A strategy to determine the source of intermittent faecal contamination of Tukurua Stream

December 2014

Client Report 14013
Envirolink Number 1548-TSDC108

Prepared for: Tasman District Council
Prepared by: Wendy Williamson
Project Leader: Elaine Moriarty
Reviewed by: Brent Gilpin

DISCLAIMER

This report or document ("the Report") is given by the Institute of Environmental Science and Research Limited ("ESR") for the benefit of Tasman District Council and other Third Party Beneficiaries.

Neither ESR nor any of its employees makes any warranty, express or implied, or assumes any legal liability or responsibility for use of the Report or its contents by any other person or organisation.

Table of Contents

Table of Figures

List of Tables

1 Introduction

An Envirolink project has been funded to assist Tasman DC developing a strategy to determine the source of intermittent faecal contamination of Tukurua Stream in Golden Bay (Figure 1A). Council has already spent a reasonable amount of time and money sampling this stream (including microbial source tracking (MST)) over the last 4 years. Some progress in improving faecal indicator bacteria concentrations has been made but there are still unexplainable peaks typically in late Jan-Feb.

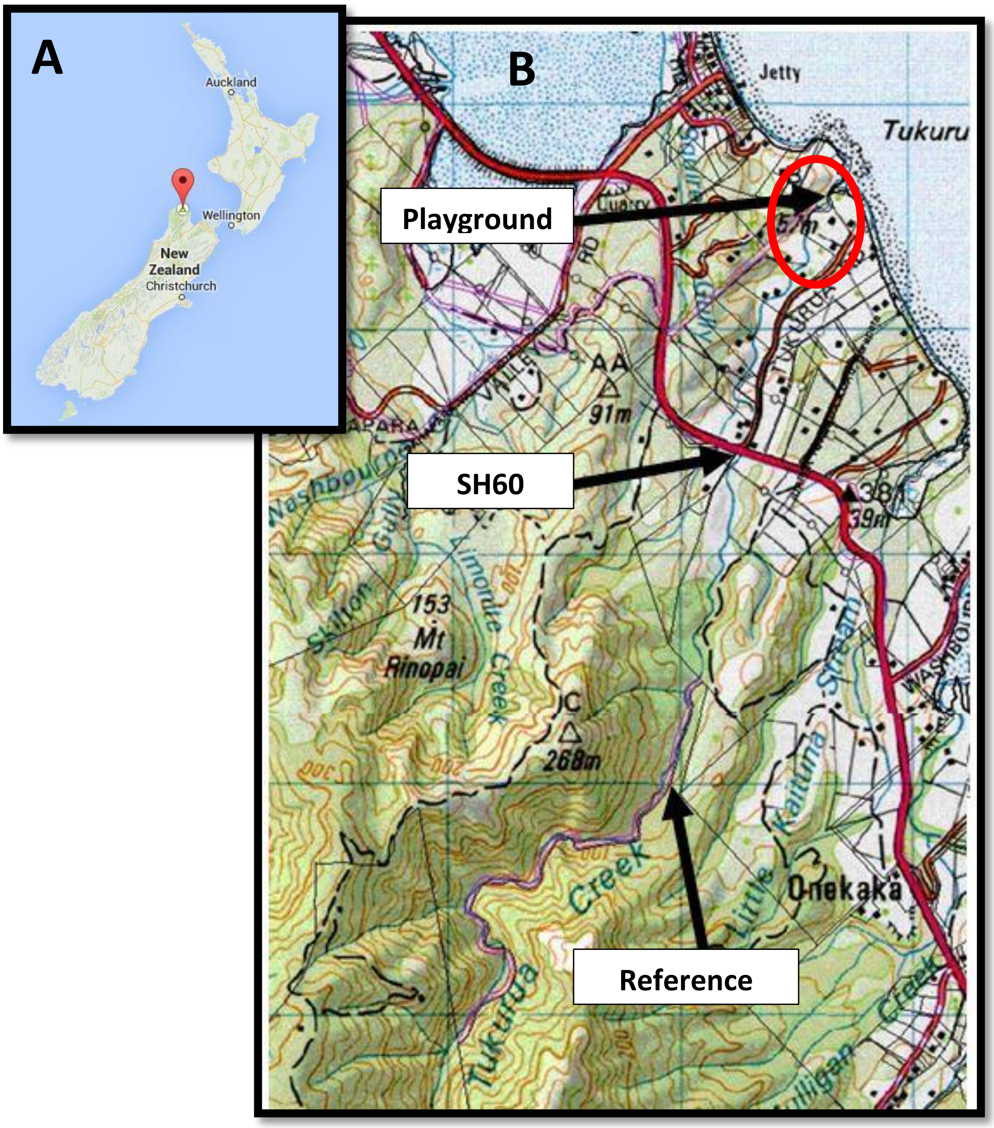


Figure 1 Map showing location of (A) the Golden Bay Kiwi holiday Park and (B) the sampling locations of the Playground and SH60 sampling locations and an upstream Reference location (provided by Tasman DC). The location of the Golden Bay Kiwi holiday Park is shown by the red outline.

Recent MST results show that the faecal contamination is neither from human or ruminant animal (the two major obvious sources of contamination in the catchment). The two primary sites were sampled were at SH60 and where the creek flowed through the playground of the camping ground (Figure 1B). A third site, as a reference (Figure 1B), was sampled occasionally but insufficiently to include in a comparison between the two primary sites.

Results from the data analyses are presented in two sections (Section 3 and Section 4). Section 3 uses the water quality data (E. coli data) for each site individually against rainfall for each date a sample was collected at that site. This allowed the influence of rainfall to be assessed for each sampling site independently. In comparison, Section 4 analyses used a restricted set of the water quality data: only for SH60 and Playground sites and only where *E. coli* data were collected from both sites on the same day. This allowed the influence of water quality at one site on the water quality at the other site to be assessed against rainfall.

2 Methods

2.1 Statistics

The statistical package used to analyse the data was Statistix 10 (Analytical Software 2013) with significant difference at $P<0.050$. Summary statistics of mean, median and percentile were calculated. Time series and scatter plots were used to visually assess the water quality between the two principle sites. Where data permitted, correlation analysis was used to assess the relationship between rainfall and water quality and between the water quality at hydrologically linked sites.

2.2 Data

While multiple sites were identified as being suitable for sampling along the Tukurua Creek from the reference site to past the Playground site in the camping ground, only two sites, at SH60 and the Playground, were comprehensively sampled (Figure 1B). Three variables were measured by Tasman DC:

- Water quality by *E. coli* concentration per 100 mL of water.
- Rainfall was measured (in mm) as an average per 48 hours for the day of and day prior to sampling.
- Air temperature was in °C, where Tasman DC used the temperature of 20°C to fill data spaces when actual data were not available.

These variables were not necessarily measured on the same dates. The site data were analysed for each site for mean, median, and 95^{th} percentile and for correlation with rainfall and between sites.

A new data set was created by matching the sampling dates for the SH60 and Playground sites. This new data set (date-matched data set) was used to test the difference in water quality between the two sites (two-sample t test) and the relationship between water quality at SH60 and at the Playground (correlation analysis).

3 Statistics from Reference, SH60 and the Playground Sampling Sites

3.1 Reference site data analysis

Six samples were taken at the reference site from 29 January 2010 to 9 February 2011. The *E. coli* were normally distributed (W>0.80), however for such a small data set rainfall was not normally distributed and data-transformations could not improve this (Table 1).

The median concentration of *E. coli* for the reference site over the 1 year sampling was 8 *E. coli* per 100 mL, however there were insufficient data to calculate the 95th percentile (Table 2). Due to insufficiency of data for the reference site, no correlation analysis for *E. coli* and rainfall was performed.

Table 1 Testing for data normality at the Reference site using the Shapiro-Wilk normality test (W). Rainfall data (mm averaged over 48 h on the day before and the day of sampling) were sine-transformed.

Variable	N	W
Reference water quality (*E. coli* per 100 mL)	6	0.9015
Rainfall (mm)	6	0.7013
Sine Rainfall (mm)	6	0.7446

Table 2 Summary statistics for Reference site data as provided by Tasman DC for *E. coli* data (per 100 mL) and its date-matched rainfall (mm averaged over 48 h on the day before and the day of sampling) and air temperature.

Variable	Mean	Median	Minimum	Maximum	95th Percentile
Reference Water quality (*E. coli* per 100 mL)	7	8	2	15	Insufficient data
Rainfall	2.25	0.00	0.00	9.00	ND

3.2 SH60 site data analysis

SH60 water quality and rainfall data were not normally distributed (Shapiro-Wilk test normality W≤0.80). Data transformation approximated a normal distribution sufficiently for parametric analyses: $\log_{10}$ and sine-function transformations for *E. coli* and rainfall, respectively (Table 3).

The median concentration of *E. coli* for SH60 over the 4 year sampling time (21 January 2010 to 18 February 2014, n=65) was 306 per 100 mL with the 95th percentile at 2001 *E. coli* per 100 mL and a data range from 2 to 10,001 *E. coli* per 100 mL (Table 4).

Table 3 SH60 site data normality and transformations using the Shapiro-Wilk normality test (W) for the all data provided for SH60 site for *E. coli* data (per 100 mL) and its date-matched rainfall (mm averaged over 48 h on the day before and the day of sampling).

Variable	N	W
SH60 (*E. coli* per 100 mL)	65	0.3434
Log_{10} SH60 (*E. coli* per 100 mL)	65	0.9230
Rainfall (mm)	65	0.3496
Sine Rainfall (mm)	65	0.8043

Table 4 Summary statistics for SH60 site data as provided by Tasman DC for *E. coli* data (per 100 mL) and its date-matched rainfall (mm averaged over 48 h on the day before and the day of sampling) and air temperature. ND = not determined.

Variable	Mean	Median	Minimum	Maximum	95^{th} Percentile
SH60 (*E. coli* per 100 mL)	600	306	2	10001	2001
Log_{10} SH60 (*E. coli* per 100 mL)	2.48	2.49	0.30	4.00	3.30
Rainfall (mm per 48 h)	11.94	0.00	0.00	264.00	ND
Air (°C)	20.1	20.0	12.0	25.0	ND

Scatter plots of the transformed *E. coli* and rainfall and correlation analysis (Table 5) showed no linear relationship (P=0.469). A strong cluster of *E. coli* concentrations from very low to very high were recorded at zero rainfall, showing that rainfall did not sufficiently explain the elevated levels of *E. coli* at SH60 from 2010 to 2014 (Figure 2).

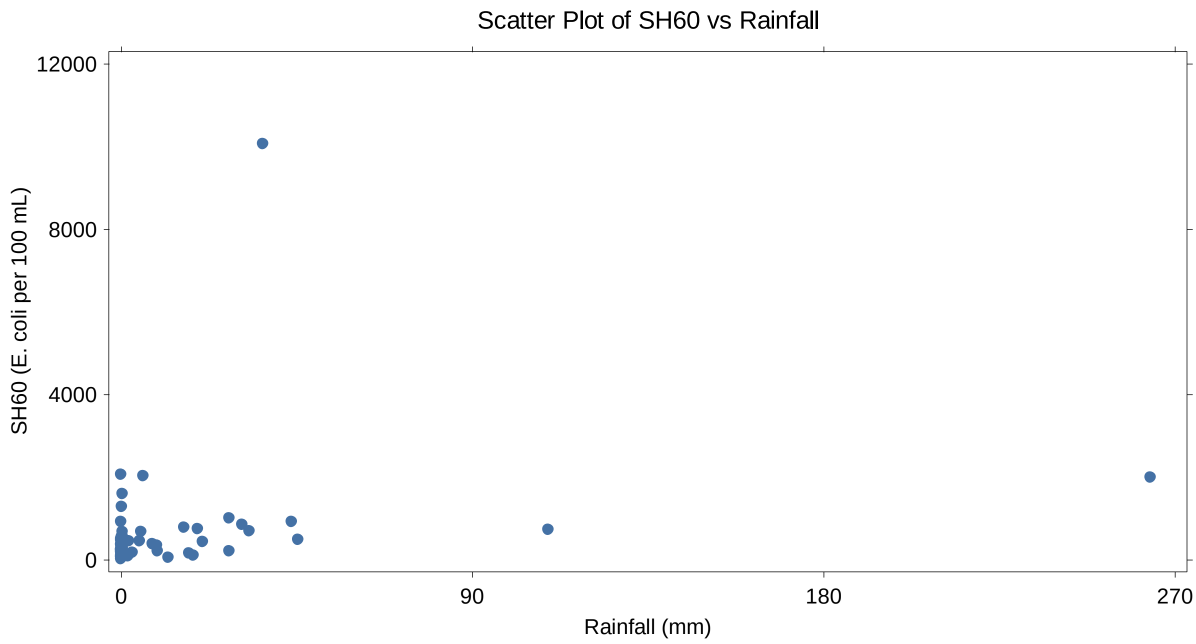


Figure 2 SH60 site vs rainfall scatter plot to visualise the relationship between water quality *(E. coli* data per 100 mL) and rainfall (mm).

Table 5 SH60 site Pearson correlation between water quality *(E. coli* data per 100 mL) and rainfall (mm averaged over 48 h on the day before and the day of sampling), where transformed data were used for the analysis. Transformations applied to improve data normality (aim for W>0.80) were: for *E. coli* data $\log_{10}$ function and rainfall data the sine function transformation.

	Log_{10} SH60	Sine Rainfall
Log_{10} SH60 (*E. coli* per 100 mL)	1.0000	
p-value	0.0000	
Sine Rainfall (mm)	-0.0913	1.0000
	0.4694	0.0000
Cases Included 65 Missing Cases 0		

Water quality data (and in particular 95th percentiles) are most impacted by the very high values. Table 6 categorises the data first according to whether ≤540 *E. coli*/100ml and then whether ≤1000 *E. coli*/100ml. As shown, 74% (48/65) of the samples had less than 540 *E. coli*/100ml, and 91% less than 1000 *E. coli*/100ml. Targeting the source of the samples with very high levels of *E. coli* would have the greatest impact on water quality (and managing an associated health risk), and should be a priority.

Table 6 SH60 site data binned according to whether ≤540, >540, ≤1000, or >1000, for a total of 65 data points.

	≤540	>540	≤1000	>1000
Median	233	885	254	2001
Mean	245	1604	341	3151
min	2	560	2	1300
max	531	10001	1000	10001
count	48	17	59	6

Heavy rainfall (>15mm) is moderately predictive of *E. coli* levels >540 *E. coli*/100 ml (Table 7). Consistent with the previous rainfall analysis, a significant number of samples showed elevated *E. coli* in the absence of rainfall events. Targeting sampling during and immediately after rainfall events would increase the probability of acquiring samples with >540 *E. coli*, which may allow better resolution of the sources of contamination contributing to elevated *E. coli* numbers at SH60. However, given the number of samples with very high levels of *E. coli* where no rainfall event occurred suggests more than one mechanism of contamination.

Table 7 48 hour rainfall at SH60 site.

	Binned at 540 *E. coli*	**48 Hour Rainfall**		**Binned at 1000 *E. coli***	**48 Hour Rainfall**	
		>5mm	>15mm		>5mm	>15mm
***E. coli*/100ml**	≤540	8	4	≤1000	17	12
***E. coli*/100ml**	>540	12	10	>1000	3	2
Positive Predictive Value		60%	71%		15%	14%
Likelihood Ratio		4.2	7.1		1.7	1.6
% of samples	>540	71%	59%	>1000	50%	33%

These interpretations are supported by examination of the data from SH60 sites when >1000 *E. coli*/100mL were recorded (Table 8). Although limited data are available, it appears that when a rainfall event occurred, both sites had elevated *E. coli*, while in the absence of rainfall, only SH60 site showed significantly elevated levels of *E. coli*.

Table 8 SH60 and Playground sites' data where water quality showed levels greater than 1000 *E. coli* per 100 mL.

Date	SH60 (*E. coli*)	Rainfall (mm)	Playground (*E. coli*)
12/02/2013	1300	0	384
10/01/2012	1601	0	504
26/11/2013	2000	5.9	2001
4/02/2014	2001	0	42
22/02/2012	2001	264	1100
24/03/2010	10001	36.5	

3.3 Playground site data analysis

Playground water quality and rainfall data were not normally distributed (Shapiro-Wilk test normality W≤0.80). Data transformation approximated a normal distribution sufficiently for parametric analyses: $\log_{10}$ and sine-function transformations for *E. coli* and rainfall, respectively (Table 9).

The median concentration of *E. coli* for the Playground over the 4 year sampling time (15 January 2010 to 18 February 2014, n=68) was 230 *E. coli* per 100 mL with the 95th percentile at 2001 *E. coli* per 100 mL and a data range from 10 to 2,001 *E. coli* per 100 mL (Table 10).

Table 9 Playground site data normality testing using the Shapiro-Wilk normality test (W) and transformations for the all data provided for the Playground site for *E. coli* (per 100 mL) and its date-matched rainfall (mm averaged over 48 h on the day before and the day of sampling).

Variable	N	W
Playground (*E. coli* per 100 mL)	68	0.7648
Log_{10} Playground (*E. coli* per 100 mL)	68	0.9678
Rainfall (mm per 48 h)	68	0.3823
Sine Rainfall (mm)	68	0.7954
Air (°C)	68	0.8715

Table 10 Summary statistics for Playground site data as provided by Tasman DC for *E. coli* data (per 100 mL) and its date-matched rainfall (mm averaged over 48 h on the day before and the day of sampling) and air temperature. ND = not determined.

Variable	Mean	Median	Minimum	Maximum	95th Percentile
Playground (*E. coli* per 100 mL)	511	230	10	2001	2001
Log_{10} Playground (*E. coli* per 100 mL)	2.42	2.36	1.00	3.30	3.30
Rainfall (mm per 48 h)	12.78	0.00	0.00	264.00	ND
Air (°C)	20.1	20.0	12.0	25.0	ND

Scatter plots of the *E. coli* and rainfall (Figure 3) and correlation analysis (Table 11) showed no linear relationship (P=0.512). A strong cluster of *E. coli* concentrations from very low to very high were recorded at zero rainfall, showing that rainfall did not sufficiently explain the elevated levels of *E. coli* at Playground from 2010 to 2014.

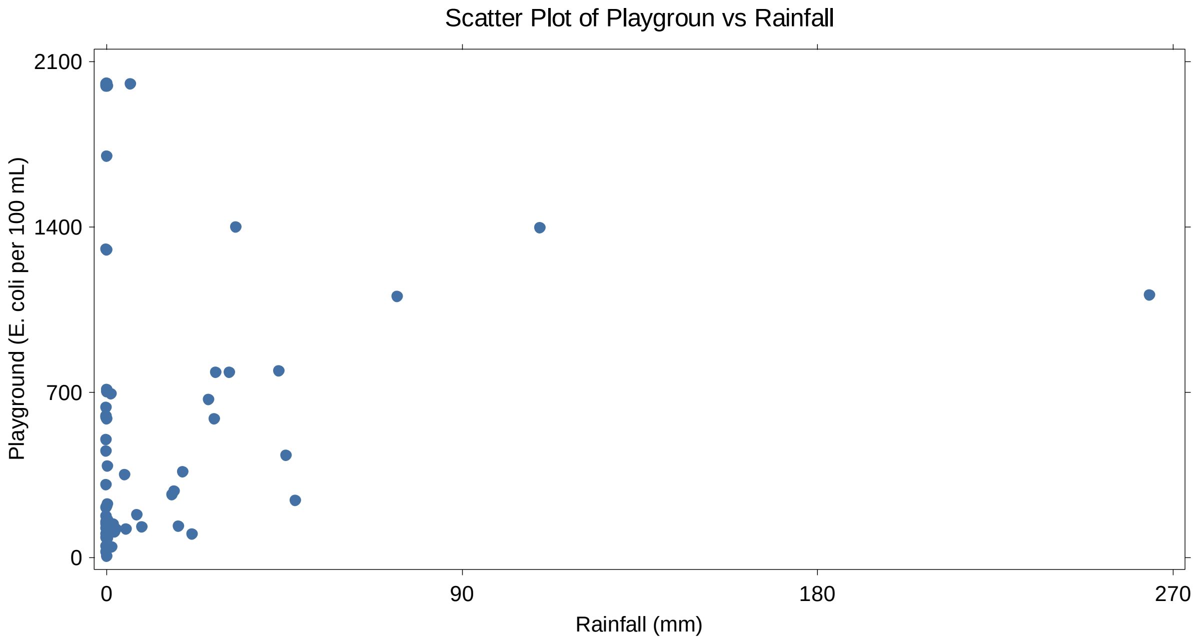


Figure 3 Playground site *E. coli* data vs rainfall scatter plot to visualise the relationship between water quality (*E. coli* data per 100 mL) and rainfall (mm).

Table 11 Playground site Pearson correlation between water quality (*E. coli* data per 100 mL) and rainfall (mm averaged over 48 h on the day before and the day of sampling), where transformed data were used for the analysis. Transformations applied to improve data normality (W>0.80) were: for *E. coli* data log_{10} function and rainfall sine function.

	Log_{10} Playground	Sine Rainfall
Log_{10} Playground (*E. coli* per 100 mL)	1.0000	
p-value	0.0000	
Sine Rainfall (mm)	-0.0809	1.0000
p-value	**0.5120**	0.0000
Cases Included 68 Missing Cases 0		

With the data binned according to levels of *E. coli*, 65% of the samples were less than 540 *E. coli*/100mL, and 82% less than 1000 *E. coli*/100mL (Table 12).

Table 12 Playground site data binned according to whether ≤540, >540, ≤1000, or >1000 *E. coli*/100mL, for 68 data points.

	≤540	>540	≤1000	>1000
Median	124	941	157	1550
Mean	166	1141	275	1609
min	10	591	10	1100
max	504	2001	782	2001
count	44	24	56	12

Compared with SH60, even less of the high levels occurred during rainfall (Table 13), suggesting that while rainfall may increase levels, there are significant inputs occurring that are unrelated to rainfall.

Table 13 48 hour rainfall at the Playground site.

	Binned at 540 *E. coli*	48 Hour Rainfall		Binned at 1000 *E. coli*	48 Hour Rainfall	
		>5mm	**>15mm**		**>5mm**	**>15mm**
***E. coli*/100ml**	≤540	10	7	≤1000	15	12
***E. coli*/100ml**	>540	10	9	>1000	5	4
Positive Predictive Value		50%	56%		25%	25%
Likelihood Ratio		1.8	2.4		1.6	1.6
% of samples>540	>540	42%	38%	>1000	42%	33%

Evaluation of the 12 samples with > 1000 *E. coli*/100ml confirmed rainfall was associated with only some of these samples, and little correlation with SH60 sites in the absence of rainfall.

Table 14 Playground site data from water quality above 1000 *E. coli*/100mL.

Date	Playground	SH60	Rainfall
	***E. coli*/100ml**	***E. coli*/100ml**	**48h mm**
19/01/2011	1100		74
22/02/2012	1100	2001	264
7/02/2013	1300	306	0.3
3/12/2013	1300	238	0
3/10/2011	1400	700	110
3/11/2011	1400	700	32.5
26/02/2013	1700	137	0
31/01/2012	2000	230	0
12/01/2011	2001	215	0
17/01/2012	2001		0
17/01/2012	2001		0
26/11/2013	2001	2000	5.9

4 Data Analysis Using Date-Matched Data

4.1 Normality testing and summary statistics

For the date-matched data set, assessing data normality indicated that normality could be improved by a $\log_{10}$ transformation of the *E. coli* values, which increased W from 0.75 to 0.89 for SH60 and from 0.76 to 0.97 for the Playground site (Table 15). For the date-matched data set, the rainfall was not normally distributed and data normality was improved by a sine transformation. Only transformed data were used for assessment of the correlation between sites for water quality.

Table 15 Assessment of data normality using the Shapiro-Wilk normality test (W) for the sampling date-matched data set for water quality, as estimated by *E. coli* data (per 100 mL), rainfall and air temperature. Transformations of the *E. coli* data ($\log_{10}$) and rainfall data (sine) were carried out to improve data normality.

Variable	N	W
Playground (*E. coli* per 100 mL)	59	0.7577
Log_{10} Playground (*E. coli* per 100 mL)	59	0.9678
SH60 (*E. coli* per 100 mL)	59	0.7464
Log_{10} SH60 (*E. coli* per 100 mL)	59	0.8931
Rainfall (mm)	59	0.3456
Sine Rainfall (mm)	59	0.7847
Air (°C)	59	0.8565

For the date-matched data set, SH60 had a median *E. coli* concentration of 306 per 100 mL with the 95^{th} percentile at 2000 *E. coli* per 100 mL. The Playground had a median *E. coli* concentration of 207 per 100 mL and the same 95^{th} percentile as SH60 (Table 16).

Table 16 Summary statistics for the date-matched data set for water quality (non-transformed and $\log_{10}$ transformed), rainfall and air temperature.

Variable	Mean	Median	Minimum	Maximum	95th Percentile
Playground (*E. coli* per 100 mL)	470	207	10	2001	2000
Log_{10} Playground (*E. coli* per 100 mL)	2.40	2.32	1.00	3.30	3.30
SH60 (*E. coli* per 100 mL)	471	306	2	2001	2000
Log_{10} SH60 (*E. coli* per 100 mL)	2.47	2.49	0.30	3.30	3.30
Rainfall (mm)	12.19	0.00	0.00	264.00	ND
Air (°C)	20.2	20.0	12.0	25.0	ND

4.2 Time plots

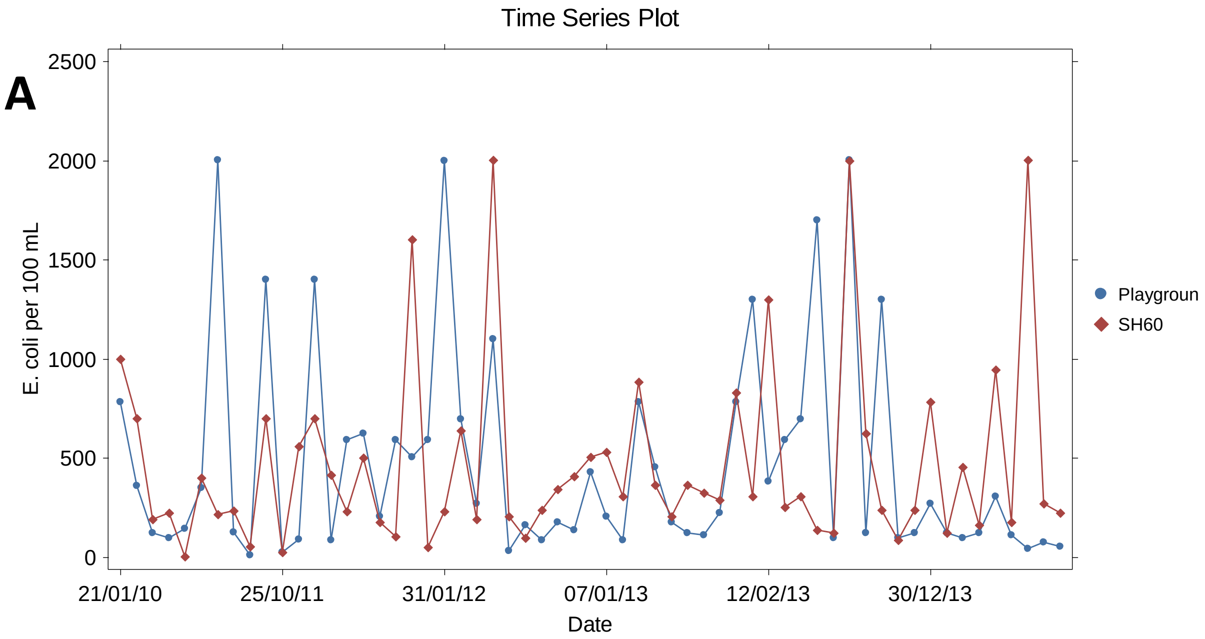
Time-series plots of the date-matched data sets showed broad similarity between the levels of *E. coli* (Figure 4). There are however, times when high levels of *E. coli* are recorded at one site and not the other.

4.3 Two-sample T-test

As expected from the preceding discussion, there was no significant difference (pooled P=0.425) between the concentration of *E. coli* at the Playground or SH60 locations (Table 17).

4.4 Correlation analysis

Scatter plots were used to assess the linearity of the relationships between the concentrations of *E. coli* (per 100 mL water) at the two sites (Figure 5). There was a weak (0.34) but significant positive association was present between the water quality at SH60 and that at the Playground (P=0.008, Table 18). There was no significant relationship (P>0.050) between rainfall or air temperature variables and *E. coli* for the date-matched variables (Table 18). However, the water quality at SH60 explained only 34% of the water quality at the Playground site. This suggests that even though the sites are hydrologically linked over about 1 km, the sites are independently receiving inputs that are contributing *E. coli* at the individual sites.



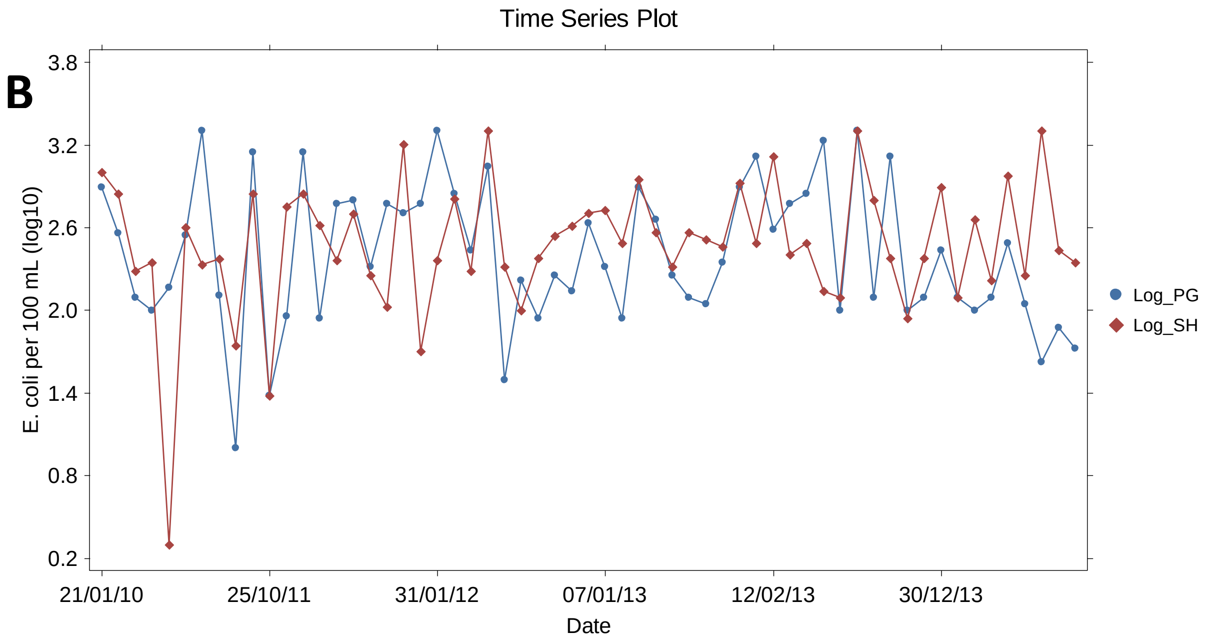


Figure 4 Date-matched data set time-series plot for *E. coli* concentration (per 100 mL) at the sampling locations (Playground and SH60 sites) from (A) non-transformed and (B) $\log_{10}$ transformed. NB: The Playground label is truncated to "Playgroun".

Table 17 For the date-match data set, a two sample T-test was used to compare the mean *E. coli* concentration at the two sampling locations (Playground and SH60 sites) using the $\log_{10}$ transformed data ($\log_{10}$PG and $\log_{10}$SH).

Variable	N	Mean	SD	SE
Log_SH60	59	2.4736	0.4859	0.0633
Log_Playground	59	2.3995	0.5182	0.0675
Difference		0.0741	0.5023	0.0925

T-Tests for Mean Difference

Null Hypothesis: difference = 0

Alternative Hyp: difference ≠ 0

Method	Variances	DF	T	P	Lower 95% C.I.	Upper 95% C.I.
Pooled	Equal	116	0.80	0.4246	-0.1091	0.2573

Cases Included 118 Missing Cases 0

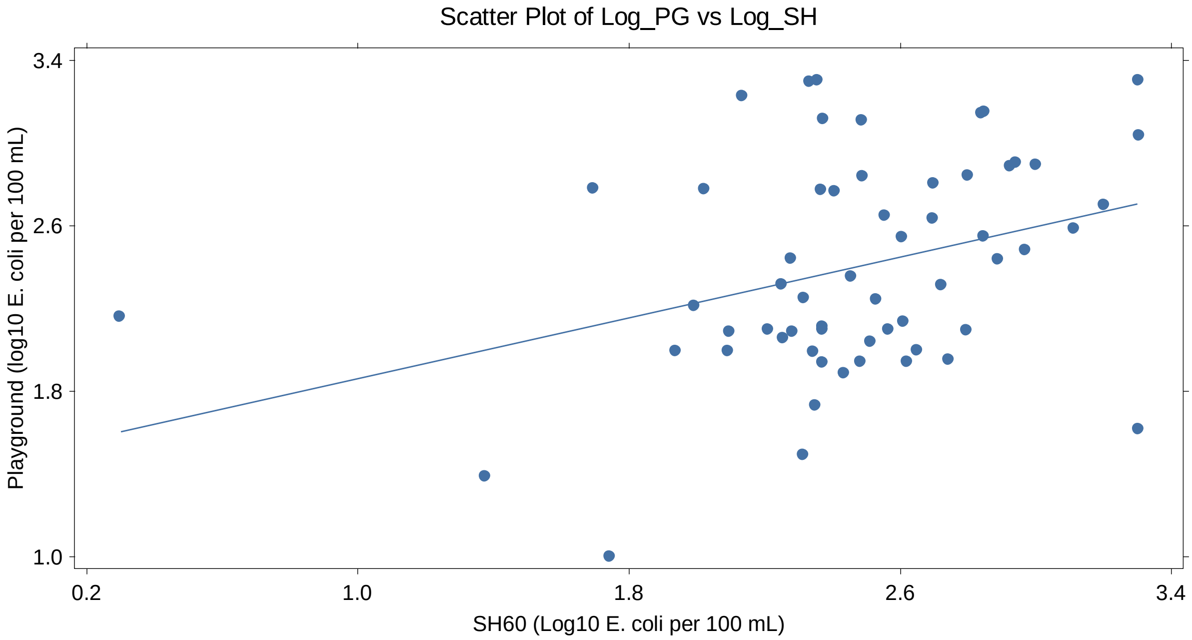


Figure 5 For the date-matched data set, scatter plot with linear fit that was used to visually assess the relationship between *E. coli* ($\log_{10}$ per 100 mL) at the SH60 and Playground sites.

Table 18 Date-matched data set for Pearson correlation between the different variable (*E. coli* data were also assessed as $\log_{10}$ transformed). The significant relationship between sites is highlighted in yellow.

	Log_{10} Playground	Log_{10} SH60	Sine Rainfall	Air
Log_{10} Playground (*E. coli* per 100 mL)	1.0000			
p-value	0.0000			
Log_{10} SH60 (*E. coli* per 100 mL)	0.3438	1.0000		

p-value	**0.0077**	0.0000		
Sine Rainfall (mm)	-0.0183	-0.0120	1.0000	
p-value	0.8906	0.9280	0.0000	
Air temperature (°C)	0.1151	0.2434	0.1818	1.0000
p-value	0.3856	0.0633	0.1682	0.0000

As shown in Table 19, when >10 mm of rainfall occurred, there was often elevated *E. coli* at both sites, although not always, and as previously shown many of the highest recorded levels of *E. coli* are unrelated to rainfall.

Table 19 Microbial Water Quality measurements when there is rainfall of >10 mm in previous 48 hours at Aorere@Devils.

Date	Playground	SH60	Rainfall
	***E. coli*/100ml**	***E. coli*/100ml**	**48h mm**
27/01/2012		65	12
30/12/2013	271	782	16.7
14/02/2012	271	192	17
2/01/2014	124	124	18.8
22/01/2010	360	700	20
29/11/2011	87	415	21.5
6/12/2011	591	230	28
21/01/2010	782	1000	28
5/02/2013	782	831	31.5
3/11/2011	1400	700	32.5
24/03/2010		10001	36.5
16/01/2013	782	885	43.7
3/01/2013	429	504	45.5
3/10/2011	1400	700	110
22/02/2012	1100	2001	264

5 Conclusion

Water quality monitoring data for two sites on the Tukurua Creek (SH60 site upstream of the Playground site) were provided to ESR by Tasman DC. These data were statistically analysed to assess whether there were trends or association that could help explain elevated *E. coli* concentrations frequently found at the Playground location. Rainfall was suggested by Tasman DC as the most likely influence on water quality, and it was this hypothesis that the statistical analysis tested.

The samples taken from a reference site upstream of the SH60 site showed good water quality. However, there were insufficient reference site data (n=6) to include the reference site in the analysis or to determine catchment-wide influences.

Comparison between water quality at SH60 and the Playground were made for days where data were available for both these sites. Based on these *E. coli* concentrations, there was no significant difference between the water quality at the two sampling locations (P>0.05). The temporal pattern of *E. coli* presence was broadly similar between the two sites, with fluctuations from less than 10 to greater than 2000 *E. coli* per 100 mL occurring at both sites, but not necessarily on the same day.

While the correlation analysis found a significant association (P<0.05) between the two sites, the scatter of elevated *E. coli* in the absence of rain or with low rainfall levels meant that factors other than rainfall were influencing water quality in the Tukurua Creek.

Therefore, the data analysis does not support the rainfall-linked hypothesis for explaining most of the occurrences of elevated *E. coli* and suggests point-source inputs. Furthermore, there was only a weak association between the water quality at SH60 and the water quality at the downstream Playground sampling location. The weakness of the latter association is surprising; it would be expect that over the distance of only about one kilometre and in the absence of additional inputs, downstream water quality would be strongly influenced by upstream water quality.

Hence, while rainfall-associated *E. coli* presence explains some of the elevated *E. coli* levels at the Playground and SH60, it does not explain the high levels found in the absence of rain, nor the apparent independence between the Playground water quality and that of SH60 at zero or low rainfall. Therefore, the data analysed suggest probable faecal inputs are entering at or upstream of the Playground sampling site, but somewhere downstream from the SH60 sampling site.

Our recommendations are:

1. Carrying out a detailed **sanitary survey**, including intensive sampling, along the Tukurua Creek from above SH60 to below the Playground to identify possible inputs that could contribute to the presence of faecal indicator bacteria.
2. That based on the outcome of the sanitary survey, **faecal source tracking** is used to identify the likely sources of the contamination.

Both these recommendations would need to be done on several occasions and at multiple locations so that with and without recent rainfall contexts can be assessed.

Future studies on the source of faecal pollution in the stream could employ filtration of water samples in house by TDC. The methods for these and consumable can be provided by ESR. TDC could then review the results of the *E. coli* monitoring and chosen which samples they wished to investigate. This would remove the need to resample to “catch” a high *E. coli* count for FST analysis and allows for the budget for investigations to be spent wisely.