water

MDPI

Article

Climate and Land Use Influences on Bacteria Levels in Stormwater

Kaifeng Xu [1], Caterina Valeo [1,*], Jianxun He [2] and Zhiying Xu [1]

[1] Department of Mechanical Engineering, University of Victoria, Victoria, BC V8W 2Y2, Canada; xukaifengian@gmail.com (K.X.); sunny520leo@gmail.com (Z.X.)
[2] Civil Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada; jianhe@ucalgary.ca
* Correspondence: valeo@uvic.ca; Tel.: +1-250-721-8623

Received: 31 October 2019; Accepted: 18 November 2019; Published: 22 November 2019

Abstract: The influence of climatic variables and land use on fecal coliform (FC) levels in stormwater collected from outfalls throughout southern Vancouver Island between 1995 and 2011 are examined through statistical analyses, Fourier analysis, Multiple Linear Regression (LR) and Multivariate Logistic Regression (MLR). Kendall's τ-b demonstrated that FC levels were significantly and positively correlated with the amount of residential area within a drainage catchment generating the runoff, and that FC levels were location dependent. Climatic variables of temperature and antecedent dry period length were significantly and positively correlated with FC levels at both the sampling location level and across the region overall. Precipitation and flowrates were negatively correlated with FC levels. Fourier analysis showed that monthly FC levels shared the same 12 month cycle (peaking in July) as precipitation and temperature. MLR modelling was applied by aggregating the LogFC data by order of magnitude. The MLR model shows that the data are subject to different influences depending on the season and as well, the month of the year. The land use and climate analyses suggest that future climate change impact studies attempted on nearshore bacterial water quality should be conducted at the urban catchment scale.

Keywords: stormwater quality; fecal coliforms; Vancouver Island; nearshore areas; bacteria loading; multinomial logistic regression; periodicity analysis; land use impacts; climate impacts

1. Introduction

Contaminates transported through stormwater runoff to coastal waters can pose a potential risk to public health and the environment [1]. Furthermore, the stormwater drainage system can also be contaminated by sewage through infiltration or unintended connections with sewer systems and poorly maintained in-ground sewage disposal systems [2]. Contaminated stormwater runoff will often contain excessive levels of bacterial contaminants, which are directly related to disease outbreaks and adverse impacts to aquatic life [3,4]. Therefore, many municipalities will monitor and manage stormwater quality for both health and environmental concerns. Fecal coliforms (FC) have historically been used as a fecal indicator to indicate the presence of microbial contamination in surface and ground waters [5,6]. Since they are considered as an indicator of surface water quality and safety, FC is often selected for monitoring in those water quality monitoring programs concerned with microbial loading. Microbially contaminated water can be a serious source of intestinal disease through ingestion, or exposure through bathing or by consuming contaminated shellfish [7]. Factors that could affect fecal indicator bacteria levels have been investigated in the literature [8–10]. Sewage overflow, wildlife and stormwater runoff from urban and agricultural land use are important sources of fecal coliform affecting water quality [11,12]. Non-point urban and agricultural land use zones have significant

impacts on water quality and produce a large number of fecal bacteria conveyed to water bodies [13,14]. As well, climatic variables like precipitation and wet seasons in a region are suggested to have a positive correlation with the concentration of fecal bacteria in the surface waters [15–17].

There are a variety of studies looking at stormwater generated microbial loadings and how they are influenced by climatic variables. Henry et al. [18] found rainfall within the 24 h preceding sampling was found to be correlated with incidence of fecal matter, however, absence of rainfall was also significantly tied the incidence of fecal matter depending on the location. The length of the period over which precipitation was computed for determining correlations with fecal contamination, impacted whether the correlations were positive, negative, significant, or had no impact [19–23]. Seasonal variations in the correlations computed for fecal indicators versus temperature, precipitation and antecedent dry period length were observed in several studies [21,24,25] but generally, temperature was positively correlated with indicator bacteria concentrations [26], and precipitation was also positively correlated [20,22]. McCarthy et al. [27] found that *E. coli* (EC) levels were highly correlated to antecedent climatic parameters but like [26], found they were less correlated to hydrologic parameters such as runoff and suggested that EC concentrations "were not prone to displaying a first flush effect", which is in contrast to other studies in which fecal coliforms peaked after storm events [28]. Solar intensity was also seen to have a negative correlation with EC [29]. The additional influence of spatial location in analyses examining climatic influences on fecal contamination was observed in several studies [30,31] with Vermeulen and Hofstra [31] suggesting that the great variability among location characteristics helped to explain variations in *E. coli* levels in their study. All of these studies vary extensively by how strongly or weakly correlations are seen between bacteria loading and climate variables, as well as to the way the variable is constructed. That is, depending on whether precipitation is measured during a rain event, on the day of sampling, or averaged over a 3 day period over which the sampling took place, produced variations in results. This suggests that the scale of the climate variable matters as much in the research as the climatic variable itself (whether temperature, precipitation, solar radiation, etc.).

Spatial influence studies in the literature include several works stating that the contamination in coastal water quality is caused by the combined effects of human activity and environmental factors in coastal areas [28,32]. The accumulated fecal coliform can be transported into nearshore ocean regions from direct runoff or sewage overflow during storm events [33,34] in combined sewer systems. Mallin et al. [35] showed that watershed population and watershed size were significantly related to average fecal coliform levels, with the strongest relationships for percentage of impervious surface and fecal coliform levels. Jent et al [19] found that in non-dry periods, fecal contamination was significantly correlated with agricultural land use, while Tiefenthaler et al. [21] showed that mean EC were significantly greater in developed watersheds than undeveloped watersheds. Similarly, Vitro et al. [36] found that road network density was associated with increasing fecal coliform levels but housing unit density had a significantly negative relationship with FC levels; Paule-Mercado et al. [25] found the highest FC concentrations were also found in urban areas versus agricultural areas. Delpla and Rodriguez [37] showed that FC were significantly and positively correlated with urban and agricultural land, but forests were negatively correlated. FC concentration was found to be significantly and positively correlated with urban land use in even low percentages of urbanization [38]. Additional studies [20,22] found higher fecal indicator bacteria concentrations in urban sites over forested sites. The literature suggests that urban areas provide higher bacterial loads in comparison to undeveloped green spaces, and agricultural lands also contribute higher bacterial loads than green spaces.

Other studies have gone a step further in attempts to mathematically model bacteria levels as a function of land-use or climatic variables in order to provide insight into the relationships. Wu et al. [23] used logistic regression modelling and showed that EC *presence* was significantly and positively associated with developed area but negatively associated with agricultural area. Multivariable logistic regression models were used to show that increases in the number of heavy rain days significantly increased the likelihood of the presence of EC and that EC was also likely to be detected with increases in mean temperature. Cha et al. [24] developed a model using Bayesian regression that used

meteorological and land-use characteristics to estimate FC concentrations. The modelling was used to provide predictions in FC increases given temperature increases. Galfi et al. [39] used PCA and cluster analysis to find correlation patterns and found that significant climatic variables for influencing indicator bacteria levels varied among catchments as well as seasonally. St. Laurent and Mazumder [38] used a classification tree to better understand the influence of land use on FC levels. A common aspect of all of these modelling studies is that they use data-driven, stochastic methods for modelling. This is reasonable as causal mechanisms are difficult to capture and physically based models may be onerous to implement when simply attempting to determine climatic influences on FC concentrations. Examples of sophisticated, physically-based models in the literature that predict bacterial levels as a function of environmental parameters include [29,40], but both look at transport in river systems. The latter did note that the sparsity in the time series of bacteria concentrations adversely affected the modelling [40].

The Capital Regional District (CRD) of southern Vancouver Island is a government body representing 13 municipalities and three electoral areas. The core area of CRD has approximately 270,000 people and 27,600 hectares including residential, industrial, commercial, institutional and agriculture regions. The area has approximately 8350 properties with onsite sewage disposal and with others connected to a sewer system [2]. The CRD assists these municipalities in developing their stormwater management plans and infrastructure, as well as water quality monitoring for these municipalities. Coastal water quality has been a critical issue worldwide for nearshore inhabited regions and the CRD is no exception. Coastal water quality is known to be affected by natural and human factors includes runoff, sewage wastewater, land reclamation and climate change [41]. The implementation of monitoring programs is necessary and critical in managing coastal water contamination [42,43]. The CRD collects pollutant levels (including fecal coliforms) within stormwater pipes, streams and nearshore areas throughout the CRD, with an interest in identifying hotspots and remediating those areas of highest priority. This process includes collecting water samples and analyzing for fecal coliform bacteria in the sample. The collected data are used to estimate and analyze the distribution of microbial contamination and any possible public health concerns. This allows the jurisdictions involved to undertake remedial measures where most needed. Currently, the sampling frequency is regulated primarily by cost and capacity but increased sampling strategies are advised for locations exceeding a certain threshold. Given that fecal coliform contamination in stormwater runoff is directly influenced by climate and watershed properties [10,13,44], a regular sampling scheme that does not consider climate and weather will likely miss peaks in contamination.

The literature shows some consensus on the impact of land use on bacteria levels but the influence of climate variables remains contentious, particularly where the study focuses on stormwater runoff generation. Much of what is observed is likely due to site specificity and the scale of the sampling and analysis [40,45]. If agricultural areas are absent from the drainage area, bacteria loads in stormwater runoff generation are believed to be low and are thus, highly affected by sampling plans, which tend to be sparse and intermittent. Thus, stormwater specific data sets tend to be limited in terms of sampling period, length of database in time and extent over space, and therefore, the insight they can provide on environmental influences is limited. This often drives researchers to turn to data-driven methods [23,24,38,39,46,47] to provide insight but these methods are not unaffected by the temporal and spatial scale of the sampling and analysis.

Given the extensive monitoring network developed by the CRD, the objectives of this work are to determine if the fecal coliform data collected by the CRD during its regular monitoring program are influenced by local weather and land use; what insights arise from the scale of the collection and variable (fecal coliforms in this research) studied, and in particular, whether data-driven methods are able to provide insights into the causal mechanisms behind the observations. With the outcomes, the monitoring program can be modified to improve sampling statistics and possibly indicate the impacts of climate change on bacterial loads in the future.

2. Materials and Methods

2.1. Study Area

The study area is located on the southeastern core area of the CRD between latitudes 48°29′57.3″ N and 48°24′02.2″ N, and longitudes 123°26′26.3″ W and 123°15′40.4″ W. Shown in Figure 1, the study area forms part of the "core area" of the CRD and is approximately 15,000 hectares in size and spanning six municipalities, two First Nations, federal, provincial, regional, and municipal parkland. Land use in the core includes residential, commercial, industrial, institutional, and agricultural activities and is shown in Figure 2b. The region lies within the *Coast Mountains and Islands* physiographic region, and contains a number of watersheds of variable size, some natural and some urbanized, draining to the coastline through stormwater drainage systems or creeks and rivers. Within the *Georgia Depression* Ecozone, the Biogeoclimatic zone is *Coastal Douglas Fir*, which experiences mean annual temperatures of almost 10 °C and mean annual precipitation of over 1000 mm annually [48]. The mean coldest month is January with average temperatures of around 3 °C and the mean warmest month is in July, which averages roughly 17 °C. This mild climate experiences on average 204 frost-free days per year. While weather systems in British Columbia are highly impacted by elevation changes, the elevation change in the study region is less than 50 m. Rainfall systems for this area tend to form in the north Pacific and move easterly across the mountain range running down the center of the island. The process leads to greater rainfall on the western side of the Island and less on the leeward side (eastern side) resulting in a rain shadow for the study area because it is on the eastern side of the Island [48]. Rainfall is higher in the winter months and leads to stormwater runoff draining through the natural portions of the watershed (see Figure 2a) or through the stormwater minor drainage system and pipe system to end up discharged to the coastal harbors shown in Figure 1. Soils in the overburden above rock formations in pervious areas are generally Distric Brunisol [48] soil type within the Luvisolic soils group, which are typically well draining.

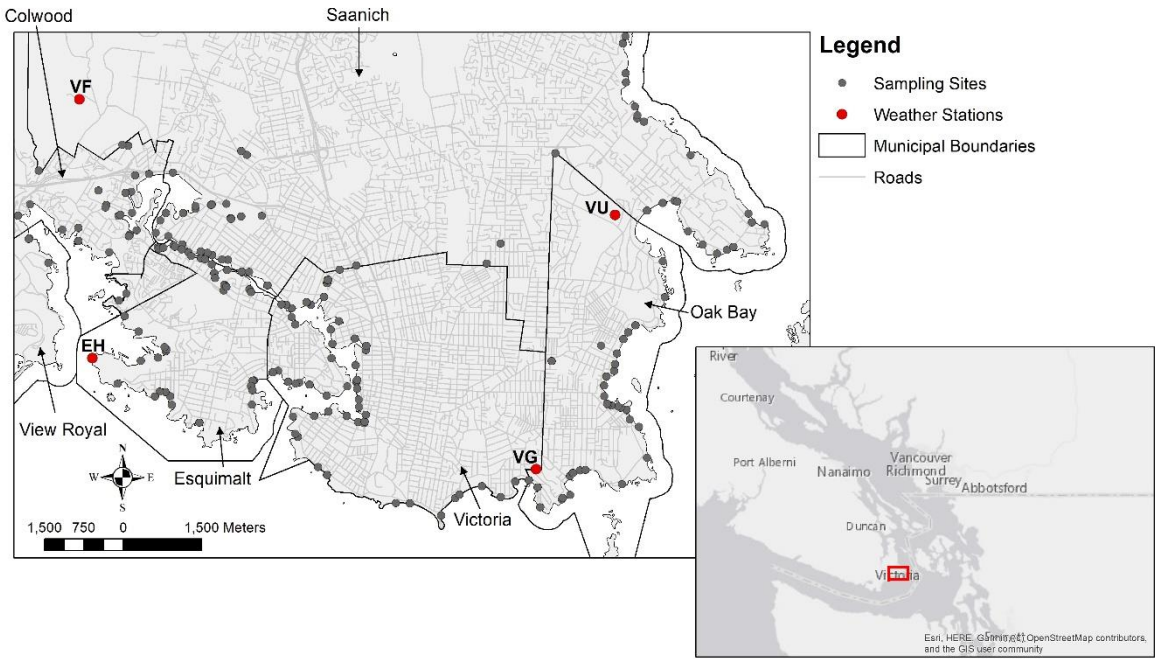
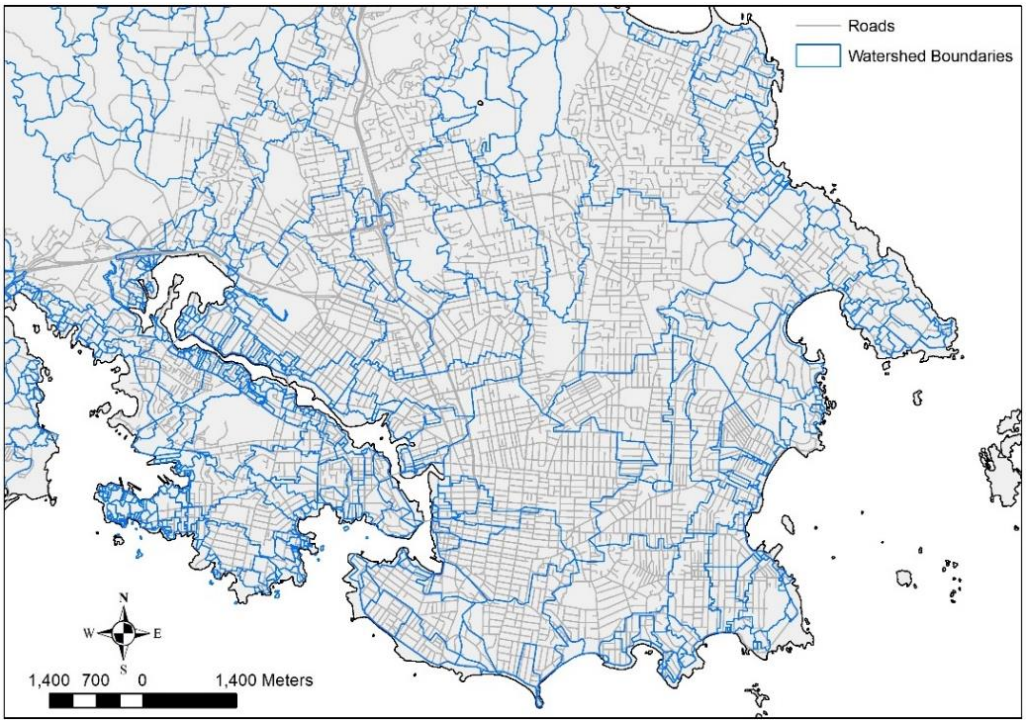
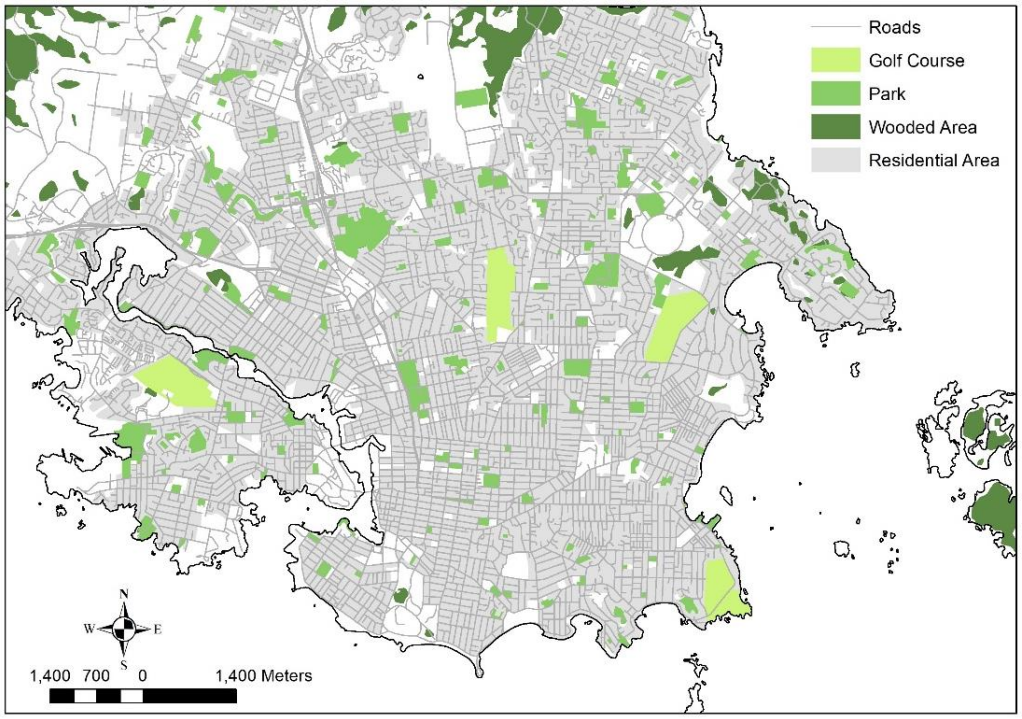


Figure 1. Location of sampling sites and meteorological stations along Nearshore Areas of Southern Vancouver Island. Municipal boundaries are marked with a solid dark line.



(**a**)



(**b**)

Figure 2. (**a**) Watershed map in Capital Region District; (**b**) Land-use map in Capital Region District.

2.2. Database Construction

The fecal coliform samples were collected from stormwater outfalls along the coastline of Esquimalt, Victoria, Oak Bay, and Saanich, and analyzed by the Stormwater, Harbours and Watersheds Program (SHWP) from 1995 to 2011. The database contains in total 6112 sampling data (N = 6112) from over 480 unique sampling locations. Stormwater samples for most stations were collected from each discharge

point once during January to April (wet season) and once during June to September (dry season); although, there are data in all months of the year for most years in the database. Stormwater flows were sampled by land or boat at the point of discharge before going into the ocean to avoid unwanted flows. The sampling process attempts to avoid first flush conditions in order to reduce the chances of an unusual result. The measurements are also compared to historical results where the discharge is flagged for resampling outside the usual sampling period in the case of an unusually high observation. Thus for some years, there are data points for all 12 months [2].

A series of statistical tests were conducted on the data to determine which methods should be used in the analysis. The Kolmogorov–Smirnov test for normality showed that the fecal coliform data are not normally distributed and are highly right-skewed. In order to reduce the skewness and kurtosis, the fecal coliform (*FC*) data were log base10 transformed but the Kolmogorov–Smirnov test again rejected a normal distribution assumption for Log*FC*. All further analyses are conducted on *FC* data that are above a recreational water quality guideline of 200 CFU/100 mL [49] because data lower than this limit are not of concern in the environment. Restricting the database to values equal to, and above this limit resulted in a database of 2749 individual data observations over the 17 year period.

Contamination by sewage is considered an important factor causing high fecal coliform levels in waterbodies in many previous studies [11,12]. The report from SHWP also comments on the presence of sewage odors when a sample is taken, or other related problems. As a large amount of fecal bacteria exist in sewage wastewater, sewage cross-connections or infiltration can cause seriously adverse impacts. Therefore, non-parametric ANOVA is used to investigate if a significant difference exists in fecal coliform levels between the "sewage odor" identified stormwater samples and no odor stormwater samples. Sewage presence analysis was performed using the Kruskal Wallis Test and showed that the group of sewage-related samples (from odor detection) were significantly different from the group of regular samples with p-value = 0.000; the sewage-related samples have much higher levels of FC. There were 384 sewage-related samples. Since the focus of this study is stormwater that is not impacted by sewage, these data were removed from the final database used in the analysis. The frequency histogram of the final database is shown in Figure 3.

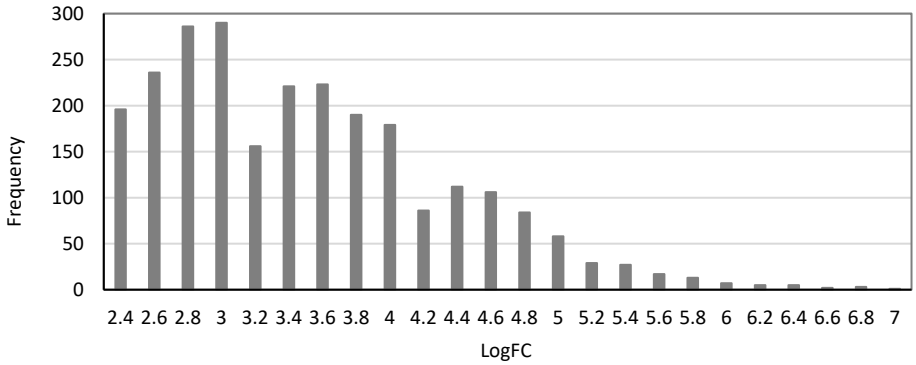


Figure 3. Histogram of cleaned Log*FC* data. The bin value indicates upper limit of that bin.

Daily historical data for this study were obtained from Environment Canada. The four weather stations selected in the CRD area, and shown in Figure 1, include the Esquimalt Harbour weather station, Victoria University CS weather station, Victoria Gonzales CS weather station and Victoria Francis Park weather station. Daily temperature and precipitation measured from 1995 to 2011 are used to check the relationship between bacterial contamination and meteorological variables. An average of the data from the four stations was computed and used in the analyses. Monthly climatic data averaged over the 17 years of collection were also generated for use in the periodicity analysis. In addition, cloud cover ratio data for the Victoria area were derived from Environment Canada, 1981–2010 station data for the study area. The watershed information shown in Figure 2a was obtained from the CRD and the land use information (Figure 2b) was obtained from Natural Resources Canada.

ArcGIS was used to visually explore the logarithm of the geometric mean of fecal coliform data above the regulated threshold of 200 CFU/100 mL collected within and adjacent to nearshore areas in the study region. The spatial distribution of bacterial contamination is visualized using the Kernel Density [50] which calculates the density of points in a region for each cell multiplied by the unit area. This map can provide a visual summary of the spatial distribution of fecal coliform data across the region as well as suggest any "hot spots". Figure 4 shows the areas determined as hot spots with high fecal coliform levels. The areas with deeper shading, indicating higher levels of FC are seen near the Victoria downtown, Esquimalt and the eastern shore area of Oak Bay. The southern shore of Oak Bay and the northeastern shores of Victoria have relatively lower pollution levels. The density map was also used to select hot-spot stations for further analysis (also shown in Figure 4). The five stations shown were selected to span the visually obvious areas of high FC levels, but also were stations that had consistent data collection over the period. The hot spots are located in the densely populated residential areas, which suggest that the source of fecal coliform is likely from urban activities. These five stations are further studied in the temporal analysis section.

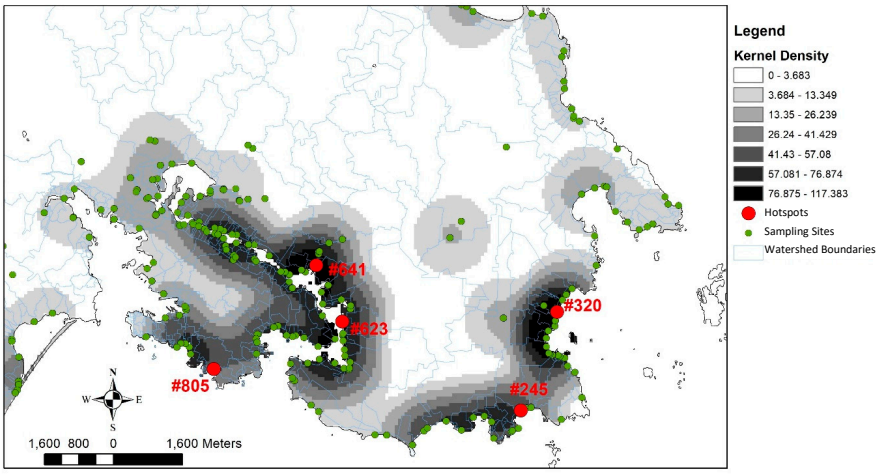


Figure 4. Density map based on the Log*FC* value of samples. Density values shown in the legend are in units of km^2.

For determining relationships between these data and climatic variables, correlations are computed between the data and several independent variables: 7, 3, 2 and 1-day rainfall totals preceding, or on the day of sampling; 7, 3, 2 and 1-day mean temperature; maximum temperature and minimum temperature preceding, or on the day of sampling; mean monthly cloud cover ratio; antecedent dry period length preceding the day of sampling, and flowrate measured at or close to the sampling location. Spatially distributed variables considered are watershed drainage area leading to the sampling point and land use in that drainage area, which is divided into residential use and green spaces.

2.3. Data Exploration and Analysis Methods

IBM SPSS, MATLAB and MS Excel are used in the statistical analyses and exploration of the data. To investigate trends and periodicity, dependent variables such as Log$FC_{i,j}$, are distinguished here from independent variables (such as air temperature and precipitation), where i represents the date collected and j indicates the station number. Other dependent variables used in this research include, the monthly (m) average of the log-transformed fecal coliform values, Log$FC_{y,m}$ in each year y over the 17 year period for all stations combined; and the 17 year average by month, log-transformed fecal coliform values (LogFC_m) is also considered. The geometric mean of the $FC_{i,j}$ values in the averaging period are computed first, then log transformed.

2.3.1. Periodicity and Fourier Analysis

Periodicity investigations [51] can provide insight into seasonal influences on the temporal distribution of bacterial contamination, and seasonality has already been shown to influence results in the literature. The data are imported into MATLAB for periodicity analysis in which Fourier analysis is conducted to investigate peaks, periodic cycles and potential models. In a periodicity analysis using Fourier models, the following equation is fitted to the Log$FC_{y,m}$ versus a year-month index spanning from 1 to n = 204 (equal to 17 years of 12 month data), and to LogFC_m versus m (n = 12) with the following equation:

$$f(x_t) = a_0 + \sum_{i=1}^{n} a_i \cos(i\omega t) + \sum_{i=1}^{n} b_i \sin(i\omega t) \quad (1)$$

where a_0; a_i; b_i are Fourier coefficients, and $f(x_t)$ is the predicted value of variable x_t at time t. The fundamental frequency ω is equal to $2\pi/T$ in which T is the period of the signal. The climate data are also investigated with this function in order to determine if the periods in the climate data are similar to the periods in Log$FC_{y,m}$ and LogFC_m.

2.3.2. Temporal Trends and Correlations between Variables

The overall correlations between Log$FC_{i,j}$ and independent variables are analyzed to determine the influence of each variable toward the level of bacterial contamination using Kendall's τ-b for non-normal distributions. While, Kendall's τ-b and the commonly used Spearman's ρ produce similar values much of the time, Kendall's test provides better explanations for non-linear correlations as compared to Spearman's test [52]. The data observed at the hot-spot stations are also tested for correlations with the independent variables. Correlations are computed between Log$FC_{i,j}$ and the following independent variables: 7-day total precipitation ($P7$), 3-day total precipitation ($P3$), 2-day total precipitation ($P2$), precipitation (P), 7-day average temperature ($T7$), 3-day average temperature ($T3$), 2-day average temperature ($T2$), mean temperature (T_{mean}), maximum temperature (T_{max}), minimum temperature (T_{min}), antecedent dry period length (t_{dry}), watershed area (WA), residential area (RA), greenspace area (GA), cloud cover (CC), and flow rate at discharge (FR).

2.3.3. Regression Modelling

Depending on the outcomes of the correlation analysis and periodicity analysis with Fourier modelling, multivariate linear regression (LR) is attempted to provide the best (in terms of coefficient of determination R^2) model possible given all the temporal data. The equations for an LR model are well known and not repeated here. In addition, Multinomial Logistic Regression modelling (MLR) is also used to explore other potential prediction models based on the limitations of LR. MLR is a method similar to LR but has the added advantage of not being constrained by several assumptions including that the independent variables must be statistically independent (although collinearity should not be great). In addition, the dependent variable may be a categorical or nominal value [53]. This was intentionally chosen because of the incredible variance in the data of FC_{ij} across the entire database of 17 years, which is also true of LogFC_{ij}. By using categories of LogFC_{ij} determined by the order of magnitude of the LogFC_{ij} value, this method can provide insight into how well individual parameters work to influence observed levels of LogFC, and if predictions on the likelihood of a LogFC value being observed at a certain level are possible based on climatic observations. This method is considered a classification method that relates a linear prediction function to the "logit" $g(\boldsymbol{x})$ with n number of predictors $\boldsymbol{x}$ ($\boldsymbol{x} = x_1, x_2, \ldots, x_n$) such that [53]:

$$g(\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \quad (2)$$

and the conditional probability that the outcome Y is present is $\Pr(Y = 1|\boldsymbol{x}) = \pi$ and,

$$\pi = \frac{e^{g(\boldsymbol{x})}}{1 + e^{g(\boldsymbol{x})}} \tag{3}$$

Thus, the model attempts to compute the odds, or likelihood of observing a value associated with a certain level or class. The application of the MLR model was conduct with SPSS which also provides measures of (1) model fitting information (in the form of a X^2 test that compares the model to a −2 Log Likelihood model) such that the model is significant if the p-value is <0.05; (2) goodness of fit (also at 0.05 level) relative to a baseline null model; (3) the likelihood ratio tests in which the significance of each predictor is rated as having a significant influence on the classification if the p-value is <0.05; (4) model coefficients; (5) performance in each class or level between the observed and predicted values (as a percent) as well as the overall percent performance; and (6) three Pseudo R^2 values that provide a measure of model performance [54].

In such a model, the choice of levels or classes is important. If too few classes are chosen that are too gross with little discretization, the model will perform perfectly as there are few "choices" to make, even though the model fit may not be significant. If there are too many classes, the model performance is poor. The application of this model will be on the order of magnitude of the LogFC values instead of the actual values because the actual values would create too many classes. MLR models will be developed for all of Log$FC_{i,j}$, Log$FC_{y,m}$ and for values of Log$FC_{i,j}$ at hotpsots but the "values" will now be one of 11 possible classes created as the order of the LogFC_{ij} value: 2, 2.7 (representing 500 CFU/100 mL), 3, 3.7, 4, 4.7, 5, 5.7, 6, 6.7, 7, etc. The $FC_{i,j}$ values were rounded up to the nearest order of magnitude (thus, values greater than 500 CFU/100 mL, for example, would be rounded up to 1000 CFU/100 mL), then log transformed before placing them into a class. For Log$FC_{i,j}$, no intermediate classes beyond the order (i.e., no value of 2.7 for example) are chosen so there are only six classes.

It should be noted that this method has no equivalent to R^2 such as that which can be computed for a linear regression model; thus, researchers have devised what are known as Pseudo-R^2 values, with SPSS reporting three different types. Since all three provide similar values, the first reported, which is the Cox and Snell value, is given in this study and is computed as:

$$\text{Pseudo} - \text{R}^2 = 1 - \left(\frac{L_0}{L_M}\right)^{2/N} \tag{4}$$

where N are the number of observations in the dataset, L_0 is likelihood of a model with no predictors and L_M is the likelihood of the model being assessed. Great care must be taken when using this value to make interpretations. Firstly, a perfect prediction does not produce a value of 1 as in R^2, but instead for Cox and Snell's Pseudo-R^2, a perfect model performance would give a value less than 1. The literature has reported that a value close to 0.2 is considered a "good model" [55]. Since great care must be taken in the interpretation of this performance metric, MLR model performances are *compared* as opposed to examining the Pseudo-R^2 value in isolation. This is sufficient for understanding the *relative* influence of climate variables on predictions in an effort to gain knowledge where significant correlation analysis is lacking.

3. Results and Discussion

3.1. Periodicity and Fourier Models

Figure 5a,b show the time series of Log$FC_{y,m}$ and the average precipitation and average temperature in the corresponding month, along with the best fitted series for each variable. Figure 6 shows three panels with the average monthly data (LogFC_m is shown in Figure 6c) over all 17 years. Fourier series models describing the fitted curves are also shown. Table 1 shows the Fourier coefficients and the RMSE and R^2 values for the best fit models. Higher order harmonics were attempted until R^2 values

ceased to increase. For $P_{y,m}$ and Log$FC_{y,m}$, a simple, one harmonic best fits the data although the RMSE is high. For temperature, $T_{y,m}$, an 8 harmonic Fourier model fit the temperature function for this area over the 17 year period extremely well. The lower harmonics did not produce nearly as good an R^2 value.

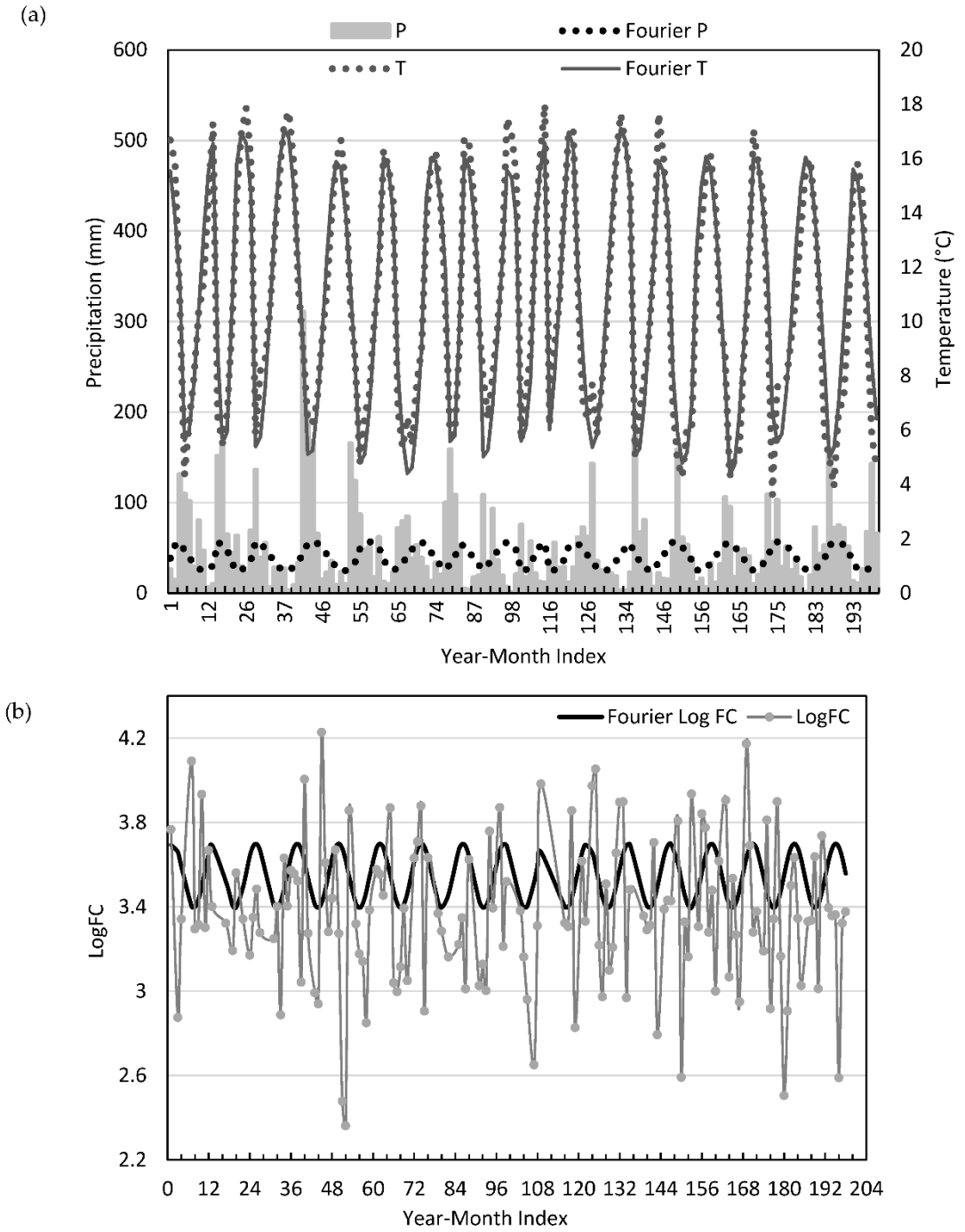


Figure 5. (**a**) Monthly climate variables and best fit Fourier series; and (**b**) Log$FC_{y,m}$ data with fitted Fourier series model.

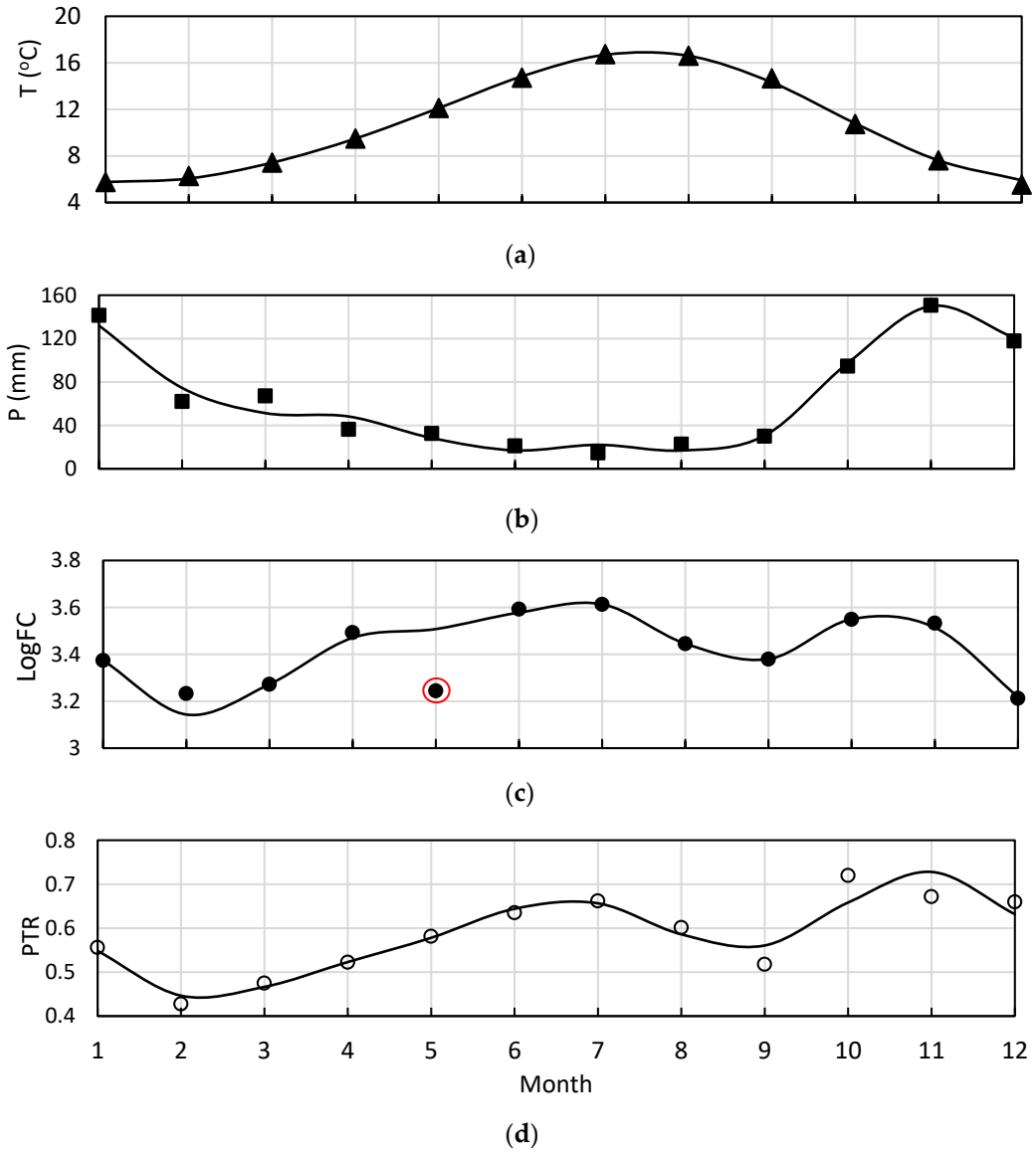


Figure 6. (**a**) Monthly temperature averaged over the entire 17 year period with fitted Fourier model shown as the solid line; (**b**) monthly averaged precipitation with Fourier model; (**c**) LogFC_m with fitted Fourier model and (**d**) PTR versus month with fitted Fourier model.

The non-uniform sampling schedule can lead to an inconsistent number of samples collected in each month, which can lead to a skewed data set in favour of times in the year when samples are taken [51]. As in [51], the positive test ratio (*PTR*) is considered in order to screen out the influence of the number of tests taken in any given month or season. *PTR* is the number of cases above the regulation limit 200 CFU/100 ml divided by the number of total cases in each month of the 17 years. Figure 6 shows the P_m, T_m, *PTR* and LogFC_m data with fitted Fourier models. Information on the fitted models are given in Table 1.

Table 1. Fourier Coefficients.

Model	Variable and Fourier Model	Adj. R^2	RMSE
1	Monthly temperature ($t = y, m$): = 10.72 + 0.03cos(0.07t) + 0.41sin(0.07t) − 0.19cos(0.14t) − 0.16sin(0.14t) − 0.12cos(0.21t) + 0.05sin(0.21t) − 0.21cos(0.28t) + 0.12sin(0.28t) + 0.10cos(0.35t) + 0.08sin(0.35t) − 0.14cos(0.42t) − 0.07sin(0.42t) − 0.15cos(0.49t) − 0.21sin(0.49t) + 4.261cos(0.56t) + 3.648sin(0.56t)	0.93	1.13 °C
2	Monthly precipitation ($t = y, m$): = 40.6 − 9.273cos(0.48t) + 13.32sin(0.48t)	0.92	15.17 mm
3	Log$FC_{y,m}$ = 3.547 + 0.11cos(0.52t) + 0.11sin(0.52t)	0.92	0.13
4	Mean monthly temperature ($t = m$) = 10.96 − 3.41cos(0.56t) − 4.371sin(0.56t) −0.52cos(1.12t) + 0.26sin(1.12t)	0.99	0.16 °C
5	Mean monthly precipitation ($t = m$) = 59.06 + 48.6cos(0.59t) + 3.97sin(0.59t) + 24.07cos(1.18t) + 7.43sin(1.18t) + 13.69cos(1.77t) + 5.3sin(1.77t)	0.92	13.95 mm
6	Log FC_m = 3.44 − 0.6cos(0.6t) − 0.12sin(0.6t) + 0.11cos(1.2t) + 0.01sin(1.2t) + 0.09cos(1.8t) + 0.02sin(1.8t)	0.69	0.08
7	PTR = 0.59 + 0.01cos(0.55t) + −0.07sin(0.55t) + 0.08cos(1.1t) − 0.02sin(1.1t) + 0.02cos(1.65t) − 0.03sin(1.65t)	0.67	0.05

In interpreting these values, what's important is the fundamental frequency ω in the dominant portion of the series. Cosine and sine terms are dominant if their coefficients are higher than other cosine and sine terms. For example, in Table 1, models 2 and 3 for precipitation and Log$FC_{y,m}$, respectively, have very similar fundamental frequencies (0.48, 0.52, respectively), which translate into both signals having a roughly 12 month periodicity in the signal. The fitted model shown in Figures 5a and 5b, respectively show the simple wave with a high R^2 value suggesting that while the peaks and troughs could not be captured in either model, both signals have this frequency. With regard to model 1, which is the 8 term Fourier model of temperature, the curves as seen in Figure 5a is nearly perfectly captured but model 1 suggests that the fundamental frequency is 0.07. However, the dominant terms in this expression are the last sine and cosine terms as they have the largest coefficients. The frequency of those terms is 0.56, again illustrating the annual periodic signal in the temperature. Since all fundamental frequencies are essentially equal, the analysis does not preclude the suggestion that Log$FC_{y,m}$ signal characteristics arise from precipitation and temperature signals.

Similar profiles are seen in Figure 6a,b for mean monthly temperature and precipitation over the entire 17 year period. Models 4 and 5 show fundamental frequencies of roughly 0.6, corresponding to a roughly annual periodicity that is clearly shown in the two panels. Models 6 and 7, for LogFC_m, and PTR, respectively, also show periodic signals of roughly 12 months; however, the signal peaks do not mimic either of those made by precipitation or temperature. Figure 6c shows two peaks—one in July and one in October. While both the PTR and LogFC_m curves show a nearly identical curve in the fitted Fourier models, there is a glaring discrepancy in the month of May (shown circled in red in Figure 6c). Recall that the sampling program is focused on two seasons: January to April and June to September. May is not usually sampled. The PTR curve seems to suggest that the data point shown in May is an outlier requiring further consideration in the data. The models also suggest that more than just precipitation or just temperature are responsible for the oscillations seen in the curves.

The climate in southern Vancouver Island can be summarized as having less precipitation with moderate temperatures in the summer and more precipitation with cooler temperatures in the winter.

Both the minimum and maximum temperature over most of the year are not sufficient to kill the fecal coliforms since the fecal coliform could grow at temperatures between 4 °C to 35 °C and grow quickly around 20 °C [56]. The general inspection indicates temperature is one of the most important factors that cause high fecal coliform levels in the summer. During the winter time, the temperature is relatively low but the precipitation and cloud cover ratio is high and these provide moisture for fecal coliform growth and less possibility of mortality by sunlight.

The results also suggest a different seasonal pattern in fecal coliform levels from a regular monsoon climate pattern in which the level of fecal coliforms is high in summer and low in winter. The high peaks of fecal coliform concentration appear twice during the year; once is in summer around July and once in winter around October. The lowest level of fecal coliform is found in February as well as for the positive test ratio. It is apparent that the mild Mediterranean climate in Victoria could result in a relatively unusual distribution of bacterial contamination compared to other studies conducted in the world. The positive test ratio of high fecal coliform concentration is even found to be highest in October instead of July. There is also a high level of fecal coliform appearing in the samples collected in October. This finding indicates the monitoring schedule should expand outside its two season monitoring approach.

3.2. Correlation Analysis with Temporally and Spatially Related Variables in $LogFC_{i,j}$

Kendall's τ-b test results are listed in Table 2 and only those tests that returned a p-value < 0.05 are shown. In general, temperature variables and antecedent dry period have a positive correlation with the $LogFC_{i,j}$. The result shows a relatively high temperature would help fecal coliform growth in the summer and accumulate in the drainage area while no precipitation occurred. Conversely, precipitation variables have a negative correlation with the $LogFC_{i,j}$, which is unexpected as moisture helps bacteria grow and much of the literature showed an opposite correlation. The reason could be that heavy precipitation continuously washes off the bacteria and dilutes the density of fecal coliform during the wet season. The flow rate at discharge is largely negatively correlated with the fecal coliform levels, which, like precipitation suggests that higher flow rates of stormwater are diluting contaminants to the discharge. The data collection methods by the CRD intentionally avoided first flush timing, which would mean that only the late stages of the event are sampled.

Table 2. Non-parametric correlation Kendall τ-b of $LogFC_{i,j}$ for all stations and at each hot spot. The symbol '*' indicates the numeric value of τ-b has a p-value less than 0.05 and '**' indicates a p-value less than 0.01. Symbols '+' and '−' indicate the correlation was positive or negative, respectively, but is not significant.

Variable	All Stations	#805	#641	#623	#320	#245
T_{mean}	0.04 **	0.24 *	+	+	0.54 **	+
T_{min}	0.04 **	0.24 *	+	+	0.54 **	+
T_{max}	0.05 **	+	+	+	0.51 **	0.15 *
$T2$	0.06 **	+	+	+	0.51 **	0.16 *
$T3$	0.06 **	+	+	+	0.51 **	0.22 **
$T7$	0.07 **	+	0.10 *	+	0.51 **	0.17 *
P	−0.01 **	+	0	−	−0.19 *	−0.20 *
$P2$	−0.05 **	−	−	−	−0.21 *	−0.18 *
$P3$	−0.06 **	−0.23 *	-	-	−0.25 **	−0.25 **
$P7$	−0.06 **	−0.33 **	-	-	−0.38 **	−0.16 *
t_{dry}	0.06 *	+	+	-	0.26 **	0.22 *
FR	−0.05 **	−0.26 *	0.10 *	-	−0.23 **	−0.26 **
WA	+					
RA	0.11 **					
GA	+					

With regard to the land use, the residential area has a positive correlation with Log$FC_{i,j}$ which confirms the previous hypothesis that the bacteria contamination is mainly caused by human activities. The literature suggests that the watersheds with a high percentage of urban and agricultural land use mostly have high contamination levels [13]. The watershed area and the greenspace area both showed a positive but insignificant correlation suggesting that greenspace and the watershed as a whole are also contributing to the amount of fecal coliform observed. The GIS analysis shows the heavily contaminated areas are in Victoria downtown, Esquimalt and the southeastern shore of Oak Bay are densely populated and contain high percentages of impervious land use. It is reasonable to suggest that in this region, high levels of bacterial contamination are likely to be observed in populated residential areas with high percentages of impervious land use.

Table 2 also shows the correlations for selected hotspots. Station #320 showed a very strong correlation with all variables; both temperature and precipitation being significant variables with many p-values around 0.000. Thus, at least for this station, LogFC is greatly affected by meteorological changes. This seems to be also true for station #805. The correlations shown conform to the correlation analysis of all samples from all stations: the temperature and antecedent dry period are positively correlated to LogFC while the precipitation has a negative correlation. However, station #641 showed only one significant positive relationship with temperature and the opposite relationship with flowrate. While station #623 showed correlations that were in keeping with the general trend for all stations, none of them were significant. Since this sampling station is located in the nearshore of Victoria's downtown area, the distribution of fecal coliform could be dominated by spatial variables (land use) rather than temporal variability. Station #320 shows that all temperature variables are positively correlated with LogFC at this sampling station and in fact, all of the relationships are significant and strong. All the temperature variables are correlated so it is not surprising that all the temperature correlations are negative, near 0.5 and all very significant (p-value near zero). However, there is next to no difference amongst the correlations and thus, no one variable distinguishes itself in its influence on LogFC than any other temperature variable. The precipitation shows a negative correlation with LogFC while the antecedent dry period has a positive correlation. The 7-day total precipitation provides the strongest negative correlation suggesting that long-term precipitation could greatly reduce the level of fecal coliforms in the area draining to this station. Station #245 shows similar trends but now the 3-day precipitation variable $P3$ has the strongest correlation for all the precipitation variables studied (i.e., P, $P2$, $P3$, $P7$). This variability in correlation strength across temperature or precipitation variables, and the lack of significance in the group of hotspot stations selected, suggests incredible variation by location in the database and that analysis on climatic/land use influences on FC or potentially prediction of FC levels should be done on a station by station basis.

The literature on fecal bacteria growth and die-off rates in the environment [56,57] suggest that the summer temperature in the study area (under 20 °C) is considered to be ideal for fecal bacteria growth. Since the maximum temperature is rarely over 30 °C, the die-off rate of fecal coliform is also low in the study area. McCarthy et al. [27] suggested that temperature was an important variable such that E. coli could be persisting or growing in the catchment during periods of warmer temperature, and this may be applicable in this study region.

The correlation analysis shows that precipitation is negatively correlated with fecal coliform concentration, but the relationship is weaker than that for temperature. The negative correlation contradicts some of the literature [58] but seasonal dependence is important [26,30]. The 7-day total precipitation has the strongest negative correlation coefficient among the precipitation variables in two of the stations, which may suggest that long-term rainfall could further decrease the fecal coliform level of the area. The southern Vancouver Island region usually has light and long-term precipitation during the wet season. With regard to the positive correlation of LogFC and antecedent dry period, this could be due to long inter-event periods allowing significant accumulation of fecal coliform in the drainage area. When storms occur, because they are not sampled at the first flush, there will be dilution of the fecal coliform concentration. However, one needs to note that the chances of obtaining a

sample with over the limit bacteria concentrations is even higher during the winter (as noted by the curve of PTR in Figure 6d) even if the average value of fecal coliform is lower than in the summer. This indicates that the minimum temperature (approximately 3 °C) in the study area is not sufficient to inactivate fecal bacteria, and the bacterial contaminants could be transported by stormwater to the discharge in the cooler, wetter season.

Table 3 shows the correlations when comparing Log$FC_{y,m}$ and the available temporal datasets including cloud cover *CC*. This table shows that all the temperature variables and the cloud cover variable have a significant influence on Log$FC_{y,m}$ with T_{min} having the strongest correlation of the variables, albeit, the difference is very small. This would suggest that temperature could be a useful predictor of Log*FC* for general planning purposes on an annual basis, or in climate change studies.

Table 3. Non-parametric correlation Kendall τ-*b* of Log$FC_{y,m}$ and *p*-values.

Statistics	T_{min}	T	T_{max}	P	CC
τ-*b* correlation	0.122	0.121	0.119	−0.092	−0.117
p-value	0.024	0.025	0.027	0.088	0.037

3.3. Climate Related Modelling

3.3.1. Multiple Linear Regression

Through the previous correlation analysis process, the fecal coliform value has shown a significant relationship with meteorological variability, although the correlations are weak. The periodicity modelling showed that there are issues raised by PTR versus the month averaged over all 17 years. Figure 6c is the data with a Fourier model superimposed on the LogFC_m data showing a circled May data point. This data point in Figure 6d corresponds to the PTR data which shows no uncharacteristic dip but follows the 3 term Fourier model much like the LogFC_m Fourier model. Since the data sampling period strongly skews data in favour of the months of January to April (henceforth, referred to as Season 1) and June to September (Season 2), the month of May does not have a lot of data points but interestingly, the data point in Figure 6c for May seems uncharacteristically out of place. The months of October to December have some data but are also sparse over the 17 year period. Therefore, any further climate analysis examines the data for Log$FC_{i,j}$ with the months of May, and October to December removed. The database of the full 17 year period now spans 2181 data points down from 2365 (after the sewage suspected samples were removed as indicated earlier).

The goal of this section is to develop a general function for predicting the Log*FC* with selected meteorological variables. The significant but low correlations suggest that attempting to linearly regress all of Log$FC_{i,j}$ with one or more climate variables may likely show a significant relationship but the R^2 would likely be poor. Figures 7a, 7b and 7c show the data as a function of mean temperature for all Log$FC_{i,j}$, Log$FC_{y,m}$ and LogFC_m, respectively. These scatterplots indicate the variance in the data that a linear regression model must attempt to explain. Thus, LR was only attempted on LogFC_m.

The LR modelling was conducted through SPSS and uses Akaike information criterion to avoid potential overfitting and only chose the most significant variables. The variables tested were all those listed in Table 2 (with the exception of the watershed information, and the flowrate when it was not possible) and cloud cover (*CC*) was also included. The variables providing the best fit are T_{min} and *CC* with an R^2 of 0.55 and an adjusted R^2 of the model shown in Equation (5) as 0.45. The RMSE = 0.109 and the significance of the T_{min} and *CC* coefficients are 0.02, and 0.07, respectively.

$$\text{Log}FC = 2.05 + 0.08T_{min} + 1.41CC. \quad (5)$$

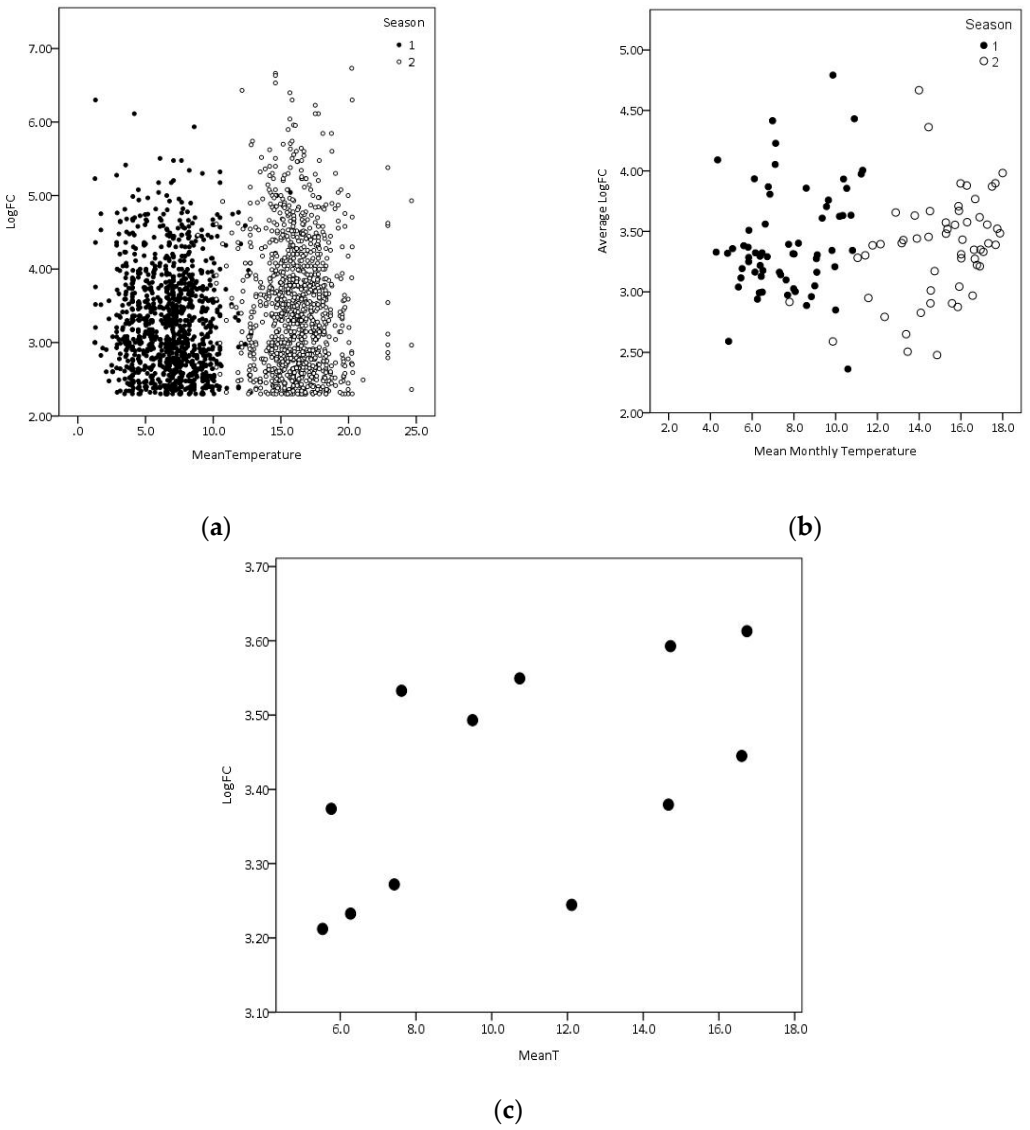


Figure 7. Scatterplots of (**a**) Log$FC_{i,j}$ versus temperature of the day of sampling at each location; (**b**) Log$FC_{y,m}$ versus mean monthly temperature; and (**c**) $LogFC_m$ versus mean monthly temperature over the entire period in each month.

Notice that in Equation (5), the coefficient multiplied by *CC* is positive not negative. When no weather or obstructions to visibility occur, sky conditions are provided reflecting the observation of total cloud amount. The value of *CC* based on the amount (in tenths) of cloud covering the dome of the sky with 0 being clear and 1.0 being completely clouded over sky. While higher cloud clover might suggest greater opportunity for precipitation, which is negatively correlated with FC in this region, in the absence of precipitation, cloud cover would prevent solar radiation from promoting bacteria die-off. This concurs with [29] that stated that higher solar radiation lead to lower bacteria levels.

While the R^2 is only fair, the model coefficient for T_{min} is significant and nearly significant for *CC*. This is not surprising given the results of the correlation analysis. In addition, some of the principles behind the non-parametric correlation analysis are similar to some of the assumptions underlying multiple linear regression. The Fourier analysis suggested a seasonal influence on the data and the differences in correlations in Table 2 suggest that location in space is important. Thus, averaging across space and over the year, as is done for LogFC_m in this analysis, would produce large errors in any prediction model. Figure 8 shows plots of the residuals in the predict Log$FC_{y,m}$ values. The residuals for T_{min} are scattered over the temperature range experienced over the year and there is no trend that

warmer or cooler temperatures lead to better predictions. This is also true of cloud cover in that the residuals are scattered around 0 for the full range of *CC*.

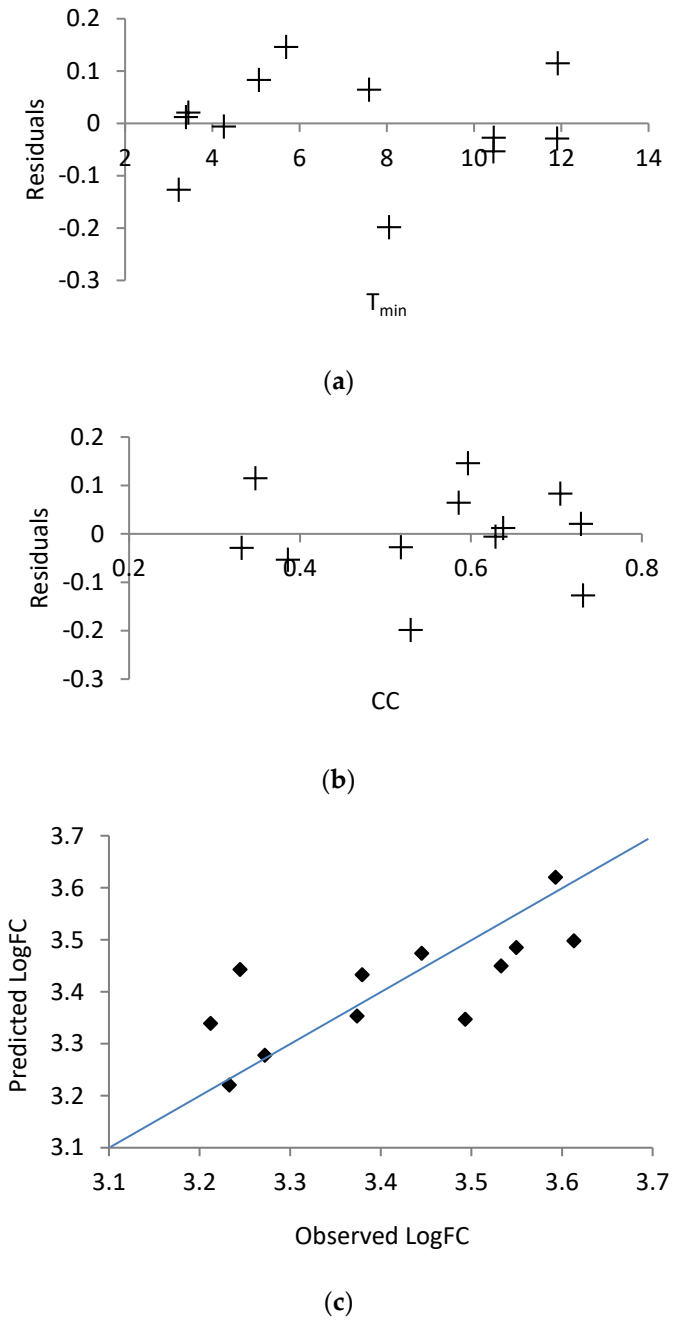


Figure 8. Residuals as a function of (**a**) T_{min}, and (**b**) *CC*, and (**c**) trend of predicted versus observed LogFC_m with one to one line shown in blue.

3.3.2. Multinomial Logistic Regression in Log$FC_{y,m}$

Tables 4–6 show the results of the MLR on Log$FC_{i,j}$, Log$FC_{i,j}$ at the hotspots and LogFC_m, respectively. Each variable is tested for its significance in explaining the likelihood of LogFC belonging to a certain order of magnitude (class), and then the percentage of model performance for each class is computed. The shaded cells indicate that the predictor had a significant influence on the overall model performance. The model performances are calculated as a weighted performance in each class. If no value of performance is shown for a class that means that there were no observations for that class in

the data set. The number of data points in the set are indicated by N and whether or not the model fit is significant or not is shown with shading (no shading indicates it is not significant).

Table 4. MLR modelling results for Log$FC_{i,j}$ for all stations. A shaded box indicates a significant relationship with no shading indicating the p-value was greater than 0.05.

	Period	**Season**					**Month**					
		All	**1**	**2**	**1**	**2**	**3**	**4**	**6**	**7**	**8**	**9**
	N	2181	1038	953	307	424	216	91	157	419	357	210
Qual-ity	Model Fit											
	Pseudo-R^2	0.06	0.10	0.06	0.17	0.13	0.31	0.49	0.34	0.19	0.21	0.32
Predictor Significance	Intercept											
	Flow Rate											
	T											
	T_{min}											
	T_{max}											
	$T2$											
	$T3$											
	$T7$											
	P											
	$P2$											
	$P3$											
	$P7$											
	t_{dry}											
Class of LogFC	Class	% Performance										
	2	41	57	42	59	71	68	58	53	45	55	65
	3	64	52	70	56	47	51	67	72	69	62	35
	4	0	0	4	10	0	16	13	128	16	17	47
	5	0	0	0	0	0	0	100	0	0	5	33
	6	0	100	0		100	100		100	0	100	100
	Overall	41	46	42	52	50	51	56	48	44	47	50

In examining Table 4 closely, for all the 2181 records of Log$FC_{i,j}$, flowrate and seven day average temperature and flowrate were found to significantly influence the likelihood of observing an order of LogFC but the overall model performance was only 41%. When examining the difference introduced by seasonality, Season 1 data provided a better model performance than Season 2 and was able to perfectly predict the values of order of magnitude 6 and higher. However, it could only predict values in order 2 and 3, which were the bulk of the database with only roughly 50% performance. Here again flowrate was significant but now so was temperature T and $T2$ and P and $P2$. Season 2 showed a very good performance for order 3—better than in Season 1 but now only $T3$ and $T7$ were significantly influencing results indicating that only temperature matters in the warmer season. This is not surprising as this is also the drier season. However, flowrate was still significantly influencing predictions in Season 2 as well. This however changed when focusing on the monthly performance.

The analysis by month showed that flowrate was significant in August and September—two of the four months of the warm season but the peak in FC levels occurs in July which does not have flowrate significantly influencing the model performance. The month of February showed absolutely no significant influence of any variable. The peak month of July showed that maximum observed temperature and precipitation on the day or within two days of sampling were significantly influencing the model results. July also had the lowest performance as compared to the other months in Season 2, which were all able to predict occurrences of order 6 and higher in Log$FC_{i,j}$ at 100%. Interestingly, September had the best spread of performances in each class for Season 2 and here flowrate and seven day accumulation of precipitation were key to that performance. The best performance overall was for the month of April in which, flowrate, $T3$, $T7$, and t_{dry} were key. The Pseudo-R^2 values were only above 0.2 in the months of March, April, June, August and September, again suggesting that monthly time scales for examining climate influences is warranted.

Table 5. MLR modelling results for Log$FC_{i,j}$ for the hotspot stations. A shaded box indicates a significant relationship with no shading indicating the *p*-value was greater than 0.05. The letter "B" indicates the relationship is not significant but the *p*-value was very close to 0.05.

Period	**Season**			**Month**							
	All	**1**	**2**	**1**	**2**	**3**	**4**	**6**	**7**	**8**	**9**
N	210	99	111	33	28	30	8	22	33	29	27
Model Fit		B								B	
Pseudo R^2	0.52	0.60	0.66	0.96	0.94	0.93	0.86	0.97	0.94	0.92	0.54
Intercept											
Flow Rate										B	
T											
T_{min}											
T_{max}											
*T*2					B						
*T*3											
*T*7											
P											
*P*2											
*P*3											
*P*7			B								
t_{dry}											
Class	**Performance in (%)**										
2	100	100	100	100		0					0
2.70	6	20	17	100	100	67		100	0	0.5	100
3	38	50	39	100	100	71	100	100	100	100	100
3.70	54	57	66	100	83	25	100	100	83	92	71
4	21	44	33	100	100	83	100	100	88	50	25
4.70	50	47	48	100	100	86		100	80	80	100
5	38		36					100	100	100	67
5.70	100		50					100			100
6	75	100	100			100			100		
6.70	100		100					100			
Overall	41	50	48	100	96	70	100	100	84	86	74

Table 6. MLR modelling results for Log$FC_{y,m}$. A shaded box indicates a significant relationship and no shading indicates the *p*-value was greater than 0.05.

	All	**Season 1**	**Season 2**
N	116	59	57
Model Fit			
Pseudo R^2	0.40	0.29	0.52
Intercept			
T_{min}			
T			
T_{max}			
P			
CC			
Class	% Performance (*N* in class)		
2.7	0	100 (1)	0 (2)
3	77	95 (41)	52 (21)
3.7	60	27 (11)	77 (26)
4	8	0 (6)	33 (6)
4.7	100		100 (2)
Overall Performance	63	73	61

Table 5 shows the results for the hotspot stations. Recall from Table 2, only two to three of the five hotspot stations had climatic variables and land use variables with significant effects on Log$FC_{i,j}$. In Table 5, the overall performance for all the data, Seasons 1 and 2, are similar to the entire dataset but now, for the whole set, flowrate and precipitation are influencing results as opposed to flowrate and temperature. The only variable influencing performance in Season 1 is seven day precipitation and only marginally in Season 2, in which the model fit was not significant. A more granular analysis by month showed that the two months with high levels of FC, June and July were nicely modelled with flowrate and $T3$ providing a perfect model performance at all classes in the month of June. However, the month of July, had a poorer performance but flowrate was not a factor but other temperature variables, precipitation variables and dry days were statistically relevant. The Cox and Snell Pseudo-R^2 values are all greater than 0.2 suggesting good model performances in all variables used in the model.

Table 6 shows the results for the 116 data points without May, October–December values seen in Figure 5b. Again, the data were also examined by season but data by month were sparse, which would lead to falsely inflated values of model performance and thus, were not examined.

The model performance overall is the best seen for all data sets which is not surprising. As well, while Season 1 had the best performance, the fit was not significant and no single predictor had a significant influence. The overall model performance in Season 2 was 61% with good performance in the middle and higher order classes. Here all the variables were significant with the exception of cloud cover, which proved useful in the multiple linear regression modelling for the data set of Figure 6c (monthly averages over the entire database LogFC_m). In the logit model of Equation (3), the model coefficients β are specific to each class.

To illustrate the model implications on the relationship between increases or decreases in fundamental variables, two classes are created for this dataset: Log$FC < 3.7$ assigned to order 3 and those ≥ 3.7 are assigned to order 4. The results are shown in Table 7. In this analysis, category 4 was the reference category, and P and T were the selected variables. Unlike the LR modelling of Section 3.3.1, CC had no impact on the results. As T_{min} and T_{max} are certainly related to T, and to understand the general influence of T and P as two distinct climatic variables on Log$FC_{y,m}$, these two variables were selected for the analysis in this illustration.

Table 7. MLR modelling results for Log$FC_{y,m}$ with only classes 3 and 4 and variables, T and P.

N	116
Model Fit Significant?	Y
Goodness of Fit Significant?	Y
Pseudo R^2	0.12
Intercept β_0	−0.614
β of T	0.078
β of P	−0.013
3	68 (65)
4	67 (51)
Overall Performance	67%

In terms of the logistic model, these coefficients suggest that,

$$\text{logit}(\pi) = -0.614 + 0.078T - 0.013P \quad (6)$$

In Equation (6) each coefficient is the expected change in the log odds of observing a LogFC of magnitude 3 for a unit increase in a predictor variable when the other predictor variables are fixed. The value of e^{β} for a specific predictor is the change in odds (in the multiplicative scale) for a unit increase in that predictor. The intercept, in this case, is the log odds of observing a value of order

of magnitude 3 if temperature and precipitation are zero because that was chosen as the reference category in this analysis. This fitted model says that if precipitation is held fixed, the coefficient for temperature suggests that for a unit increase in temperature (1 °C) would result in an increase in the odds of observing a Log*FC* with order of magnitude of 4 by 8% (since $e^{0.078} = 1.08$). Conversely, if temperature is held fixed, a unit increase in precipitation (1 mm) would result in a decrease in the odds of observing an order of magnitude 4 by 2%.

This analysis reaffirms the suggestion that in this study region, temperature increases are likely to increase Log*FC* levels and precipitation likely lowers Log*FC* levels. The results in Tables 6 and 7 all involve a spatial averaging over the study area, which has already suggested possible errors in modelling. Vermuelen and Hofstra [31] noted that temperature was a significant variable in four of six locations studied yet leaving it out of their overall model did not make a "great difference" for a model with an adjusted R^2 of 0.49. The implication of spatial influences was further iterated by Mallin et al. [35] who found that multiple regression models failed to explain any more of the variability in bacteria than simply using percent impervious surface cover alone. Seasonality implications were clearly visible in the results. In Selvakumar and Borst [59], the authors found seasonal variations in the data in which early season storms generally had higher levels of bacteria than late season storms, both within and between watersheds. But when spatial implications were ignored and all watersheds were lumped together, bacteria levels decreased with increasing cumulative annual rainfall. This suggests that if detailed modelling or planning is going to be conducted, it should not only be conducted with attention to location, but attention to season.

4. Conclusions

This study provides insight into the temporal and spatial distribution of bacterial contamination in nearshore areas of southern Vancouver Island and explains the potential relationships in fecal coliform levels in stormwater with land-use and climate variables. Non-parametric correlation analysis, Fourier analysis, multiple linear regression and multivariate logistic regression were used to determine the significance and relative influence of climatic variables on Log*FC* levels measured between 1995 and 2011 at three temporal scales: at the time of sampling, the monthly average of Log*FC* across all stations in the region, and the average across space and time for each month.

The Fourier model demonstrates an annual periodicity of fecal coliform with two peaks in July and October. Despite the extreme values observed in the 17 years' historical data, the data follow the general trend of this periodicity as do the precipitation and temperature. No matter the temporal scale used, the Fourier analysis always produced a periodicity of 12 months.

The correlation analysis showed that fecal coliform concentrations for all stations throughout the period have a significant correlation with many climatic variables including positive correlations with temperature, antecedent dry period and land use in the watershed, and negative correlations with precipitation and flowrate. Drainage area showed no significant correlations with bacterial contamination. But drainage area, amount of residential area in the drainage area as well as green space were all positively correlated with FC levels but only the amount of residential, or urban area, was significantly correlated. This suggests that the bacteria loading mostly originates from residential activity. The incredible variability in whether relationships were significant or not suggest that modelling or planning should be done on a location basis and not necessarily averaged across the region. Seasonal effects were also apparent and planning and prediction should be conducted differently in the two seasons. In particular, the drier, hotter season had higher levels of *FC* and therefore, should be the focus of future planning.

Linear regression developed a prediction model for Log*FC* in each month with a low adjusted R^2 of under 0.5, but with significant or marginally significant variables of minimum monthly temperature and cloud cover. Multivariate logistic regression was used with classes representing the order of magnitude of Log*FC*. This type of regression showed very good model performance, which improved when considering the wet season separately from the dry season, and when looking at a individual months.

The authors recommend that when examining the influences of climate change on fecal coliforms, and possibly, on other types of bacterial contaminants in this region, that variables be distinguished by season, and possibly by month, as well as by location. In addition, the authors recommend that the CRD monitor FC more intensively at every month of the year, and in the fall season that includes October. The analysis should be updated whenever new data are made available. Since the land use was only represented by only two different types of land use (residential versus greenspace), the study should be expanded to other areas of the region that include forested and agricultural lands. The MLR model could be easily applied on the updated and expanded datasets.

Author Contributions: Conceptualization, C.V.; Methodology, C.V. and K.X.; Formal Analysis, K.X. and Z.X.; Investigation, K.X.; Data Curation, K.X. and Z.X.; Writing-Original Draft Preparation, K.X.; Writing-Review & Editing, C.V. and J.H.; Supervision, C.V. and J.H.

Funding: This research received no external funding.

Acknowledgments: The Authors would like to thank Natalie Bandringa and Dale Green of the Capital Regional District for the support of this work; Barri Rudolph of the CRD for supplying data; and Joachim Carolsfeld of World Fish Organization for his support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bowen, R.; Depledge, M. Rapid Assessment of Marine Pollution (RAMP). *Mar. Pollut. Bull.* **2006**, *53*, 631–639. [CrossRef]
2. Stormwater, Harbours and Watersheds Program Environmental Sustainability. In *Core Area Stormwater Quality 2012 Annual Report*; Capital Regional District: Victoria, BC, Canada, 2013; Available online: http://www.crd.bc.ca/about/what-we-do/stormwater-wastewater-septic/monitoring-stormwater (accessed on 1 January 2014).
3. Curriero, F.; Patz, J.; Rose, J.; Lele, S. The Association Between Extreme Precipitation and Waterborne Disease Outbreaks in the United States, 1948–1994. *Am. J. Public Health* **2001**, *91*, 1194–1199. [CrossRef]
4. Gaffield, S.; Goo, R.; Richards, L.; Jackson, R. Public Health Effects of Inadequately Managed Stormwater Runoff. *Am. J. Public Health* **2003**, *93*, 1527–1533. [CrossRef] [PubMed]
5. Ahmed, W.; Goonetilleke, A.; Gardner, T. Human and bovine adenoviruses for the detection of source-specific fecal pollution in coastal waters in Australia. *Water Res.* **2010**, *44*, 4662–4673. [CrossRef] [PubMed]
6. Frenzel, S.; Couvillion, C. Fecal-indicator bacteria in streams along a gradient of residential development. *J. Am. Water Resour. Assoc.* **2002**, *38*, 265–273. [CrossRef]
7. Campos, C.; Cachola, R. Faecal Coliforms in Bivalve Harvesting Areas of the Alvor Lagoon (Southern Portugal): Influence of Seasonal Variability and Urban Development. *Environ. Monit. Assess.* **2007**, *133*, 31–41. [CrossRef] [PubMed]
8. Stocker, M.; Rodriguez-Valentin, J.; Pachepsky, Y.; Shelton, D. Spatial and temporal variation of fecal indicator organisms in two creeks in Beltsville, Maryland. *Water Qual. Res. J. Can.* **2016**, *51*, 167–179. [CrossRef]
9. Davis, K.; Anderson, M.; Yates, M. Distribution of indicator bacteria in Canyon Lake, California. *Water Res.* **2005**, *39*, 1277–1288. [CrossRef] [PubMed]
10. Sibanda, T.; Chigor, V.; Okoh, A. Seasonal and spatio-temporal distribution of faecal-indicator bacteria in Tyume River in the Eastern Cape Province, South Africa. *Environ. Monit. Assess.* **2012**, *185*, 6579–6590. [CrossRef]
11. Hunter, C.; Perkins, J.; Tranter, J.; Gunn, J. Agricultural land-use effects on the indicator bacterial quality of an upland stream in the Derbyshire peak district in the UK. *Water Res.* **1999**, *33*, 3577–3586. [CrossRef]
12. Crowther, J.; Kay, D.; Wyer, M. Faecal-indicator concentrations in waters draining lowland pastoral catchments in the UK: Relationships with land use and farming practices. *Water Res.* **2002**, *36*, 1725–1734. [CrossRef]
13. Tong, S.; Chen, W. Modeling the relationship between land use and surface water quality. *J. Environ. Manag.* **2002**, *66*, 377–393. [CrossRef] [PubMed]

14. Traister, E.; Anisfeld, S. Variability of Indicator Bacteria at Different Time Scales in the Upper Hoosic River Watershed. *Environ. Sci. Technol.* **2006**, *40*, 4990–4995. [CrossRef] [PubMed]
15. Bolstad, P.; Swank, W. Cumulative impacts of landuse on water quality in a southern Appalachian watershed. *J. Am. Water Resour. Assoc.* **1997**, *33*, 519–533. [CrossRef]
16. Chu, Y.; Tournoud, M.; Salles, C.; Got, P.; Perrin, J.; Rodier, C.; Caro, A.; Troussellier, M. Spatial and temporal dynamics of bacterial contamination in South France coastal rivers: Focus on in-stream processes during low flows and floods. *Hydrol. Process.* **2013**, *28*, 3300–3313. [CrossRef]
17. Wu, J.; Rees, P.; Dorner, S. Variability of *E. Coli* Density and Sources in an Urban Watershed. *J. Water Health* **2011**, *9*, 94–106. [CrossRef]
18. Henry, R.; Schlang, C.; Coutts, S.; Kolotelo, P.; Prosser, T.; Crosbie, N.; Grant, T.; Cottam, D.; O'Brien, P.; Deletic, A.; et al. Into the deep: Evaluation of SourceTracker for assessment of faecal contamination of coastal waters. *Water Res.* **2016**, *93*, 242–253. [CrossRef]
19. Jent, J.R.; Ryu, H.; Toledo-Hernández, C.; Santo Domingo, J.W.; Yeghiazarian, L. Determining Hot Spots of Fecal Contamination in a Tropical Watershed by Combining Land-Use Information and Meteorological Data with Source-Specific Assays. *Environ. Sci. Technol.* **2013**, *47*, 5794–5802. [CrossRef]
20. Liang, Z.; He, Z.; Zhou, X.; Powell, C.A.; Yang, Y.; He, L.M.; Stofella, P.J. Impact of Mixed Land-Use Practices on the Microbial Water Quality in a Subtropical Coastal Watershed. *Sci. Total Environ.* **2013**, *449*, 426–433. [CrossRef]
21. Tiefenthaler, L.; Stein, E.D.; Schiff, K.C. Levels and Patterns of Fecal Indicator Bacteria in Stormwater Runoff from Homogenous Land Use Sites and Urban Watersheds. *J. Water Health* **2011**, *9*, 279–290. [CrossRef]
22. Walters, S.P.; Thebo, A.L. *Water Res.* **2010**, *45*, 1752–1762. [CrossRef] [PubMed]
23. Wu, J.; Yunus, M.; Islam, M.S.; Emch, M. Influence of Climate Extremes and Land Use on Fecal Contamination of Shallow Tubewells in Bangladesh. *Environ. Sci. Technol.* **2016**, *50*, 2669–2676. [CrossRef] [PubMed]
24. Cha, Y.; Park, M.H.; Lee, S.H.; Kim, J.H.; Cho, K.H. Modeling Spatiotemporal Bacterial Variability with Meteorological and Watershed Land-Use Characteristics. *Water Res.* **2016**, *100*, 306–315. [CrossRef] [PubMed]
25. Paule-Mercado, M.A.; Ventura, J.S.; Memon, S.A.; Jahng, D.; Kang, J.H.; Lee, C.H. Monitoring and Predicting the Fecal Indicator Bacteria Concentrations from Agricultural, Mixed Land Use and Urban Stormwater Runoff. *Sci. Total Environ.* **2016**, *550*, 1171–1181. [CrossRef]
26. Cho, K.H.; Pachepsky, Y.A.; Kim, J.H.; Guber, A.K.; Shelton, D.R.; Rowland, R. Release of *Escherichia Coli* from the Bottom Sediment in a First-Order Creek: Experiment and Reach-Specific Modeling. *J. Hydrol.* **2010**, *391*, 322–332. [CrossRef]
27. McCarthy, D.T.; Hathaway, J.M.; Hunt, W.F.; Deletic, A. Intra-Event Variability of *Escherichia Coli* and Total Suspended Solids in Urban Stormwater Runoff. *Water Res.* **2012**, *46*, 6661–6670. [CrossRef]
28. Malin, M.; Cahoon, L.; Parsons, D.; Ensign, S. Effect of Nitrogen and Phosphorus Loading in Coastal Plain Blackwater Rivers. *J. Freshw. Ecol.* **2001**, *16*, 455–464. [CrossRef]
29. Cho, K.H.; Cha, S.M.; Kang, J.H.; Lee, S.W.; Park, Y.; Kim, J.W.; Kim, J.H. Meteorological Effects on the Levels of Fecal Indicator Bacteria in an Urban Stream: A Modeling Approach. *Water Res.* **2009**, *44*, 2189–2202. [CrossRef]
30. Jung, Y.; Park, Y.; Lee, K.; Kim, M.; Go, K.; Park, S.; Kwon, S.; Yang, J.; Mok, J. Spatial and seasonal variation of pollution sources in proximity of the Jaranman-Saryangdo area in Korea. *Mar. Pollut. Bull.* **2017**, *115*, 369–375. [CrossRef]
31. Vermeulen, L.C.; Hofstra, N. Influence of Climate Variables on the Concentration of *Escherichia Coli* in the Rhine, Meuse, and Drentse Aa during 1985–2010. *Reg. Environ. Chang.* **2014**, *14*, 307–319. [CrossRef]
32. Martinez-Urtaza, J.; Saco, M.; de Novoa, J.; Perez-Pineiro, P.; Peiteado, J.; Lozano-Leon, A.; Garcia-Martin, O. Influence of Environmental Factors and Human Activity on the Presence of Salmonella Serovars in a Marine Environment. *Appl. Environ. Microbiol.* **2004**, *70*, 2089–2097. [CrossRef] [PubMed]
33. Patz, J.; Vavrus, S.; Uejio, C.; McLellan, S. Climate Change and Waterborne Disease Risk in the Great Lakes Region of the U.S. *Am. J. Prev. Med.* **2008**, *35*, 451–458. [CrossRef] [PubMed]
34. Arnone, R.; Walling, J. Waterborne pathogens in urban watersheds. *J. Water Health* **2007**, *5*, 149–162. [CrossRef] [PubMed]

35. Mallin, M.A.; Williams, K.E.; Esham, C.; Lowe, R.P. Effect of Human Development on Bacteriological Water Quality in Coastal Watersheds. *Ecol. Appl.* **2000**, *10*, 1047–1056. [CrossRef]
36. Vitro, K.A.; BenDor, T.K.; Jordanova, T.V.; Miles, B. A Geospatial Analysis of Land Use and Stormwater Management on Fecal Coliform Contamination in North Carolina Streams. *Sci. Total Environ.* **2017**, *603–604*, 709–727. [CrossRef] [PubMed]
37. Delpla, I.; Rodriguez, M.J. Effects of Future Climate and Land Use Scenarios on Riverine Source Water Quality. *Sci. Total Environ.* **2014**, *493*, 1014–1024. [CrossRef]
38. St. Laurent, J.; Mazumder, A. The influence of land-use composition on fecal contamination of riverine source water in southern British Columbia. *Water Resour. Res.* **2012**, *48*. [CrossRef]
39. Galfi, H.; Österlund, H.; Marsalek, J.; Viklander, M. Indicator Bacteria and Associated Water Quality Constituents in Stormwater and Snowmelt from Four Urban Catchments. *J. Hydrol.* **2016**, *539*, 125–140. [CrossRef]
40. Bravo, H.; McLellan, S.; Klump, J.; Hamidi, S.; Talarczyk, D. Modeling the fecal coliform footprint in a Lake Michigan urban coastal area. *Environ. Model. Softw.* **2017**, *95*, 401–419. [CrossRef]
41. Kuppusamy, M.; Giridhar, V. Factor analysis of water quality characteristics including trace metal speciation in the coastal environmental system of Chennai Ennore. *Environ. Int.* **2006**, *32*, 174–179. [CrossRef]
42. Simeonov, V.; Stratis, J.; Samara, C.; Zachariadis, G.; Voutsa, D.; Anthemidis, A.; Sofoniou, M.; Kouimtzis, T. Assessment of the surface water quality in Northern Greece. *Water Res.* **2003**, *37*, 4119–4124. [CrossRef]
43. Singh, K.; Malik, A.; Mohan, D.; Sinha, S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—A case study. *Water Res.* **2004**, *38*, 3980–3992. [CrossRef] [PubMed]
44. Huang, J.; Ho, M.; Du, P. Assessment of temporal and spatial variation of coastal water quality and source identification along Macau peninsula. *Stoch. Environ. Res. Risk Assess.* **2010**, *25*, 353–361. [CrossRef]
45. Marmontel, C.V.F.; Lucas-Borja, M.E.; Rodrgues, V.A.; Zema, D.A. Effects of land use and sampling distance on water quality in tropical headwater springs (Pimenta cree, Sao Paulo State, Brazil). *Sci. Total Environ.* **2018**, *622*, 690–701. [CrossRef] [PubMed]
46. Khan, U.T.; Valeo, C. Comparing a Bayesian and fuzzy number approach to uncertainty quantification in short-term dissolved oxygen prediction. *J. Environ. Inform.* **2017**, *30*, 1–16. [CrossRef]
47. He, J.; Chu, A.; Ryan, M.C.; Valeo, C.; Zaitlyn, B. Abiotic influences on dissolved oxygen in a riverine environment. *Ecol. Eng.* **2011**, *37*, 1804–1814. [CrossRef]
48. Pike, R.G.; Redding, T.E.; Moore, R.D.; Winkler, R.D.; Bladon, K.D. *Compendium of Forest Hydrology and Geomorphology in British Columbia*; B.C. Ministry of Forests and Range, Government Publications Services: Victoria, BC, Canada, 2010; ISBN 978-0-7726-6332-0.
49. *Guidelines for Canadian Recreational Water Quality*, 3rd ed.; Catalogue No H129-15/2012E; Health Canada: Ottawa, ON, Canada, 2012; Available online: http://www.healthcanada.gc.ca/waterquality (accessed on 1 January 2014).
50. Bailey, T.C.; Gatrell, A.C. *Interactive Spatial Data Analysis*; Longman Scientific Technical: Essex, UK, 1995.
51. Valeo, C.; Checkley, S.; He, J.; Neumann, N. Rainfall and microbial contamination in Alberta well water. *J. Environ. Eng. Sci.* **2016**, *11*, 18–28. [CrossRef]
52. Hauke, J.; Kossowski, T. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. *Quaest. Geogr.* **2011**, *30*. [CrossRef]
53. Hosmer, D.W.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*, 3rd ed.; John Wiley Sons, Inc.: Hoboken, NJ, USA, 2013; ISBN 978-0-470-58247-3.
54. Cleophas, T.; Zwinderman, A. *SPSS for Starters and 2 Levelers*; Springer International Publishing: Cham, Switzerland, 2016; ISBN 978-3-319-20600-4.
55. Tabachnick, B.G.; Fidell, L.S. *Using Multivariate Statistics*, 6th ed.; Pearson: New York, NY, USA, 2012; p. 983. ISBN 978-0205849574.
56. Guber, A.; Fry, J.; Ives, R.; Rose, J. *Escherichia coli* Survival in, and Release from, White-Taid Deer Feces. *Appl. Environ. Microbiol.* **2014**, *81*, 1168–1176. [CrossRef]
57. Selvakumar, A.; Borst, M.; Struck, S. Microorganisms Die-Off Rates in Urban Stormwater Runoff. *Proc. Water Environ. Fed.* **2007**, *5*, 214–230. [CrossRef]

58. Molina, M.; Hunter, S.; Cyterski, M.; Peed, L.A.; Kelty, C.A.; Sivagensan, M.; Mooney, T.; Prieto, L.; Shanks, O.C. Factors Affecting the Presence of Human-Associated and Fecal Indicator Real-Time Quantitative PCR Genetic Markers in Urban-Impacted Recreational Beaches. *Water Res.* **2014**, *64*, 196–208. [CrossRef] [PubMed]
59. Selvakumar, A.; Borst, M. Variation of Microorganism Concentration in Urban Stormwater Runoff with Land Use and Seasons. *J. Water Health* **2006**, *4*, 109–124. [CrossRef] [PubMed]