include soil drainage class as an input parameter so it is unknown if this parameter would be selected in a RF model.

4.1.2 Hydrology

The runoff generation calculations in all of the process-based models start with rainfall and then use potential evapotranspiration (PET) and a soil water balance to calculate runoff. They then typically use a curve number approach to calculate the overland flow proportion of the total runoff/drainage. Therefore, all these parameters appear to be important. However, none of these hydrology parameters were consistently identified as important in the RF models. As a potential explanation for these results, the process-based models are calculating the load (or mass) of *E. coli* in the runoff, whereas the RF models are attempting to predict the instream *E. coli* concentrations. It is possible that runoff volumes and loads of *E. coli* correlate and result in similar in-stream concentrations. It is also possible that some of the variables that appear in the RF models, such as elevation and rainfall variation, correlate with rainfall.

Interestingly, the one hydrology parameter that was identified in 13 of the 17 RF models is upstream rainfall variation (usrvar). Upstream rainfall variation is the coefficient of variation in the annual rainfall and was negatively correlated with *E. coli* concentrations i.e. increased rainfall variation was associated with a decrease in *E. coli* concentrations. This effect may be more related to its effect on farm systems than on rainfall-generated runoff. In farm systems, consistent rainfall from year to year makes it easier to budget feed for animals. If the rainfall from year to year varies appreciably, it is challenging to feed the animals in the drier years and hence these farm systems tend to have lower stocking rates. The alternative is (where available) to import expensive extra feed or install irrigation systems to maintain higher stocking rates in the dry years. Therefore, if two landscapes had the same long-term annual rainfall amount, a landscape with higher annual rainfall variation would typically have a lower stocking intensity.

4.1.3 Land use

There are some clear influences of land use that are represented by the models. The RF modelling studies did not use the same land use layers for each of the 5 studies. For the one RF modelling study that used up-stream pastoral land use as a predictor variable, this land use was the first or second most important predictor variable for all four *E. coli* statistic (individual RF models). The other four studies used upstream intensive agricultural land use as a predictor variable and this was in the top 10 of all 13 RF models. Furthermore, three of these RF studies included an additional predictor variable of “stock unit density total” and this layer was in the top 10 of all 12 RF models. These last three RF studies (12 RF Models) also included input layers for the proportion of sheep or dairy but these factors were not consistently in the top 10 most important predictor variables. From these RF modelling studies we can see that intensive pastoral land use is a significant driver of in-stream *E. coli* concentrations. It also appears that total stock numbers are driving this effect, rather than a single animal species.

The reference site study by McDowell et al. (2013) also showed a significant positive relationship between the proportion of area in pasture in a catchment and the *E. coli* concentrations in the river.

Of the six CLUES modelling studies (Muirhead, 2022), land use could only be significantly separated into urban, pasture and non-pasture areas in five of these studies. In one study of the Waikato/Waipa catchment, the pastoral land use could be significantly sub-divided into dairy, intensive sheep and beef (S&B) and hill & high country S&B land uses. Interestingly, in this study the highest *E. coli* losses were from the hill & high country S&B land use, followed by dairy and then intensive S&B. The general results of the land uses from the CLUES models are that urban land use has the highest yields of *E. coli* followed by pastoral and then non-pastoral land use. The *E. coli* yield differences between pastoral and non-pastoral land uses ranged from six times higher in the Northland study to 176 time higher in a national scale study.

None of the process-based models specifically addressed the question of *E. coli* losses from different land uses. Due to a lack of data on *E. coli* excretion from sheep at the time, the earliest work by Collins and Rutherford (2004) assumed that sheep and cattle faecal concentrations were equivalent and, therefore, their model showed no differences in runoff from sheep or cattle grazed areas. The model developed by Wilkinson et al. (2011) used very simplified farm inputs and assumed faecal loadings only from cows. Hong et al. (2018) modelled the Toenepi catchment in the Waikato and assumed all land was dairy grazing. Dymond et al. (2016) and Srinivasan et al. (2021) both calculated overland flow volumes generated from the landscape and then multiplied this by an *E. coli* concentration to calculate the load of *E. coli* from the land use. Due to a lack of data on *E. coli* concentrations in overland flow from a range of land uses, both these models used a single concentration value for all pastoral land use and assumed zero *E. coli* concentration in overland flow from non-pastoral land uses.

4.1.4 Other factors

The process-based models included other factors that could affect in-stream *E. coli* concentrations, such as direct inputs from animals, farm dairy effluent (FDE) management and stream proximity. These individual factors are generally not contained in land information as direct stock access to streams and FDE are farm specific management actions. However, these managements are also land use specific. For example, almost all dairy farms have now fully fenced off streams and deferred FDE irrigation to land is the preferred management system across NZ. Therefore, these management actions can be assumed for any land use mapped as dairying.

4.2 Development of the *E. coli* runoff risk matrix

4.2.1 Land use

The land use factor was divided into five classes for use in the matrix: urban, pastoral, horticulture, arable and forestry/other. The urban land use was ranked as the highest risk. The CLUES modelling consistently identifies urban as a large *E. coli* source and this is consistent with data from water quality monitoring stations (MfE, 2017). This urban ranking is applied to all urban land regardless of climate or landscape features as there is no published evidence to support varying this risk.

Forests (native and harvested) and other non-productive land cover categories are given a ranking of 1 as the lowest potential *E. coli* risk to water. At a national scale we recognise that there is a temperature effect in that warmer northern regions of NZ are associated with higher *E. coli* concentrations than in cooler southern climates (RF and CLUES models). This temperature effect could be related to higher productivity in warmer climates leading to them containing larger populations of animals and birds that are shedding *E. coli* into the environment and impacting microbial water quality (Cookson et al. 2022). However, for a risk matrix to support decision making at the FMU scale we believe that any temperature effect would be too small to include as an individual factor. However, some effect of temperature at an FMU scale will be included in the elevation factor, as described below.

There is no data on *E. coli* concentrations in runoff from arable or horticulture land. We would expect higher volumes of runoff from these categories than from forested land, but lower *E. coli* concentrations than from pasture due to the minimal number of farm animals in these arable and horticultural systems. The risk rankings are therefore between those for forests and pastoral land use categories. The arable land use is given a lower ranking than horticultural land for two reasons. Firstly, arable land use usually involves frequent cultivation which will increase infiltration rates (reducing runoff volumes) and will remove any faecal material from the surface and mix this into the soil. A study in a pastoral area showed that cultivating the soils had a significant effect on *E. coli* concentrations in runoff (Muirhead et al. 2006). Secondly, horticultural land is likely to have a higher bird population and is more likely to use organic fertilizers that contain faecal material which will remain on the soil surfaces. Therefore, the arable land is given a risk ranking of 2 and horticultural land a risk ranking of 3. Given the low loading of faecal material in these systems, the risks should be relatively low regardless of the climate conditions. The CLUES models predicted large differences in the *E. coli* losses from non-pastoral versus pastoral land uses.

Pastoral land use was lumped into a single classification rather than separate classes for sheep & beef land or dairying. This decision is based on multiple lines of evidence. Firstly, the RF and CLUES models intensive land use category was consistently identified as a significant driver of *E. coli* contamination, rather than any animal species difference. Secondly, a recent study that compared sheep versus cow grazing, at equivalent stocking rates, showed that *E. coli* runoff from the sheep pasture was four times greater than from cow pasture (Muirhead, 2023). This is due to the extremely high concentrations of *E. coli* in sheep faeces. Therefore, although S&B farms are often less intensively managed than observed for dairy farms, it seems best to give all pastoral land use the same *E. coli* risk ranking until more science is conducted to provide more refined advice.

4.2.2 Elevation

Elevation was a more consistent predictor variable of *E. coli* losses in the RF models than slope, and slope was not significant in the CLUES models. Elevation is however related to farm intensity, with stocking rates reducing as farm elevation increases due to the reducing amount of feed that can be grown in elevated regions where temperatures are usually cooler. We have therefore chosen to split the elevation factor into 2 classes of high (>350 m) and low (<350 m) altitude. This 350 m elevation threshold was chosen from two lines of evidence. Firstly, based on the marginal response curves from the RF models, the mid-point of the elevation predictor variable curves was typically in the 300 to 400 m altitude range. Secondly, we conducted a GIS analysis of the proportion of farmed area below "X m" elevation using the 2017 AgriBase farm data and NZSoS version 1 DEM (15 m/pixel). This spatial analysis identified that 90% of dairy farmed land was < 340m, 90% of arable land use was below 312 m, 70% of beef farmed land was below 365 m, 50% of sheep farmed land was below 337 m and 50% sheep and beef farmed land was below 376 m. We therefore felt that a threshold altitude of 350 m would capture the more intensive land used below this altitude. We acknowledge that other factors of temperature and slope, which are not used in this risk matrix, are strongly correlated with elevation.

4.2.3 Wetness

Wetness is included in the *E. coli* risk matrix due to its strong influence on both overland flow generation and stock carrying capacity of the land. Rainfall was a key driver in the process-based models and was a significant driver in the CLUES models. We have adopted the rainfall categories as used in Srinivasan et al. (2021) but have combined the irrigated and moist categories into one risk level. This is because irrigated pasture will have a similar stocking rate to the moist land and due to irrigation may generate similar amounts of overland flow i.e. when rainfall does occur the soil moisture levels in the irrigated land will be higher and, therefore, generate more runoff than from the equivalent dryland farm. Wet areas are expected to generate the most overland flow and dry areas the least; these thus remain the same classes as documented in Srinivasan et al. (2021). These wetness classes as developed by Monaghan et al. (2021) and used in Srinivasan et al. (2021) are reproduced in the Appendix.

4.2.4 Soil Drainage

Soil drainage features are known to have a large effect on overland flow generation. Therefore, we used the same three soil drainage categories as used in Monaghan et al. (2022). The CLUES models all identified soil drainage characteristics as a significant driver of *E. coli* losses and soil drainage parameters are used in all the process-based models. Surface runoff is usually the dominant pathway of *E. coli* losses from land (Monaghan et al. 2016). Well drained soils allow for greater infiltration of water and will trap a larger proportion of *E. coli* in the soil pores (Smith et al. 1985, Muirhead et al. 2006; McLeod et al. 2008), resulting in the lowest risk factor. Poorly drained soils have high losses of *E. coli* due to high volumes of overland flow (Dymond et al. 2016; Srinivasan et al. 2021). Farm production on poorly drained soils can be increased by installing artificial drainage systems but still result in high *E. coli* losses through the drains (Monaghan et al. 2016). The light soil classification in Monaghan et al. (2022) was given an intermediate risk level due to the relatively low plant available water holding capacities of this soil group. This feature results in potentially greater volumes of drainage through these soils and an associated elevated risk of microbial bypass flow (McLeod et al. 2008; McLeod et al. 2014). These soil drainage classes as developed by Monaghan et al. (2021) and used in Srinivasan et al. (2021) are reproduced in the Appendix.

4.2.5 Matrix construction and the proposed matrix

The *E. coli* runoff risk ranking matrix was based on the 4 most important factors identified in the modelling assessments described above, namely: land use, soil drainage, wetness and elevation. These 4 factors were subdivided into classes as shown in Table 3. The risk ranking process was undertaken by starting with forestry/other as a ranking of 1, arable as 2 and horticultural as 3. We then determined that the lowest pastoral land use risk should be the combination of high elevation, well drained soils and a dry environment; this category was given the next ranking of a 4. Within an elevation class, as the drainage class moved from well drained to light to poorly drained soils we increased the risk index by 1 point per step. Likewise, within a soil drainage class, as the elevation decreased from high to low we increased the risk ranking by 1 point. Further, as the wetness class increased from dry to irrigated/moist to wet, we increased the risk ranking by 1 point per step. This resulted in a pastoral land use risk ranking ranging from 4 for a well-drained soil type in a dry environment at high elevation through to a risk ranking of 9 for a poorly drained soil type in a wet environment at low elevations. Urban land use was then assigned the next highest ranking of 10.

The underlying assumption in this approach is that the relative impact of the 3 factors of elevation, soil drainage and wetness are similar. As a result of this risk ranking, we have created a scenario where pastoral land use on poorly drained soils in a wet environment at a high elevation has a similar risk ranking to that of a well-drained soil at a low elevation in the same wet environment or a poorly drained soil in a dry environment at low elevation. There is no data on *E. coli* losses to support or disagree with this approach. However, this tool is most likely to be used to evaluate land use change or land retirement away from pasture. In this situation it would seem reasonable that retiring pastoral land on poorly drained soils in a wet environment will have a much larger effect on reducing *E. coli* concentrations in a FMU than retiring pasture on a well-drained soil in a dry environment. Similarly, retiring land at a high elevation will have less impact on *E. coli* losses than retiring land at a low elevation.

Table 3. The proposed *E. coli* runoff risk ranking matrix.

	Wetness →	Dry			Irrigated and/or Moist			Wet		
	Drainage →	Well	Light	Poor	Well	Light	Poor	Well	Light	Poor
LU ↓	Elevation ↓									
Urban	All	10	10	10	10	10	10	10	10	10
Pastoral	Low	5	6	7	6	7	8	7	8	9
	High	4	5	6	5	6	7	6	7	8
Horticulture	All	3	3	3	3	3	3	3	3	3
Arable	All	2	2	2	2	2	2	2	2	2
Forestry/other*	All	1	1	1	1	1	1	1	1	1

* includes other non-productive landuses

The GIS layer is available through the Data Supermarket at https://landuseopportunities.nz/

5 Recommendations and further steps

This *E. coli* risk ranking is based on expert opinion following multiple lines of evidence and as such represents the best available knowledge of impacts on microbial water quality statistics at this time, and could be further developed. This ranking matrix would be best suited to use at a FMU scale.

The validity of this risk matrix could be evaluated using the empirical modelling approach currently under development for nitrogen (Snelder et al. 2023). This approach may enable the development of a more quantitative risk matrix. This current *E. coli* risk matrix is focused on the risk related to in-stream *E. coli* concentrations. Further work could investigate the development of a risk matrix for catchment loads that are more relevant to defining impacts on receiving water bodies. Furthermore, the "potential pasture growth" layers now available in the Data Supermarket could be explored as an additional or alternative factor in the *E. coli* risk matrix.

6 Acknowledgements

The study was funded by the Our Land & Water National Science Challenge through the Future Landscapes Theme. The review of the NZ *E. coli* models was conducted with funding from the Mapping Freshwater Contaminants programme and the *E. coli* risk ranking matrix was developed through the Land Use Opportunities programme.

7 References

DJ Ballantine, RJ Davies-Colley (2013). Nitrogen, phosphorus and *E. coli* loads in the Sherry River, New Zealand. New Zealand Journal of Marine and Freshwater Research, 47(4): 529-547. DOI: 10.1080/00288330.2013.815640

R Collins, K Rutherford (2004) Modelling bacterial water quality in streams draining pastoral land. Water Research 38: 700–712.

AL Cookson, JC Marshall, PJ Biggs, LE Rogers, RM Colis, M Devane, R Stott, DA Wilkinson, J Kamke, G Brightwell (2022). Whole-genome sequencing and virulome analysis of *Escherichia coli* isolated from New Zealand environments of contracting observed land use. Applied and Environmental Microbiology 88(9): DOI:10.1128/aem.00277-22.

R Davies-Colley, E Lydiard, J Nagels (2008) Stormflow-dominated loads of faecal pollution from an intensively dairy-farmed catchment. Water Science & Technology 57:10, 1519-1523. DOI: 10.2166/wst.2008.257

JD Drummond, T Aquino, RJ Davies-Colley, R Stott, S Krause (2022). Modeling contaminant microbes in rivers during both baseflow and stormflow. Geophysical Research Letters, 49, e2021GL096514. https://doi.org/10.1029/2021GL096514

JR Dymond, D. Serezat, A-G.E. Ausseil, R. Muirhead (2016) Mapping of *Escherichia coli* sources connected to waterways in the Ruamahanga catchment, New Zealand. Environmental Science and Technology 50:1897-1905.

E-M Hong, Y Park, R Muirhead, J Jeong, YA Pachepsky (2018) Development and evaluation of the bacterial fate and transport module for the Agricultural Policy/Environmental eXtender (APEX) model. Science of the Total Environment 615: 47-58

RW McDowell, TH Snelder, N Cox, DJ Booker, RJ Wilcock (2013) Establishment of reference or baseline conditions of chemical indicators in New Zealand streams and rivers relative to present conditions. Marine and Freshwater Research, 64: 387–400. http://dx.doi.org/10.1071/MF12153

M McLeod, J Aislabie, J Ryburn, A McGill (2008) Regionalizing potential for microbial bypass flow through New Zealand soils. Journal of Environmental Quality 39: 1959-1967.

M McLeod, J Aislabie, A McGill, P Rhodes, S Carrick (2014) Leaching of *Escherichia coli* from stony soils after effluent application. Journal of Environmental Quality 43: 528–538 doi:10.2134/jeq2013.06.0256

MfE (2017) Clean water: 90% of rivers and lakes swimmable by 2040. MfE Publication number ME1283. Published February 2017 by the Ministry for the Environment. Wellington, NZ.

RM Monaghan, LC Smith, R Muirhead (2016) Pathways of contaminant transfers to water from an artificially-drained soil under intensive grazing by dairy cows. Agriculture, Ecosystems and Environment 220: 76-88.

R Monaghan, A Manderson, L Basher, C Smith, D Burger, E Meenken, R McDowell (2021) Quantifying contaminant losses to water from pastoral landuses in New Zealand I. Development of a spatial framework for assessing losses at a farm scale. New Zealand Journal of Agricultural Research 64(3): 344-364, DOI:10.1080/00288233.2021.1936572

R Muirhead, RP Collins, PJ Bremer (2006) The association of *E. coli* and soil particles in overland flow. Water Science and Technology 54(3): 153-159.

R Muirhead, AH Elliot, RM Monaghan (2011) A model framework to assess the effect of dairy farms and wild fowl on microbial water quality during base-flow. Water Research 45(9): 2863-2874.

R Muirhead, RM Monaghan (2012) A two reservoir model to predict *Escherichia coli* losses to water from pastures grazed by dairy cows. Environment International 40(2): 8-14.

R Muirhead (2015) A farm-scale risk-index for reducing fecal contamination of surface waters. Journal of Environmental Quality 44: 248-255. Doi:10.2134/jeq2014.07.0311

R Muirhead, G Doole (2017) A farm-scale *E. coli* model for Gisborne District Council. AgResearch Report for Gisborne District Council. RE450/2017/025. 38 pages. Available from: https://www.mfe.govt.nz/sites/default/files/media/Fresh%20water/A-Farmscale-Ecoli-Model-for-GDC-Final.pdf

R Muirhead (2019) The effectiveness of streambank fencing to improve microbial water quality: a review. Agricultural Water Management 223: 105684.

R Muirhead (2022) OL&W: Transport risk factors for *E. coli* impacts on surface waters: Bibliography Milestone. AgResearch client report prepared for Our Land and Water. Client Report Number: RE450/2022/029. 10 pages.

RW Muirhead (2023) *Escherichia coli* runoff from sheep and dairy cow grazed pasture: a plot scale simulation. Journal of Environmental Quality: http://doi.org/10.1002/jeq2.20448.

DM Oliver, KDH Porter, YA Pachepsky, R Muirhead, SM Reaney, R Coffey, D Kay, DG Milledge, E Hong, SG Anthony, T Page, JW Bloodworth, P-E. Mellander, P Carboneau, SJ McGrane, RS Quilliam (2016) Predicting microbial water quality with models: over-arching questions for managing risk in agricultural catchments. Science of the Total Environment 544: 39-47.

Y Pachepsky, M Stocker, M Olmeda-Saldaña, D Shelton (2017) Enrichment of stream water with fecal indicator organisms during baseflow periods. Environmental Monitoring and Assessment 189: 51. DOI 10.1007/s10661-016-5763-8

MS Smith, GW Thomas, RE White, D Ritonga (1985) Transport of *Escherichia coli* through intact and disturbed soil columns. Journal of Environmental Quality 14(1): 87-91.

T Snelder, T Cox, C Fraser, S Elliott, T Kerr (2023) Quantifying catchment nutrient modelling parameters. LWP Client Report 2023-03. May 2023.

MS Srinivasan, RW Muirhead, SK Singh, RM Monaghan, R Stenger, ME Close, A Manderson, JJ Drewry, LC Smith, D Selbie, R Hodson (2021) Development of a national-scale framework to characterise transfers of N, P and *Escherichia coli* from land to water. New Zealand Journal of Agricultural Research 64(3): 286-313.

IA Thomas, P Jordan, P-E Mellander, O Fenton, O Shine, DO hUallachain, R Creamer, NT McDonald, P Dunlop and PNC Murphy (2016) Improving the identification of hydrologically sensitive areas using LiDAR DEMs for the delineation and mitigation of critical source areas of diffuse pollution. Science of the Total Environment 556: 276-290. http://dx.doi.org/10.1016/j.scitotenv.2016.02.183

A Whitehead (2018) Spatial modelling of river water-quality state. NIWA report prepared for Ministry for the Environment. December 2018. 41 pages

RJ Wilkinson, LA McKergow, RJ Davies-Colley, DJ Ballantine, RG Young (2011). Modelling storm-event *E. coli* pulses from the Motueka and Sherry Rivers in the South Island, New Zealand, New Zealand Journal of Marine and Freshwater Research 45(3): 369-393 DOI: 10.1080/00288330.2011.592839

Appendix

The wetness and soil drainage classes developed by Monaghan et al. (2021) and used in Srinivasan et al. (2021) are reproduced below:

*(2) **Wetness** was used to categorise variability in contaminant loss due to the influence of water availability and transport. Wetness levels were predicated on the known influence that surplus rainfall directly exerts on the transport of contaminants through (Cichota et al. 2012) or over (McDowell et al. 2005) soil. Four classes of wetness were distinguished based on the following criteria:*
(2a) 'Irrigated' – farms where >50% of the farm area is irrigated.
(2b) 'Wet' – farms where mean annual rainfall exceeded 1700mm.
(2c) 'Moist' farms were categorised based on the upper 50th percentile for calculations of annual surplus rainfall (rainfall minus evapotranspiration), as derived from rainfall input settings in the Overseer farm files not assigned to 'Irrigated' or 'Wet' categories. The surplus rainfall threshold between 'Moist' and 'Dry' classes approximated to an annual rainfall total of 1100 mm.
(2d) 'Dry' farms were assigned based on the lower 50th percentile category for calculations of annual surplus rainfall...

*(3) **Soil Drainage** was used to capture the effects of two fundamental processes that influence the vulnerability of soil to nitrate leaching. The first effect is N displacement from soils that have contrasting abilities to store water and nutrients (Wild 1981; Addiscott 2011; Boy-Roura et al. 2016); Plant Available Water (PAW) holding capacity was selected as the soil attribute that best represented this aspect of leaching vulnerability (Cichota et al. 2012; Horne and Scotter 2016). The second effect considered was soil denitrification where nitrate is reduced and removed from anoxic soil via gaseous forms (Addiscott and Powlson 1992; Cameron et al. 2013). Soil drainage class, as defined in the Land Resource Information Systems (LRIS) soil map layers (Newsome et al. 2008), was selected as the attribute that best represented these aspects of N leaching vulnerability. This hydrological feature has also been used by others to improve models that link landscape vulnerability to nitrate contamination of groundwater (e.g. Nolan and Hitt 2006; Burow et al. 2010). Three classes of soil drainage were thus determined:*
(3a) 'Light' soils, defined as having PAW60cm contents less than 85 mm (and representing 20% of the soil types documented in Dairybase file information);
(3b) 'Well-drained' soils, classified as 'well' or 'moderately well' drained in the LRIS mapping system; and
(3c) 'Poorly-drained' soils, classified as having 'imperfect', 'poor' or 'very poor' soil drainage classes (Newsome et al. 2008).