Freshwater analysis support to Environment Southland (June 2020)

30th June 2020

Prepared by:

Beverley Horn

PREPARED FOR: **Environment Southland**

CLIENT REPORT No: **CSC20015**

REVIEWED BY: **Bridget Armstrong and Sarah Coxon**

DISCLAIMER

NOTE

The report has been written as a tool for ES staff to assist in understanding the analyses that have been conducted by ESR and provide some possible ways/text to communicate the findings of the analyses. It is not intended to be a stand alone report suitable for the reading by the general public, though the aim is that some sections may be adapted in other communications for a wider audience.

CONTENTS

1. INTRODUCTION

This document provides a description of the data analyses undertaken for Environment Southland in June 2020 to assist them to improve their understanding of *E. coli* attribute states and how these relate to sites around the Southland Region.

This report includes

1. An introduction to percentile calculation methods and how the choice of percentile affects the calculated value.
2. Calculation and summary of the *E.coli* attribute states for NOF monitoring sites and how these relate to the distribution of the *E.coli* concentration data.
3. Univariate check for seasonal effects on *E.coli* concentrations.
4. An investigation into rainfall effect on *E. coli* levels for a number of recreational water use sites.

The analyses have been conducted using the R statisitical software and associated packages. The analysis and graphics/table output source code has been provided to Environment Southland along side this report. Functions written by the report authors that relate to specific parts of the report are provided in blue italics, e.g. *ByMonth_plot()*

The report has been written as a tool for ES staff to assist in understanding the analyses that have been conducted by ESR and provide some possible ways/text to communicate the findings of the analyses. It is not intended to be a stand alone report suitable for the reading by the general public, though the aim is that some sections may be adapted in other communications for a wider audience.

2. PERCENTILE METHODS

Summary:

- There are a number of methods to calculate a percentile value. The choice of method impacts on the calculated percentile values of *E. coli* concentration. For five NOF sampling sites, the choice of method affects the assignment of the *E. coli* attribute band.
- The Hazen method is the method required to be used in the 2003 Microbiological Water Quality Guidelines for Marine and Freshwater Recreational Areas (MfE) and provides a mid-point approach to calculating percentiles from sample data. The authors suggest this approach should be continued to be used.
- The 95th percentile lies between the 57th and 58th sample *E. coli* concentrations, when ordered smallest to largest. The two largest concentrations recorded at a site are not taken into account in allocating an *E. coli* attribute band.

2.1 WHAT IS A PERCENTILE VALUE?

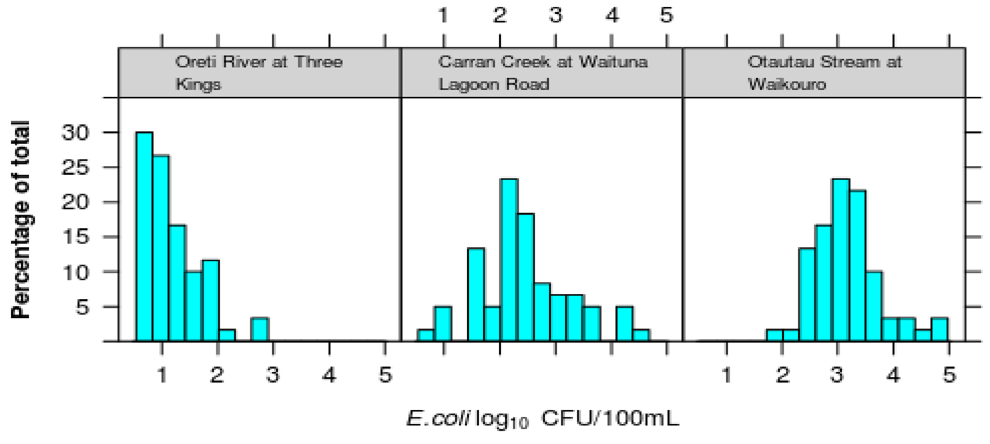
A 95th percentile value, is the value at which 95% of the data values fall below. In the freshwater sampling context; 95% of the samples from a site have concentrations less than or equal to the 95th percentile of *E.coli*/100mL.

The site median concentration is the 50th percentile of the site data.

2.2 HOW TO CALCULATE THE PERCENTILE VALUE?

If we plot a histogram of the frequency of different *E.coli* concentrations at a sampling site, we can see that the general shape of the distributions are not same at all sites. Figure 1 gives some examples of different distributions observed in the Southland data. A plot of all the distributions can be found in Appendix A.

Figure 1: Log$_{10}$ *E. coli* concentration distribution from three Southland NOF sampling sites to show different distribution shapes. *[R function: compare_somedist_plot()]*



If the shape of the *E. coli* concentration distribution was consistent across sampling sites, established distribution types could be used to model the data. Such distributions are defined by parameters and have defined approaches to calculating percentile values. For example, the parameter standard deviation defines the shape of the distribution for the Normal Distribution.

As shown above, the *E.coli* concentration distribution shapes are not consistent across sampling sites and in some cases there are censored data due to the limit of detection or limit of quantification of the sampling method. This means we need an approach that does not expect the data to follow a specific pattern (non-parametric approach) and can deal with censored values.

The percentile is calculated using a Ranking approach:

Step 1: Order the concentrations (CFU/mL) from smallest to largest
[4, 200, 2, 1, 7, 10] → **[1, 2, 4, 7, 10, 200]**

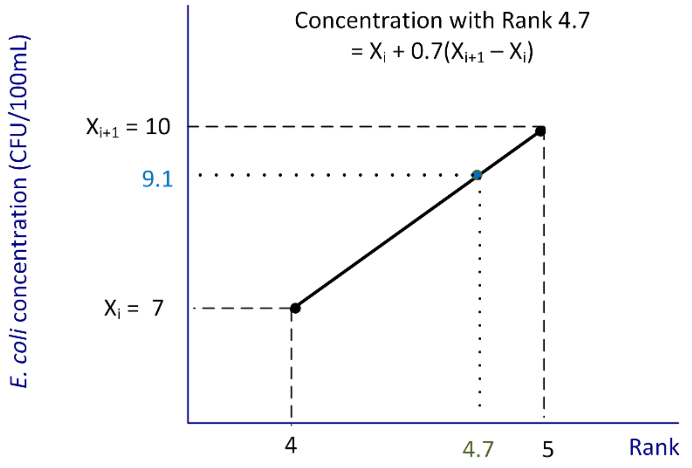
Step 2: Calculate the **Percentile Rank** you wish to calculate. This can be calculated in a number of different ways which is discussed in the next section.

The Rank number can be split into two parts; the whole (or integer) part and the fraction part. For example, if Rank is **4.7** then **4** is the integer (R_{int}) and **0.7** is the fraction (R_{frac}).

Step 3: **If the Percentile Rank only has an integer part** (e.g. **4.0** or **5.0**), the Rank indicates the sample concentration that represents the required percentile.

In the above example, a 75th percentile has a Hazen Rank of **5.0**, which indicates the 5th ordered value, which is **10 CFU/mL**.

If the Percentile Rank has an integer and fraction part (e.g. **4.7**), the percentile sample concentration is found by linearly interpolating between the data points adjacent to the rank. In this example a Rank of **4.7** corresponds to a concentration of **9.1 CFU/100mL**.



2.3 HOW TO CALCULATE THE PERCENTILE RANK?

Percentiles have a range from 0 to 100% and in the context of this work we assume the data points are positioned at equally spaced percentiles.

For example:

Percentile:	**0**	**17**	**33**	**50**	**67**	**83**	**100 %**
Data point	**?**	**1**	**2**	**4**	**7**	**10**	**200 CFU/100mL**

In this example we assume the maximum possible value is 200 CFU/mL, but do not set the minimum possible value, though one option could be to set this to zero.

Given we are sampling a small number of samples from the environment, it could equally be put forward that relationship between the data and percentiles should be

Percentile:	**0**	**17**	**33**	**50**	**67**	**83**	**100 %**
Data point	**1**	**2**	**4**	**7**	**10**	**200**	**? CFU/100mL**

This time we set the lowest observed value to be at the 0^{th} percentile, but do not define the maximum value. These two examples are the possible extremes of the relationship between the data points and the percentile placement.

A better option to use a relationship that is somewhere between these two options.

Some possible options are given below, along with the sample size required to determine the required percentile.

METHOD	PERCENTILE RANK	SAMPLE SIZE REQUIRED FOR GIVEN *p* AND *q=1-p*
Hazen	$R = \frac{1}{2} + pn$	$n \geq \frac{1}{2q}$ and $n \geq \frac{1}{2p}$
Excel PERCENTILE.INC	$R = 1 + p(n-1)$	$n \geq 1$
Excel PERCENTILE.EXC (Weibull)	$R = p(n+1)$	$n \geq \frac{p}{q}$ and $n \geq \frac{q}{p}$
Tukey	$R = \frac{1}{3} + p\left(n + \frac{1}{3}\right)$	$n \geq \frac{1+p}{3q}$ and $n \geq \frac{1+q}{3p}$

n = number of data points, ***p*** is the percentile value divided by 100 and ***R*** is the calculated Rank

The Hazen method is the mid-point of the two examples given at the beginning of section and the method currently used by the 2003 Microbiological Water Quality Guidelines for Marine and Freshwater Recreational Areas (MfE).

These methods calculate percentiles values from lowest to highest in the following order:

PERCENTILE.INC **<** Hazen **<** Tukey **<** Weibull, for percentiles greater than 50%.

The methods in the table also have the advantage that the median value aligns with the 50^{th} percentile.

2.4 HOW MUCH EFFECT DOES PERCENTILE METHOD HAVE ON SOUTHLAND DATA GRADING?

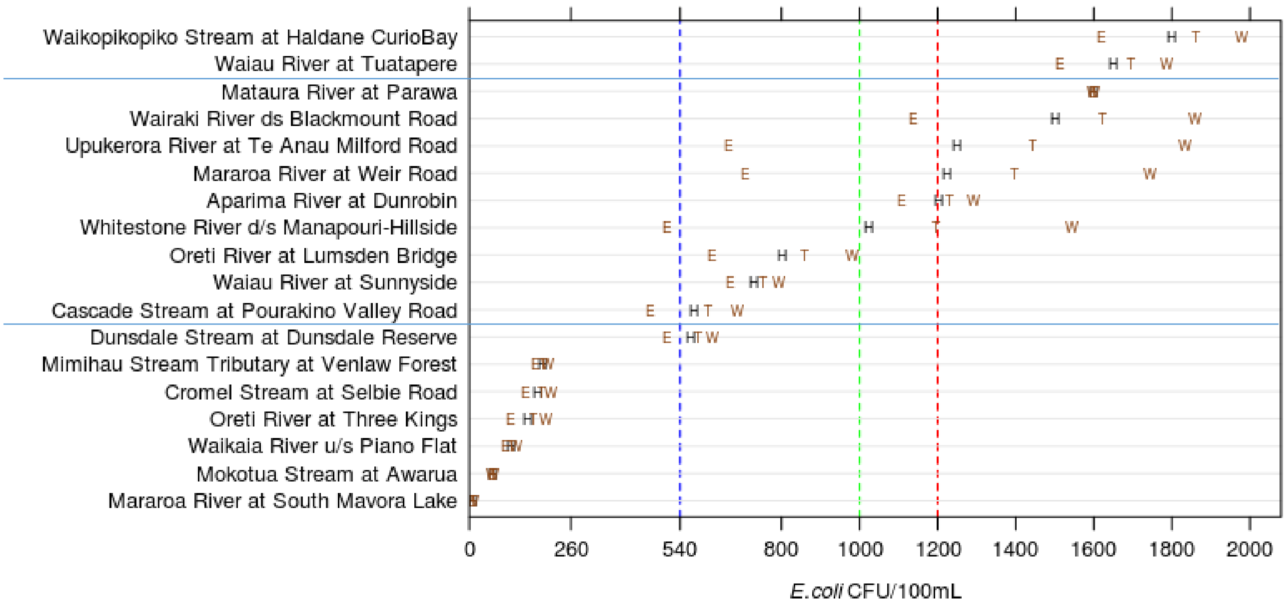
The 95th percentile (P_{95}) is one of the fours state measures considered in the *E.coli* attribute band table.

A (Blue)	B (Green)	C (Yellow)	D (Orange)	E (Red)
$P_{95} \leq 540$	$P_{95} \leq 1000$	$P_{95} \leq 1200$	$P_{95} > 1200$	$P_{95} > 1200$

The data from 60 NOF sampling sites was used to look at how the choice of percentile method affected the 95th percentile estimate of the *E. coli* concentration at a site. For each site, a target of 60 samples were collected over 5 years. Where 60 samples where not collected during the 5 years, earlier samples were included to give a total of 60. *[R function: number_samples()]*

Figure 2 shows the 95th percentile estimates from the Excel, Hazen, Tukey and Weibull methods for the 18 best performing NOF sites *[R functions: RWQ_stats_dataframe() and compare_percentile_plot()]*.

Figure 2: Comparison of 95th percentile of *E. coli* concentration samples derived by the E: Excel PERCENTILE.INC(), H: Hazen, T: Tukey and W: Weibull methods for the 18 best performing NOF sites. Coloured lines are at 540 (blue), 1000 (green) and 1200 (red) E. coli/100mL.



From considering all sites and Figure 2, the following can be observed:

- The 6 best performing sites (using 95th percentile) have data well below the 540 CFU/100mL value and allocation of attribute band is not affected by percentile method.
- For the 45 worst performing sites (using 95th percentile), the 95th percentile calculated by all the methods resulted in a value greater than 1200 CFU/100mL, so method choice does not affect the allocation of the attribute band.
- For eight of the nine remaining sites, the choice of percentile calculation method does affect the location within the 95th percentile bands. In the most extreme case, the

Whitestone River d/*s Manapouri-Hillside* site can be classed as less than 540 or more than 1200 CFU/100ml depending on the approach used.

- Comparing the Hazen to the Weibull (Excel PERCENTILE.EXC) 95th percentiles shows there is a difference between the overall band allocation for two sites; *Aparima River at Dunrobin* and *Whitestone River d/s Manapouri-Hillside*. Both sites would have an overall classification of **Orange** if using the Weibull method, and **Yellow** if using the Hazen method.

2.5 WHY CAN THERE BE SO MUCH DIFFERENCE BETWEEN PERCENTILE METHODS SOMETIMES?

A minimum of 60 samples from each site is used to calculate the site *E. coli* concentration attribute statistics. If 60 samples are considered …

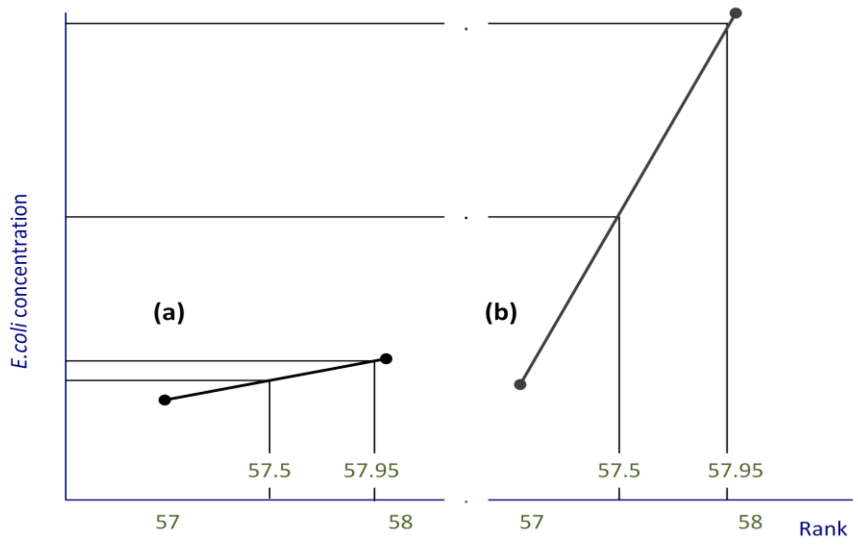
For Hazen 95th percentile, the Rank = ½ + np = ½ + 60x0.95 = 57.50

For Weibull 95th percentile, the Rank = p (n+1) = 0.95 (60+1) = 57.95

So for this scenario, the value of the 57th and 58th ordered *E.coli* concentrations will be used to calculate the 95th percentile. The two highest concentrations sampled are not considered when determining the attribute band.

The size of the difference between the 57th and 58th concentrations will drive how different the results from the different methods are, as shown in Figure 3.

Figure 3: Sensitivity of 95th percentile estimation to Hazen (Rank 57.5) and Weibull (Rank 57.95) methods when the 57th and 58th ordered values of *E. coli* concentration are close together (a) or very different (b).



3. *E. COLI* BANDS (REGIONAL OVERVIEW)

3.1 DEFINING THE SITE ATTRIBUTE BAND

To calculate the Human Contact Attributable Band for *E. coli,* four different measures of the distribution must be calculated as shown in Table 1 *[R function: RWQ_stats_dataframe()]*. The attribute band is allocated by choosing the band associated with the highest risk state observed in the four measures *[R function: RWQ_Ecoli_compare_Table11()]*.

For example, consider a site had 7% of samples over 540 *E. coli*/100mL (B Band), 10% samples over 260 *E. coli*/100mL (A Band), median 120 *E. coli*/100mL (A Band) and 95th percentile of 1500mL (D Band). This site would be classified as Band D (Orange).

Table 1*:* *E. coli* Attribute Table for Human Contact

Attribute Band	Numeric Attribute State			
	% exceedances over 540 *E. coli*/100mL	**% exceedances over 260 *E. coli*/100mL**	**Median concentration *E. coli*/100mL**	**95th percentile of *E. coli*/100mL**
A (Blue)	< 5%	< 20%	≤ 130	≤ 540
B (Green)	5 – 10%	20 – 30%	≤ 130	≤ 1000
C (Yellow)	10 – 20%	20 – 34%	≤ 130	≤ 1200
D (Orange)	20 – 30%	> 34%	> 130	> 1200
E (Red)	> 30%	> 50%	> 260	> 1200

Increasing risk ↓

3.2 CALCULATION OF NUMERIC ATTRIBUTE STATES

The numeric attribute states for the 60 ES NOF sites have been calculated using 60 samples with the latest month of sampling; July 2019. For some sites one or more samples are missing during the five year period, and data closest to the defined 5 year period have been added to ensure at least 60 samples are used in the calculations *[R function: number_samples()]*.

The 95th percentile is calculated using the Hazen approach. The numeric states are presented in Table 2 as numeric values, and the cell shaded with the colour associated with the lowest risk band that the data complies with. In the above example, the 95th percentile was a value 1500 *E. coli*/100mL. This is shaded orange, as this is the lowest risk that meets the condition '> 1200' *[R function: HumanCont_AttTable()]*.

3.3 JULY 2019 ATTRIBUTE STATES

Table 2 provides the attribute numerical states and band for each of the sixty NOF sampling sites, ordered by band and then alphabetical order of the site name.

Ten of the sites fall into the A or B bands, 2 in the C band, 19 in the D band and 29 in the E band. There are 10 sites in the C and D bands where the 95th percentile is the state controlling the band allocation. For the remaining C/D sites the 95th percentile is one of two or more states which control band allocation. There are nine D band sites where the numerical states are consistently in the D and C ranges.

Table 2: *E.coli* Human Contact Attribute Table for NOF sites

Site	Numeric Attribute State				Attribute Band
	% exceedances over 540 *E. coli* /100mL	% exceedances over 260 *E. coli* /100mL	Median concentration *E. coli* /100mL	95th percentile of *E. coli* /100mL	
Cromel Stream at Selbie Road	0.0	1.7	14	175	A
Mararoa River at South Mavora Lake	0.0	0.0	5	10	A
Mimihau Stream Tributary at Venlaw Forest	0.0	1.7	30	185	A
Mokotua Stream at Awarua	1.7	1.7	5	60	A
Oreti River at Three Kings	0.0	3.3	10	150	A
Waikaia River u/s Piano Flat	0.0	0.0	20	105	A
Cascade Stream at Pourakino Valley Road	5.0	23.3	130	575	B
Dunsdale Stream at Dunsdale Reserve	5.0	23.3	120	565	B
Oreti River at Lumsden Bridge	8.3	16.7	60	800	B
Waiau River at Sunnyside	8.3	15.0	35	730	B
Aparima River at Dunrobin	8.3	18.3	60	1,200	C
Whitestone River d/s Manapouri-Hillside	5.0	10.0	10	1,025	C
Aparima River at Thornbury	15.0	26.7	125	3,150	D
Carran Creek at Waituna Lagoon Road	30.0	41.7	215	16,500	D
Hamilton Burn at Affleck Road	10.0	26.7	120	2,650	D
Irthing Stream at Ellis Road	11.7	20.0	90	3,050	D
Lill Burn at Lill Burn-Monowai Road	18.3	36.7	160	3,000	D
Mararoa River at The Key	11.7	15.0	32	2,400	D
Mararoa River at Weir Road	6.7	15.0	35	1,225	D
Mataura River at Mataura Island Bridge	25.0	48.3	240	7,500	D
Mataura River at Parawa	16.7	25.0	145	1,600	D
Mokoreta River at Wyndham River Road	26.7	50.0	260	3,450	D
Oreti River at Wallacetown	16.7	25.0	140	2,350	D
Upukerora River at Te Anau Milford Road	6.7	8.3	30	1,250	D
Waiau River at Tuatapere	16.7	25.0	80	1,650	D
Waikaia River at Waikaia	20.0	35.0	170	3,900	D
Waikaia River at Waipounamu Bridge Road	20.0	31.7	135	4,050	D
Waikopikopiko Stream at Haldane CurioBay	15.0	35.0	160	1,800	D
Waimea Stream at Mandeville	26.7	48.3	260	13,500	D
Wairaki River ds Blackmount Road	11.7	11.7	50	1,500	D
Waituna Creek at Marshall Road	28.3	45.0	235	2,900	D

Site	Numeric Attribute State				Attribute Band
	% exceedances over 540 *E. coli* /100mL	% exceedances over 260 *E. coli* /100mL	Median concentration *E. coli* /100mL	95th percentile of *E. coli* /100mL	
Bog Burn d/s Hundred Line Road	68.3	88.3	925	5,800	E
Dipton Stream at South Hillend-Dipton Road	33.3	55.0	290	10,250	E
Hedgehope Stream 20m u/s Makarewa Confl	53.3	73.3	590	8,500	E
Longridge Stream at Sandstone	31.7	60.0	315	15,500	E
Makarewa River at Lora Gorge Road	46.7	63.3	460	8,000	E
Makarewa River at Wallacetown	36.7	58.3	345	26,000	E
Mataura River 200m d/s Mataura Bridge	73.3	86.7	1,050	18,500	E
Mataura River at Gore	36.7	63.3	385	9,500	E
Mimihau Stream at Wyndham	35.0	55.0	310	3,200	E
Moffat Creek at Moffat Road	36.7	48.3	215	5,700	E
North Peak Stream at Waimea Valley Road	31.7	41.7	220	5,050	E
Opouriki Stream at Tweedie Road	55.0	80.0	600	11,000	E
Orauea River at Orawia Pukemaori Road	33.3	48.3	250	10,000	E
Otamita Stream at Mandeville	26.7	53.3	275	4,100	E
Otapiri Stream at Otapiri Gorge	35.0	60.0	365	8,000	E
Otautau Stream at Otautau-Tuatapere Road	60.0	85.0	850	12,700	E
Otautau Stream at Waikouro	75.0	91.7	1,400	22,500	E
Otepuni Creek at Nith Street	81.7	90.0	1,700	9,000	E
Oteramika Stream at Seaward Downs	53.3	75.0	655	5,450	E
Pourakino River at Traill Road	31.7	65.0	370	2,400	E
Sandstone Stream at Kingston Crossing Rd	45.0	56.7	440	15,000	E
Tokanui River at Fortrose Otara Road	26.7	50.0	270	9,000	E
Tussock Creek at Cooper Road	63.3	86.7	800	28,000	E
Waihopai River u/s Queens Drive	31.7	56.7	330	4,450	E
Waikaka Stream at Gore	31.7	46.7	225	20,000	E
Waikawa River at Progress Valley	51.7	78.3	585	18,500	E
Waikiwi Stream at North Road	40.0	71.7	420	6,500	E
Waimatuku Stream at Lorneville Riverton Hwy	33.3	70.0	380	3,350	E
Winton Stream at Lochiel	73.3	91.7	900	13,000	E

3.4 SEASONAL TRENDS

To examine possible seasonal trends, the River Water Quality (RWQ) data for each sampling site are plotted in Appendix B. The *E. coli* concentration is plotted against the sampling day of the year. So the 3rd of January in any year is plotted as day 3 for example. The colour of header panel relates to the Attribute Band of the site. *E. coli* data greater than 2000 is indicated by an X in the plot.

There are no strong seasonal components for counts which result in higher risk attribute bands, across catchments. There may be a lower risk in the third quarter of the year in the highest upstream sampling sites for the Mataura catchment, but given the variability in sampling results, this may just be an artefact of the natural variability of the data. This will be considered in more detail in the rainfall section below.

4. SUMMARY BY CATCHMENT

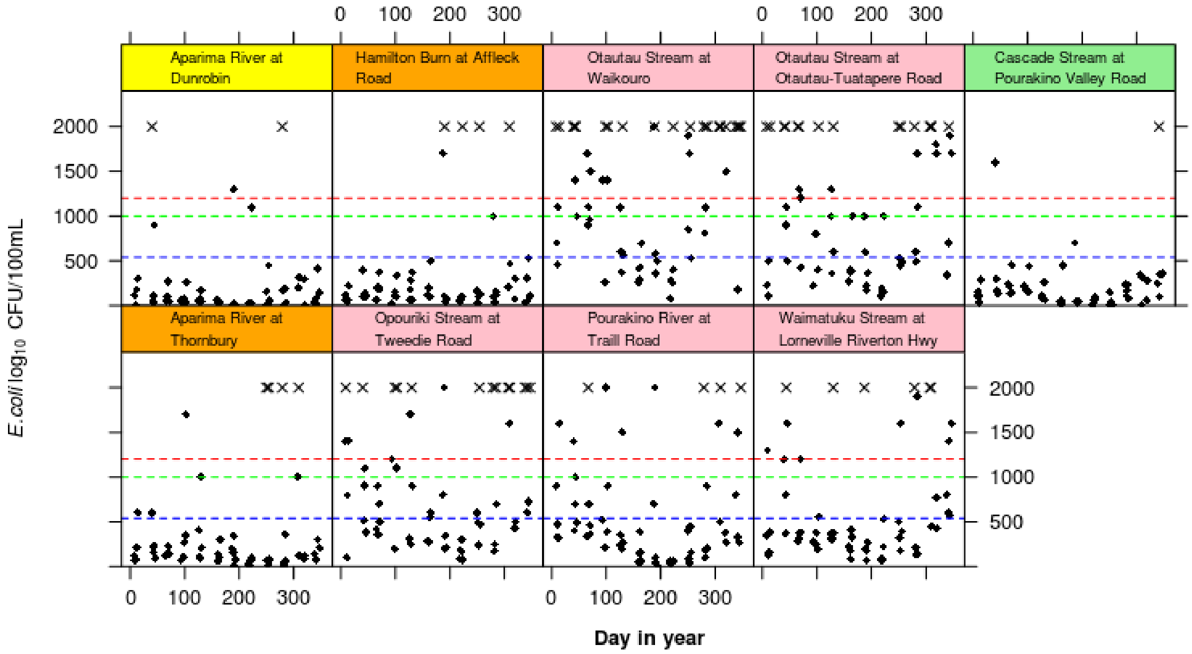
This section provides the *E. coli* catchment specific data from 60 samples by:

- Attribute state and bands
- Sample *E.coli* plotted against day of the year to investigate. Any seasonal trends in the higher *E. coli* concentrations.
- Distribution of *E. coli* concentrations.

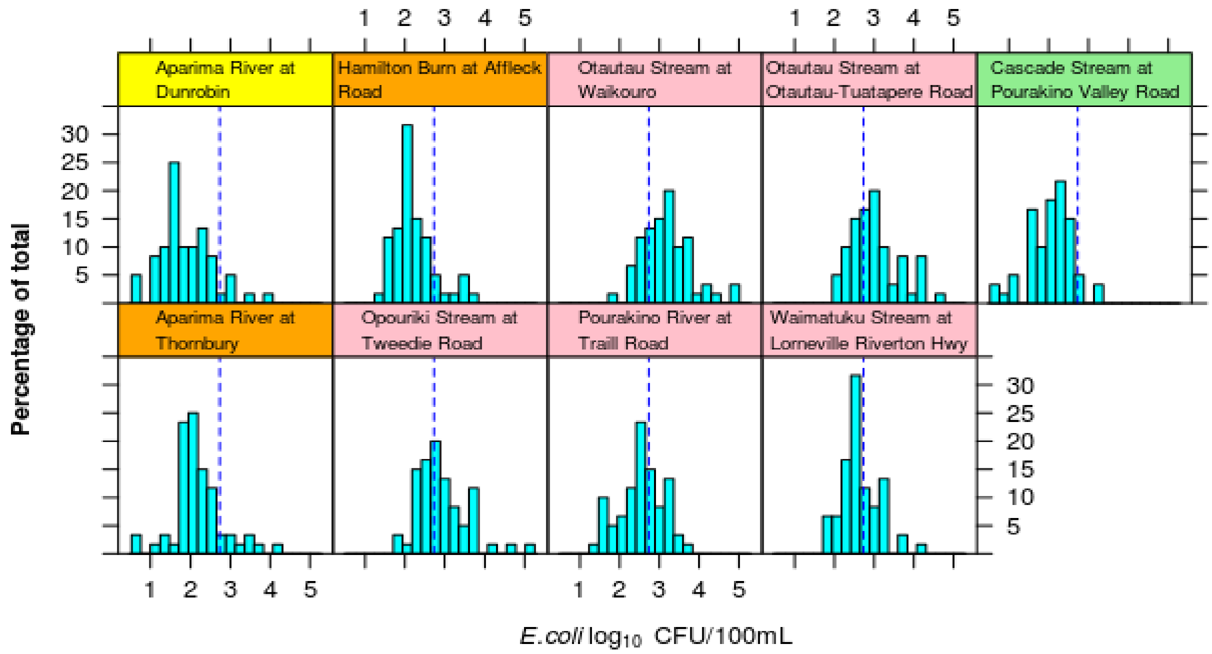
4.1 APARIMA

Site	Numeric Attribute State				Attribute Band
	% exceedances over 540 *E. coli* /100mL	% exceedances over 260 *E. coli* /100mL	Median concentration *E. coli* /100mL	95th percentile of *E. coli* /100mL	
Cascade Stream at Pourakino Valley Road	5.0	23.3	130	575	B
Aparima River at Dunrobin	8.3	18.3	60	1,200	C
Aparima River at Thornbury	15.0	26.7	125	3,150	D
Hamilton Burn at Affleck Road	10.0	26.7	120	2,650	D
Opouriki Stream at Tweedie Road	55.0	80.0	600	11,000	E
Otautau Stream at Otautau-Tuatapere Road	60.0	85.0	850	12,700	E
Otautau Stream at Waikouro	75.0	91.7	1,400	22,500	E
Pourakino River at Traill Road	31.7	65.0	370	2,400	E
Waimatuku Stream at Lorneville Riverton Hwy	33.3	70.0	380	3,350	E

Plot of *E. coli* concentration against the sampling day of the year to investigate possible seasonal trends in the data. Plots in order of distance to from the sea (top left – furthest away). Colour of header panel relates to the Attribute Band of the site. *E. coli* data greater than 2000 is indicated by an X in the plot. Coloured lines are at 540 (blue), 1000 (green) and 1200 (red) *E. coli*/100mL.



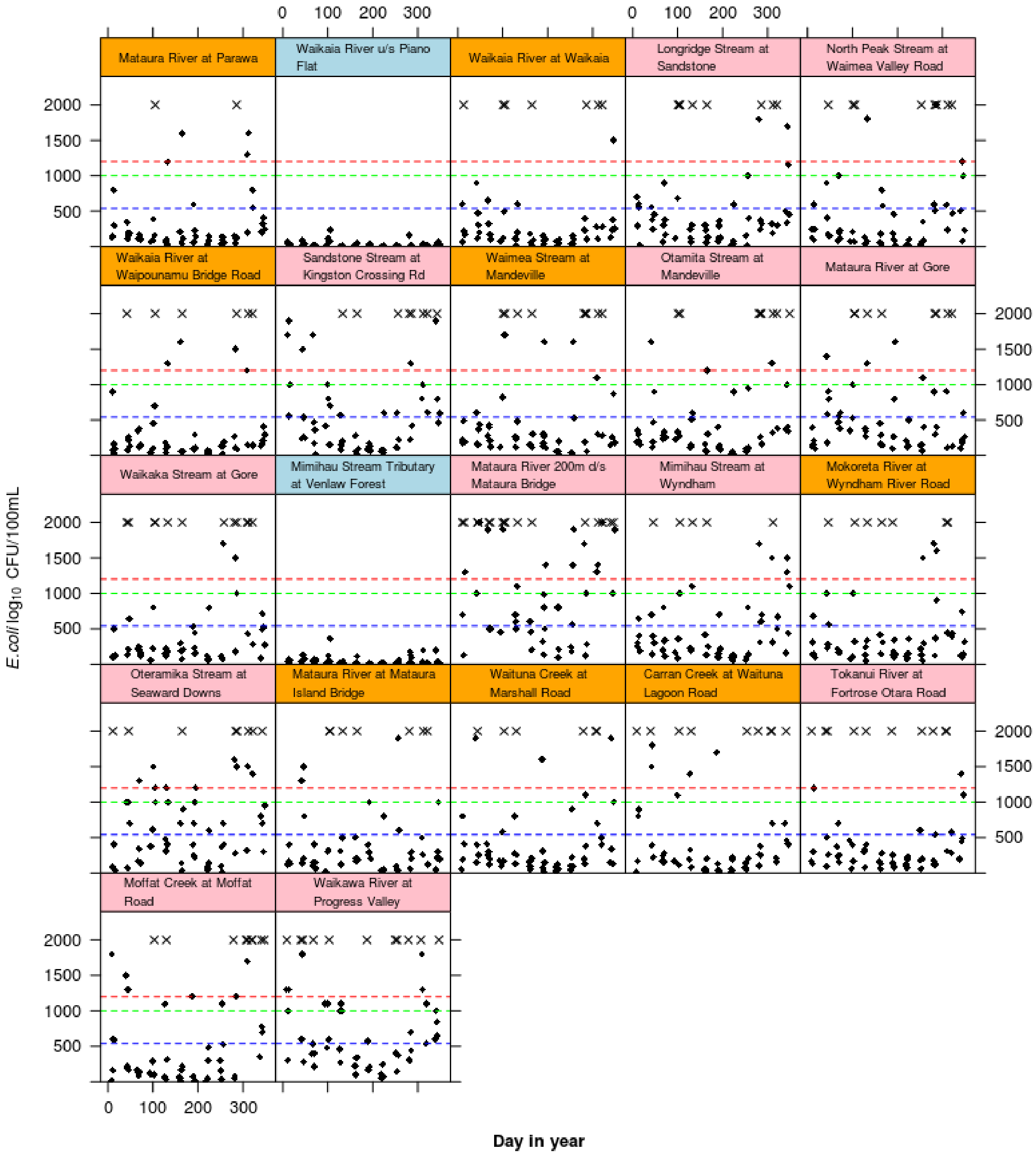
Plot of *E. coli* concentration distributions. Plots in order of distance to from the sea (top left – furthest away). Colour of header panel relates to the Attribute Band of the site. Blue dotted line is at 540 *E.coli/*100mL.



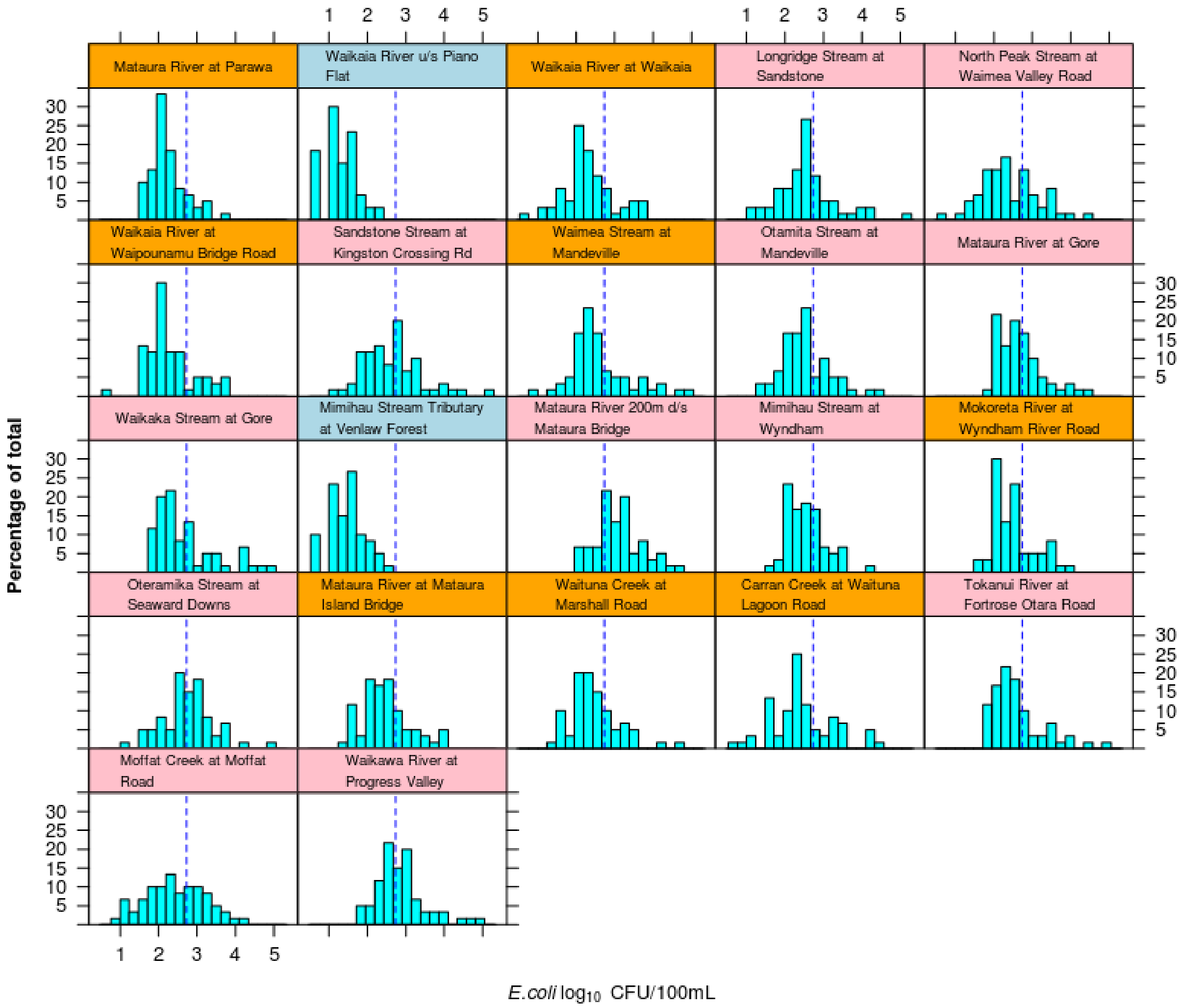
4.2 MATAURA

Site	Numeric Attribute State				Attribute Band
	% exceedances over 540 *E. coli* /100mL	% exceedances over 260 *E. coli* /100mL	Median concentration *E. coli* /100mL	95th percentile of *E. coli*/100mL	
Mimihau Stream Tributary at Venlaw Forest	0.0	1.7	30	185	A
Waikaia River u/s Piano Flat	0.0	0.0	20	105	A
Carran Creek at Waituna Lagoon Road	30.0	41.7	215	16,500	D
Mataura River at Mataura Island Bridge	25.0	48.3	240	7,500	D
Mataura River at Parawa	16.7	25.0	145	1,600	D
Mokoreta River at Wyndham River Road	26.7	50.0	260	3,450	D
Waikaia River at Waikaia	20.0	35.0	170	3,900	D
Waikaia River at Waipounamu Bridge Road	20.0	31.7	135	4,050	D
Waimea Stream at Mandeville	26.7	48.3	260	13,500	D
Waituna Creek at Marshall Road	28.3	45.0	235	2,900	D
Longridge Stream at Sandstone	31.7	60.0	315	15,500	E
Mataura River 200m d/s Mataura Bridge	73.3	86.7	1,050	18,500	E
Mataura River at Gore	36.7	63.3	385	9,500	E
Mimihau Stream at Wyndham	35.0	55.0	310	3,200	E
Moffat Creek at Moffat Road	36.7	48.3	215	5,700	E
North Peak Stream at Waimea Valley Road	31.7	41.7	220	5,050	E
Otamita Stream at Mandeville	26.7	53.3	275	4,100	E
Oteramika Stream at Seaward Downs	53.3	75.0	655	5,450	E
Sandstone Stream at Kingston Crossing Rd	45.0	56.7	440	15,000	E
Tokanui River at Fortrose Otara Road	26.7	50.0	270	9,000	E
Waikaka Stream at Gore	31.7	46.7	225	20,000	E
Waikawa River at Progress Valley	51.7	78.3	585	18,500	E

Plot of *E. coli* concentration against the sampling day of the year to investigate possible seasonal trends in the data. Plots in order of distance to from the sea (top left – furthest away). Colour of header panel relates to the Attribute Band of the site. *E. coli* data greater than 2000 CFU/100mL is indicated by an X in the plot. Coloured lines are at 540 (blue), 1000 (green) and 1200 (red) *E. coli*/100mL.



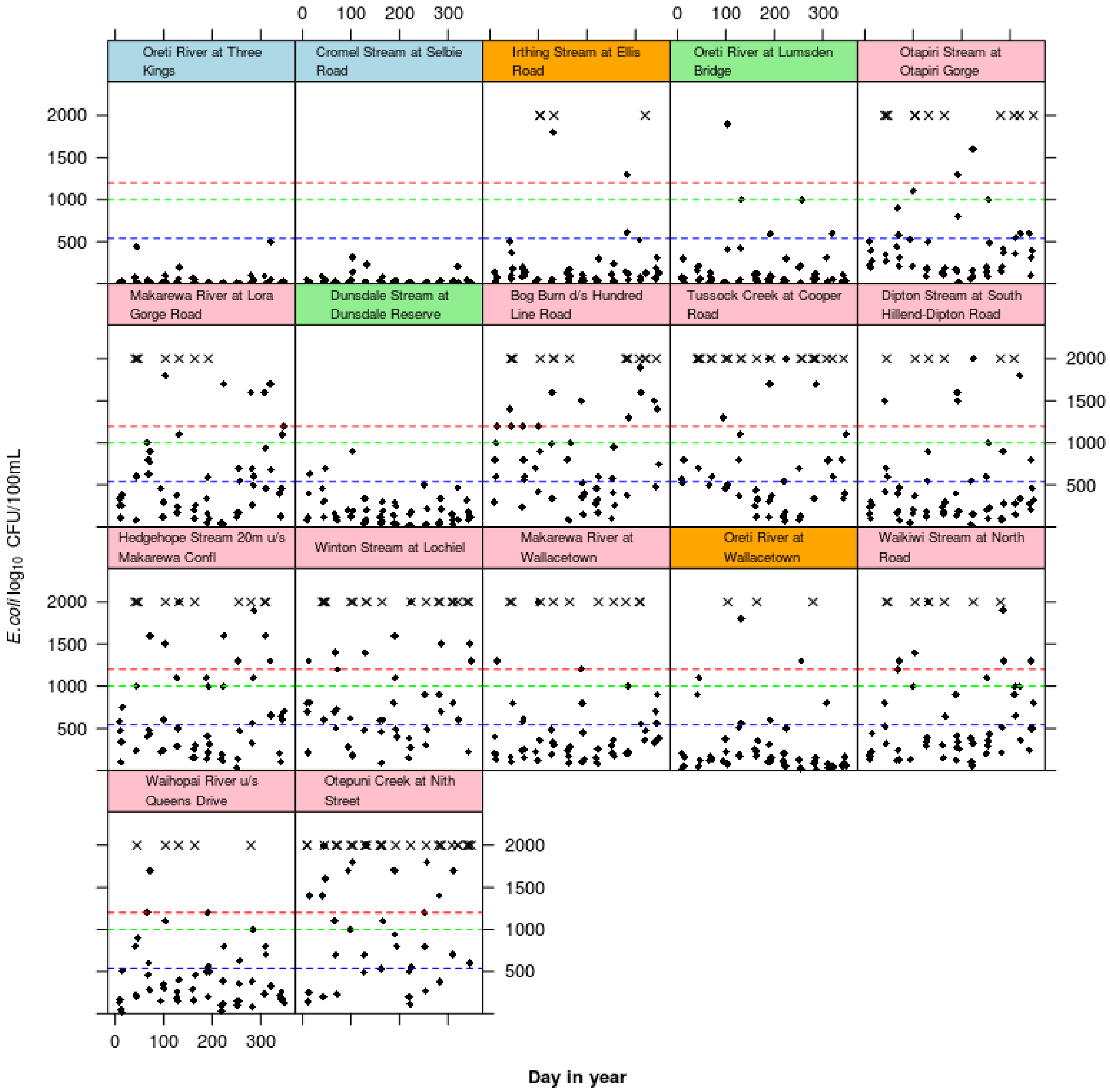
Plot of *E. coli* concentration distributions. Plots in order of distance to from the sea (top left – furthest away). Colour of header panel relates to the Attribute Band of the site. Blue dotted line is at 540 *E.coli*/100mL



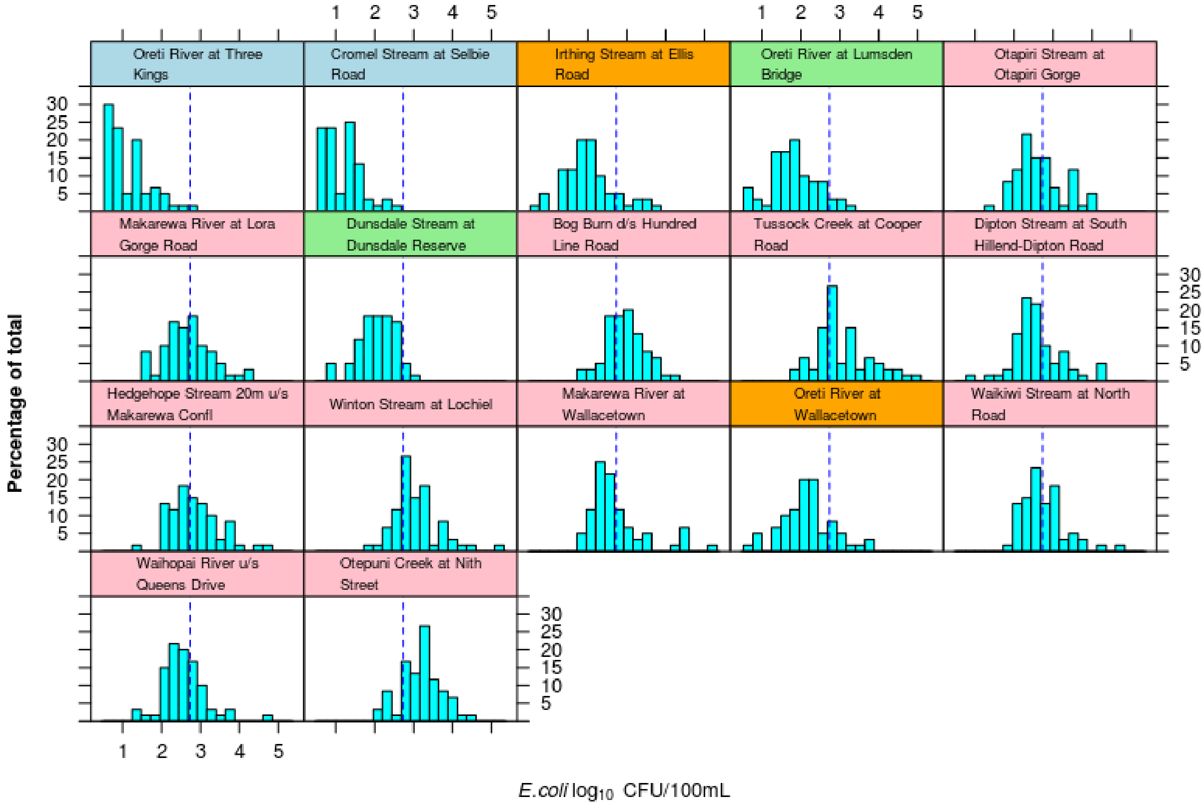
4.3 ORETI

Site	Numeric Attribute State				Attribute Band
	% exceedances over 540 *E. coli* /100mL	% exceedances over 260 *E. coli* /100mL	Median concentration *E. coli* /100mL	95th percentile of *E. coli*/100mL	
Cromel Stream at Selbie Road	0.0	1.7	14	175	A
Oreti River at Three Kings	0.0	3.3	10	150	A
Dunsdale Stream at Dunsdale Reserve	5.0	23.3	120	565	B
Oreti River at Lumsden Bridge	8.3	16.7	60	800	B
Irthing Stream at Ellis Road	11.7	20.0	90	3,050	D
Oreti River at Wallacetown	16.7	25.0	140	2,350	D
Bog Burn d/s Hundred Line Road	68.3	88.3	925	5,800	E
Dipton Stream at South Hillend-Dipton Road	33.3	55.0	290	10,250	E
Hedgehope Stream 20m u/s Makarewa Confl	53.3	73.3	590	8,500	E
Makarewa River at Lora Gorge Road	46.7	63.3	460	8,000	E
Makarewa River at Wallacetown	36.7	58.3	345	26,000	E
Otapiri Stream at Otapiri Gorge	35.0	60.0	365	8,000	E
Otepuni Creek at Nith Street	81.7	90.0	1,700	9,000	E
Tussock Creek at Cooper Road	63.3	86.7	800	28,000	E
Waihopai River u/s Queens Drive	31.7	56.7	330	4,450	E
Waikiwi Stream at North Road	40.0	71.7	420	6,500	E
Winton Stream at Lochiel	73.3	91.7	900	13,000	E

Plot of *E. coli* concentration against the sampling day of the year to investigate possible seasonal trends in the data. Plots in order of distance to from the sea (top left – furthest away). Colour of header panel relates to the Attribute Band of the site. *E. coli* data greater than 2000 CFU/100mL is indicated by an X in the plot. Coloured lines are at 540 (blue), 1000 (green) and 1200 (red) *E. coli*/100mL.



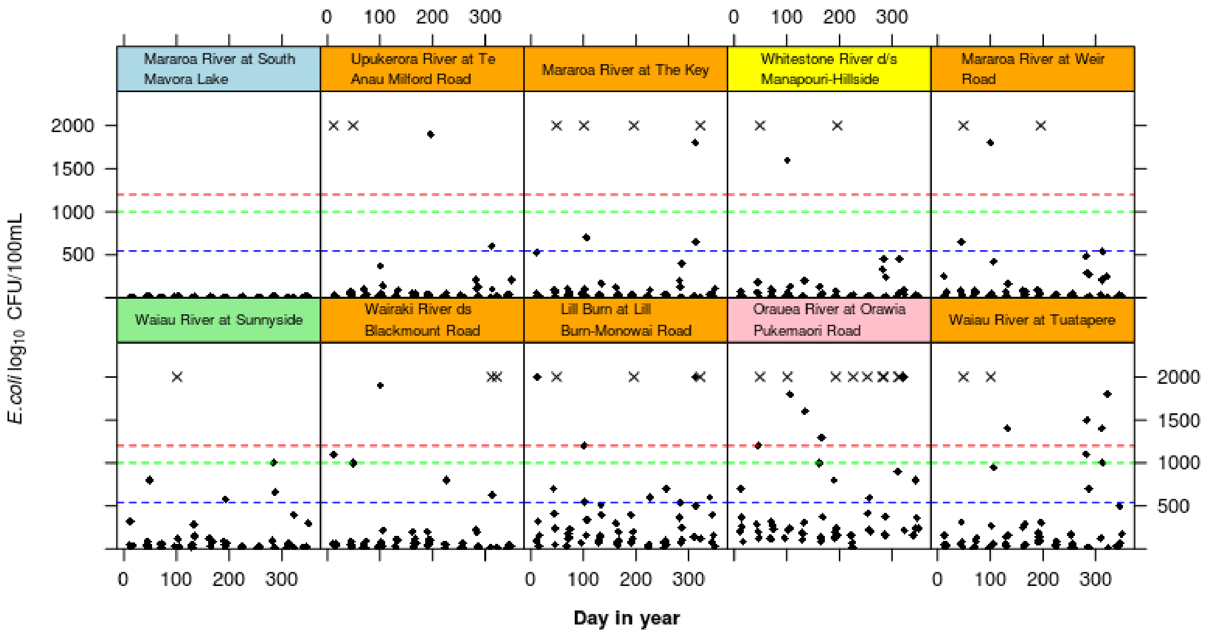
Plot of *E. coli* concentration distributions. Plots in order of distance to from the sea (top left – furthest away). Colour of header panel relates to the Attribute Band of the site. Blue dotted line is at 540 *E.coli*/100mL.



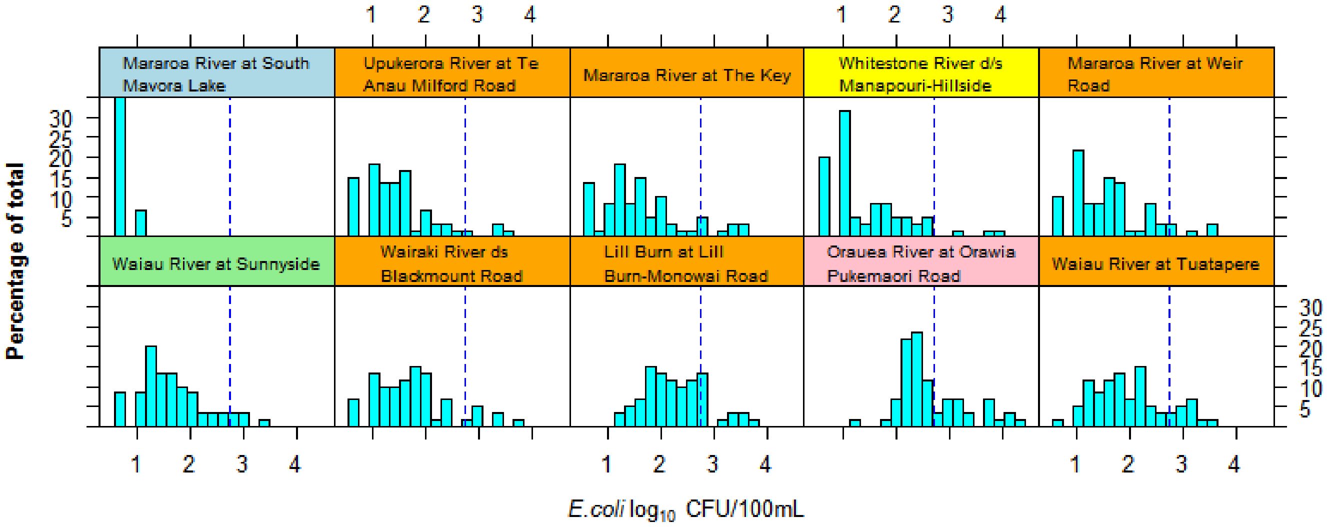
4.4 WAIAU

Site	Numeric Attribute State				Attribute Band
	% exceedances over 540 *E. coli* /100mL	% exceedances over 260 *E. coli* /100mL	Median concentration *E. coli* /100mL	95th percentile of *E. coli* /100mL	
Mararoa River at South Mavora Lake	0.0	0.0	5	10	A
Waiau River at Sunnyside	8.3	15.0	35	730	B
Whitestone River d/s Manapouri-Hillside	5.0	10.0	10	1,025	C
Lill Burn at Lill Burn-Monowai Road	18.3	36.7	160	3,000	D
Mararoa River at The Key	11.7	15.0	32	2,400	D
Mararoa River at Weir Road	6.7	15.0	35	1,225	D
Upukerora River at Te Anau Milford Road	6.7	8.3	30	1,250	D
Waiau River at Tuatapere	16.7	25.0	80	1,650	D
Wairaki River ds Blackmount Road	11.7	11.7	50	1,500	D
Orauea River at Orawia Pukemaori Road	33.3	48.3	250	10,000	E

Plot of *E. coli* concentration against the sampling day of the year to investigate possible seasonal trends in the data. Plots in order of distance to from the sea (top left – furthest away). Colour of header panel relates to the Attribute Band of the site. *E. coli* data greater than 2000 CFU/100mL is indicated by an X in the plot. Coloured lines are at 540 (blue), 1000 (green) and 1200 (red) *E. coli*/100mL.



Plot of *E. coli* concentration distributions. Plots in order of distance to from the sea (top left – furthest away). Colour of header panel relates to the Attribute Band of the site. Blue dotted line is at 540 *E.coli*/100mL.



5. WAIAU AT TUATAPERE – RAIN EFFECTS

Plots in the following sections were generated using the *ES_RWQ_rainfall_script.R* and *ES_RWQ_rainfall_functions.R* files. Unless otherwise stated the data is from the period 1998 to March 2020.

RWQ and Recreational sampling site: Waiau River at Tuatapere

Rainfall site: Clifden (Approximately 16 km upstream from water sampling site)

5.1 *E.COLI* SAMPLING WITH NO RAIN RECORDED PREVIOUS 72 HOURS

This section considers the baseline *E. coli* concentrations in the river when there has been no recent rain at the Clifden rain gauge. Figure 4 and Figure 5 show histograms of *E. coli* concentration from samples where there has been less than 1mm of rain ("No Rain") and 1mm or more ("Rain"), for the RWQ and Recreational water sampling. *RWQrain_QC_dataframe() Ecoli_with_no_rain()*

Figure 4: Distribution of *E. coli* concentrations for samples collected at "Waiau at Tuatapere" when rain greater than 1mm was recorded at Clifden in the 72 hours prior to sampling. Data collected between April 1998 and March 2020. Blue line is 540 *E.coli*/100mL and the red line is 1200 *E.coli*/100mL.

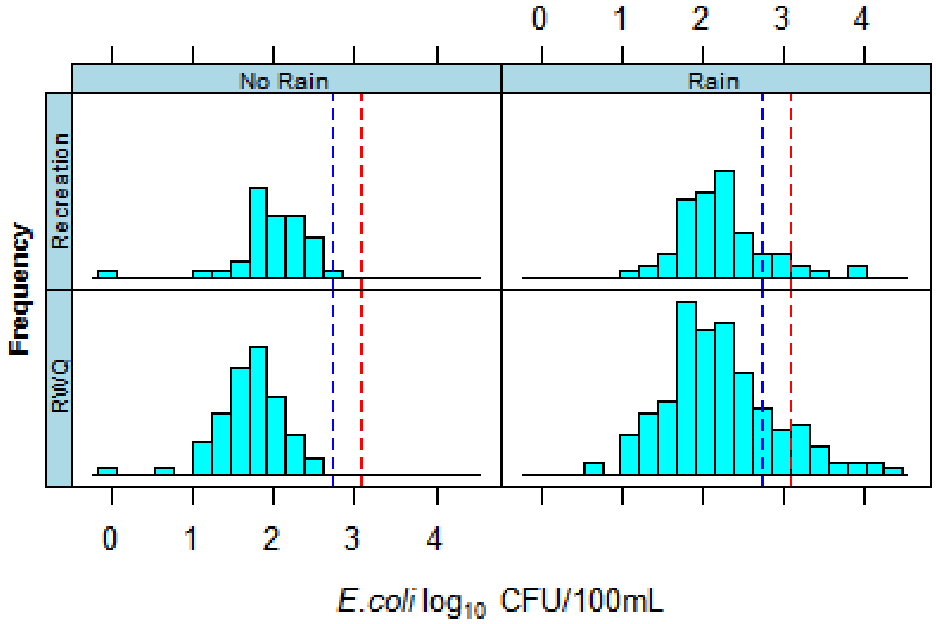
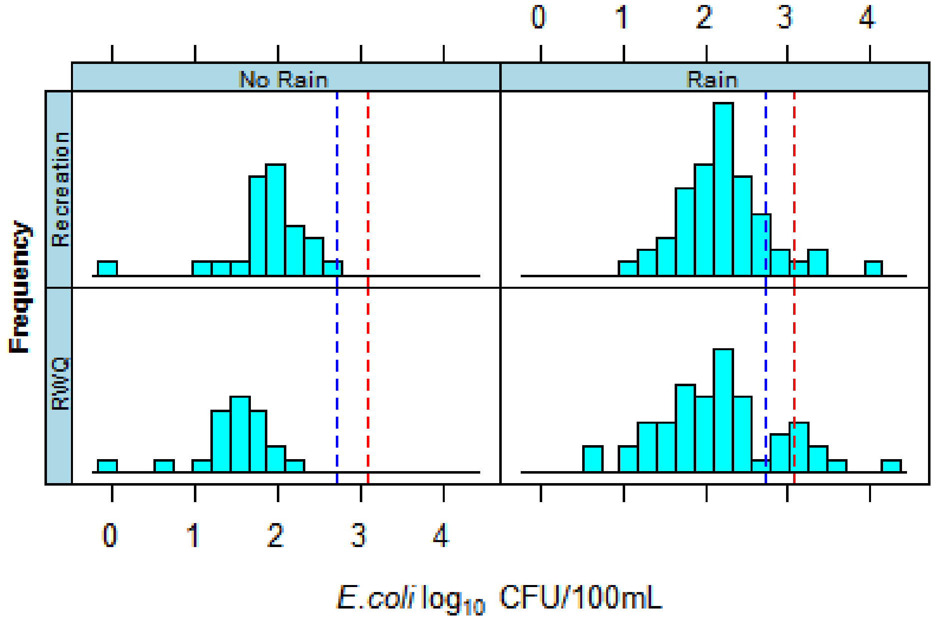


Figure 5: Distribution of *E. coli* concentrations for samples collected at "Waiau at Tuatapere" when rain greater than 1mm was recorded at Clifden in the 72 hours prior to sampling. Data collected between January 2014 and March 2020. Blue line is 540 *E.coli*/100mL and the red line is 1200 *E.coli*/100mL.



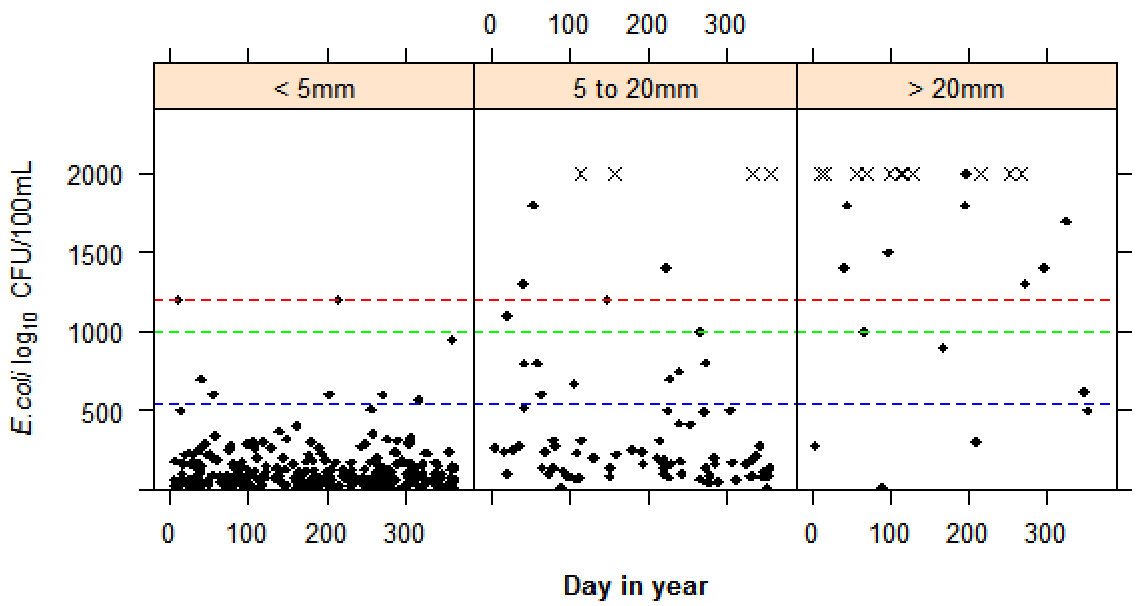
The above plots suggest:

- Rainfall may be a factor in determining the Attribute Band at this site and the relationship should be explored further.
- The data from the last five years is similar to the whole data set (1998 to 2020)
- The RWQ *E. coli* concentration distribution is similar to the summer recreation site distribution when rain has been observed in the 72 hours before sampling.
- The summer recreation site has slightly more samples with greater than 100 *E. coli*/100mL, than the RWQ samples when rain has not been observed.

5.2 SEASON AND RAIN

There is no strong seasonal alignment with the *E. coli* concentrations that exceed the numeric attribute state limits, when considered in isolation or when also considering rainfall at Clifden. Figure 6 gives a plot of one example of splitting the data by rainfall, other options were explored using the *ByDay_plot()* function, but no season patterns were detected.

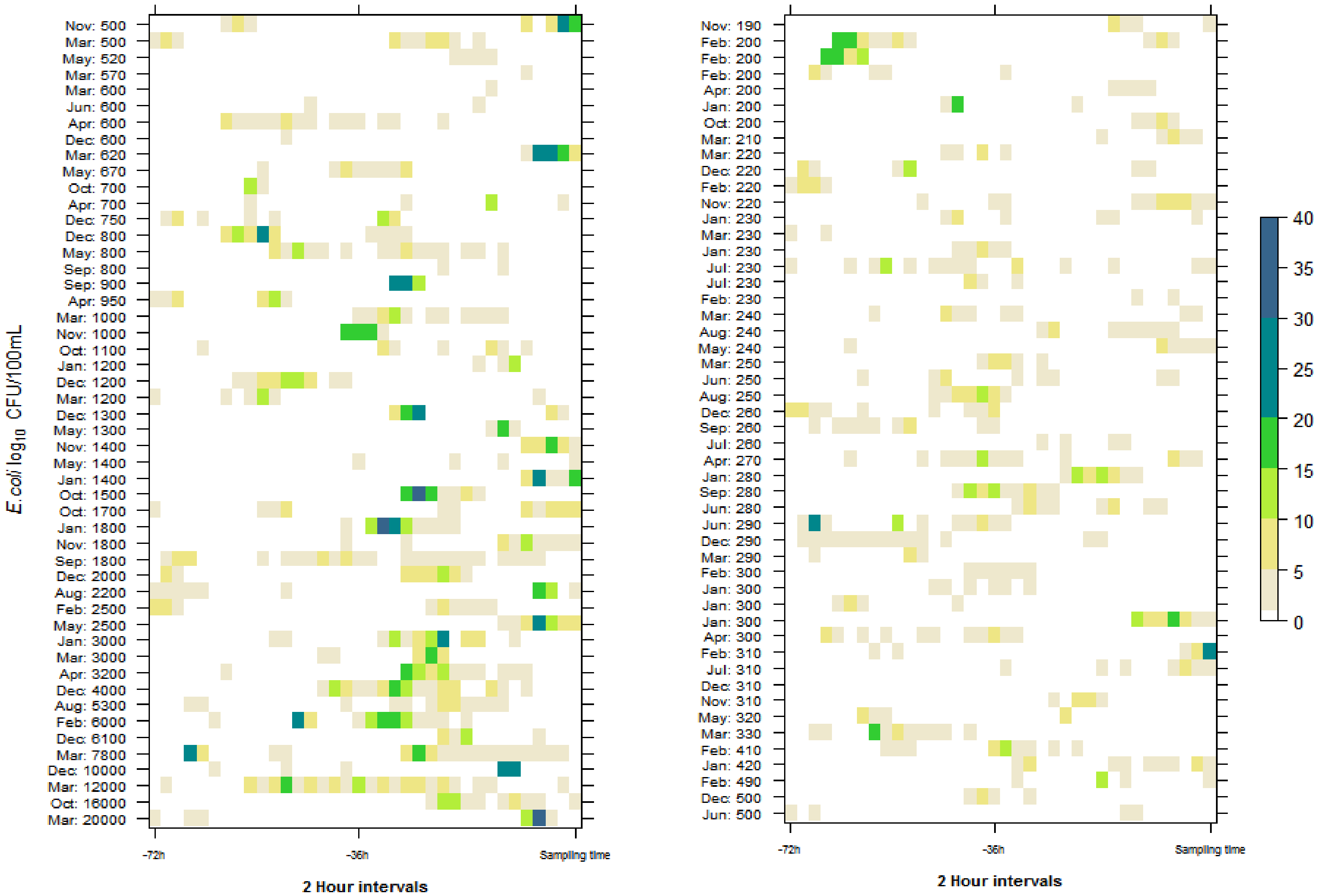
Figure 6: *E. coli* concentrations for samples collected at "Waiau at Tuatapere", plotted by sampling day of the year, and grouped by the total rainfall at Clifden in the 36 hours prior to water sampling. Coloured lines are at 540 (blue), 1000 (green) and 1200 (red) *E. coli*/100mL.



5.3 HOW DOES RAIN QUANTITY AND TIME PERIOD EFFECT *E. COLI* CONCENTRATIONS?

Figure 7 shows the two hourly rainfall recorded at Clifden in the hours preceding the water sampling, for the samples with the 100 highest *E. coli* concentrations. The plot suggests sustained and/or heavy rain in the 36 hours prior to sampling is aligned to the *E. coli* concentrations which control the Attribute band for the water sampling site. However, there are a number of samples with concentrations in the range 500 to 1400 *E. coli*/100mL that do not align to the rainfall at Clifden. *Ecoli_rain_map_single()*

Figure 7: Rainfall (mm) recorded at the Clifden measuring site during the 72 hours prior to water sampling at "Waiau at Tuatapere". Month of sampling provided on vertical axis. Plot gives the data for the 100 highest concentrations of *E. coli* from the combined RWQ and Recreational monitoring datasets.



6. APARIMA AT THORNBURY

Plots in the following sections were generated using the *ES_RWQ_rainfall_script.R* and *ES_RWQ_rainfall_functions.R* files. Unless otherwise stated the data is from the period 1998 to March 2020.

RWQ and Recreational sampling site: Aparima at Thornbury

Rainfall sites: Hamilton Burn at Mount Hamilton Road

Ohai at Wether Hill Station

Wairio at Otautau Nightcorps Rd

Scotts Gap

Central Plains Aquafer at Heddon Bush

Lower Aparima Aquifer at Riverton

Hamilton Bush and Ohai are in the head waters of the catchment and have similar recordings for rainfall in the data provided. Similarly, the central weather stations; Wairio, Scotts Gap and Central Plains are also geographically grouped and have similar rainfall patterns. The Lower Aparima rainfall site is the closest and downstream from the Aparima at Thornbury water sampling site.

6.1 *E.COLI* SAMPLING WITH NO RAIN RECORDED PREVIOUS 72 HOURS

This section considers the baseline *E. coli* concentrations in the river when there has been no recent rain at the Hamilton Bush rain gauge (Figure 8) and at Lower Aparima aquifer at Riverton (Figure 9) Histograms of *E. coli* concentration from samples where there has been less than 1mm of rain ("No Rain") and 1mm or more ("Rain"), for the RWQ and Recreational water sampling are plotted. *RWQrain_QC_dataframe() Ecoli_with_no_rain()*

The plots suggest:

- Rainfall may be a factor in determining the Attribute Band at this site and the relationship should be explored further.
- The RWQ *E. coli* concentration distribution is similar to the summer recreation site distribution when rain has been observed in the 72 hours before sampling.

Similar plots for the other sites can be generated with the *Ecoli_with_no_rain()* function, and were similar to those observed in the figures below.

Figure 8: Distribution of *E. coli* concentrations for samples collected at "Aparima at Thornbury" when rain greater than 1mm was recorded at Hamilton Burn in the 72 hours prior to sampling. Data collected between Jan 2014 and March 2020. Blue line is 540 *E.coli*/100mL and the red line is 1200 *E.coli*/100mL.

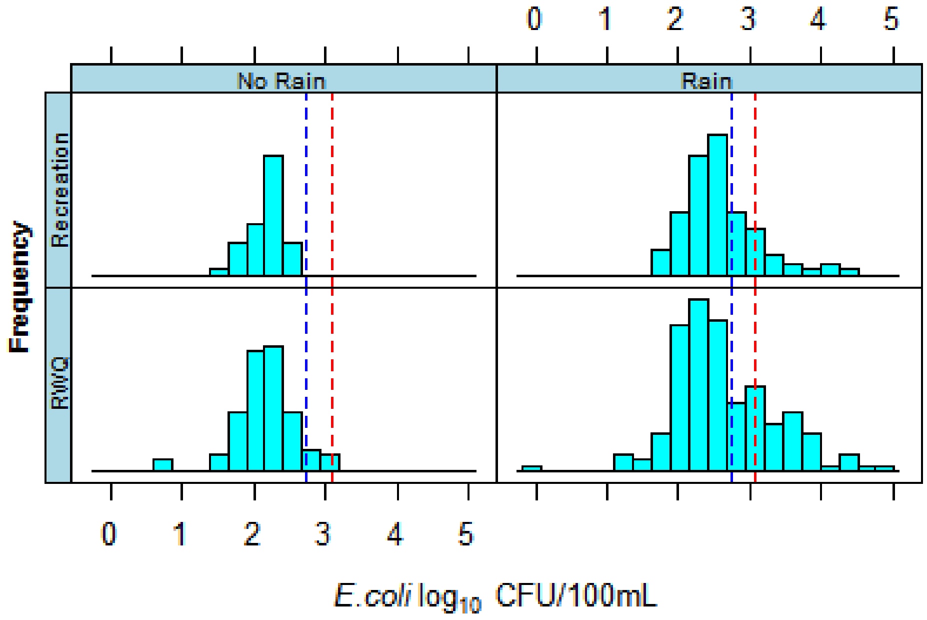
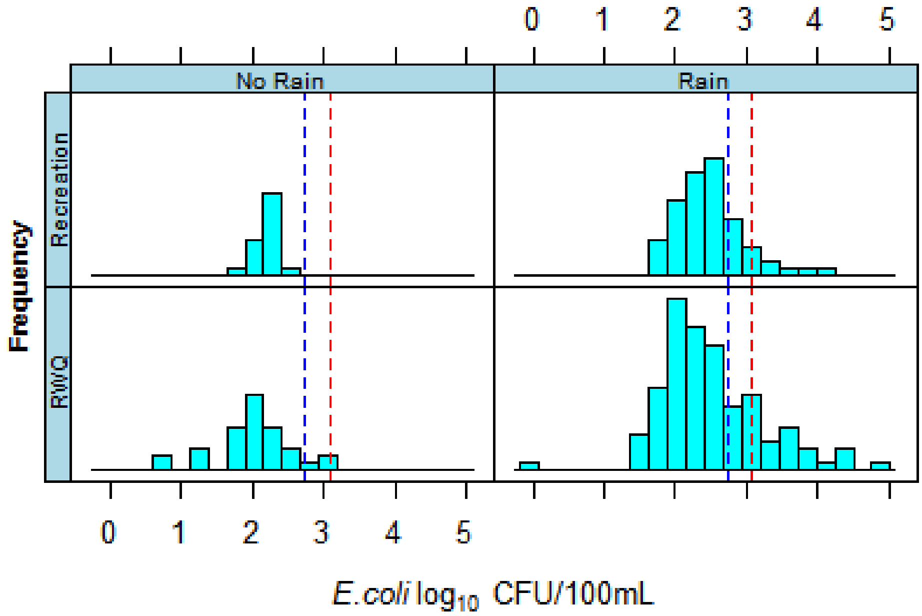


Figure 9: Distribution of *E. coli* concentrations for samples collected at "Aparima at Thornbury" when rain greater than 1mm was recorded at Lower Aparima Aquifer in the 72 hours prior to sampling. Data collected between Jan 2014 and March 2020. Blue line is 540 *E.coli*/100mL and the red line is 1200 *E.coli*/100mL.



6.2 SEASON AND RAIN

There is no strong seasonal alignment with the *E. coli* concentrations that exceed the numeric attribute state limits, when considered in isolation or when also considering rainfall at the six rain gauge sites individually. Figure 10 gives a plot of one example of splitting the data by rainfall, other options were explored using the *ByDay_plot()* function, but no season patterns were detected.

The Wairio rainfall grouping did appear to have some predictive capability for *E.coli* concentration. However, given the inherent variability of the data, this may need further investigation to confirm.

While the plot is useful to screen for seasonal trends it may be easily misinterpreted for multiple rainfall sites by someone not familiar with lattice type plots. It would be easy to think the plotted data is the water sample data at these sites, rather than the stratification of the rainfall data. Each row of plots represents the complete set of *E.coli* samples which had rainfall data available from the site given in the green boxes.

Figure 10: *E. coli* concentrations for samples collected at "Aparima at Thornbury", plotted by sampling day of the year, and grouped by the total rainfall at different rainfall monitoring sites in the 48 hours prior to water sampling. Coloured lines are at 540 (blue), 1000 (green) and 1200 (red) *E. coli*/100mL.

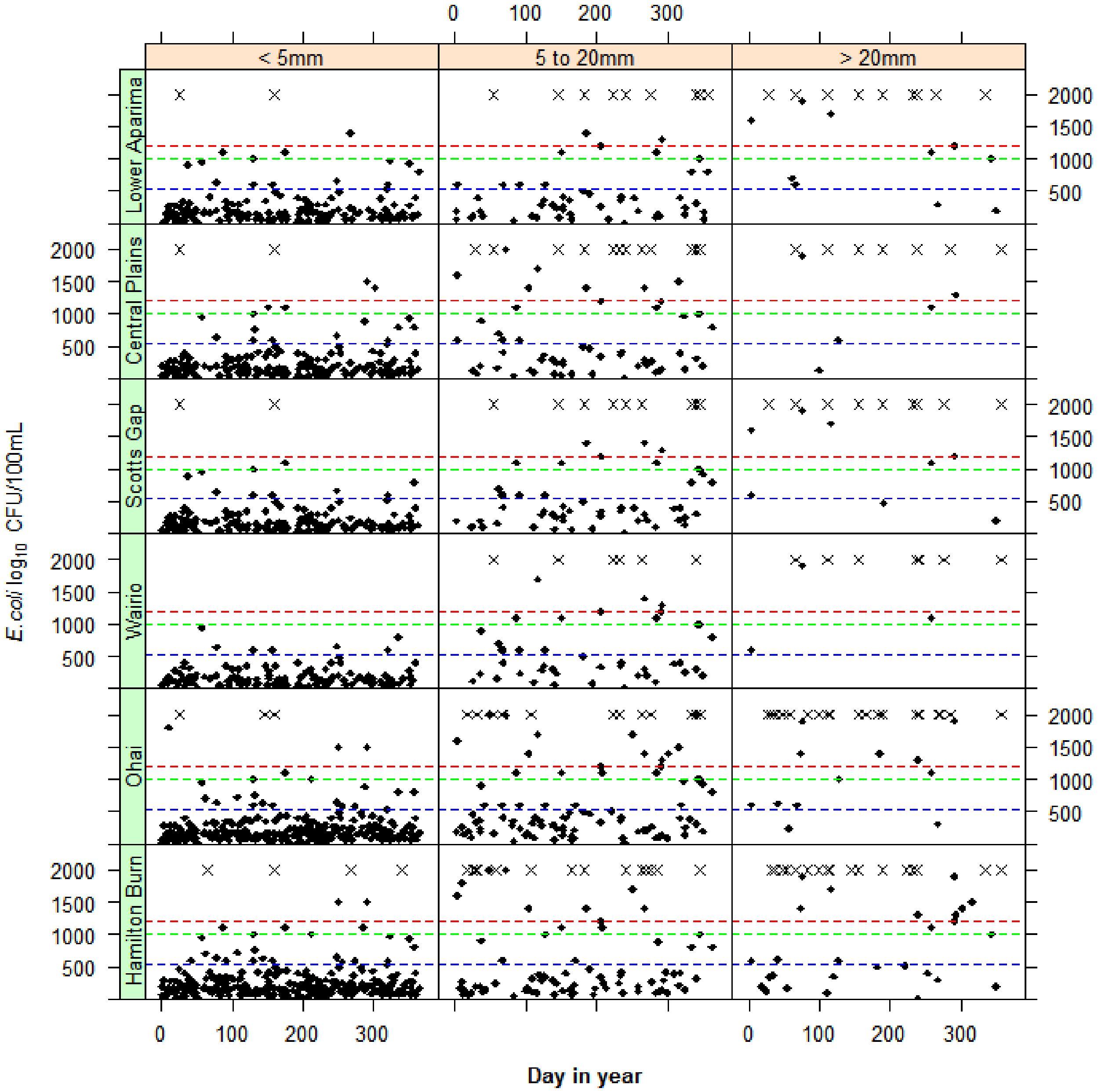
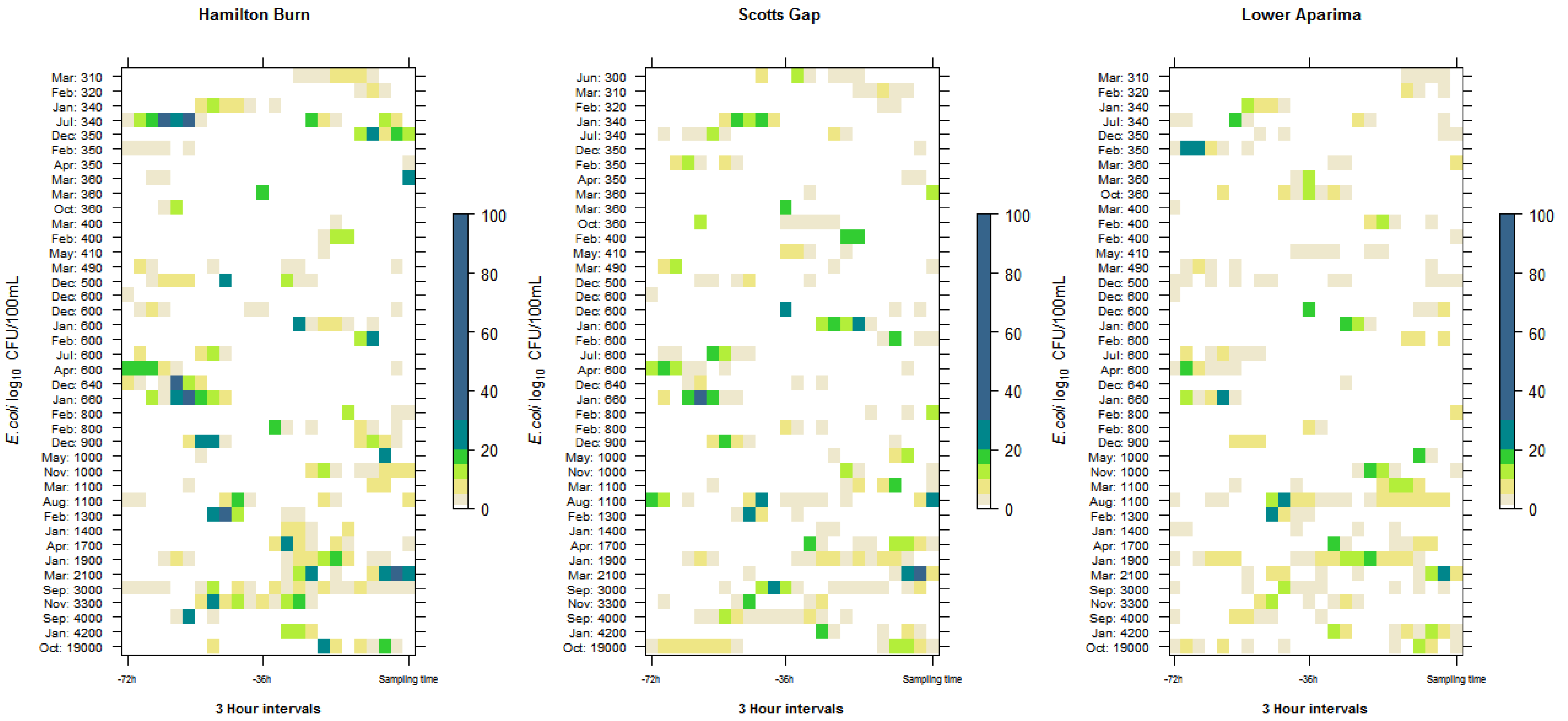


Figure 11: Rainfall (mm) recorded at three sites during the 72 hours prior to water sampling at "Aparima at Thornbury". Month of sampling provided on vertical axis. Plot gives the data for the 40 highest concentrations of *E. coli* from the combined RWQ and Recreational monitoring datasets, which also had rainfall data available.



6.3 HOW DOES RAIN QUANTITY AND TIME PERIOD EFFECT *E. COLI* CONCENTRATIONS?

Figure 11 shows the 3 hourly rainfall recorded at Hamilton Burn, Scotts Gap and Lower Aparima aquifer in the 72 hours preceding the water sampling for the samples.

The data with the 40 highest *E. coli* concentrations is presented. More of the data can be explored using the *Ecoli_rain_map_multi()* function that was used to generate this figure.

This data suggests that rain in the catchment for sustained periods in the 40 hours before sampling can be associated with elevated *E. coli* concentrations in the river water at Thornbury, which in turn will affect the Attribute Band for the site.

7. MATAURA AT GORE

Plots in the following sections were generated using the *ES_RWQ_rainfall_script.R* and *ES_RWQ_rainfall_functions.R* files. Unless otherwise stated the data is from the period 1998 to March 2020.

RWQ and Recreational sampling site: Mataura River at Gore

Rainfall sites:

- Upper Waikaia at Hyde Rock
- Waikaia River at Piano Flat
- Glenlapa at Round Hill
- Balfour at Glenure St Patricks Road
- Wendon Valley at Waikaka
- Riversdale Aquifer at Liverpool Street
- Waimea Stream at Mandeville
- Chatton Aquifer at Cunningham Road

Waimea and Riversdale are located upstream form Gore and have similar rainfall patterns. Similarly the rainfall sites north of Gore; Wendon Valley and Chatton Aquifer also record similar rainfall patterns. The Waimea and Chatton Aquifer rainfall sites are the closest to the Mataura at Gore water sampling site.

7.1 *E.COLI* SAMPLING WITH NO RAIN RECORDED PREVIOUS 72 HOURS

This section considers the baseline *E. coli* concentrations in the river when there has been no recent rain at the Upper Waikaia and Waimea rainfall monitoring sites, using data recorded between January 2014 and March 2020. Histograms of *E. coli* concentration from samples where there has been less than 1mm of rain ("No Rain") and 1mm or more ("Rain"), for the RWQ and Recreational water sampling are plotted (

Figure 12 and Figure 13). *RWQrain_QC_dataframe() Ecoli_with_no_rain()*

The plots suggest:

- Rainfall is likely to be contributing to attribute band allocation, especially the increase in *E.coli* concentrations much greater than 1200 CFU/100mL mainly occurring in the "Rain group". However, concentrations above 540 CFU/100mL are observed in both the "Rain" and "No Rain" groups. Suggesting rain in the 72 hours preceding sampling is not the only factor associated with elevated *E.coli* concentrations in the river.
- The RWQ *E. coli* concentration distribution is similar to the summer recreation site distribution when rain has been observed in the 72 hours before sampling.

Figure 12: Distribution of *E. coli* concentrations for samples collected at "Mataura at Gore" when rain greater than 1mm was recorded at Upper Waikaia monitoring site in the 72 hours prior to sampling. Data collected between Jan 2014 and March 2020. Blue line is 540 *E.coli*/100mL and the red line is 1200 *E.coli*/100mL.

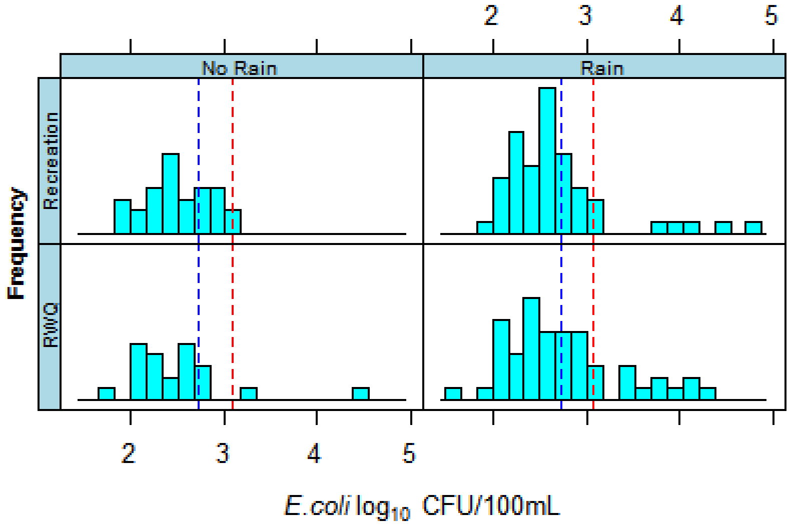
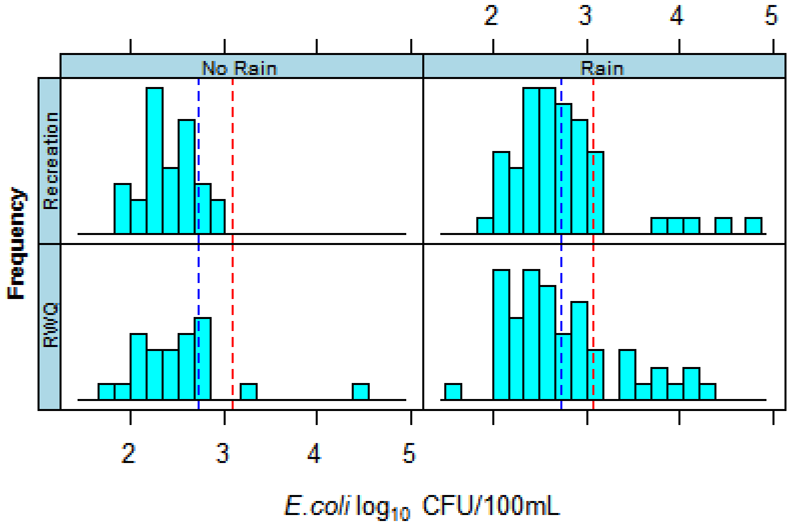


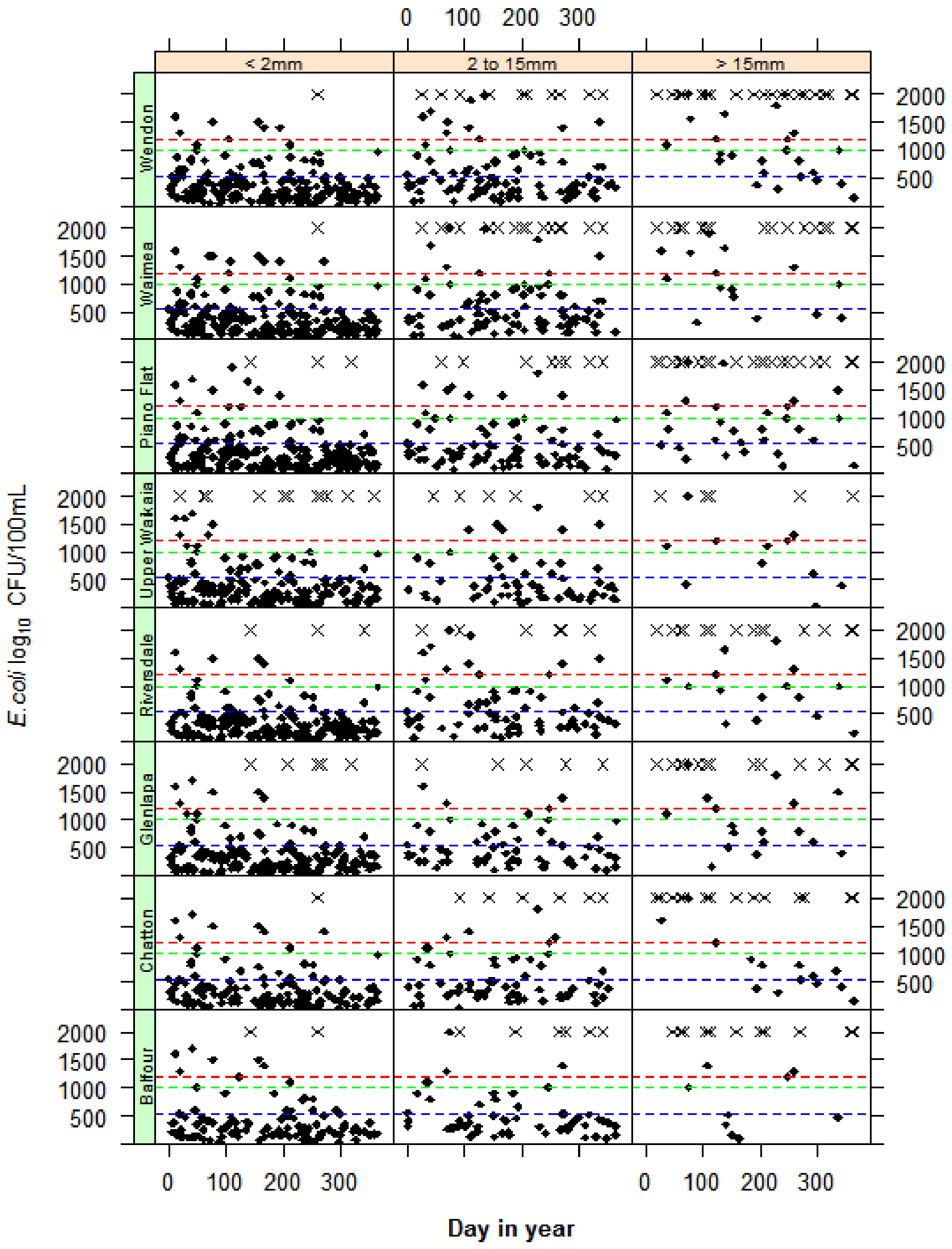
Figure 13: Distribution of *E. coli* concentrations for samples collected at "Mataura at Gore" when rain greater than 1mm at the Waimea rainfall monitoring site in the 72 hours prior to sampling. Data collected between Jan 2014 and March 2020. Blue line is 540 *E.coli*/100mL and the red line is 1200 *E.coli*/100mL.



7.2 SEASON AND RAIN

There is no strong seasonal alignment with the *E. coli* concentrations that exceed the numeric attribute state limits, when considered in isolation or when also considering rainfall at the eight rain gauge sites individually. Figure 14 gives a plot of one example of splitting the data by rainfall, other options were explored using the *ByDay_plot()* function. There is a possible seasonal difference in the < 2mm rainfall group, with less *E. coli* concentrations over 540 CFU/100mL observed in the last quarter of the year.

Figure 14: *E. coli* **concentrations for samples collected at "Mataura at Gore", plotted by sampling day of the year, and grouped by the total rainfall at different rainfall monitoring sites in the 48 hours prior to water sampling. Coloured lines are at 540 (blue), 1000 (green) and 1200 (red)** ***E. coli*****/100mL.**



7.3 HOW DOES RAIN QUANTITY AND TIME PERIOD EFFECT *E. COLI* CONCENTRATIONS?

An alternative way of presenting the rainfall patterns to those shown in Figure 11, is presented in Figure 15. Initially the method used for Figure 11 was tried. It was difficult to compare the rainfall patterns at different sites with increasing concentrations of *E. coli*, due to missing data at some of the rainfall sites for the some of the river water samples. So the rows across the plots did not align with the same *E.coli* concentration.

Instead, Figure 15 shows individual plots specific to a water sample. Each plot shows the rainfall, at the different rainfall sites at different intervals before the sampling time. This type of plot may be easier for people to interpret than the type in Figure 11, when multiple rain sites are being considered.

Each plot was created using the *Ecoli_rain_map_SES()* function. The function can be called for different; water samples, time duration from sampling, and time intervals for rainfall measuring. To change the colour scale intervals or the colours, the *Ecoli_rain_map_SES()* function will need to updated within the *levelplot* function (*at* and *col.regions* arguments).

Figure 15 shows there are multiple mechanisms contributing to the elevated *E.coli* levels in the Mataura River at Gore. While there are some plots showing rainfall occurring prior to sampling, a number of samples relating to high *E. coli* concentrations show no or little rain at the rainfall monitoring sites in the 72 hours prior to sampling. For example, the third and fourth highest recorded *E. coli* concentrations (28,000 and 26,000 CFU/100mL) related to very limited rainfall in the catchment.

Further examination of these plots for lower concentrations of *E.coli* may help establish how much rain and in what periods prior to sampling, the rain can impact the attribute bands. This in turn, may help to inform which areas of the catchment are contributing to the rain associated elevated *E.coli* concentrations.

Figure 15 (Part1): Rainfall (mm) patterns associated with high *E.coli* concentrations measured from water sampling at "Mataura River at Gore". Each plot is the data associated with a single water sample, with the measured *E. coli* concentration and month of sampling provided in the plot header.

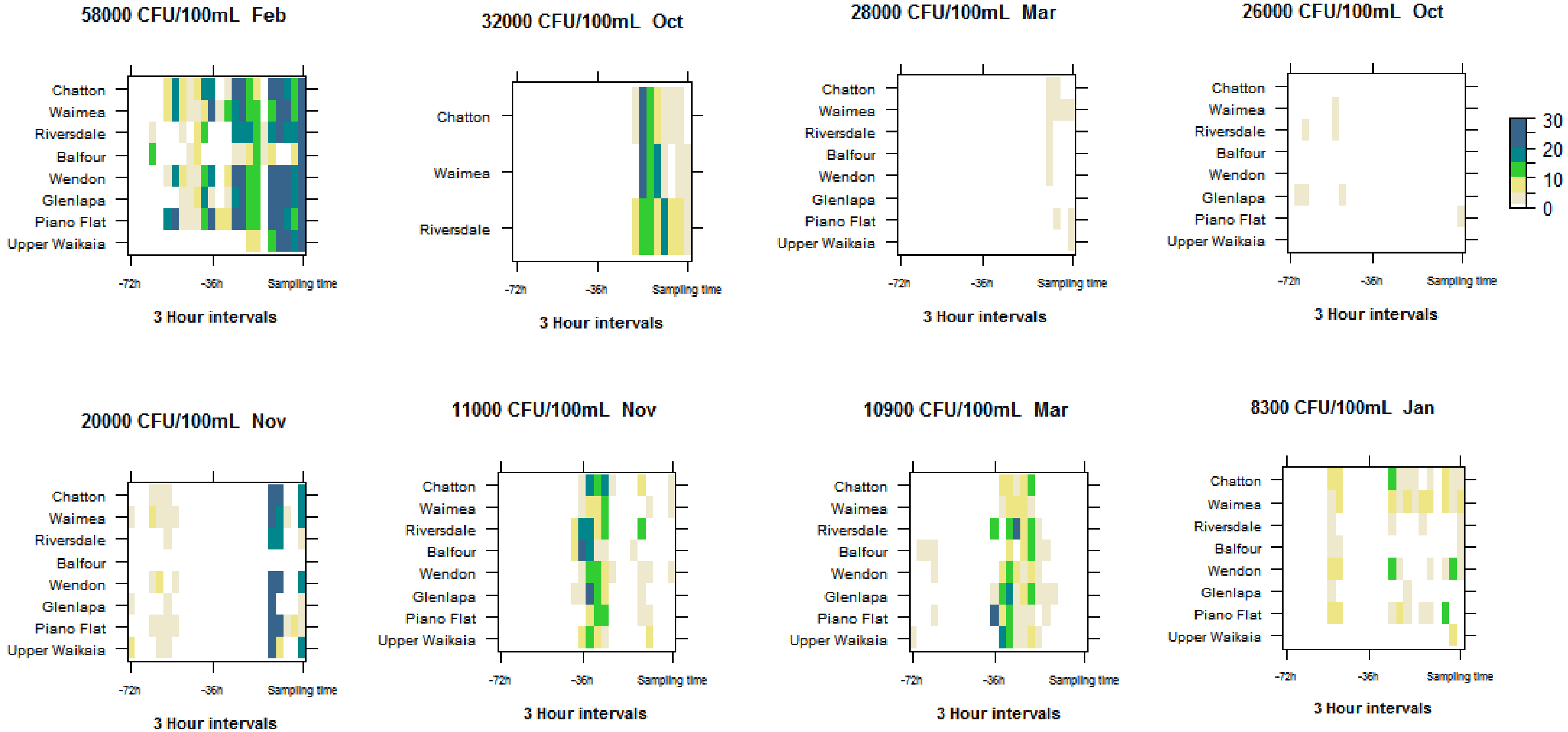
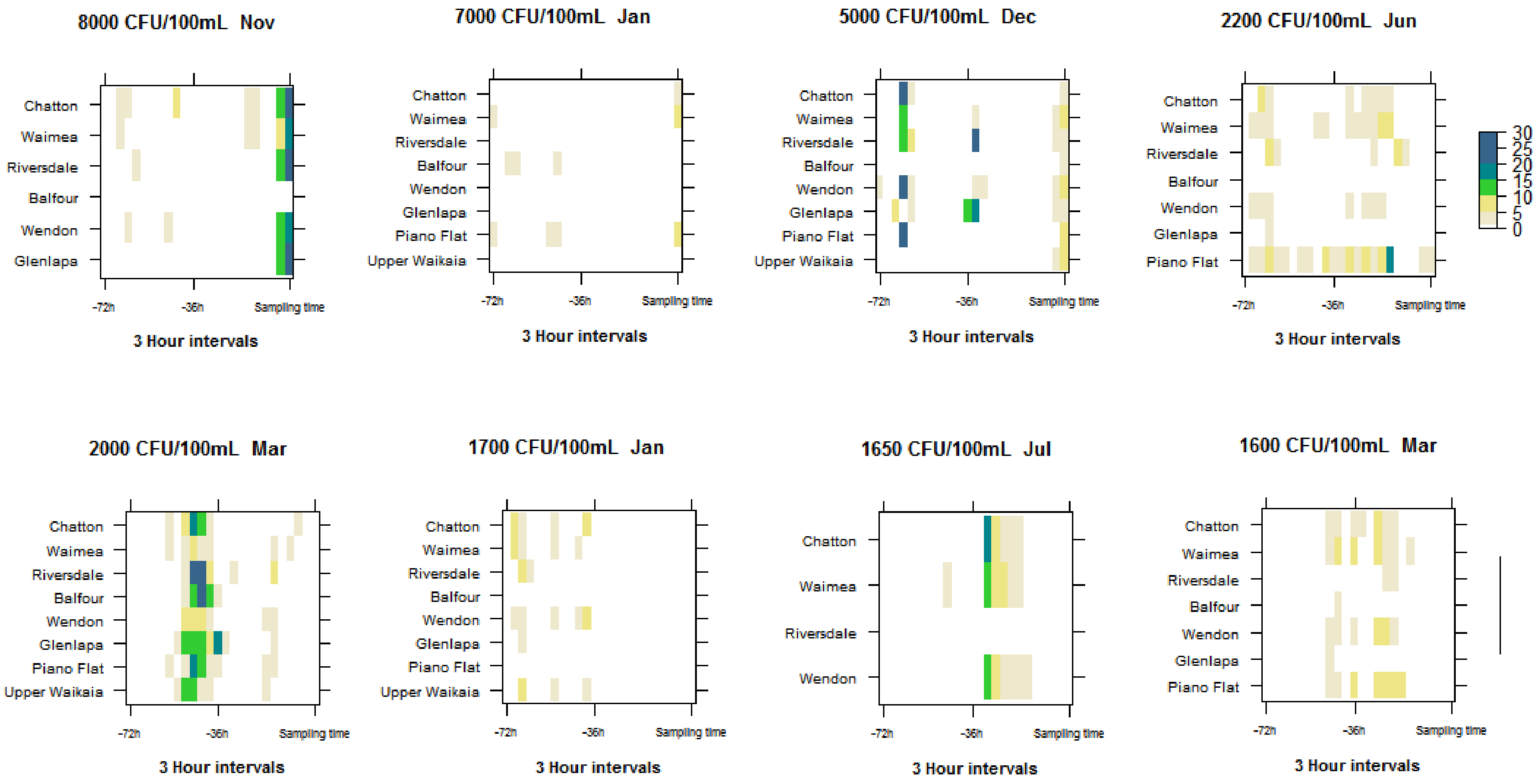
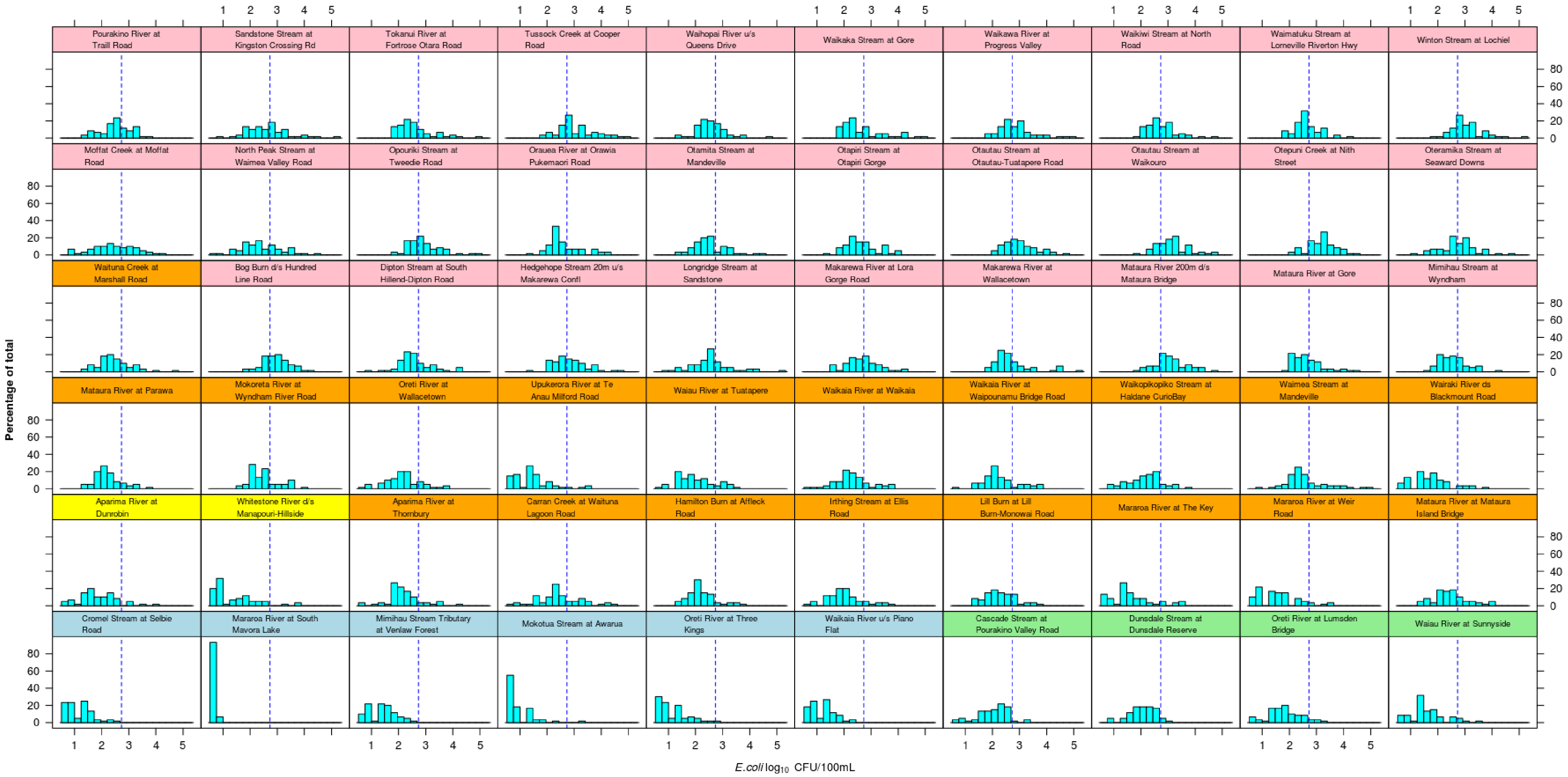


Figure 15 (part 2): Rainfall (mm) patterns associated with the high recorded river water *E. coli* samples from water sampling at "Mataura River at Gore". Each plot is the data associated with a single water sample, with the measured *E. coli* concentration and month of sampling provided in the plot header.



APPENDIX A: DISTRIBUTION OF *E. COLI* CONCENTRATION BY NOF SITE



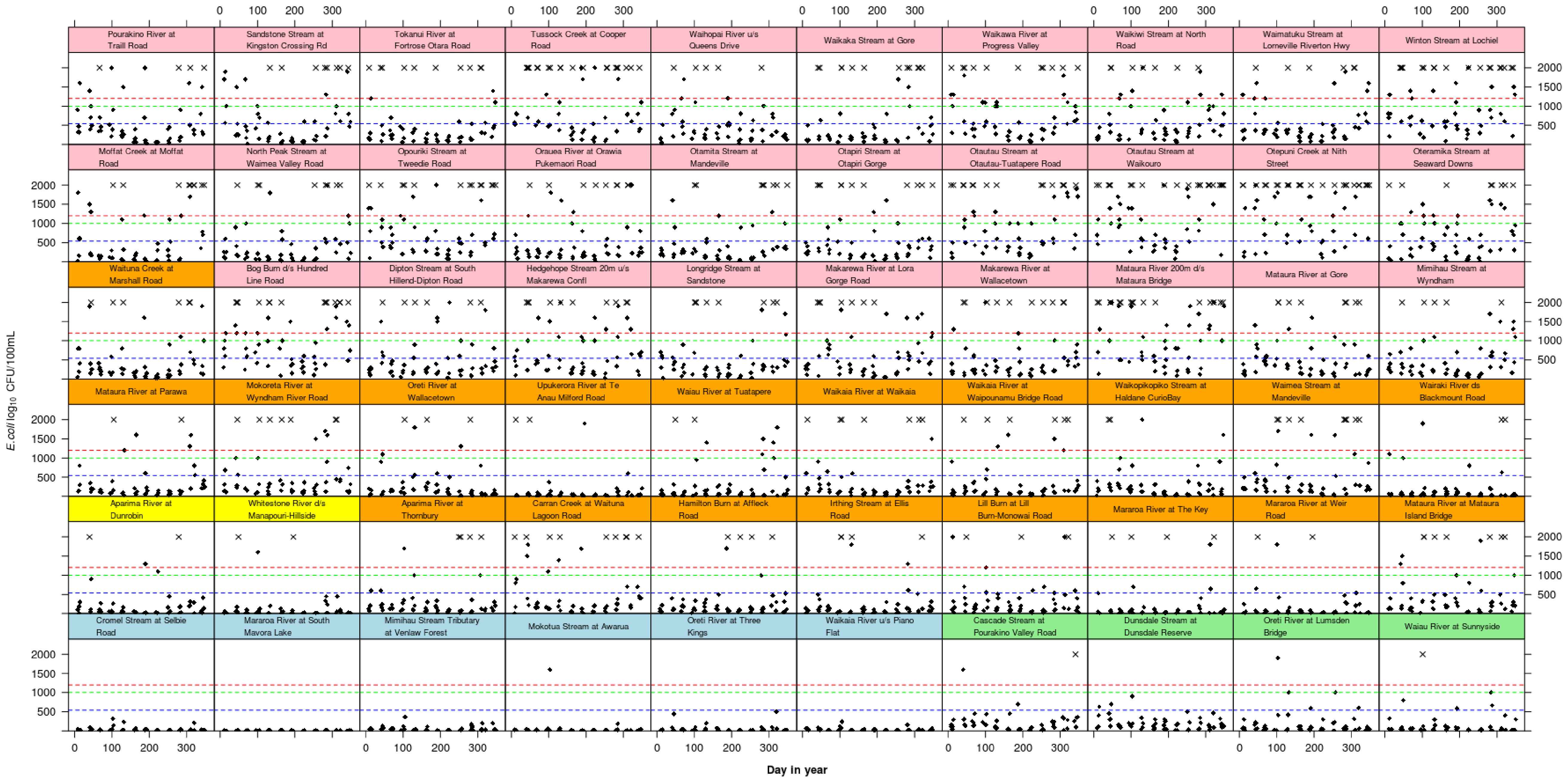
Notes: Distributions all based on the 60 samples previous to and including July 2019, any samples with values less than 5 CFU/mL have been set to 5 CFU/mL for plotting purposes, given most data has a LOD of 10.

The graph header colours relate to the Consultation NPS for Freshwater Management Report, Table 11: *E. coli* attribute bands.

The blue dashed line is a concentration of 540 CFU/100mL

Plot created with R function *compare_dist_plot()*

APPENDIX B: *E. COLI* CONCENTRATION BY DAY OF YEAR AND NOF SITE



Notes: Distributions all based on the 60 samples previous to and including July 2019, any samples with values more than 2000 CFU/mL have been set to 2000 CFU/mL for plotting purposes and are indicated by an 'x'

The graph header colours relate to the Consultation NPS for Freshwater Management Report, Table 11: *E. coli* attribute bands.

The blue dashed line is a concentration of 540 CFU/100mL, green dashed line is 1000 CFU/100mL and red dashed line is 1200 CFU/100mL

Plot created with R function *ByMonth_plot()*

REFERENCES

McBride GB (2005) Using statistical methods for water quality management: issues, problems and solutions. Statistics in Practice Series. John Wiley & Sons. ISBN0-471-47016-3

Hunter PR (2002) Does the calculation of the 95th percentile of microbiological results offer an advantage over percentage exceedance in determining compliance with bathing water quality standards? *Letters in Applied Microbiology* 34: 283-286.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Science for Communities

INSTITUTE OF ENVIRONMENTAL
SCIENCE AND RESEARCH LIMITED

Kenepuru Science Centre
34 Kenepuru Drive, Kenepuru, Porirua 5022
PO Box 50348, Porirua 5240
New Zealand
T: +64 4 914 0700 **F:** +64 4 914 0770

Mt Albert Science Centre
120 Mt Albert Road, Sandringham, Auckland 1025
Private Bag 92021, Auckland 1142
New Zealand
T: +64 9 815 3670 **F:** +64 9 849 6046

NCBID – Wallaceville
66 Ward Street, Wallaceville, Upper Hutt 5018
PO Box 40158, Upper Hutt 5140
New Zealand
T: +64 4 529 0600 **F:** +64 4 529 0601

Christchurch Science Centre
27 Creyke Road, Ilam, Christchurch 8041
PO Box 29181, Christchurch 8540
New Zealand
T: +64 3 351 6019 **F:** +64 3 351 0010

www.esr.cri.nz