An Analysis of Spatial and Temporal Variation in Fecal Indicator Concentrations in Singapore

by
Suejung Shin

B.S. Environmental Engineering
Stanford University, 2011

Submitted to the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements of the Degree of

Master of Engineering
in Civil and Environmental Engineering
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

Signature of Author_____
Suejung Shin
Department of Civil and Environmental Engineering
May 15, 2012

Certified By_____
Peter Shanahan
Senior Lecturer of Civil and Environmental Engineering
Thesis Supervisor

Accepted By_____
Heidi M. Nepf
Chairman, Departmental Committee for Graduate Students

An Analysis of Spatial and Temporal Variation in Fecal Indicator Concentrations in Singapore

by
Suejung Shin

Submitted to the Department of Civil and Environmental Engineering
on May 15, 2012 in Partial Fulfillment of the
Requirements for the Degree of Master of Engineering in
Civil and Environmental Engineering

ABSTRACT

This study used extensive measurements of indicator concentrations to describe spatial and temporal patterns of four fecal indicators: *E. coli*, enterococci, total coliform, and human factor. Twenty twelve-hour time series were examined, with indicator concentrations measured every hour from 8 am to 7 pm. Six stations in Singapore were evaluated, across three land-use categories (high-density residential, low-density residential, commercial), and two sewer-age categories (new, old). The distributions of *E. coli*, enterococci, and total coliform were roughly lognormally distributed, showing that bacterial indicator concentration distributions are described by the same statistical model in tropical climates as in temperate climates, even with a wide and varied data set. Human factor indicator, for which there is limited preceding literature, was found to be roughly lognormally distributed. There was no obvious time pattern in measured concentrations, except that one hour's concentration roughly correlated with the next 1 to 3 hours' concentrations. This differs from findings from a study by the Public Utilities Board of Singapore that reported a diurnal pattern in total coliform and enterococci concentrations at a single sampling station.

This study found, for all indicators but total coliform, that older sewers had a significantly higher indicator concentration than newer sewers. This suggests that sewer leakage likely contributes to fecal contamination. Leaking sewers can explain some of the high indicator concentrations, but lack of diurnal pattern suggests there are more factors at play than just older sewers dispersing contaminants. There are mixed findings with regard to land use. For enterococci, low-density residential areas exhibit significantly different concentrations than high-density residential and commercial areas. For human factor, commercial areas exhibit significantly different concentrations than high-density and low-density residential areas. As human factor might be the best indicator of true pathogen concentrations, this suggests that land use plays a role in differences in concentrations, in agreement with previous studies.

Thesis Supervisor: Peter Shanahan
Title: Senior Lecturer of Civil and Environmental Engineering

Acknowledgments

I would like to thank Dr. Peter Shanahan for his warm support throughout the academic year and for making the Master of Engineering program what it is. I will never forget your selfless commitment to your students and incomparably amiable disposition.

I would also like to thank Professor Lloyd Chua for his support from overseas and for being a solid source of support back in Singapore.

Next, my deep thanks to PhD candidate Eveline Ekklesia, from whom I learned incredible amounts and whose hard work ethic and soft heart made this project simultaneously productive and enjoyable.

And to Jean Pierre Nshimyimana and Syed Alwi Bin Sheikh Bin Hussein Alkaff, thank you for all your steadfast assistance in data analysis and collection.

A special thanks to Chiara Lepore for volunteering her time to provide such keen comments and important insight in the framing of my research project.

To my fellow students of the Master of Engineering program and LIS Solutions (Janhvi Doshi, Laurie Kellndorfer, and Shobhna Kondepudi), thank you for every great laugh and memory. It wouldn't have been the same without you.

And finally, thank you to my family for helping me pursue my academic dreams. It means so much to have your support.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1. Background

This section has been written in collaboration with Janhvi Doshi, Laurie Kellndorfer, and Shobhna Kondepudi.

The Public Utilities Board (PUB) wishes to expand recreational activities within Singapore's reservoirs. Singapore has limited land area for recreation, and making use of selected waterways and waterbodies is an integral part of PUB's plan to meet public recreational needs. Singapore has been working to enhance the accessibility, usability, and aesthetics of green spaces and parks, especially near waterways and drainage (Soon et al. 2009). The PUB wishes to open more of Singapore's surface waters to recreational activities, under the Active, Beautiful, and Clean Waters Program (ABC Waters). The goals of the ABC Waters program are to bring the people of Singapore closer to their water resources by providing new recreational space and developing a feeling of ownership and value. The program aims to develop surface waters into aesthetic parks, estates, and developments. This plan will minimize pollution in the waterways by incorporating aquatic plants, retention ponds, fountains, and recirculation to remove nutrients and improve water quality (PUB Singapore 2011). One of the greatest areas of concern with this plan is microbial pollution.

Disease-causing pathogens pose the greatest immediate threat to human health in polluted surface waters. Humans can come into contact with waterborne pathogens through drinking water supply and through recreation in contaminated surface waters. Infection in humans can be caused by ingestion of, contact with, or inhalation of contaminated waters (Hurston 2007). While the exact total number of waterborne pathogens is unknown, it is estimated that over 1,000 viral and bacterial agents in surface waters can make humans sick. Diseases from waterborne pathogens can range from mild to life-threatening forms of gastroenteritis, hepatitis, skin and wound infections, conjunctivitis, respiratory infection, and other general infections. In order to open surface waterways and reservoirs for recreation, PUB must minimize pathogenic pollution in surface waters in order to keep the public safe.

1.2. Past Work

Prior student teams have collected concentration data for 25 parameters: total coliform, *E. coli*, enterococci, temperature, conductivity, salinity, chloride, bromide, boron, cholesterol, cholestanol, DBP, DEHP, coprostanol, caffeine, acetaminophen, ibuprofen, diclofenac, acesulfame-k, sucralose, saccharin, triclosan, detergent as MBAS, orthophosphate, and human factor marker. The five sampling locations were Choa Chua Kang (CCK) Crescent, Verde, Bras Basah, Serangoon Garden, and Toa Payoh North. Data at each location was collected hourly in January and/or June/July 2011 (Ekklesia 2011).

Prior teams have also investigated the effects of land use on the concentrations of a single bacterial indicator, *E. coli*. The study included concentrations measured over 24 locations in January and 45 locations in July 2009. It found that percentage of land used for agriculture correlated with *E. coli* concentrations levels. The study also found a weak inverse relationship between percent of developed land and *E. coli* levels (Foley et al. 2010). I plan on expanding upon this study by examining how differences in land use affect the concentration of not only *E. coli*, but also enterococci, total coliform, and human factor.

1.3. Project Scope

My project continues work in determining appropriate tracers to indicate bacterial contamination levels in the stormwater runoff that fills the reservoirs. Our team used the common bacterial indicators of total coliform, *E. coli*, and enterococci, as well as a DNA-based human factor marker to correlate to bacterial concentrations in the water.

Previous work has found that the effectiveness of tracers is dependent on the sampling location. My study examines the effects of two parameters upon the suitability of each tracer for a given location: sewer age and land-use distribution. Our results show how tracer concentration correlates with sewer age and land use, using information from six sampling locations.

1.4. Objectives

The objectives of this project are four-fold. First, I will characterize the distribution of the concentrations of four fecal indicators: *E. coli*, enterococci, total coliform, and human factor. Past studies lead me to hypothesize that the distributions will be lognormal for the bacterial indicators (Aragao et al. 2007; EPA 2010; Pontius 2003). Though there are no prominent previous findings regarding the distribution of the human factor marker, given some similarities in the fate and transport of indicator species *E. coli* and human factor, I hypothesize that the human factor indicator will be lognormally distributed, as well.

My second objective is to characterize patterns in the time series of indicator concentrations. Previous work conducted by the Public Utilities Board and reported by Ekklesia (2011) leads me to hypothesize that there is a diurnal pattern in bacterial concentrations, with peaks in the morning and evening, when sewer usage is also at a peak.

My third objective is to see how indicator concentrations vary between stations. As the study is conducted across six locations, I wish to see if there is any significant difference in expected concentrations based on sampling location. This can inform what consequences might occur by aggregating data from all six stations when conducting analyses. My hypothesis is that the stations will not be significantly different from one another.

My last objective is to determine how the two catchment parameters of sewer age and land use affect indicator concentrations. I hypothesize that locations with older sewer systems will have higher indicator concentrations than locations with newer sewer systems. I also hypothesize that commercial and high-density residential areas will have higher indicator concentrations than low-density residential areas given that more urbanized sites can yield higher *E. coli* concentrations (Desai and Rifai 2010).

Chapter 2 Literature Review

2.1. Fecal Indicators

Given the tedious, difficult, and time-consuming nature of conducting microbial examinations of water samples for pathogens, it is standard practice to look for indicator microorganisms whose presence indicates the probable presence of pathogens, instead. Indicators like coliform bacteria occur in the intestines of all warm-blooded animals and are excreted in feces. As coliform bacteria generally outlive pathogenic bacteria, their presence indicates disease-causing bacteria may be present and the water is unsafe to drink (Gerba and Pepper 2005). Common concentrations of indicator organisms in raw sewage are listed in Table 1. One-day grab sample (2-Feb-12) indicator concentrations from Bedok Garden in Singapore are also provided (Ekklesia 2012).

Table 1 – Estimated indicator organism concentrations in raw sewage

Organism	$\log_{10}$ Concentration[1]	Reference
Fecal coliforms	6 - 7	(Gerba and Pepper 2005)
	6.9	(Ekklesia 2012)
Enterococci	4 - 5	(Gerba and Pepper 2005)
	5.6	(Ekklesia 2012)
E. coli	8.7	(Evison and James 1973)
	6.6	(Ekklesia 2012)
Human factor (16S rRNA, Order Bacteriodales)	9.2	(Varma et al. 2009)
	8.9	(Silkie and Nelson 2009)
	8.9	(Sercu et al. 2009)
	5.3	(Savichtcheva et al. 2007)
	10	(Seurinck et al. 2005)
	8.0	(Ekklesia 2012)

[1]Units for fecal coliforms, enterococci, and *E. coli* are #cells/100mL. Units for human factor are copies/100mL.

2.1.1. Total Coliform

Total coliform includes all aerobic and facultatively anaerobic, gram-negative, non-spore forming, rod-shaped bacteria that produce gas upon lactose fermentation in culture media within 48 hours at 35 degrees C, including genera *Escherichia, Citrobacter, Enterobacter,* and *Klebsiella* (Gerba and Pepper 2005). Though total coliform was formerly used to assess recreational water quality, it fails as an indicator for a number of reasons: 1) it regrows in aquatic environments, 2) it regrows in distribution systems, and 3) it is not always indicative of a health threat (Gleeson and Gray 1997).

Total coliform has been seen to grow in environments of high organic matter and elevated temperatures in eutrophic tropical waters, water receiving pulp and paper mill effluents, wastewater, aquatic sediments, and organically enriched soil after periods of heavy rainfall (Gerba and Pepper 2005). This poses problems in Singapore, which has a tropical climate. This topic is further explored in Section 2.2.

2.1.2. Escherichia coli

Escherichia coli (*E. coli*) is a commonly used indicator that is distinguished by the way it ferments glucose. Though it works well as an indicator in temperate climates, it may be less satisfactory in tropical climates as it can grow independently of fecal sources (Bigger 1937; Evison and James 1973). Further, in some tropical climates, such as India, 30.3% of sewage samples did not contain *E. coli* (Rao et al. 1968).

2.1.3. Enterococci

The term 'enterococci' refers to the subgroup of fecal streptococci that are more specific to feces (Byappanahalli and Fujioka 1998). Enterococci are characterized by an ability to grow at both 10° and 45°C, survive at 60°C for at least 30 minutes, to grow at pH 9.6 and 6.5% NaCl, and to reduce 0.1% methylene blue in milk (Leclerc et al. 1996). The genus name *Enterococcus* includes bacteria previously named *Streptococcus faecalis* and *Streptococcus faecium*, and is generally known as gram-positive cocci with spherical cells arranged in pairs or chains, non-spore-forming, facultatively anaerobic, and homofermentative (Hardie and Whiley 1997).

Enterococci are more reliable indicators of water quality as they rarely multiply in water and they are more resistant to environmental stresses. However, enterococci are also known to be naturally present in soil and water in tropical environments like Hawaii and Guam (Fujioka et al. 1999; Hardina and Fujioka 1991) and are thereby not reliable for an environment like Singapore.

2.1.4. Human Factor Marker

The human factor marker in this report references organisms of order *Bacteroides* that have been prepared using 16S primers. *Bacteroides* is the most common microflora genus found in the human intestine (Kenzaka et al. 2001) and is found in larger concentrations than bacterial indicators (Srinivasan et al. 2011). *Bacteroides* species are obligately anaerobic, Gram negative, rod shaped, and non-endospore forming bacteria (Wexler 2007).

2.2. Indicators in Tropical Climates

Coliform bacteria have been known to grow in tropical climates. In Hawaii, fecal indicator bacteria are naturally found in most soil environments and can grow and multiply sporadically when conditions are relatively optimal (Byappanahalli and Fujioka 1998). Indicator bacteria have also been known to grow in bromeliads, flowering plants, in the rain forest of Puerto Rico, as well (Rivera et al. 1988). A number of other studies suggest coliform bacteria multiply in soils and natural surfaces and drinking water distribution systems (Fujioka et al. 1988; Gleeson and Gray 1997; Hardina and Fujioka 1991; Hazen 1988). The fact that coliforms can grow on biofilms on distribution system pipelines might be a problem as indicators can be more persistent in biofilm form. For example, *E. coli* is 2,400 times more resistant to free chlorine when attached to a surface than as free cells in water (Gerba and Pepper 2005). These studies suggest fecal indicator bacteria are not good indicators for recreational water quality standards in tropical environments.

2.3. Sewer Leakage Pathways

Human fecal contamination of the stormwater runoff that ends up in Singapore's coastal reservoirs may be attributed to leakage from sewer lines. It would be difficult to pinpoint the exact path from the sewers to the storm drains, but it is a fact that sewers can leak and by

definition carry sewage, which contains pollutants. A study of sewers in the United Kingdom found that the extent of leakage from sewer pipes depended on, among factors, the age of the system (Reynolds and Barrett 2003). Plausible methods of transport of pollutants are illustrated in Figure 1. Older pipes may have more cracks or loose lateral connections due to wear and tear, indicating that sewer age may correlate with higher contamination of nearby waters. Similarly, the length and diameter of the sewer may have an impact on the leakage rate, as more pipe can mean more connections and pipe wall, which means more possible places for leaks.

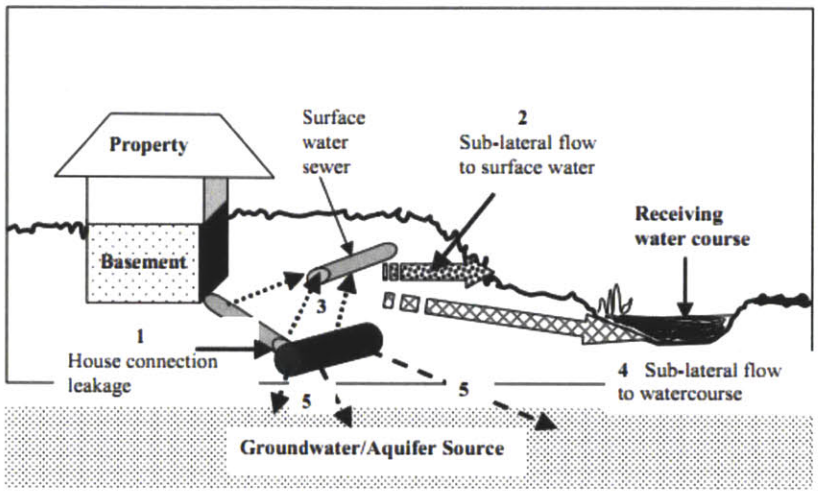


Figure 1 – Possible sewage leakage routes (Ellis et al. 2004)

There is disagreement about the rate of exfiltration from urban sewers (Rutsch et al. 2006). Some believe it is negligible (Vollertsen and Hvitved-Jacobsen 2003; Yang et al. 1999), while others found exfiltration rates on the order of 10% of dry-weather flow (Ellis et al. 2003; Wakida and Lerner 2005). My study will consider the possibility of contamination of stormwater runoff by exfiltration and will analyze how bacterial tracer concentrations vary with sewer age.

2.4. Diurnal Peaks in Sewer Flow

Sewer flow has been shown to follow a diurnal pattern, with peak usage in the morning around 9 am and in the evening around 9 pm (Enfinger and Stevens 2006). Figure 2 depicts a composite of 28 days of 24-hour hydrographs during normal dry weather conditions in a typical residential area within the United States. A repeatable diurnal pattern is visible, with differing patterns on weekdays and weekends. The curves indicate normal variation in flow expected during normal dry weather conditions. Given this pattern of sewer flow coupled with the possibility of leaking sewers introduced in Section 2.3, I would expect to see peaks in bacteria concentrations at the times of peak sewer flows.

The "Toolbox Study" conducted by the Singapore Public Utilities Board (PUB) observed similar diurnal patterns in concentrations of total coliform and enterococci (Ekklesia 2011). The study measured hourly concentrations of 23 water quality variables, including total coliform and enterococci concentrations, from 02-05 February 2009 in a storm drain in a low-density private residential area. Over the 72 hours, there was a marked diurnal pattern, especially on the second

day (03-04 February 2009), illustrated in Figure 3. From 10am to 9am, there were peaks in fecal (total) coliform and enterococci concentrations around 11am and 6pm, roughly matching the peak sewer flow times observed by Enfinger and Stevens (2006).

Diurnal patterns have also been found in a stream in Western Massachusetts (Traister and Anisfeld 2006).The highest *E. coli* levels were observed during the night and early morning, with the lowest levels in the afternoon, likely due to sunlight-induced die-off. There was less diurnal variability at more shaded sites. It was also observed that streams draining more developed watersheds generally had higher *E. coli* levels.

2.5. Land Use and Pathogens

Bacteria concentration levels have been known to correlate to land use. In Texas, urbanized areas had overall higher and less daily variability in *E. coli* concentrations than grassland areas (Desai and Rifai 2010). *E. coli* concentrations in a developed watershed were higher than in an undeveloped watershed in South Carolina, as well (Webster et al. 2003). My study will see if this is also the case in Singapore.

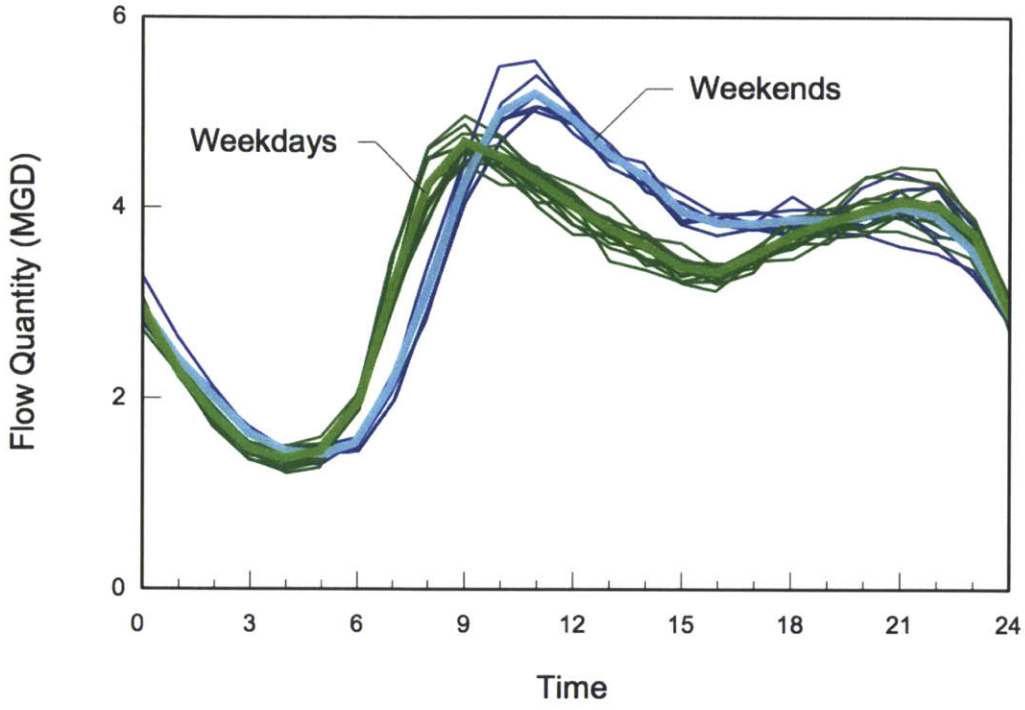


Figure 2 – Typical diurnal pattern of sewage flow (Enfinger and Stevens 2006)

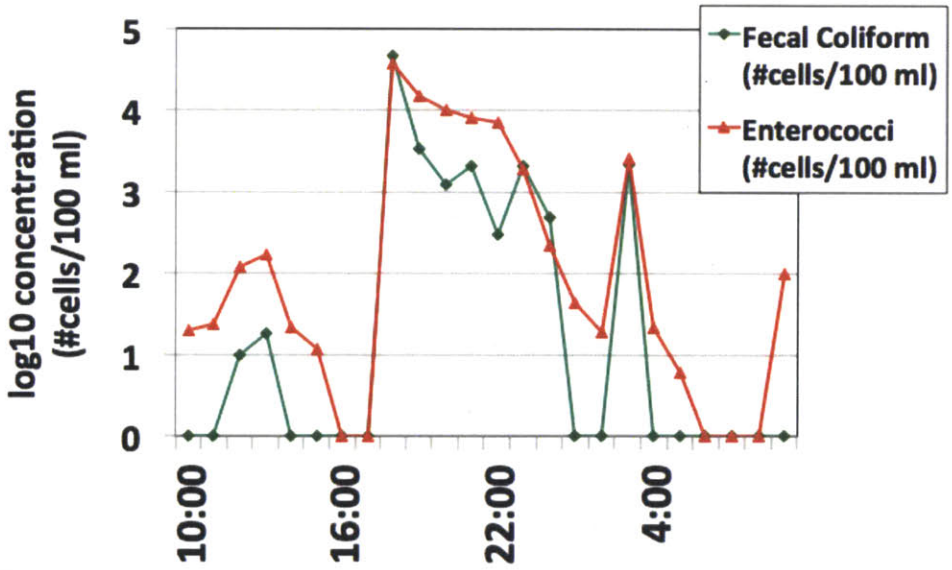


Figure 3 – Diurnal patterns in fecal (total) coliform and enterococci indicator concentrations in a storm drain in Singapore (Ekklesia 2011)

Chapter 3 Methodology

3.1. Field Data Collection

3.1.1. Sampling Locations

Water quality samples were collected from storm drains at six sampling stations with a variety of land uses and sewer ages. The locations of the six stations are depicted in Figure 4. In Figure 4, light blue represents high-density residential areas, yellow is low-density residential areas, and pink is commercial areas. Land-use designations (noted in Table 20 in Appendix II) were obtained by Ekklesia (2011) from PUB, which received the information from the Urban Redevelopment Authority (URA).

These sites were chosen based on the following criteria: relatively homogenous land use, relatively small watersheds, upstream location, and feasibility of sampling (i.e. easy access to the storm drain). Sewer age was established as old or new by Professor Lloyd Chua of Nanyang Technological University.

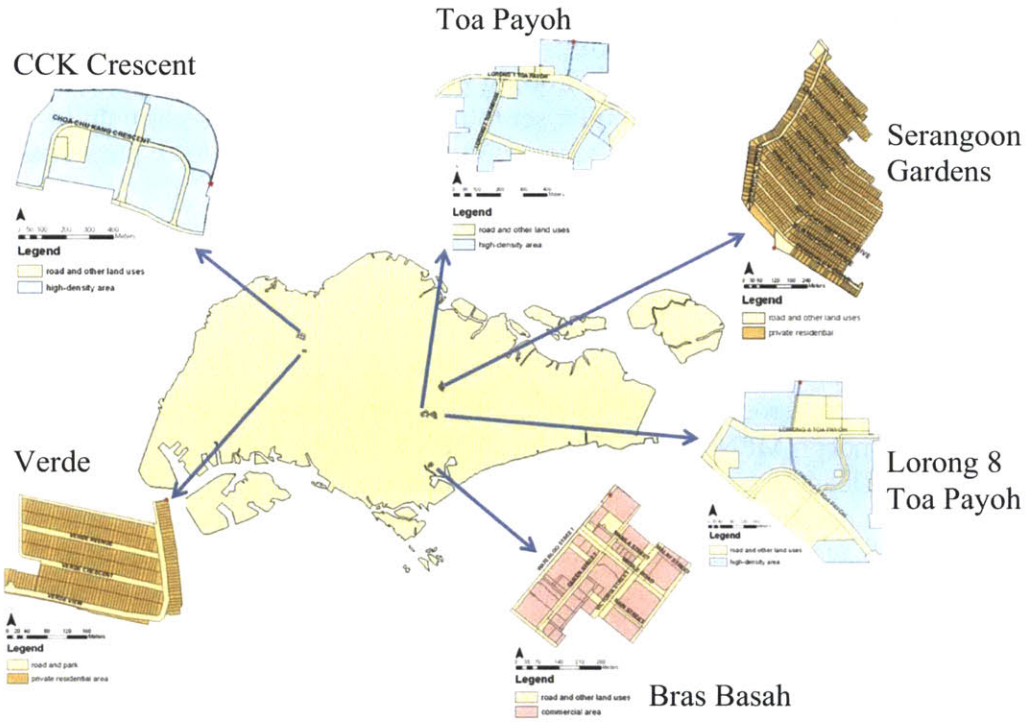


Figure 4 – Map of sampling stations. For larger maps, see Appendix II.

3.1.2. Sampling Methods

Four-liter water samples from storm drains at the above sites, excluding Lorong 8, were collected on hourly intervals from 8 am to 7 pm (Ekklesia 2011). Samples were collected by hand, dipping Whirl-Pak® (Nasco, Fort Atkinson, WI, USA) bags into the storm drain and transferring to sampling containers, or by using an extendable sampling pole (Nasco sampling pole B01367WA, Nasco, Fort Atkinson, WI, USA), as necessary. All bottles were rinsed with sample water before actual samples were collected.

One 0.25-L sample was collected in a plastic container for chloride, bromide, boron, and orthophosphate analysis (Ekklesia 2011). One 2-L sample was collected in amber glass containers for fecal sterols, plasticizers, caffeine, pharmaceutical compounds, artificial sweeteners, and triclosan analysis. One 0.5-L sample was collected in a clear glass container for surfactants (methylene blue active substances or MBAS) analysis. These three containers were sent to SETSCO Services Pte Ltd© for analysis.

One 0.25-L sample was collected in 100-mL Whirl-Pak® bags for bacterial analysis (Ekklesia 2011). One 1-L sample was collected in two 532-mL Whirl-Pak® bags for human factor analysis. One blank sample of bottled drinking water was collected in a 100-mL Whirl-Pak® bag and transported with the samples for bacterial analysis to test for contamination during transport. All samples were stored at below 4°C continuously in ice boxes on site and during transport and in the refrigerator at NTU campus. In addition to the water samples, in-situ measurements for temperature, conductivity, and salinity were also taken using a YSI Model 30 handheld meter (YSI Incorporated, Yellow Springs, OH, USA). I contributed to manual water sample collection at Choa Chu Kang Crescent on January 11 and 16, 2012 and at Verde on January 17, 2012. All other samples were collected by Eveline Ekklesia (NTU) and previous teams from the Master of Engineering program (MIT).

Water samples from station Lorong 8 were collected by an ISCO Avalanche (Teledyne Isco, Lincoln NE, USA) auto-sampler set to pump water from the storm drain into 950-mL or 5-L plastic sample bottles on the hour, every hour from 8 am to 7 pm. These samples were transported to NTU for bacterial and human factor analysis.

3.2. Laboratory Methods

3.2.1. Bacterial Tests

E. coli, enterococci, and total coliform concentrations were analyzed using the most probable number (MPN) method using IDEXX Quanti-Tray® and growth media (IDEXX 2008b). The Quanti-Tray® measures bacterial concentration, without dilution, in the range of one to 2,419 cells per 100mL. As many of our sampling locations are upstream, we use dilutions to account for bacterial counts that exceed 2,419 cells per 100mL. The three dilutions prepared were 1:1, 1:100, and 1:1,000. The diluted samples are mixed with growth reagents (Enterolert® for enterococci analysis and Colilert® for *E. coli* / total coliform analysis), producing six mixtures that are poured into six labeled Quanti-Trays®, which have 49 large wells and 48 small wells. Trays are sealed and incubated at 35°C ± 0.5°C for *E. coli* / total coliform analysis and at 41°C ±0 .5°C for enterococci analysis for 24-28 hours (IDEXX 2008a). After incubation, the numbers of large and small wells are counted. For *E. coli* / total coliform analysis, yellow wells are positive for total coliform and yellow wells that fluoresce under 365nm UV light are positive for *E. coli*. For enterococci analysis, fluorescing wells are positive for enterococci. Number of wells can be converted to a most probable number of bacterial cells using the MPN table provided by IDEXX.

I contributed to bacterial analysis of samples from Choa Chu Kang Crescent on January 11, 2012 and January 16, 2012; from Verde on January 17, 2012; and from Lorong 8 on January 9, 17, and 18, 2012.

3.2.2. Human Factor Analysis

Human factor was quantified using protocol developed by the Thompson lab at the Massachusetts Institute of Technology (Nshimyimana et al. in preparation). Human factor analyses were not completed for the earliest sampling rounds in January 2011.

3.3. Statistical Methods

3.3.1. Histograms

Histograms are commonly used to depict the distribution of data for a large sample size (Berthouex and Brown 2002). In my study, I bin indicator concentrations into equally sized intervals of concentration to see how frequently concentration values occur in each interval. A normal distribution would have a bell-shape.

3.3.2. Lognormal Probability Plots

Normal probability plots describe the distribution of the population from which the data were sampled. They have a specially scaled abscissa that yields a straight-line plot when the plotted points are normally distributed (Berthouex and Brown 2002). If the ordinate has a logarithmic scale, the plot will be a straight line if the data are lognormally distributed. Departures from the straight line indicate departures from normality, and the further the points are from the line, the greater the indication of departure from normality.

If there are just a few points that lie off the straight line, those points are likely outliers (BBN Corporation 1996). If both ends of the normal probability plot bend upwards above the straight line plot, the population from which the data were sampled might be skewed right. This is common when a variable is bounded on the left, but not on the right, so the variable tends to be closer to its minimum than maximum value (von Hippel 2010). If the ends of the plot bend down, the population may be skewed left. This indicates that a variable is usually closer to its maximum than its minimum value.

If the data forms an S-shape, with the right, upper end bending below the straight line and the left, lower end bending above the line, the population from which the data were sampled may be light-tailed (BBN Corporation 1996). In light-tailed distributions, the extreme parts of the distribution spread out less relative to the width of the center than if the distribution were normal. The probability of observing a value far from the median is less than in the case of a normal distribution.

If the data forms an S-shape, with the right, upper end bending above the straight line and the left, lower end bending below the line, the population from which the data were sampled may be heavy-tailed (BBN Corporation 1996). The probability of observing a value far from the median in either direction is greater than in the case of the normal distribution.

To create normal probability plots, I used the 'normplot' function in Matlab® with the input being indicator concentrations.

3.3.3. Artificial Time Series

Time-series plots can reveal cyclic patterns and variations of fluctuations. After determining normality, the next step of analysis was finding a pattern over time in the bacterial indicator concentrations. To do so, I lined up all 20 twelve-hour time series end-to-end to create

an artificial time series of twenty half-days. I plotted indicator concentrations against days with the 'plot' function in Matlab®.

3.3.4. Autocorrelation

To evaluate the extent of a diurnal pattern, I used autocorrelation analysis, again with the artificial time series of twenty half-days. Autocorrelation computes a correlation coefficient between observations at one hour and observations at another hour (Berthouex and Brown 2002). The distance between the observations examined for correlation is called "lag." In my case, I looked at lags of 1-12 hours to determine whether there is a diurnal pattern. The correlation coefficient, r_k, ranges in value from -1 to +1, where r_k=0 indicates complete independence and r_k=1 indicates perfect correspondence. If there is a diurnal pattern, correlation at lag 12 should be positively correlated in my half-day series. If the time series has no distinguishable pattern and the indicator concentrations are random at each time, the correlation coefficients should be near zero for all lag separations.

I completed this analysis in Microsoft Excel® (Microsoft Corporation, Redmond, WA, USA) using the 'correl' function along with an appropriate equation to call the correct concentration values for the amount of lag time in question. Correlation coefficients were then plotted against lag, in hours.

3.3.5. Correlation Matrices

For a visual representation of the autocorrelation analysis, I calculated Pearson correlation coefficients and arranged them in matrices. These coefficients quantify the relationship between two variables with 0 indicating no relationship between the two variables, -1 meaning a perfect negative relationship, and +1 representing a perfect positive relationship (Berthouex and Brown 2002). I calculated the correlation coefficients between all indicator concentrations at one hour, t, and all indicator concentrations at the next hour, t+1. Then I calculated the coefficient between concentrations at t and t+2. I continued these calculations to fill in a matrix of each hour correlated with every other hour. The matrices are shown in Tables 3 through 6. A color scale was applied such that high correlations (>0.7) are represented in the tables by shades of green, low correlations (< 0.3) by shades of red, and in-between by shades of yellow. If one hour's concentration well correlates with the next hour's concentration, the coefficient value will be high and shown in a shade of green. If there is a diurnal pattern, the correlation coefficient between any time, t, and t+12 should be high and a shade of green.

I computed Pearson correlation coefficients using the Excel® function 'pearson', which is equivalent to the function 'correl'.

3.3.6. Dot Plots

Dot plots are diagrams that help reveal the sample's distribution and variability succinctly (Berthouex and Brown 2002). They work well in depicting a large amount of data at once. In the case of the dot plots I constructed, indicator concentration was plotted on the abscissa and either time or the time-series data set number (as listed in Table 2) was plotted on the ordinate.

3.3.7. t-tests

Unpaired t-test

I used an unpaired two-tailed t-test for unequal variances to test whether there was a significant difference between old and new sewer indicator concentrations. The unpaired t-test is

used to compare means of two sets of independent samples, one from each of the two populations being compared (Fadem 2008). The null hypothesis is that the means of the two normally-distributed populations are equal. If the populations have the same mean, the probability that random sampling would lead to a difference between sample means as large as what I observed is represented by the p-value. A p-value of greater than 0.05 indicates no difference between the populations and p-value less than 0.05 shows there is a significant difference between indicator concentrations at locations with old versus new sewers. I used the Excel® function 'ttest(data1, data2, 2, 3)', where the '2' indicates two-tailed, and '3' indicates two-sample unequal variance test. Two-tailed tests are used when it is uncertain which group would have the larger mean.

Multiple paired comparisons of k averages

I used the Tukey-Kramer method to compare all possible pairs of means (Berthouex and Brown 2002). For each pair of means, the minimum significant difference (MSD) is calculated. If the observed difference between a pair of means is greater than the MSD, then the pair of means is significantly different. I conducted my analysis at the alpha = 0.05 level using an Excel® worksheet provided by (McDonald 2009).

I chose to represent results in table format. In the tables, there are two sets of numbers: one at upper right and another at lower left, with the two separated by a blank diagonal. The calculated mean significant differences (MSDs) are in the upper right triangle, while the observed differences are in the bottom left triangle. Each number in the bottom left portion of the table is compared to its mirror image MSD value in the upper right to determine if there is a significant difference. If the observed difference is significant, I marked it with an asterisk ('*') and highlighted the cell in yellow.

3.4. Limitations

The sampling and statistical methodologies had some limitations and potential sources of uncertainty, discussed below:

- In the laboratory analysis of indicator concentrations as discussed in Section 3.2.1, there was some uncertainty in the IDEXX MPN method of counting the number of fluorescing cells. It was a personal judgment call by the laboratory analyst as to which cells were clearly fluorescent and thereby indicative of bacterial presence. There is a significant difference in the most probable number depending on how many such cells are reported.
- For two concentration measurements of total coliform, the MPN reading was above the limit of 241,960 cells/100mL, so the average of the concentrations at the hour before and the hour after was taken.
- For ten concentration measurements of total coliform, the MPN yielded >24,196,000 cells/100mL, which was set to 50,000,000 cells/100mL for plotting and analysis purposes.
- During thirteen of the hours at which indicator concentrations were to be measured, grab samples were not collected and thereby lab analysis was not conducted to determine concentrations. Ten missing concentrations were due to rain (our study considered only dry-weather samples) and three can be attributed to a delayed start on one day of sampling. Times when samples were not collected were completely disregarded and not included in any calculations.

- During ten of the hours at which indicator concentrations were measured, it was raining, which can produce higher concentrations due to more sources of contamination feeding into the storm drains. This study was designed to collect concentrations during dry weather, so including these wet-weather concentration values might negatively affect results.

3.5. Summary of Data Used in This Study

I evaluated 20 twelve-hour time series, with four indicator concentrations (*E. coli*, enterococci, total coliform, and human factor) measured on the top of the hour every hour from 8 am to 7 pm. There are a total of six stations (Choa Chu Kang Crescent, Verde, Bras Basah, Serangoon, Toa Payoh, Lorong 8 Toa Payoh) across three land-use categories (high-density residential, low-density residential, commercial), and two sewer-age categories (new, old), as seen in Table 2. Sampling dates span the winter (January) and summer (June, July) in years 2011 and 2012, although seasonal differences are minor in Singapore.

Table 2 – Summary of twenty time series evaluated

	Site Name	Land-Use Category	Sewer age	Sampling date
1	Choa Chu Kang Crescent	High-density residential	New	4-Jan-11
2	Choa Chu Kang Crescent	High-density residential	New	19-Jan-11
3	Choa Chu Kang Crescent	High-density residential	New	11-Jan-12
4	Choa Chu Kang Crescent	High-density residential	New	16-Jan-12
5	Verde	Low-density residential	New	6-Jan-11
6	Verde	Low-density residential	New	12-Jan-11
7	Verde	Low-density residential	New	17-Jan-12
8	Bras Basah	Commercial	Old	10-Jan-11
9	Bras Basah	Commercial	Old	18-Jan-11
10	Bras Basah	Commercial	Old	28-Jun-11
11	Bras Basah	Commercial	Old	29-Jun-11
12	Serangoon	Low-density residential	Old	8-Jun-11
13	Serangoon	Low-density residential	Old	5-Jul-11
14	Serangoon	Low-density residential	Old	7-Jul-11
15	Toa Payoh	High-density residential	Old	7-Jun-11
16	Toa Payoh	High-density residential	Old	4-Jul-11
17	Toa Payoh	High-density residential	Old	6-Jul-11
18	Lorong 8 Toa Payoh	High-density residential	Old	9-Jan-12
19	Lorong 8 Toa Payoh	High-density residential	Old	17-Jan-12
20	Lorong 8 Toa Payoh	High-density residential	Old	18-Jan-12

Chapter 4 Data Analysis

4.1. Determining Lognormality

The first step of analysis was characterizing the distribution of the indicators *E. coli*, enterococci, total coliform, and human factor to be able to use the appropriate statistical methods. Some tests can only be performed when the data set is normally distributed. Previous study shows *E. coli*, enterococci, and total coliform concentrations to be lognormally distributed (Aragao et al. 2007; EPA 2010; Pontius 2003). To characterize the distributions, I looked at 1) histograms and 2) normal probability plots.

4.1.1. Histograms

Histograms of the indicator concentrations (Figure 5) show clearly that indicator concentrations are not normally distributed. The distributions are heavily skewed right. This can be fixed by performing a log transformation. The $\log_{10}$ concentrations of all indicators, in contrast, follow a roughly normal distribution (Figure 6). In other words, the indicators all roughly follow a lognormal distribution. *E. coli* is somewhat skewed right.

4.1.2. Lognormal Probability Plots

Probability plots also suggest lognormality in indicator concentrations. The concentration of *E. coli*, enterococci, total coliform, and human factor as depicted in Figure 7 generally follow the lognormal distribution with a few points of interest. Note again that plotting $\log_{10}$ values on the ordinate proves lognormality when the points form a straight line on a normal probability plot.

The normal probability plot shows a slightly non-linear pattern for *E. coli*, suggesting a better model can be chosen to describe the distribution of $\log_{10}$ concentrations of *E. coli*. Enterococci is a bit heavy-tailed (the upper right end of the plot bends below the straight line and the lower left end bends below it), indicating that there are more values further away from the median than in a normal distribution. In the case of total coliform, the points at the top tail of the diagram represent the ceiling of the maximum detectable level of total coliform based on the IDEXX Quanti-Tray® method. Presumably, if there were not such limitations in the test, the points would follow the straight line more closely in that region. The total coliform plot is also somewhat light-tailed (the upper right end bends below the straight line, while the lower left end bends above that line), indicating that observing a value far from the median is less likely than in a normal distribution. Despite these limitations, the indicator concentrations can still be characterized as roughly lognormally distributed, and therefore the statistical tests conducted in this report use $\log_{10}$ concentrations for each of the four indicators.

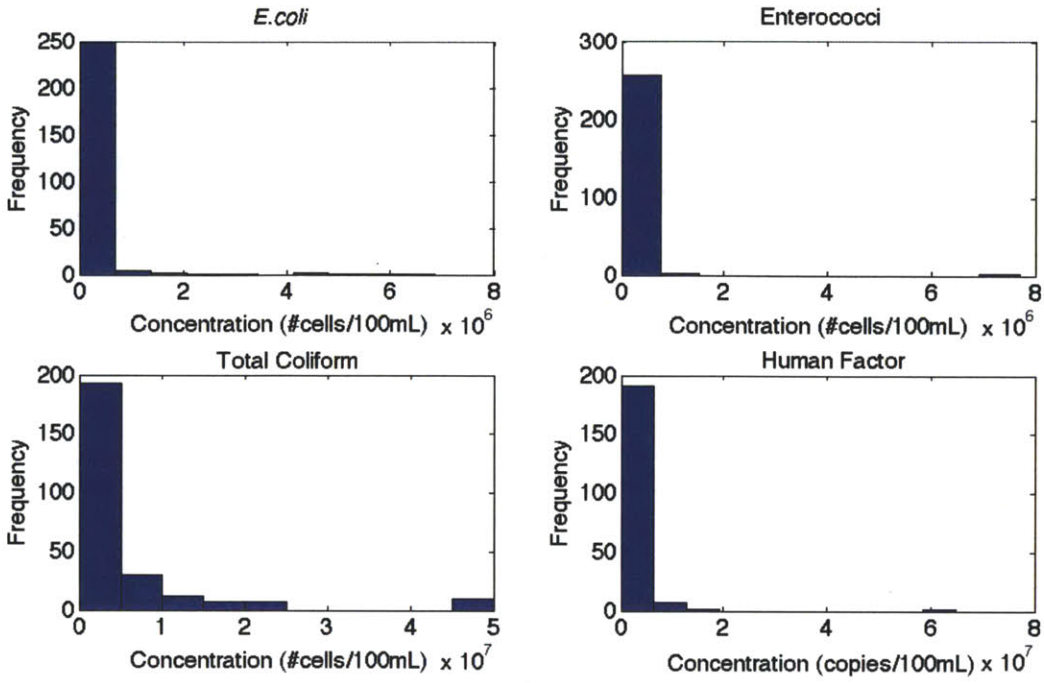


Figure 5 – Histograms of concentration, all indicators

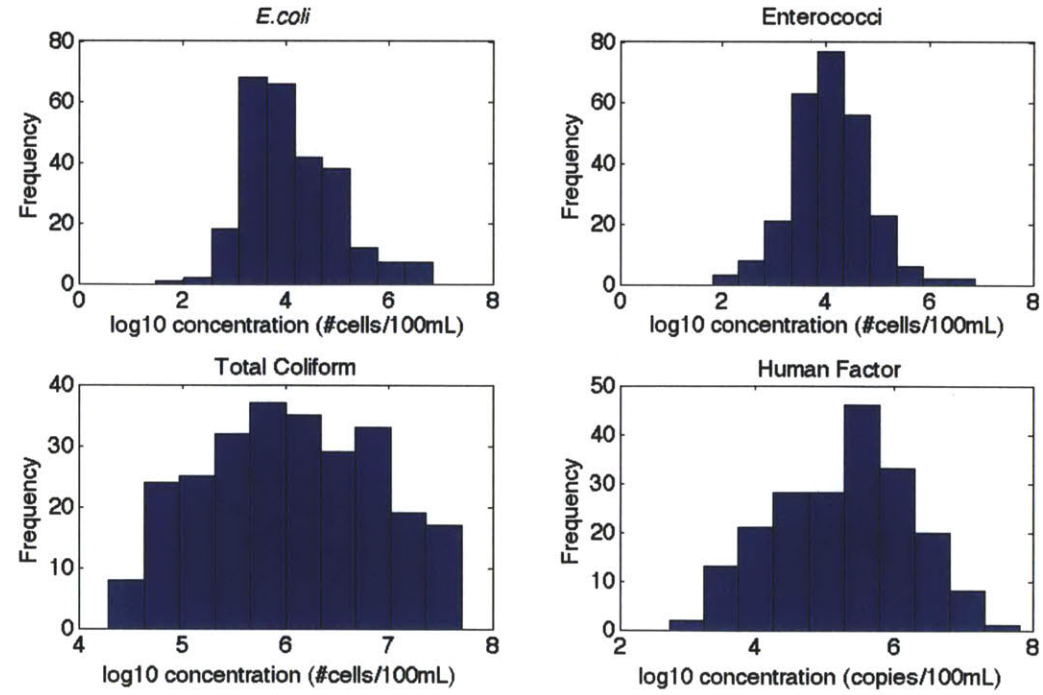


Figure 6 – Histograms of $\log_{10}$ concentration, all indicators

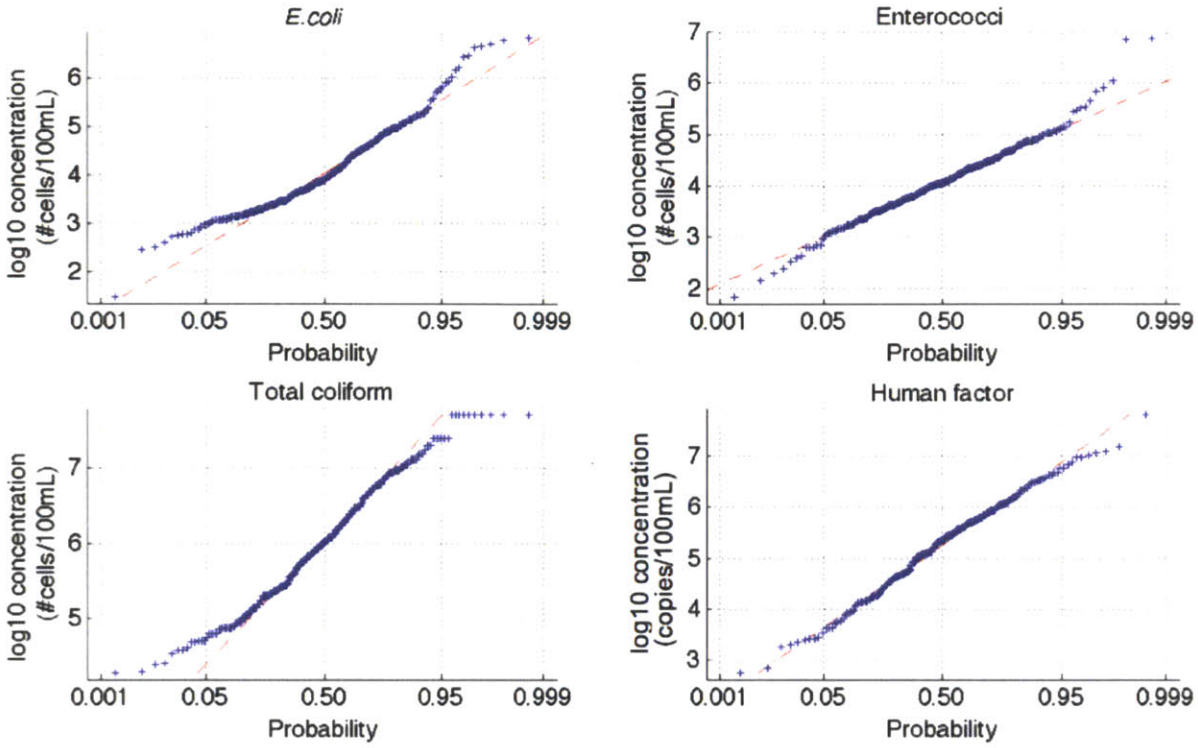


Figure 7 – Lognormal probability plot, all indicators

4.2. Determining Patterns in Time

4.2.1. Artificial Time Series

After determining normality, the next step of analysis was finding a pattern over time in the bacterial indicator concentrations. To do so, I lined up all 20 twelve-hour time series end-to-end to create an artificial time series of twenty half-days. General inspection of the time series indicates no pattern (Figure 8), negating previous findings by the PUB Toolbox Study that bacterial concentrations follow a diurnal pattern in Singapore (Ekklesia 2011).

Note again that $\log_{10}$ concentrations of the indicators were used because concentrations are lognormally distributed. Also note that human factor had fourteen time series instead of twenty as there were some days when human factor concentrations were not measured.

4.2.2. Autocorrelation

For the autocorrelation analysis, I looked at lags of 1 to 12 hours to determine whether there is a diurnal pattern (Figure 9). To confirm a diurnal pattern, I would expect concentrations at lag 12 to be positively correlated; however, this was not the case for most of the indicators, possibly with the exception of enterococci, which saw a slight peak in correlation at time lag 8. In the case of human factor, there was a slight negative correlation at lag 12, negating any presence of a diurnal pattern. One finding is a moderately high degree of autocorrelation between adjacent and near-adjacent observations, up to a lag of about 3.

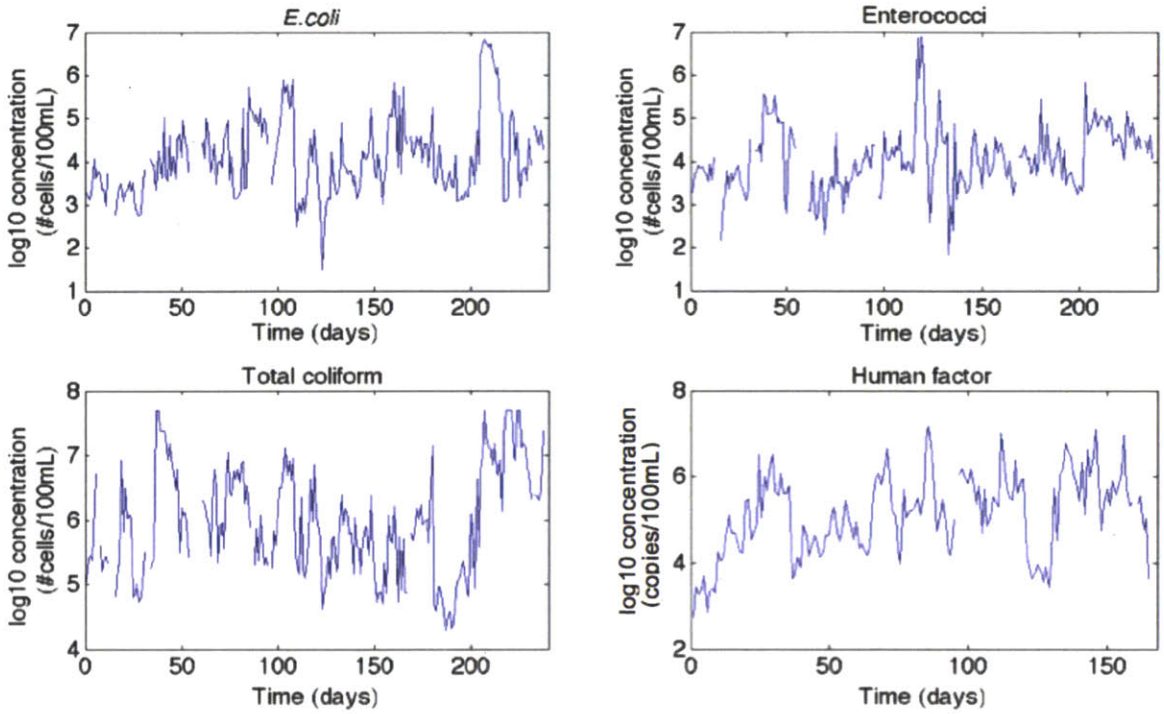


Figure 8 – Artificial time series plot, all indicators

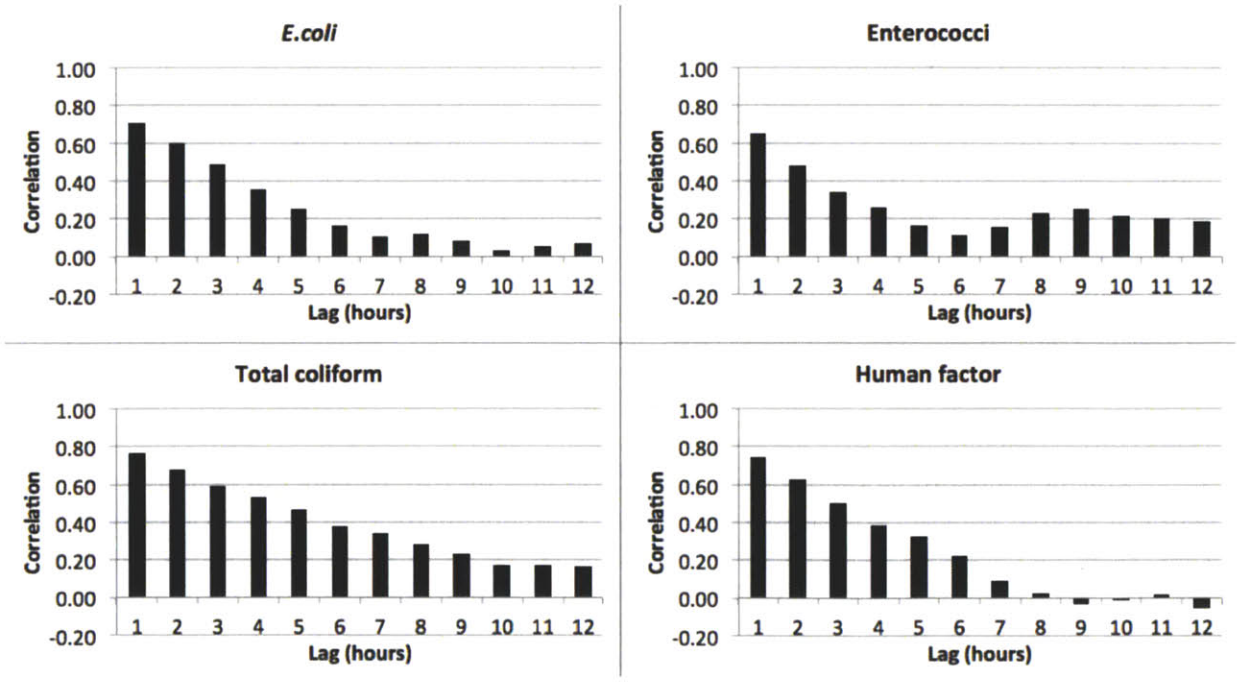
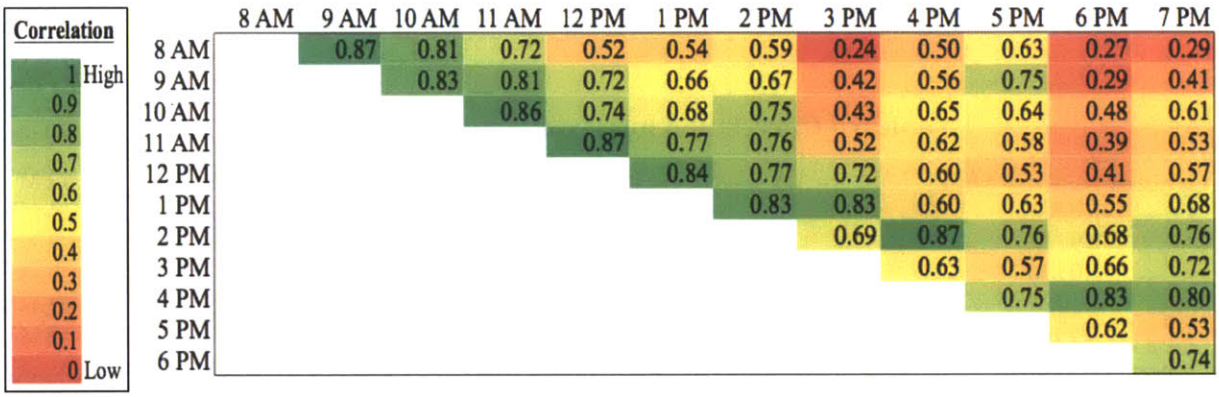


Figure 9 – Autocorrelation pattern, all indicators

4.2.3. Correlation Matrices

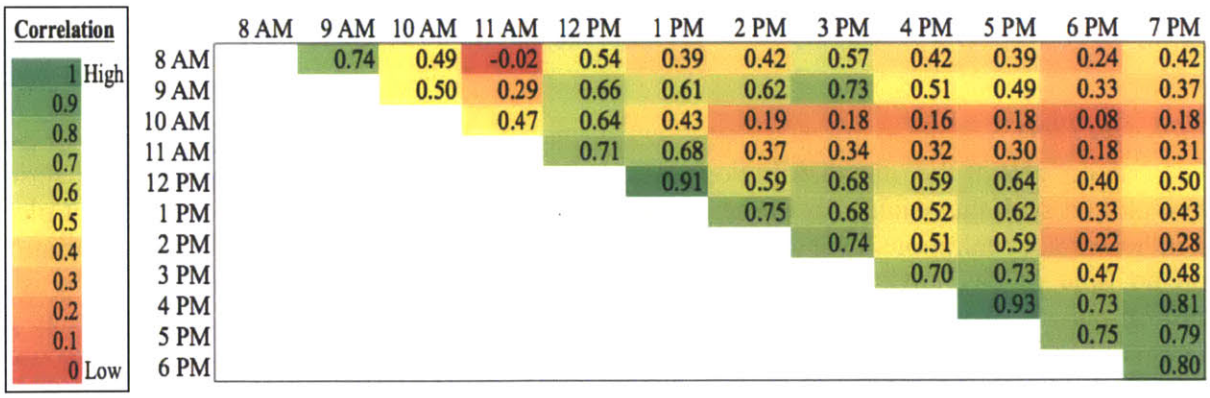
In addition to the autocorrelation analysis, for a more visual representation, I created matrices with correlation values (Tables 3 through 6). These plots confirm the autocorrelation results—that one hour's concentration well correlates with the next hour's concentration and up to about three hours out. This is represented by the predominantly green values that lie along the diagonal of the triangle, where one hour and the next hour's concentrations are correlated. Also, there is again no evidence of a diurnal pattern as concentrations at one hour, t, and concentrations at t+12 have a fairly low correlation, represented by regions of red-shaded low correlation areas in the top right corner of the matrices.

Table 3 – Correlation matrix by time, *E. coli*



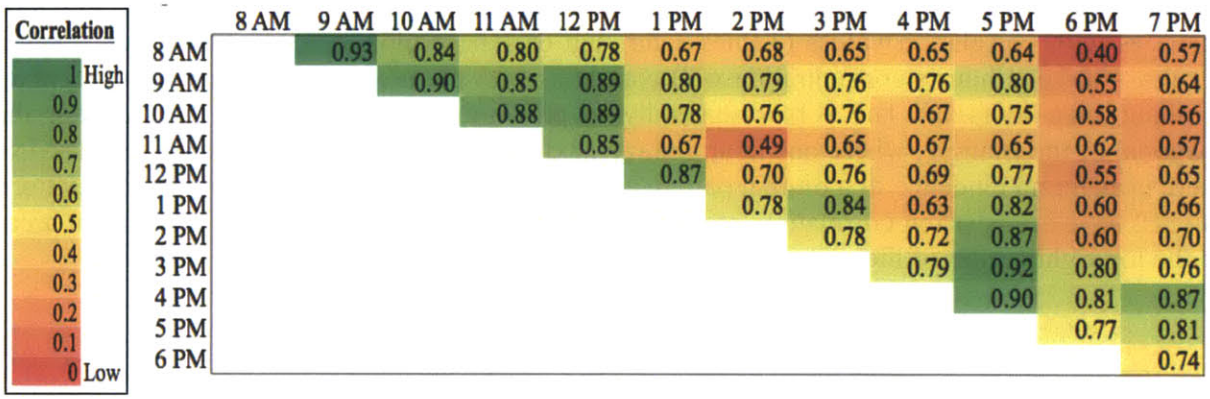
	8 AM	9 AM	10 AM	11 AM	12 PM	1 PM	2 PM	3 PM	4 PM	5 PM	6 PM	7 PM
8 AM		0.87	0.81	0.72	0.52	0.54	0.59	0.24	0.50	0.63	0.27	0.29
9 AM			0.83	0.81	0.72	0.66	0.67	0.42	0.56	0.75	0.29	0.41
10 AM				0.86	0.74	0.68	0.75	0.43	0.65	0.64	0.48	0.61
11 AM					0.87	0.77	0.76	0.52	0.62	0.58	0.39	0.53
12 PM						0.84	0.77	0.72	0.60	0.53	0.41	0.57
1 PM							0.83	0.83	0.60	0.63	0.55	0.68
2 PM								0.69	0.87	0.76	0.68	0.76
3 PM									0.63	0.57	0.66	0.72
4 PM										0.75	0.83	0.80
5 PM											0.62	0.53
6 PM												0.74

Table 4 – Correlation matrix by time, enterococci



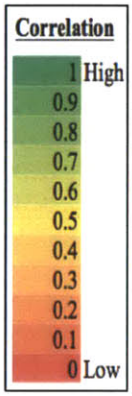
	8 AM	9 AM	10 AM	11 AM	12 PM	1 PM	2 PM	3 PM	4 PM	5 PM	6 PM	7 PM
8 AM		0.74	0.49	-0.02	0.54	0.39	0.42	0.57	0.42	0.39	0.24	0.42
9 AM			0.50	0.29	0.66	0.61	0.62	0.73	0.51	0.49	0.33	0.37
10 AM				0.47	0.64	0.43	0.19	0.18	0.16	0.18	0.08	0.18
11 AM					0.71	0.68	0.37	0.34	0.32	0.30	0.18	0.31
12 PM						0.91	0.59	0.68	0.59	0.64	0.40	0.50
1 PM							0.75	0.68	0.52	0.62	0.33	0.43
2 PM								0.74	0.51	0.59	0.22	0.28
3 PM									0.70	0.73	0.47	0.48
4 PM										0.93	0.73	0.81
5 PM											0.75	0.79
6 PM												0.80

Table 5 – Correlation matrix by time, total coliform



	8 AM	9 AM	10 AM	11 AM	12 PM	1 PM	2 PM	3 PM	4 PM	5 PM	6 PM	7 PM
8 AM		0.93	0.84	0.80	0.78	0.67	0.68	0.65	0.65	0.64	0.40	0.57
9 AM			0.90	0.85	0.89	0.80	0.79	0.76	0.76	0.80	0.55	0.64
10 AM				0.88	0.89	0.78	0.76	0.76	0.67	0.75	0.58	0.56
11 AM					0.85	0.67	0.49	0.65	0.67	0.65	0.62	0.57
12 PM						0.87	0.70	0.76	0.69	0.77	0.55	0.65
1 PM							0.78	0.84	0.63	0.82	0.60	0.66
2 PM								0.78	0.72	0.87	0.60	0.70
3 PM									0.79	0.92	0.80	0.76
4 PM										0.90	0.81	0.87
5 PM											0.77	0.81
6 PM												0.74

Table 6 – Correlation matrix by time, human factor



	8 AM	9 AM	10 AM	11 AM	12 PM	1 PM	2 PM	3 PM	4 PM	5 PM	6 PM	7 PM
8 AM		0.90	0.90	0.71	0.71	0.80	0.75	0.64	0.41	0.42	0.45	0.75
9 AM			0.83	0.63	0.54	0.61	0.66	0.60	0.34	0.31	0.38	0.77
10 AM				0.85	0.72	0.76	0.74	0.59	0.45	0.29	0.36	0.58
11 AM					0.75	0.69	0.67	0.54	0.63	0.33	0.39	0.49
12 PM						0.85	0.71	0.59	0.64	0.62	0.38	0.56
1 PM							0.89	0.80	0.67	0.76	0.62	0.74
2 PM								0.84	0.74	0.76	0.66	0.82
3 PM									0.76	0.81	0.53	0.78
4 PM										0.72	0.59	0.74
5 PM											0.58	0.81
6 PM												0.79

4.2.4. Dot Plots

A dot plot of the time distributions also shows a lack of pattern. All 20 twelve-hour series were plotted such that each hour of the day had twenty bacterial concentration values. Dot plots of indicator concentrations over time (Figures 10 through 13) show no regular cyclic variation and no time of day having consistently high or consistently low values. Further, there is a uniform pattern of variance with no tendency for data points to cluster around one average or central value. This suggests that the relative magnitude of concentration expected at a given time does not follow a predictable pattern.

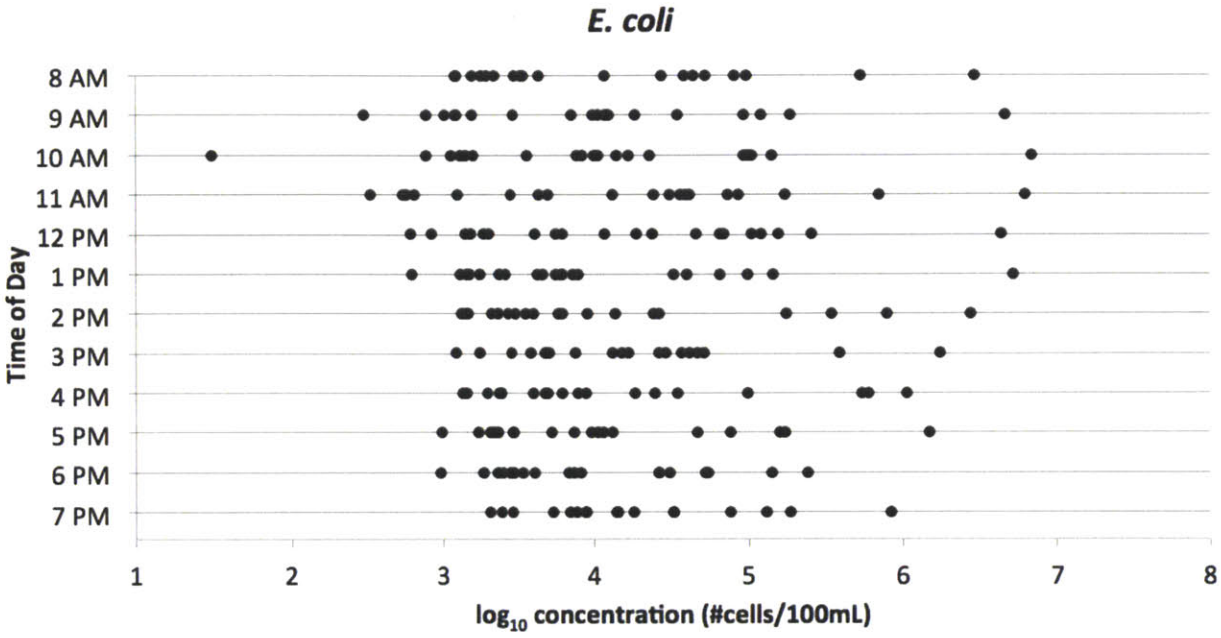


Figure 10 – Dot plot by time, *E. coli*

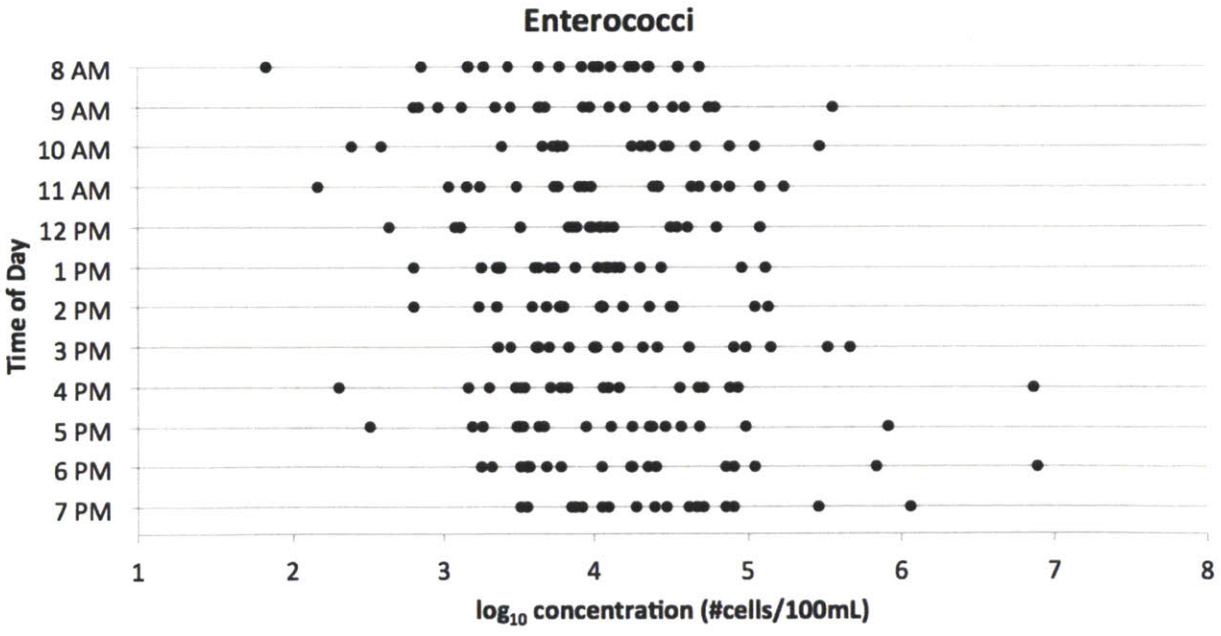


Figure 11 – Dot plot by time, enterococci

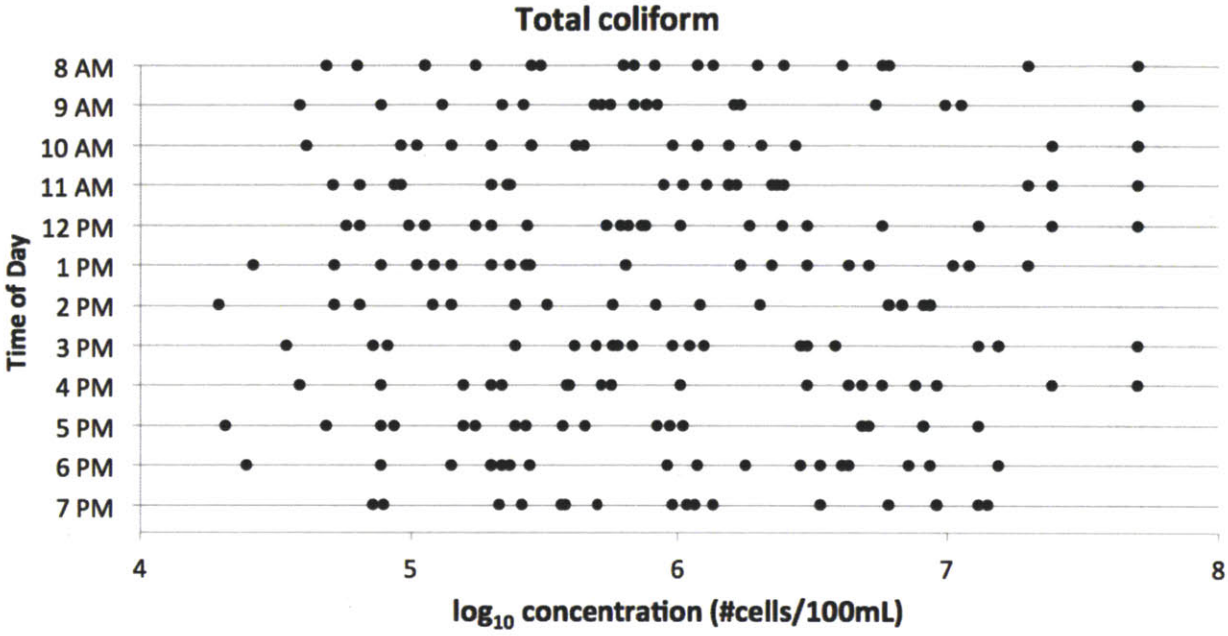


Figure 12 – Dot plot by time, total coliform

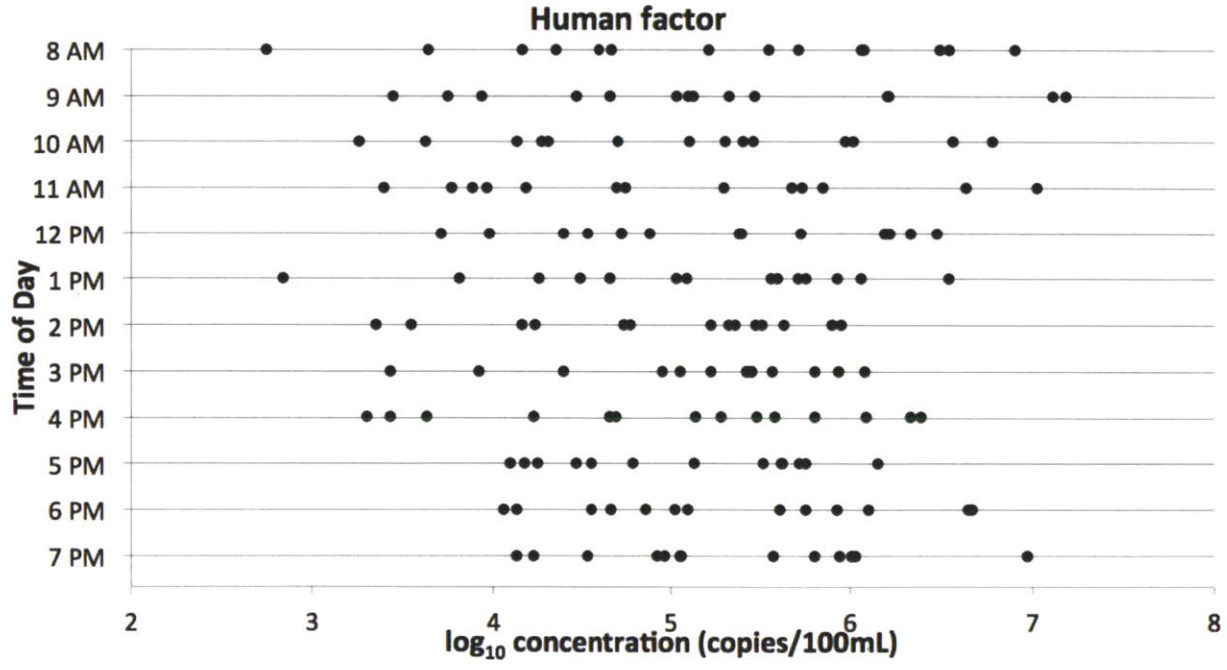


Figure 13 – Dot plot by time, human factor

4.2.5. t-tests

A multi-way t-test (Tukey-Kramer method) also confirms lack of pattern (Berthouex and Brown 2000). Results of an analysis based on observed concentrations by hour of the day are shown in Tables 7 through 10. See Section 3.3.7 for more information on the construction and interpretation of these tables. There are no significant differences between concentrations observed at different hours of the day as all the observed difference values (lower left triangle) are less than the corresponding MSDs (upper right triangle). This suggests that indicator concentrations measured at any particular hour of day are not necessarily going to be any different from concentrations measured at another hour of day.

Table 7 – t-test by time, *E. coli*

	8 AM	9 AM	10 AM	11 AM	12 PM	1 PM	2 PM	3 PM	4 PM	5 PM	6 PM	7 PM
8 AM	-	1.01	1.01	1.01	1.00	1.00	1.01	1.01	1.02	1.02	1.02	1.04
9 AM	0.10	-	1.01	1.01	1.00	1.00	1.01	1.01	1.02	1.02	1.02	1.04
10 AM	0.10	0.01	-	1.01	1.00	1.00	1.01	1.01	1.02	1.02	1.02	1.04
11 AM	0.06	0.16	0.17	-	1.00	1.00	1.01	1.01	1.02	1.02	1.02	1.04
12 PM	0.11	0.20	0.21	0.04	-	0.99	1.00	1.00	1.01	1.01	1.01	1.03
1 PM	0.13	0.03	0.03	0.19	0.24	-	1.00	1.00	1.01	1.01	1.01	1.03
2 PM	0.01	0.09	0.09	0.07	0.11	0.12	-	1.01	1.02	1.02	1.02	1.04
3 PM	0.12	0.21	0.22	0.05	0.01	0.24	0.12	-	1.02	1.02	1.02	1.04
4 PM	0.05	0.14	0.15	0.02	0.06	0.18	0.05	0.07	-	1.04	1.04	1.05
5 PM	0.04	0.06	0.06	0.10	0.15	0.09	0.03	0.15	0.09	-	1.04	1.05
6 PM	0.10	0.00	0.00	0.16	0.21	0.03	0.09	0.22	0.15	0.06	-	1.05
7 PM	0.15	0.25	0.25	0.09	0.05	0.28	0.16	0.04	0.10	0.19	0.25	-

Lower left triangle = observed differences. Upper right triangle = minimum significant difference (MSD). There is a significant difference between a pair of means if the observed difference > MSD.

Table 8 – t-test by time, enterococci

	8 AM	9 AM	10 AM	11 AM	12 PM	1 PM	2 PM	3 PM	4 PM	5 PM	6 PM	7 PM
8 AM	-	0.80	0.80	0.80	0.79	0.79	0.80	0.80	0.81	0.81	0.81	0.82
9 AM	0.11	-	0.80	0.80	0.79	0.79	0.80	0.80	0.81	0.81	0.81	0.82
10 AM	0.26	0.15	-	0.80	0.79	0.79	0.80	0.80	0.81	0.81	0.81	0.82
11 AM	0.25	0.14	0.01	-	0.79	0.79	0.80	0.80	0.81	0.81	0.81	0.82
12 PM	0.12	0.01	0.14	0.13	-	0.78	0.79	0.79	0.80	0.80	0.80	0.81
1 PM	0.07	0.04	0.19	0.18	0.05	-	0.79	0.79	0.80	0.80	0.80	0.81
2 PM	0.19	0.08	0.07	0.06	0.07	0.12	-	0.80	0.81	0.81	0.81	0.82
3 PM	0.46	0.36	0.20	0.21	0.35	0.39	0.28	-	0.81	0.81	0.81	0.82
4 PM	0.29	0.18	0.03	0.04	0.17	0.22	0.10	0.17	-	0.82	0.82	0.83
5 PM	0.23	0.12	0.03	0.02	0.11	0.16	0.04	0.24	0.06	-	0.82	0.83
6 PM	0.48	0.38	0.22	0.23	0.37	0.41	0.30	0.02	0.19	0.26	-	0.83
7 PM	0.63	0.52	0.37	0.38	0.51	0.56	0.44	0.17	0.34	0.40	0.15	-

Lower left triangle = observed differences. Upper right triangle = minimum significant difference (MSD). There is a significant difference between a pair of means if the observed difference > MSD.

Table 9 – t-test by time, total coliform

	8 AM	9 AM	10 AM	11 AM	12 PM	1 PM	2 PM	3 PM	4 PM	5 PM	6 PM	7 PM
8 AM	-	0.92	0.93	0.92	0.90	0.90	0.93	0.92	0.93	0.93	0.93	0.94
9 AM	0.07	-	0.93	0.92	0.90	0.90	0.93	0.92	0.93	0.93	0.93	0.94
10 AM	0.02	0.08	-	0.93	0.92	0.92	0.94	0.93	0.94	0.94	0.94	0.95
11 AM	0.01	0.05	0.03	-	0.90	0.90	0.93	0.92	0.93	0.93	0.93	0.94
12 PM	0.03	0.09	0.01	0.04	-	0.89	0.92	0.90	0.92	0.92	0.92	0.93
1 PM	0.13	0.20	0.11	0.15	0.10	-	0.92	0.90	0.92	0.92	0.92	0.93
2 PM	0.02	0.09	0.00	0.03	0.01	0.11	-	0.93	0.94	0.94	0.94	0.95
3 PM	0.10	0.03	0.12	0.08	0.12	0.23	0.12	-	0.93	0.93	0.93	0.94
4 PM	0.11	0.05	0.13	0.10	0.14	0.25	0.13	0.02	-	0.94	0.94	0.95
5 PM	0.21	0.28	0.20	0.23	0.19	0.08	0.19	0.31	0.33	-	0.94	0.95
6 PM	0.04	0.10	0.02	0.05	0.01	0.09	0.02	0.14	0.15	0.18	-	0.95
7 PM	0.10	0.03	0.11	0.08	0.12	0.23	0.12	0.00	0.02	0.31	0.13	-

Lower left triangle = observed differences. Upper right triangle = minimum significant difference (MSD). There is a significant difference between a pair of means if the observed difference > MSD.

Table 10 – t-test by time, human factor

	8 AM	9 AM	10 AM	11 AM	12 PM	1 PM	2 PM	3 PM	4 PM	5 PM	6 PM	7 PM
8 AM	-	1.23	1.23	1.25	1.23	1.23	1.23	1.23	1.23	1.25	1.25	1.25
9 AM	0.02	-	1.23	1.25	1.23	1.23	1.23	1.23	1.23	1.25	1.25	1.25
10 AM	0.13	0.15	-	1.25	1.23	1.23	1.23	1.23	1.23	1.25	1.25	1.25
11 AM	0.20	0.22	0.08	-	1.25	1.25	1.25	1.25	1.25	1.27	1.27	1.27
12 PM	0.00	0.02	0.13	0.20	-	1.23	1.23	1.23	1.23	1.25	1.25	1.25
1 PM	0.10	0.12	0.03	0.11	0.10	-	1.23	1.23	1.23	1.25	1.25	1.25
2 PM	0.25	0.27	0.12	0.05	0.25	0.15	-	1.23	1.23	1.25	1.25	1.25
3 PM	0.04	0.06	0.09	0.16	0.04	0.06	0.21	-	1.23	1.25	1.25	1.25
4 PM	0.19	0.21	0.06	0.02	0.19	0.09	0.06	0.15	-	1.25	1.25	1.25
5 PM	0.13	0.15	0.00	0.07	0.13	0.03	0.12	0.09	0.06	-	1.27	1.27
6 PM	0.12	0.10	0.25	0.33	0.13	0.22	0.38	0.16	0.31	0.25	-	1.27
7 PM	0.13	0.11	0.26	0.33	0.13	0.23	0.38	0.17	0.32	0.26	0.01	-

Lower left triangle = observed differences. Upper right triangle = minimum significant difference (MSD). There is a significant difference between a pair of means if the observed difference > MSD.

4.3. Determining Patterns by Station

It is worth taking a look at how much the stations differ from one another in order to understand differences between indicator concentrations in our twenty time series. To examine differences between the stations, I used dot plots for a visual representation and the Tukey-Kramer test for numerical confirmation.

4.3.1. Dot Plots

The dot plots show that most of the indicator concentration data lies within the same range of values (Figures 14 through 17). Upon visual inspection, there is no single station that has consistently higher or lower values than the others. One exception might be Lorong 8's consistently higher values of total coliform concentration. Concentrations of indicators at CCK also seemed somewhat lower than at the other stations for all four indicators.

The total coliform indicator exhibited the most spread, with the widest range of values within each station. There also seems to be almost as much variation within each station as between stations. For example, Lorong 8's *E. coli* concentrations for the first day of sampling seem overall higher than concentrations measured on its second day of sampling, as seen in Figure 14. Similarly, CCK's second day of sampling has much lower enterococci concentrations than its fourth day of sampling. These differences within each station may show that differences among stations are not as important a driver in explaining variations among the time series.

These dot plots can help determine which time series have outliers or clusters of values that may be neglected in conducting future analyses. For this project, I included all twenty time series, but future work may use trimmed data sets that exclude clouds of data that deviate from the rest.

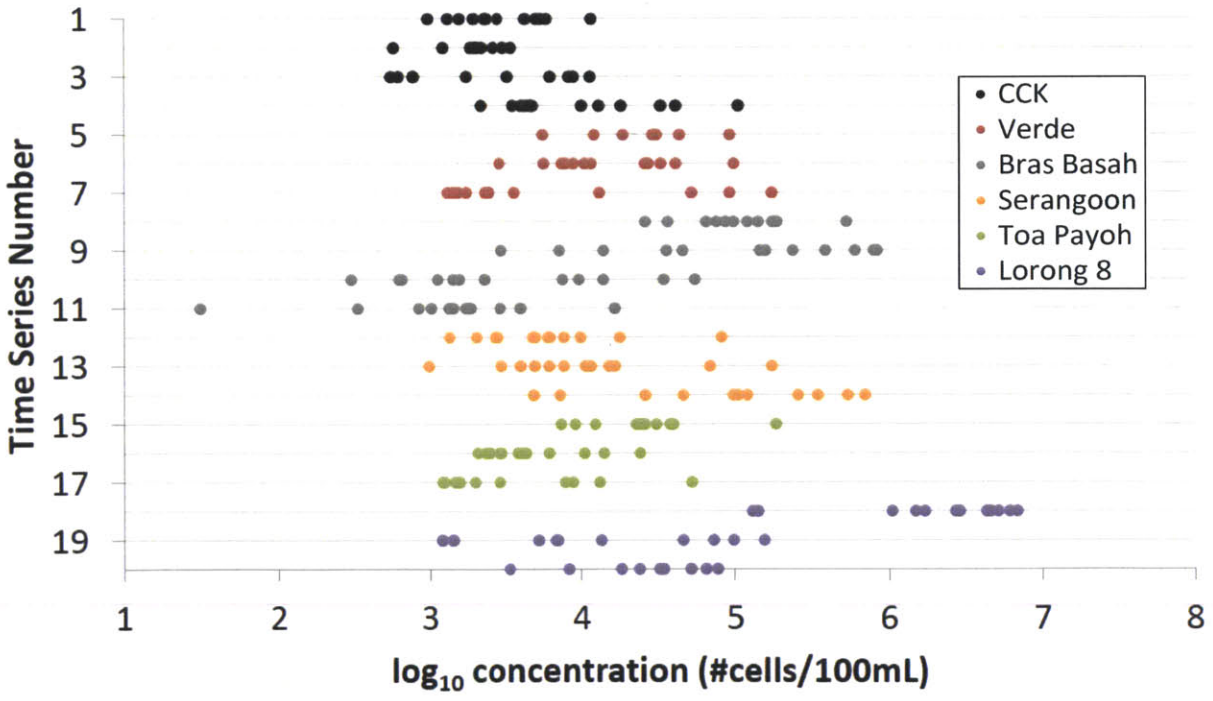


Figure 14 – Dot plot by station, *E. coli*

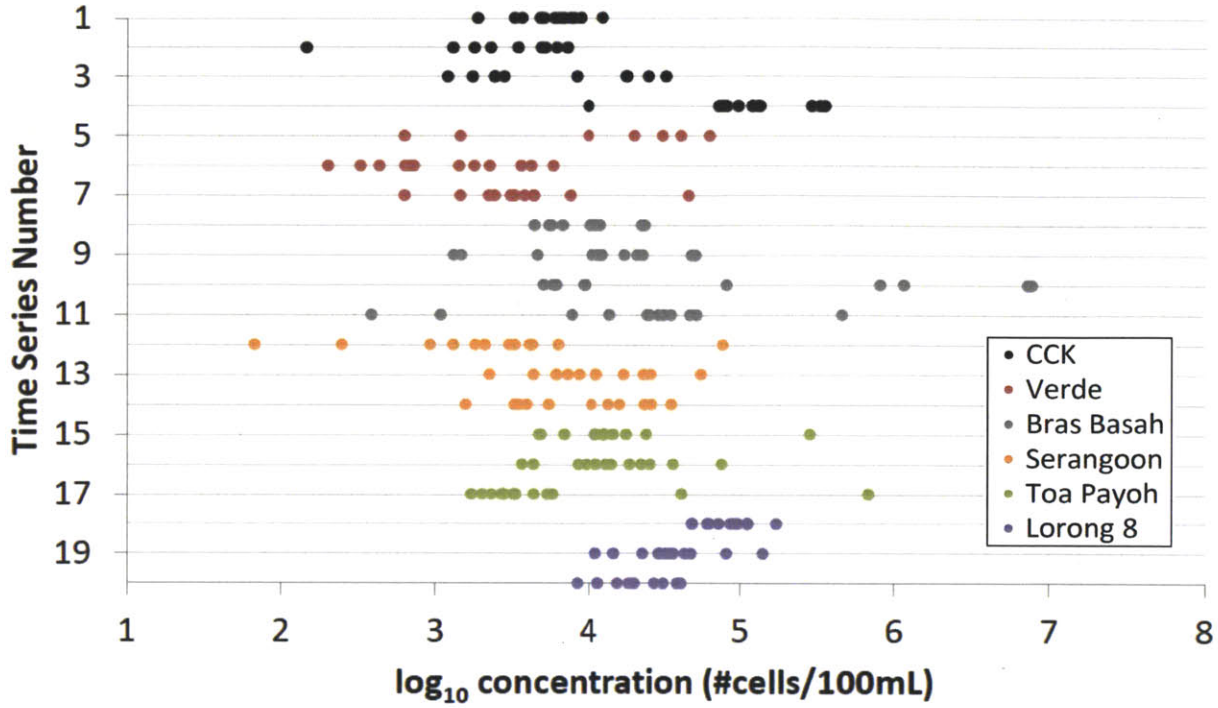


Figure 15 – Dot plot by station, enterococci

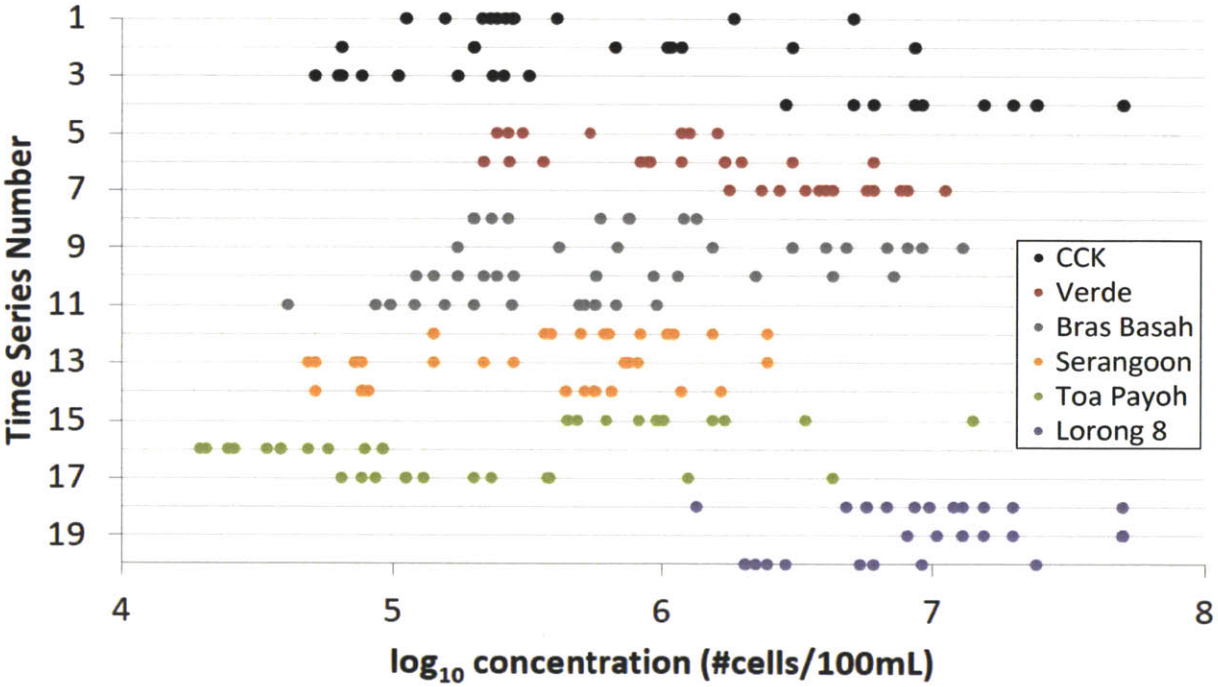


Figure 16 – Dot plot by station, total coliform

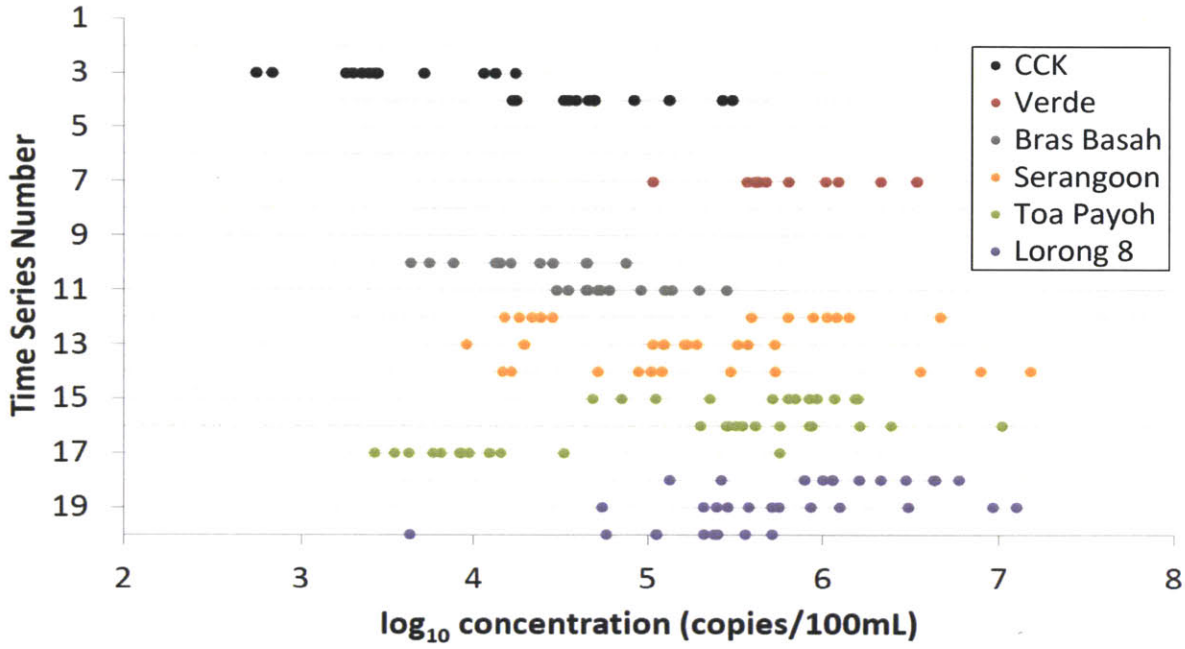


Figure 17 – Dot plot by station, human factor

4.3.2. t-tests

A multi-way t-test (Tukey-Kramer method) provides a numerical description of the difference between stations (Berthouex and Brown 2000). See Section 3.3.7 for more information on the construction and interpretation of these tables. As depicted in Tables 11 through 14, some pairs of stations had statistically significant differences in measured indicator concentrations. Lorong 8 came out consistently different from the other stations. In the case of *E. coli* and total coliform, Lorong 8 was found to be significantly different from all other stations. This could indicate that some characteristic of Lorong 8 causes it to have much different (according to the dot plots, higher) values than the other stations. Future study is recommended to find what is causing high indicator concentrations at this station. Lorong 8 has old sewers and is in a high-density residential area, both of which are characteristics that are hypothesized to yield higher concentrations.

Table 11 – t-test by station, *E. coli*

	CCK	Verde	Bras Basah	Serangoon	Toa Payoh	Lorong 8
CCK	-	0.57	0.51	0.55	0.54	0.56
Verde	0.51	-	0.56	0.59	0.59	0.60
Bras Basah	0.57*	0.06	-	0.54	0.53	0.55
Serangoon	0.66*	0.15	0.09	-	0.57	0.58
Toa Payoh	0.32	0.19	0.26	0.34	-	0.58
Lorong 8	1.4*	0.86*	0.80*	0.71*	1.1*	-

Lower left triangle = observed differences. Upper right triangle = minimum significant difference (MSD). There is a significant difference between a pair of means if the observed difference > MSD.

Table 12 – t-test by station, enterococci

	CCK	Verde	Bras Basah	Serangoon	Toa Payoh	Lorong 8
CCK	-	0.45	0.41	0.44	0.44	0.45
Verde	0.58*	-	0.45	0.48	0.47	0.48
Bras Basah	0.24	0.82*	-	0.43	0.43	0.44
Serangoon	0.29	0.29	0.53*	-	0.46	0.47
Toa Payoh	0.01	0.57*	0.25	0.28	-	0.46
Lorong 8	0.58*	1.2*	0.34	0.87*	0.59*	-

Lower left triangle = observed differences. Upper right triangle = minimum significant difference (MSD). There is a significant difference between a pair of means if the observed difference > MSD.

Table 13 – t-test by station, total coliform

	CCK	Verde	Bras Basah	Serangoon	Toa Payoh	Lorong 8
CCK	-	0.45	0.41	0.44	0.43	0.44
Verde	0.24	-	0.44	0.47	0.47	0.48
Bras Basah	0.16	0.40	-	0.43	0.42	0.44
Serangoon	0.44	0.68*	0.28	-	0.45	0.46
Toa Payoh	0.61*	0.85*	0.45*	0.17	-	0.46
Lorong 8	1.1*	0.82*	1.2*	1.5*	1.7*	-

Lower left triangle = observed differences. Upper right triangle = minimum significant difference (MSD). There is a significant difference between a pair of means if the observed difference > MSD.

Table 14 – t-test by station, human factor

	CCK	Verde	Bras Basah	Serangoon	Toa Payoh	Lorong 8
CCK	-	0.79	0.65	0.59	0.59	0.60
Verde	1.7*	-	0.79	0.75	0.74	0.75
Bras Basah	0.43	1.3*	-	0.59	0.59	0.60
Serangoon	1.2*	0.58	0.73*	-	0.53	0.54
Toa Payoh	1.0*	0.70	0.62*	0.111	-	0.54
Lorong 8	1.6*	0.11	1.2*	0.47	0.58*	-

Lower left triangle = observed differences. Upper right triangle = minimum significant difference (MSD). There is a significant difference between a pair of means if the observed difference > MSD.

4.3.3. Lognormal Probability Plots

There was also much variability in the lognormal distribution of data based on station (Figures 18 through 21). For example, in Figure 18, Lorong 8, represented by a black circle, had an *E. coli* concentration distribution that deviates substantially from a lognormal distribution. Bras Basah, represented by a magenta square, also exhibits some non-lognormal behavior.

In addition to characterizing the distribution of indicator concentrations at the stations, these plots also depict how the magnitude of concentrations differs based on station. Lorong 8 had the highest indicator concentrations for all four indicators as seen by how its plotted points are at the top of all the series plotted. On the other hand, the stations with lowest concentrations vary based on indicator. For *E. coli* and human factor, CCK appears to have the lowest concentrations, while for enterococci, Verde is lowest and for total coliform, Toa Payoh is lowest. Some stations have normal probability plots with steeper slopes than others, suggesting that there is a wider spread of concentrations found at that station.

We might suspect the order of the series from top to bottom would have older sewers stacked at the top and newer sewers in the bottom. This is roughly the case for enterococci, but not necessarily for the other indicators, suggesting other factors than sewer age are at play. This topic of the impact of sewer age is further explored in Section 4.4.

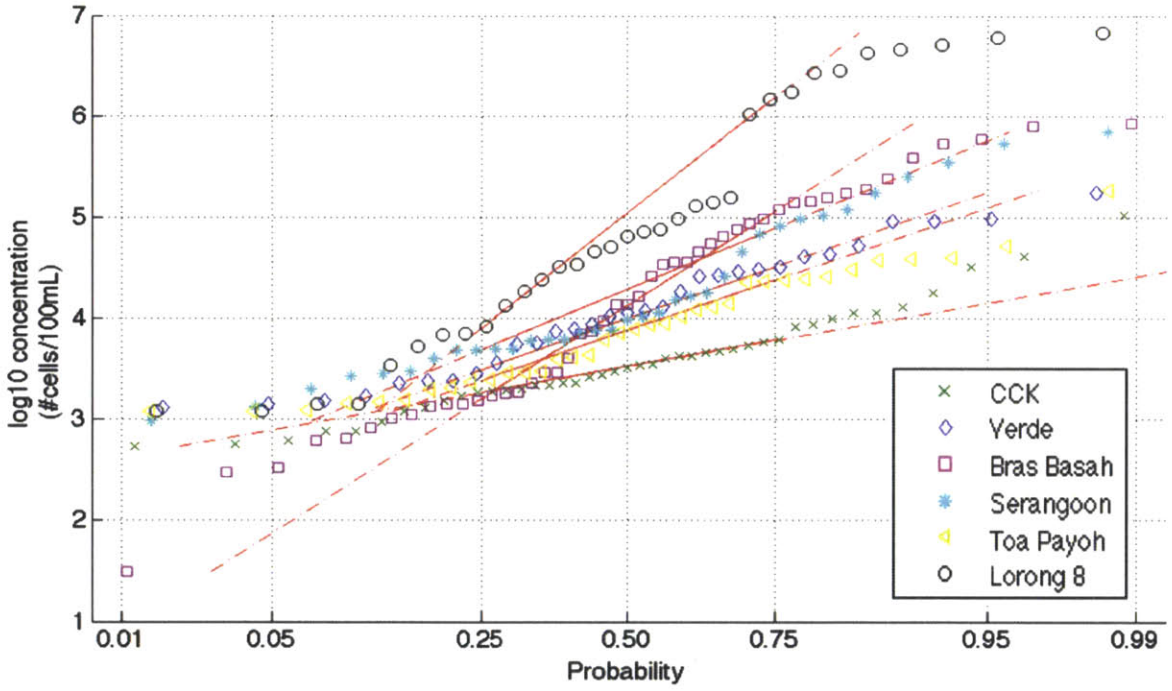


Figure 18 – Lognormal probability plot by station, *E. coli*

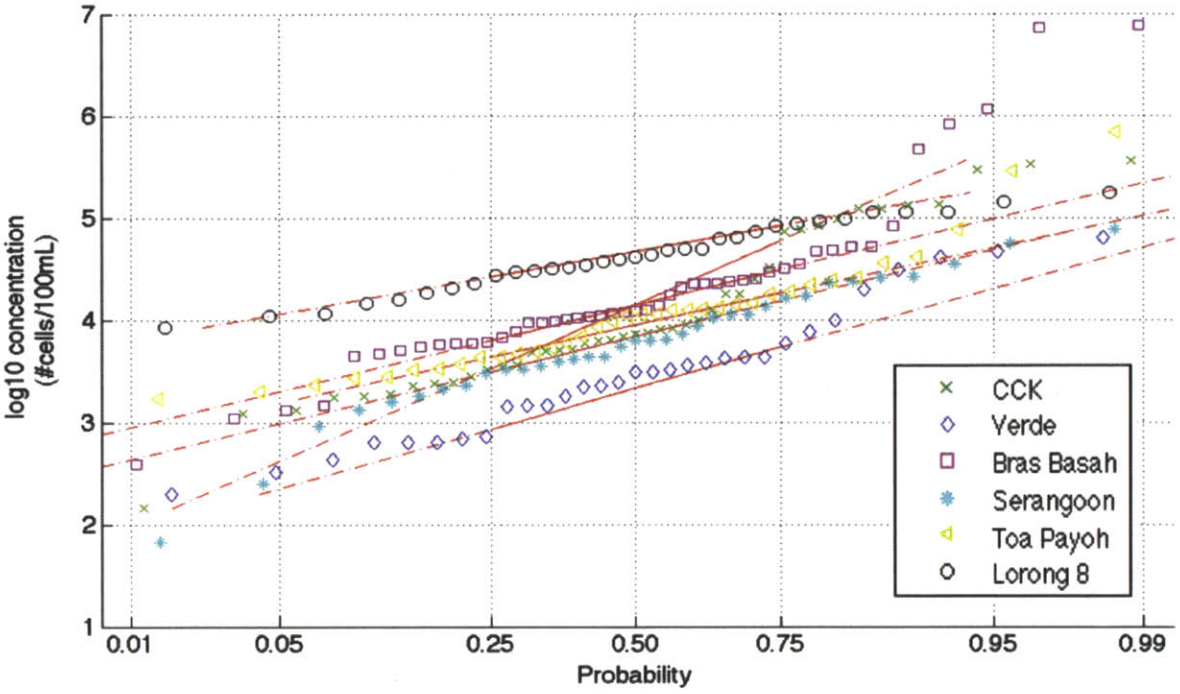


Figure 19 – Lognormal probability plot by station, enterococci

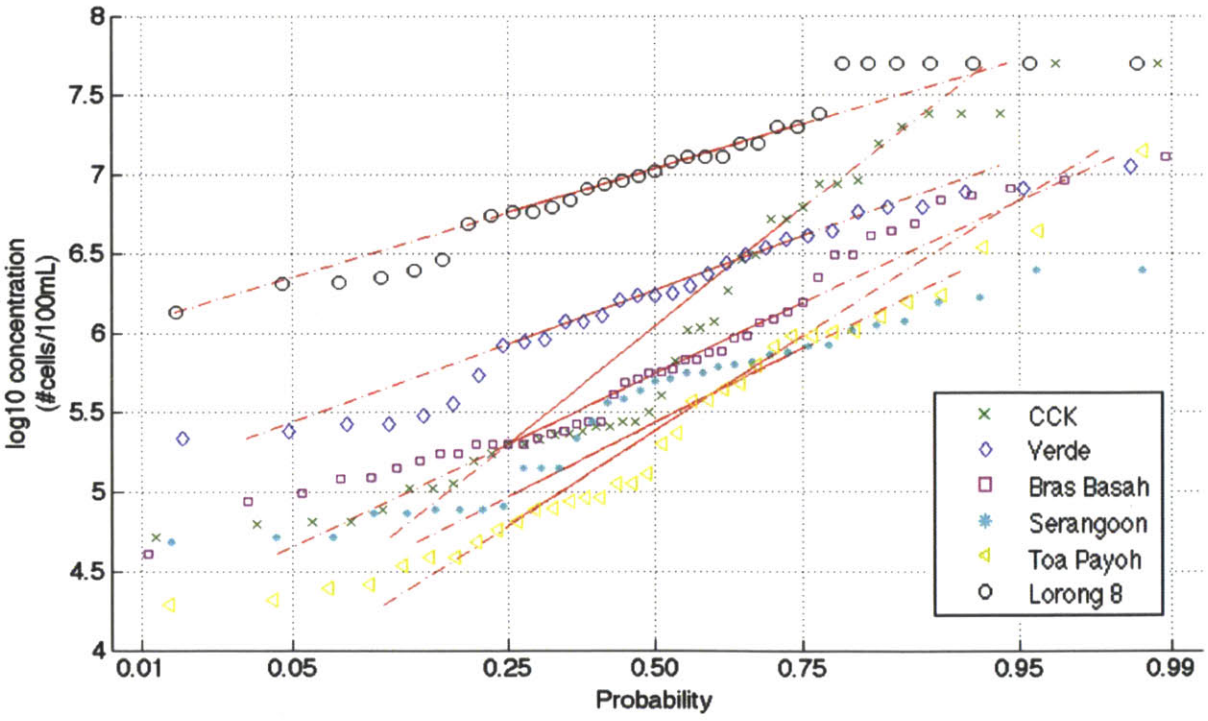


Figure 20 – Lognormal probability plot by station, total coliform

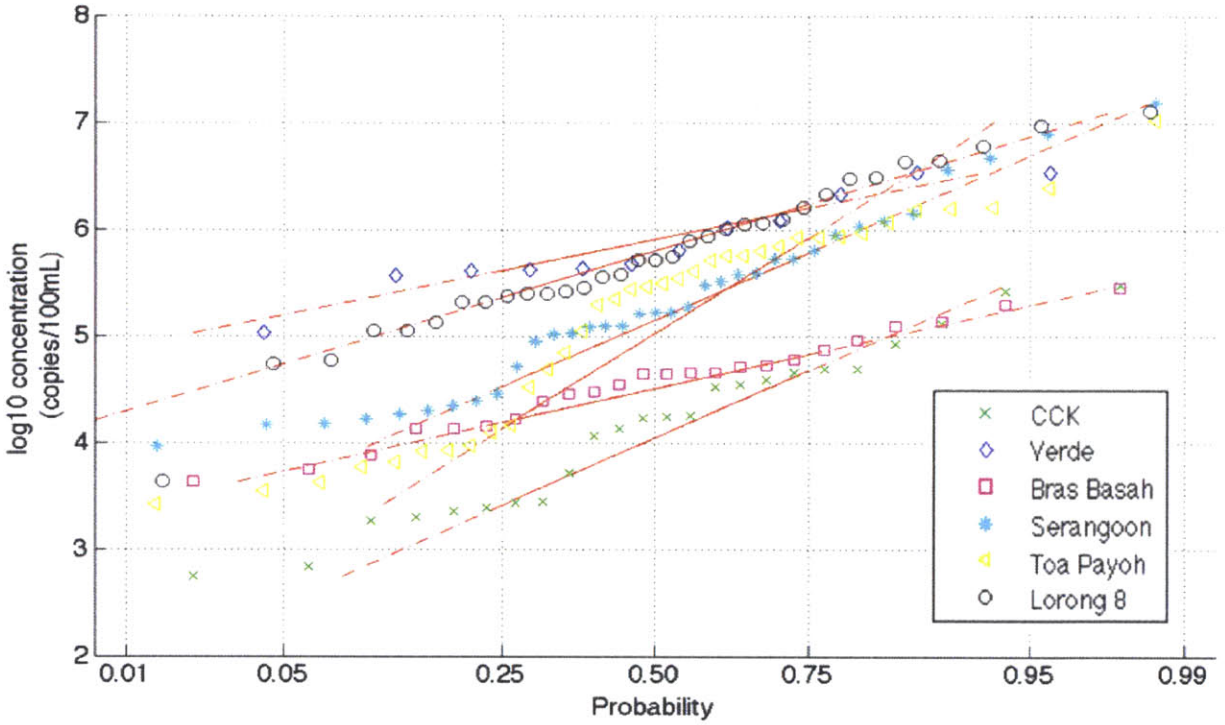


Figure 21 – Lognormal probability plot by station, human factor

4.4. Determining Patterns by Sewer Age

As discussed in Section 1.7, the literature suggests that sewer age has an effect upon indicator concentrations. To see how that catchment parameter affects the indicator concentrations, I used dot plots for a visual representation and the Tukey-Kramer test for numerical confirmation.

4.4.1. Dot Plots

Dot plots by sewer age are shown in Figures 22 through 25. By visual inspection, it appears that for older sewers the points tend to be found at higher concentrations than for newer sewers. This seems to be the case for all but total coliform, which has great variability in its distribution and no visual difference between old and new sewers.

4.4.2. t-tests

In addition to constructing the dot plots, I also conducted an unpaired t-test for unequal variances to test whether old and new sewers are significantly different. The results, in Table 15, show that aside from total coliform, all indicators tested showed a significant difference between old and new sewer concentrations, with a p-value less than 0.05. Total coliform had a p-value greater than 0.05, indicating no significant difference between the groups. However, total coliform is also the weakest of the four indicators in matching true pathogen concentrations, so the finding for total coliform does not negate the finding of a significant difference in the other three cases.

This t-test provides numerical evidence of what is visually presented in the dot plots in the previous section. I conclude that it is likely that there is a difference in bacterial concentrations in areas served by old versus new sewers with areas with older sewers showing higher concentrations in storm drains than areas with newer sewers. This is consistent with published work that older infrastructure has more potential to exfiltrate contaminants via cracks or loose lateral connections than newer sewers.

Table 15 – t-test by sewer age, all indicators

Indicator	p-value	Significant difference (p<0.05)?
E. coli	0.00000830	YES
Enterococci	0.000633	YES
Total coliform	0.153	no
Human factor	0.00878	YES

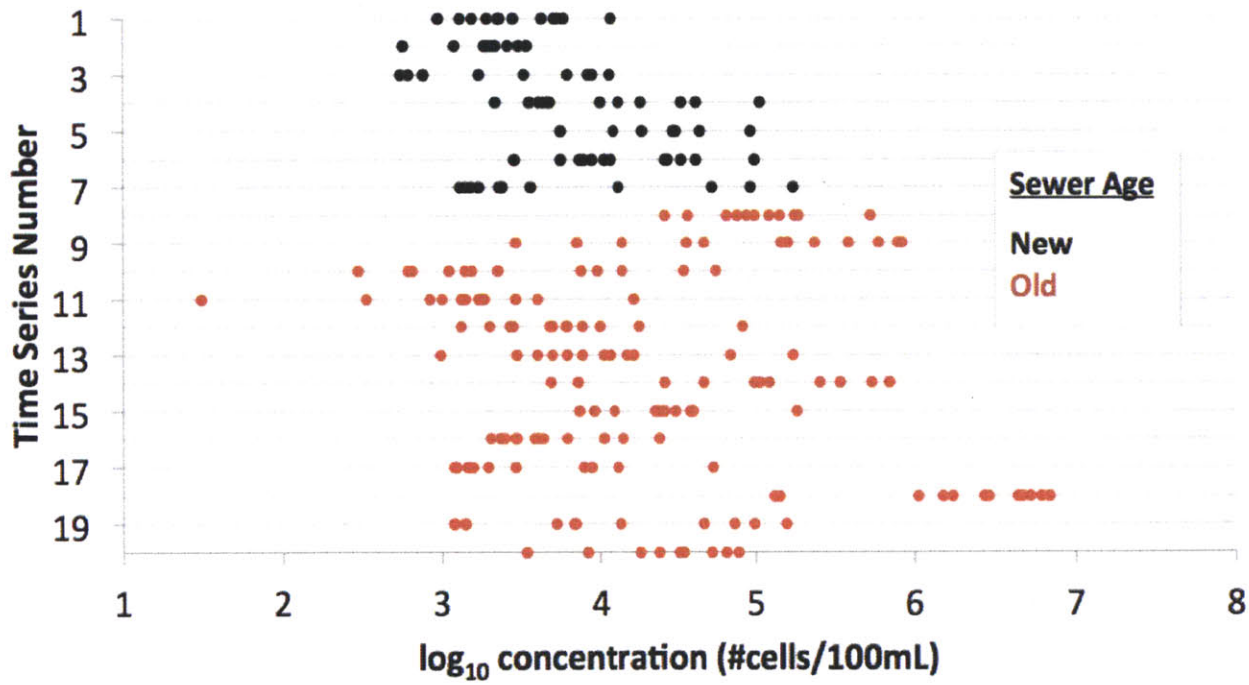


Figure 22 – Dot plot by sewer age, *E. coli*

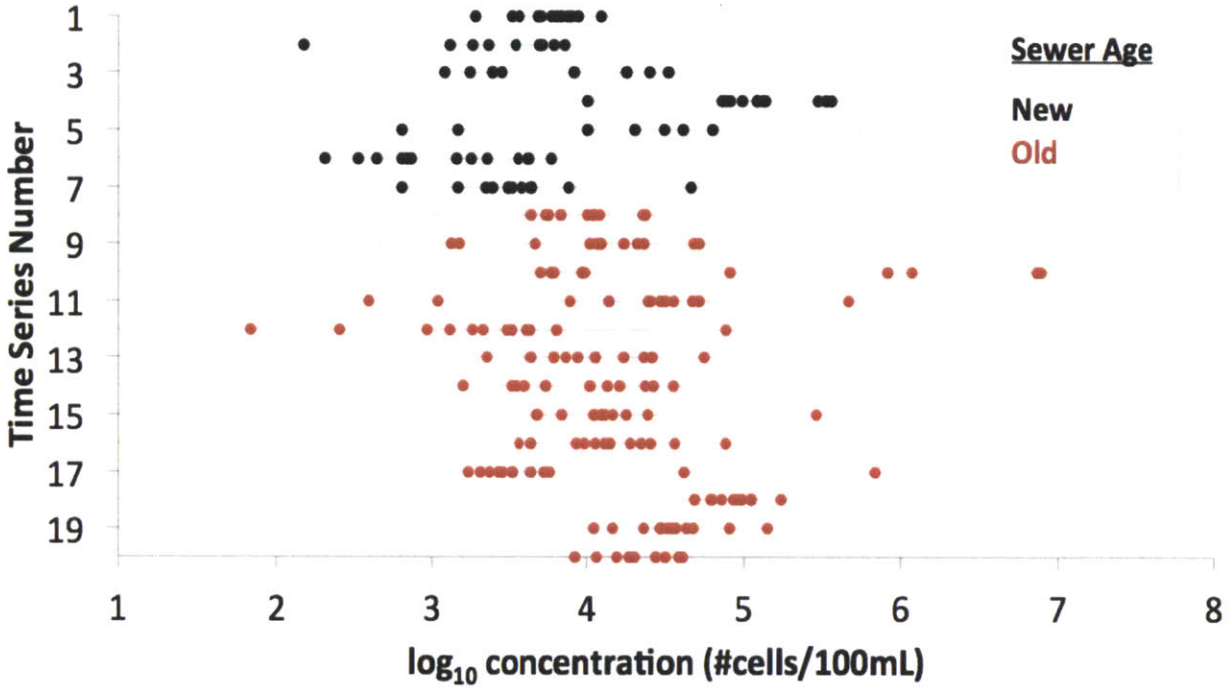


Figure 23 – Dot plot by sewer age, enterococci

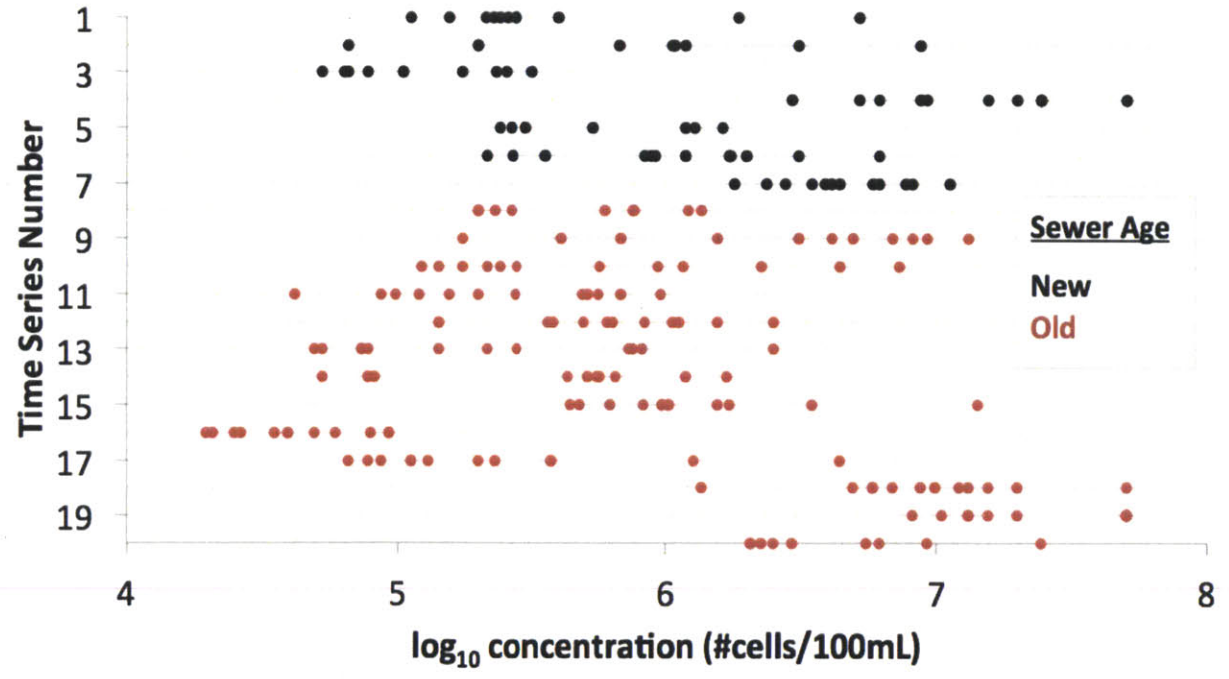


Figure 24 – Dot plot by sewer age, total coliform

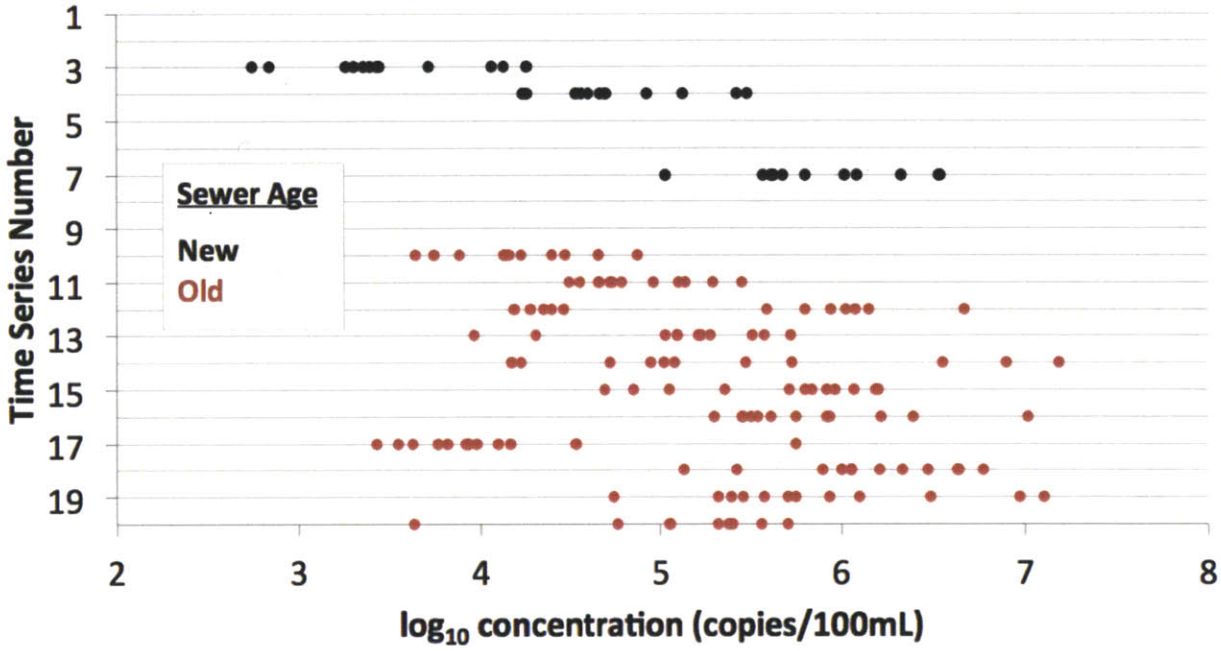


Figure 25 – Dot plot by sewer age, human factor

4.5. Determining Patterns by Land Use

4.5.1. Dot Plots

Dot plots by land use are shown in Figures 26 through 29. By visual inspection, it appears there are no obvious differences in concentration based on land use, with no land-use category having consistently high or low concentration values.

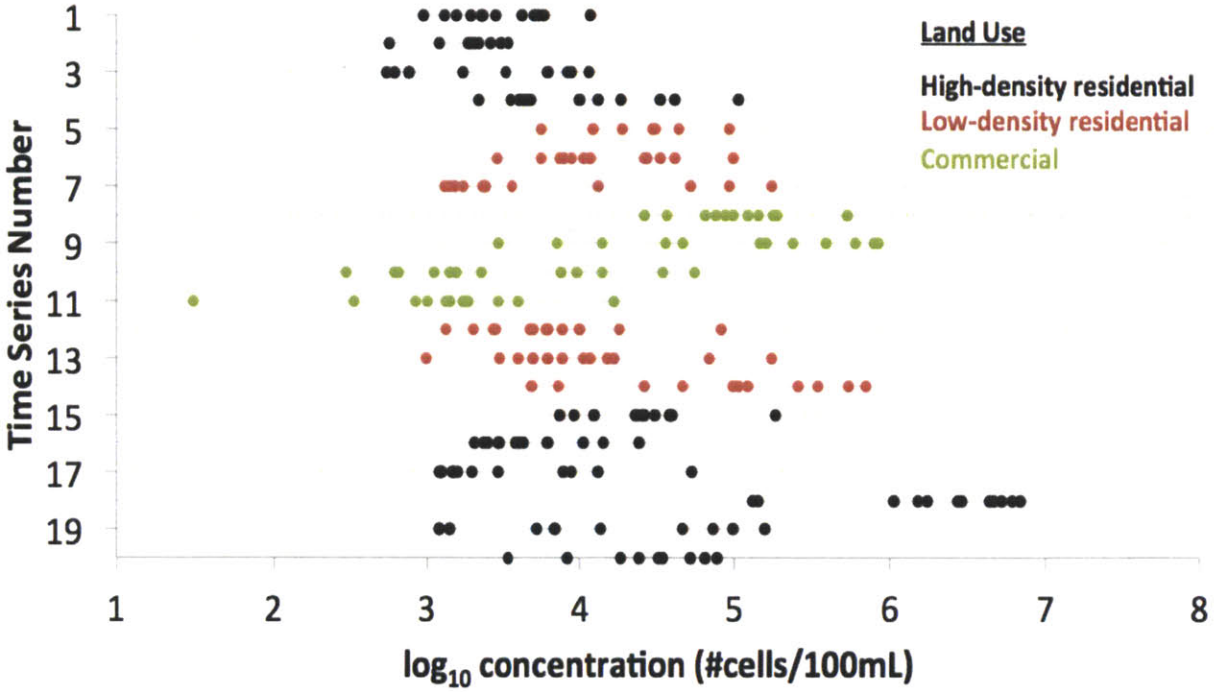


Figure 26 – Dot plot by land use, *E. coli*

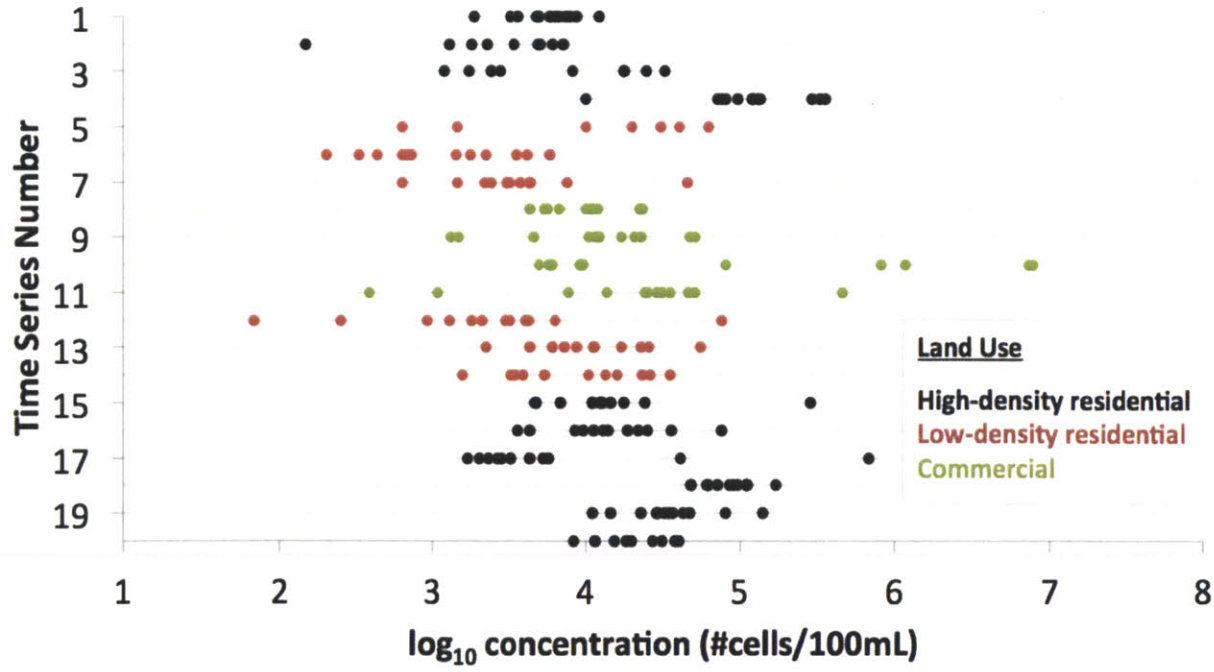


Figure 27 – Dot plot by land use, enterococci

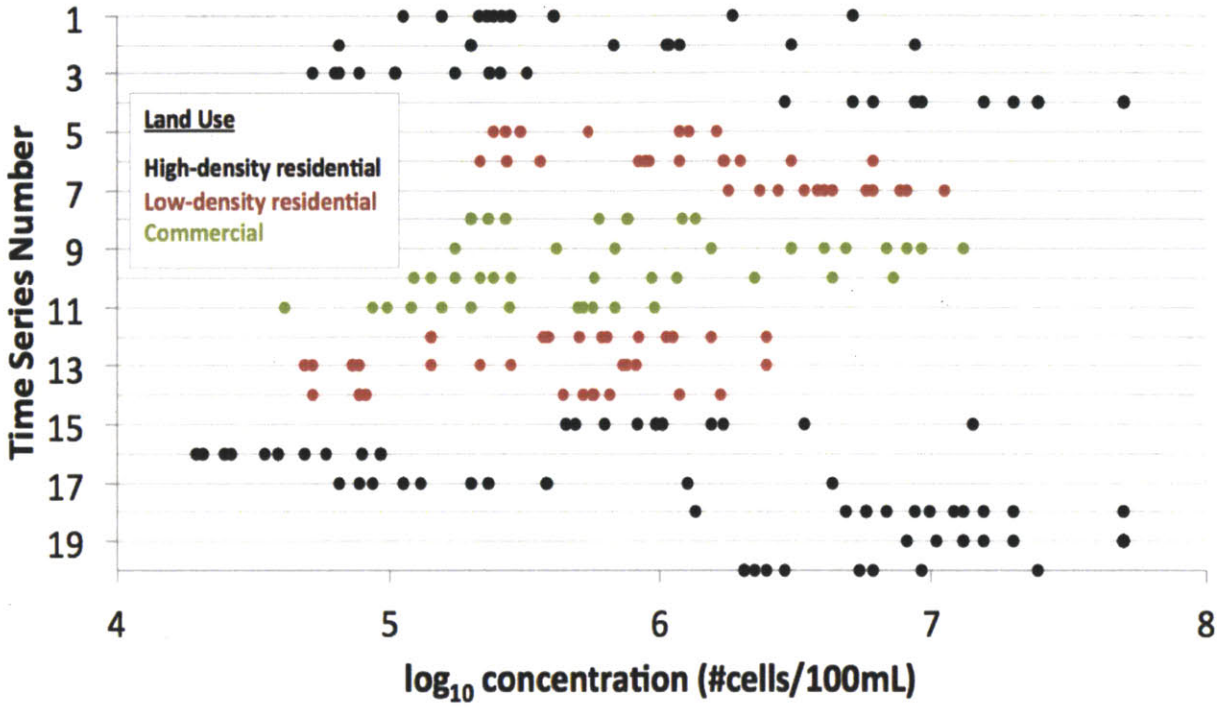


Figure 28 – Dot plot by land use, total coliform

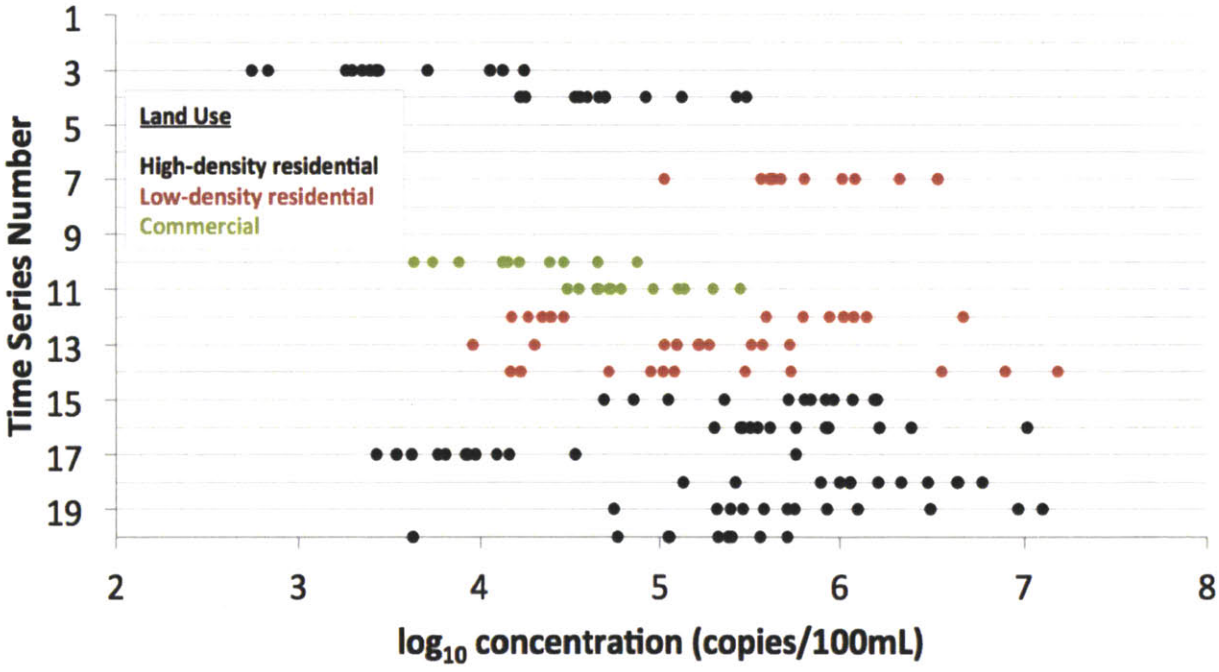


Figure 29 – Dot plot by land use, human factor

4.5.2. t-tests

A multi-way t-test (Tukey-Kramer method) was used to test whether there are significant differences in indicator concentrations based on land use (Berthoeux and Brown 2000). See Section 3.3.7 for more information on the construction and interpretation of these tables.

I found that there are mixed results, with significant differences for indicators enterococci and human factor (Tables 16 through 19). For enterococci, low-density residential areas exhibit significantly different concentrations than high-density residential and commercial areas. For human factor, commercial areas exhibit significantly different concentrations than high-density and low-density residential areas. There is almost a significant difference between high-density residential and low-density residential areas, as well, for the indicator human factor. As human factor is likely the best indicator of true pathogen concentrations, this suggests that land use is a factor in differences in concentrations.

Table 16 – t-test by land use, *E. coli*

	High-density	Low-density	Commer-cial
High-density	-	0.34	0.38
Low-density	0.08	-	0.42
Commer-cial	0.07	0.02	-

Lower left triangle = observed differences. Upper right triangle = minimum significant difference (MSD). There is a significant difference between a pair of means if the observed difference > MSD.

Table 17 – t-test by land use, enterococci

	High-density	Low-density	Commer-cial
High-density	-	0.25	0.29
Low-density	0.60	-	0.31
Commer-cial	0.07	0.67	-

Lower left triangle = observed differences. Upper right triangle = minimum significant difference (MSD). There is a significant difference between a pair of means if the observed difference > MSD.

Table 18 – t-test by land use, total coliform

	High-density	Low-density	Commer-cial
High-density	-	0.30	0.34
Low-density	0.24	-	0.38
Commer-cial	0.28	0.04	-

Lower left triangle = observed differences. Upper right triangle = minimum significant difference (MSD). There is a significant difference between a pair of means if the observed difference > MSD.

Table 19 – t-test by land use, human factor

	High-density	Low-density	Commer-cial
High-density	-	0.39	0.50
Low-density	0.33	-	0.54
Commer-cial	0.55	0.88	-

Lower left triangle = observed differences. Upper right triangle = minimum significant difference (MSD). There is a significant difference between a pair of means if the observed difference > MSD.

Chapter 5 Conclusions and Recommendations

5.1. Summary and Conclusions

This study examined 20 twelve-hour time series, with four indicator concentrations (*E. coli*, enterococci, total coliform, and human factor) measured every hour from 8 am to 7 pm. Six stations in Singapore (Choa Chu Kang Crescent, Verde, Bras Basah, Serangoon, Toa Payoh, Lorong 8 Toa Payoh) were evaluated, across three land-use categories (high-density residential, low-density residential, commercial), and two sewer-age categories (new, old). Sampling dates spanned the winter (January) and summer (June, July) in years 2011 and 2012.

The first objective of this project was to characterize the distribution of the concentrations of the four fecal indicators: *E. coli*, enterococci, total coliform, and human factor. To do so, I used histograms and lognormal probability plots. In agreement with previous studies conducted in temperate climates that found bacterial indicators are lognormally distributed (Aragao, 2007; EPA, 2010; Pontius, 2003), I found that the distributions of *E. coli*, enterococci, and total coliform were roughly lognormally distributed in this case of concentrations measured in the tropical climate of Singapore across a variety of land uses. This shows that bacterial indicator concentration distributions are described by the same statistical model in tropical climates as in temperate climates, and with a wide and varied data set. In the case of human factor, an indicator for which there is emerging research, I conclude that the human factor indicator is roughly lognormally distributed, as well. This has implications for which statistical tests can be used on the indicator concentration data, as many tests require normality. Also, the lognormal distribution can serve as a simple description for the complex temporal and spatial variation in measured indicator concentrations.

My second objective was to characterize patterns in the time series of indicator concentrations. This analysis was conducted with an artificial time series of all twenty half-days in a row. Statistical methods used in this project include a time-series plot, autocorrelation, correlation matrices, dot plots, and t-tests. Contrary to my hypothesis that there would be a diurnal pattern, I found no such pattern in the indicator concentrations using this expanded data set. The time-series plot depicted no obvious patterns. Autocorrelation and correlation matrix analysis showed correlation between adjacent and near-adjacent observations in time, but no diurnal peaks in correlation. The dot plots showed no regular cyclic variation and no time of day had consistently high or low values. The t-tests showed no correlation between one hour of day and any other.

This disagrees with findings of a diurnal pattern from the Toolbox Study conducted by the Public Utilities Board (Ekklesia 2011). The Toolbox Study consisted of a 72-hour series of hourly measurements for the indicators total coliform and enterococci. This finding is significant because it suggests that sewer leakage may not be the main contributor of fecal contamination in the stormwater. If sewer leakage was the main reason for high indicator concentrations, I would have expected peaks of bacteria concentration at times of peak sewer flow. It may be worth investigating sewer flows in Singapore to see how closely they follow the diurnal pattern shown by studies in the United States.

My third objective was to examine how indicator concentrations varied between stations. I used dot plots, t-tests, and lognormal probability tests. The dot plots showed Lorong 8 has consistently higher values for enterococci, total coliform, and human factor, and CCK had generally lower values for *E. coli*, enterococci, and human factor. In some cases, there was more variation within each station's time series than across stations; for example, CCK's second day

of sampling revealed much lower enterococci concentrations than its fourth day of sampling. These differences within stations show that variability within each station's measured concentrations can be a more important driver than variability between stations. In other words, difference between stations is not as important as I had expected. The t-tests showed that there are some significant differences in measured concentrations across stations. In many cases, Lorong 8 concentrations were significantly different from other stations in the same sewer age or land use category. The lognormal probability plots depict how the magnitude of concentrations differs based on station. The station with the highest measured concentrations for all indicators was Lorong 8. The station with the lowest measured concentrations for *E. coli* and human factor was CCK. The station with the lowest enterococci concentrations was Verde. The station with the lowest total coliform concentrations was Toa Payoh. Some stations have normal probability plots with steeper slopes than others, suggesting that there is a wider spread of concentrations at those stations. We might suspect the order of the series on the normality plots from top to bottom would have older sewers stacked at the top and newer sewers in the bottom. This is roughly the case for enterococci, but not necessarily for the other indicators, suggesting other factors than sewer age are at play. One possibility, based on field observations, could be proximity to food retail centers. The Lorong 8 and Bras Basah sampling locations are not far from eating establishments and trash collection bins, which can act as sources of contamination. Lorong 8 was also observed to have sewage odors in the perimeter drains of the nearby housing center. Such sewer odor and contamination problems are likely due to faulty sewers in specific locations.

This look at how the indicator concentrations vary between stations can inform future sampling. Stations with high variability between twelve-hour time series, such as CCK, should have more indicator measurements taken to better represent the mean value. Stations with high concentrations, such as Lorong 8, can be further investigated to track what may cause higher fecal contamination levels.

My last objective was to determine how the two catchment parameters of sewer age and land use affected indicator concentrations. I used dot plots and t-tests. First, looking at sewer age, for all indicators but total coliform, older sewers had a significantly higher indicator concentration than newer sewers. This suggests that sewer leakage likely contributes to fecal contamination. On the other hand, there are no diurnal patterns in indicator concentrations, as might be expected given the diurnal patterns in sewage flow. It may be that sewer flow pathways to storm drains eliminate any diurnal patterns. More research is recommend to evaluate the connectivity and time patterns of flow between sewers and storm drains.

Next, looking at land use, there are mixed findings. For enterococci, low-density residential areas exhibit significantly different concentrations than high-density residential and commercial areas. For human factor, commercial areas exhibit significantly different concentrations than high-density and low-density residential areas. As human factor might be the best indicator of true pathogen concentrations, this suggests that land use plays a role in differences in concentrations.

In summary, this study used extensive indicator concentration data to describe spatial and temporal patterns in four fecal indicator concentrations: *E. coli*, enterococci, total coliform, and human factor. I found the distributions for all four indicators to be approximately lognormal. There was no obvious time pattern in measured concentrations, except that one hour's concentration may roughly correlate with the next 1 to 3 hours' concentrations. Old sewers have significantly different concentrations than new sewers, and land use has some effect upon

indicator concentration. Leaking sewers can explain some of the high indicator concentrations, though lack of diurnal pattern suggests there are more factors at play than just older sewers dispersing contaminants.

5.2. Recommendations for Additional Work

Future studies can solidify these findings. First, to further investigate the impact of sewer age on fecal contamination levels, I suggest measuring sewerage flow at these locations to characterize its diurnal pattern. A weak or no diurnal sewage flow pattern could explain why I saw older sewers to have higher indicator concentrations, but no diurnal pattern in the indicator concentrations.

I also suggest collecting more concentration data at locations that show high variability between twelve-hour time series, like CCK, to get a better representation of the mean value. Future study may also try to find what is causing high indicator concentrations at Lorong 8. This location has old sewers and is in a high-density residential area, both categories which are thought to yield higher concentrations, but there may be other influencing factors. Possible sources of contamination may be nearby food retail centers or trash collection bins. Lorong 8 also had sewage odors in the drains, discovered in field observations. Once characterized, any influencing factors for higher indicator concentrations can be used to describe and predict the probable relative magnitude of bacteria concentration at any new location.

Further, expanding the study to include more sampling locations of new sewer age and commercial land use can better describe the large-scale trends in the impact of sewer age and land use on indicator concentrations. Currently, there are only two stations in the "new" sewer-age category compared to four stations in the "old" sewer-age category. Similarly, Bras Basah is the only "commercial" land-use station. The correlations based on catchment parameters of sewer age and land use can be more complete with more stations included in each sewer-age and land-use category.

The bacterial indicator concentrations I found were also on the whole quite high and in many cases approached those in raw sewage (see Table 1), including the raw sewage information found by Ekklesia (2012) for a one-day grab sample in Singapore. *E. coli* $\log_{10}$ concentrations (#cells/100mL) ranged from 0 to 8, whereas in raw sewage it would be around 8.7 (Evison and James 1973) or 6.6 (Ekklesia 2012). Enterococci $\log_{10}$ concentrations (#cells/100mL) ranged from 1 to 7, whereas in raw sewage it would be around 4-5 (Gerba and Pepper 2005) or 5.6 (Ekklesia 2012). Total coliform $\log_{10}$ concentrations (#cells/100mL) ranged from 4 to 8, while in raw sewage it would be around 7-9 (Gerba and Pepper 2005) or 6.9 (Ekklesia 2012). Human factor $\log_{10}$ concentrations (copies/100mL) ranged from 2 to 8, whereas in raw sewage it would be around 8.9 (Silkie and Nelson 2009) or 8.0 (Ekklesia 2012). The samples collected did not, at least in appearance or odor, seem to be equivalent to raw sewage, so the high indicator concentrations do not necessarily mean high pathogen concentrations. Future work might measure indicator concentrations in more samples of raw sewage in Singapore to get a fuller understanding of the actual concentrations, adding to the single value measured by Ekklesia (2012). Another suggestion would be to conduct tracing in the drain system to try to identify sewage sources. One place to start can be Lorong 8, where sewage odors can lead field investigators to a source of contamination.

Lastly, while this study looked at each of the four indicator concentrations separately, it may be worth investigating how similar or different those indicator concentration patterns or

averages are from one another. A first look at this is included in Appendix III and finds that there is low correlation between indicators.

This project has been a first look at patterns in indicator concentrations collected to this point in Singapore by Master of Engineering students from the Massachusetts Institute of Technology, in collaboration with Nanyang Technological University. Future work can continue the search for appropriate indicators for selected land uses and sewer ages, the pinpointing of sources of contamination of storm drains, and ultimately control contamination such that stormwater-filled reservoirs may be opened for public recreation.

References

Aragao, G. M. F., Corradini, M. G., Normand, M. D., & Peleg, M. (2007). Evaluation of the Weibull and log normal distribution functions as survival models of Escherichia coli under isothermal and non isothermal conditions. *International journal of food microbiology, 119*(3), 243-257.

BBN Corporation (Producer). (1996). PROPHET StatGuide: Examples of normal probability plots. Retrieved from http://www.basic.northwestern.edu/statguidefiles/probplots.html

Berthouex, P. M., & Brown, L. C. (2002). *Statistics for Environmental Engineers* (2nd ed.). Boca Raton, Florida: CRC Press.

Bigger, J. W. (1937). The growth of coliform bacilli in water. *Journal of Pathology and Bacteriology, 44*(1), 167-211.

Byappanahalli, M. N., & Fujioka, R. S. (1998). Evidence that tropical soil environment can support the growth of Escherichia coli. *Water Science and Technology, 38*(12), 171-174.

Desai, A. M., & Rifai, H. S. (2010). Variability of Escerichia coli Concentrations in an Urban Watershed in Texas. *Journal of Environmental Engineering, ASCE, 136*(12), 1347-1359.

Ekklesia, E. (2011). *Identification of Suitable Indicators for Tracing Human Fecal Contamination and Sewage in Urban Catchments.* Confirmation of Candidature Report, Department of Civil and Environmental Engineering, Nanyang Technological University, Singapore. 113 pp.

Ekklesia, E. (2012). Personal communication.

Ellis, J. B., Revitt, D. M., Blackwood, D. J., & Gilmour, D. (2004). Leaky sewers: assessing the hydrology and impact of exfiltration in urban sewers. *Hydrology: Science & Practice for the 21st Century, 2*, 266-271.

Ellis, J. B., Revitt, D. M., Lister, P., Willgress, C., & Buckley, A. (2003). Experimental studies of sewer exfiltration. *Water Science and Technology, 47*(4), 61.

Enfinger, K. L., & Stevens, P. L. (2006). Sewer Sociology - The Days of Our (Sewer) Lives. *Proceedings of the Water Environment Federation*, WEFTEC 2006, Dallas, Texas, USA. October 21-25, 2006. pp. 6962-6974.

EPA. (2010). *Sampling and Consideration of Variability (Temporal and Spatial) for Monitoring of Recreational Waters.* (EPA-823-R-10-005). Office of Water, U.S. Environmental Protection Agency. Retrieved from http://water.epa.gov/scitech/swguidance/standards/criteria/health/recreation/upload/P12-MonRept-final_508.pdf.

Evison, L. M., & James, A. (1973). A comparison of the distribution of intestinal bacteria in British and East African water sources. *Journal of Applied Bacteriology, 36*, 109-118.

Fadem, B. (2008). *High-Yield Behavioral Science (High-Yield Series).* Hagerstown, MD: Lippincott Williams & Wilkins.

Foley, K., Granger, E., Sagel, A. M., Nshimyimana, J. P., & Sudhalkar, A. (2010). *Characterization of Water Quality in Kranji Reservoir Catchment: Report on Actvities through May 2010.* Master of Engineering Program, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts. August 2010.

Fujioka, R. S., Sian-Denton, C., Borja, M., Castro, J., & Morphew, K. (1999). Soil: the environmental source of Escherichia coli and enterococci in Guam's streams. *Journal of Applied Microbiology Symposium Supplement, 85*, 83S-89S.

Fujioka, R. S., Tenno, K., & Kansako, S. (1988). Naturally occuring fecal coliforms and fecal streptococci in Hawaii's freshwater streams. *Toxicity Assessment, 3*, 613-630.

Gerba, C., & Pepper, I. (2005). *Environmental Microbiology*. Burlington, MA: Elsevier Academic Press.

Gleeson, C., & Gray, N. (1997). *The Coliform Index and Water-borne Disease*. London: E and FN Spon.

Hardie, J. M., & Whiley, R. A. (1997). Classification and overview of the genera Streptococcus and Enterococcus. *Journal of Applied Bacteriology Symposium Supplement, 83*(1S-11S).

Hardina, C. M., & Fujioka, R. S. (1991). Soil: the environmental source of E. coli and enterococci in Hawaii's streams. *Environmental Toxicology and Water Quality, 6*, 185-195.

Hazen, T. C. (1988). Fecal coliforms as indicators in tropical waters: A review *Toxicity Assessment: An International Journal, 3*(5), 461-477.

IDEXX. (2008a). Enterolert Test Kit: IDEXX Laboratories, Inc., Westbrook, Maine, USA.

IDEXX. (2008b). Quanti-Tray/2000 MPN Table: IDEXX Laboratories, Inc., Westbrook, Maine, USA.

Kenzaka, T., Yamaguchi, N., Prapagdee, B., Mikami, E., & Nasu, M. (2001). Bacterial Community Composition and Activity in Urban Rivers in Thailand and Malaysia. *Journal of Health Science, 47*(4), 353-361.

Leclerc, H., Devriese, L. A., & Mossel, D. A. A. (1996). Taxonomic changes in intestinal (faecal) enterococci and streptococci: consequences on their use as indicators of faecal contamination in drinking water. *Journal of Applied Bacteriology, 81*, 459-466.

McDonald, J. H. (2009). *Handbook of Biological Statistics* (2nd ed.). Baltimore, Maryland: Sparky House Publishing.

Nshimyimana, J. P., Ekklesia, E., Shanahan, P., Chua, L., & Thompson, J. (in preparation). *Distribution and abundance of human specific Bacteroides and relation to traditional indicators in an urban tropical catchment.* Massachusetts Institute of Technology, Nanyang Technological University.

Pontius, F. W. (2003). *Drinking Water Regulation and Health*: John Wiley & Sons.

PUB Singapore. (2011). Overview of Singapore's Drainage Management Approach. Retrieved December 9, 2011, from http://www.pub.gov.sg/general/Documents/overview_DrainageMgmt.pdf

Rao, N. U., Parhad, N. M., Rao, C. S. G., & Rao, K. S. (1968). Coliforms as indicators of fecal pollution. *Environmental Health, 10.*

Reynolds, J., & Barrett, M. (2003). A review of the effects of sewer leakage on groundwater quality. *Water and Environment Journal, 17*(1), 34-39.

Rivera, S. C., Hazen, T. C., & Toranzos, G. A. (1988). Isolation of fecal coliforms from pristine sites in a tropical rain-forest. *Applied and Environmental Microbiology, 54*(2), 513-517.

Rutsch, M., Rieckermann, J., & Krebs, P. (2006). Quantification of sewer leakage: a review. *Water Science and Technology, 54*(6-7), 135.

Savichtcheva, O., Okayama, N., & Okabe, S. (2007). Relationships between Bacteroides 16S rRNA genetic markers and presence of bacterial enteric pathogens and conventional fecal indicators. *Water Research, 41*(16), 3615-3628.

Sercu, B., Werfhorst, L. C. V. D., & Holden, P. A. (2009). Storm Drains are Sources of Human Fecal Pollution during Dry Weather in Three Urban Southern California Watersheds. *Environmental Science & Technology, 43*(2), 293-298.

Seurinck, S., Defoirdt, T., Verstraete, W., & Siciliano, S. D. (2005). Detection and quantification of the human-specific HF183 Bacteroides 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater. *Environmental Microbiology, 7*(2), 249-259.

Silkie, S. S., & Nelson, K. L. (2009). Concentrations of host-specific and generic fecal markers measured by quantitative PCR in raw sewage and fresh animal feces. *Water Research, 43*(19), 4860-4871.

Soon, T. Y., Jean, L. T., & Tan, K. (2009). *Clean, Green and Blue*. Singapore: ISEAS.

Srinivasan, S., Aslan, A., Xagoraraki, I., Alocilja, E., & Rose, J. B. (2011). Escherichia coli, enterococci, and Bacteroides thetaiotaomicron qPCR signals through wastewater and septage treatment. *Water Research, 45*(8), 2561-2572.

Traister, E., & Anisfeld, S. C. (2006). Variability of Indicator Bacteria at Different Time Scales in the Upper Hoosic River Watershed. *Environmental Science & Technology, 40*(16), 4990-4995.

Varma, M., Field, R., Stinson, M., Rukovets, B., Wymer, L., & Hauglanda, R. (2009). Quantitative real-time PCR analysis of total and propidium monoazide-resistant fecal indicator bacteria in wastewater. *Water Research, 43*(19), 4790-4801.

Vollertsen, J., & Hvitved-Jacobsen, T. (2003). Exfiltration from gravity sewers: a pilot scale study. *Water Science and Technology, 47*(4), 69.

von Hippel, P. T. (2010). Skewness. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science*. New York: Springer.

Wakida, F. T., & Lerner, D. N. (2005). Non-agricultural sources of groundwater nitrate: a review and case study. *Water Research, 39*(1), 3-16.

Webster, L. F., Thompson, B. C., Fulton, M. H., Chestnut, D. E., Van Dolah, R. F., Leight, A. K., & Scott, G. I. (2003). Identification of sources of Escherichia coli in South Carolina estuaries using antibiotic resistance analysis. *Journal of Experimental Marine Biology and Ecology, 298*(2), 179-195.

Wexler, H. M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. *Clinical Microbiological Reviews, 20*, 593-621.

Yang, Y., Lerner, D., Barrett, M., & Tellam, J. (1999). Quantification of groundwater recharge in the city of Nottingham, UK. *Environmental Geology, 38*(3), 183-198.

Appendix I – Land Use Categorization

Table 20 – Land use categorization

	High-density residential			**Low-density residential**		**Commer-cial**
	Toa Payoh Lorong 8	*Toa Payoh*	*CCK Crescent*	*Serangoon Garden*	*Verde*	*Bras Basah*
Total area (ha)	24	30	37	31	7	16
% by land use type:						
High-density residential	**54**	**77**	**84**	0	0	0
Low-density residential	0	0	0	**78**	**76**	0
Commercial (including hotel, civic & community institution, educational institution, place of worship)	23	5	4	0	0	**65**
Industrial (business)	0	0	0	0	0	0
Agricultural	0	0	0	0	0	0
Transportation (road, light rapid transit), utility	15	18	11	19	24	31
Sports and recreation	0	0	0	2	0	0
Others (including park, reserve site, open space in urban, water body)	8	0	2	1	0	4

Appendix II – Sampling Locations

Figures 30 – 35 supplement the map of the six sampling stations depicted in Figure 4. The red star marks the actual sampling location, while the rest of the map shows the contributing drainage area.

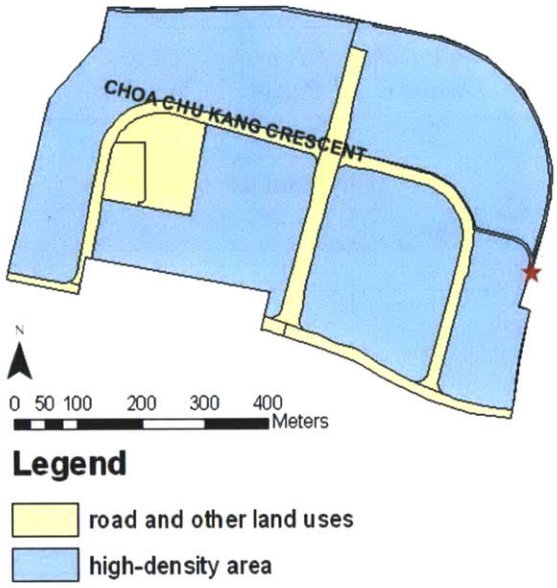


Figure 30 – Map of CCK sampling location and drainage area

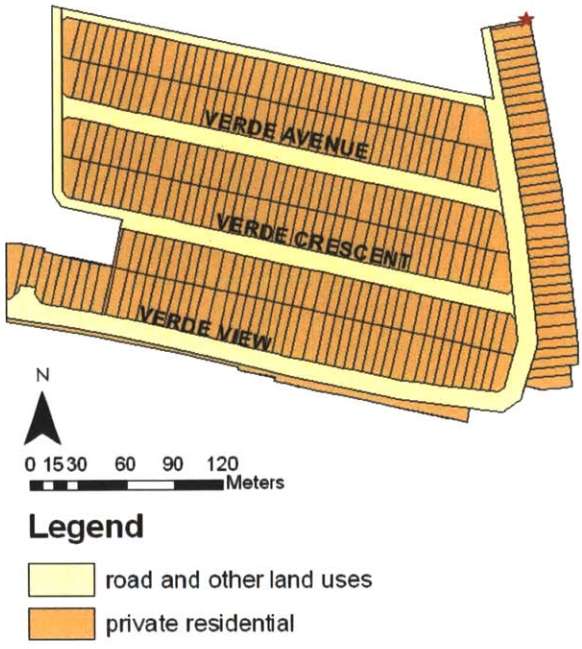


Figure 31 – Map of Verde sampling location and drainage area

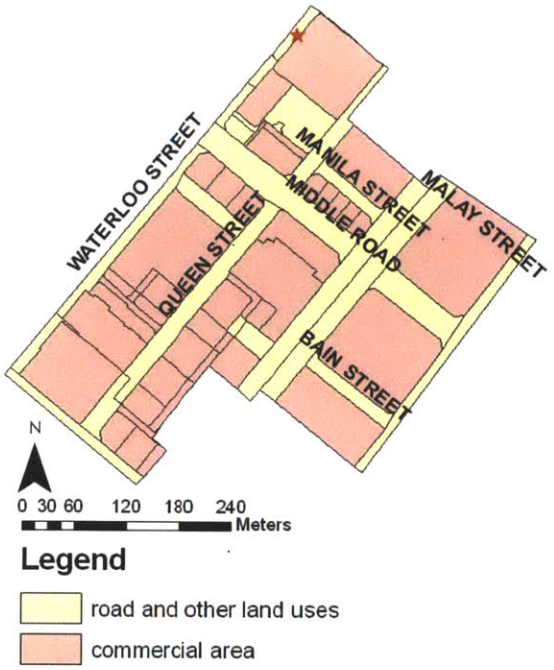


Figure 32 – Map of Bras Basah sampling location and drainage area

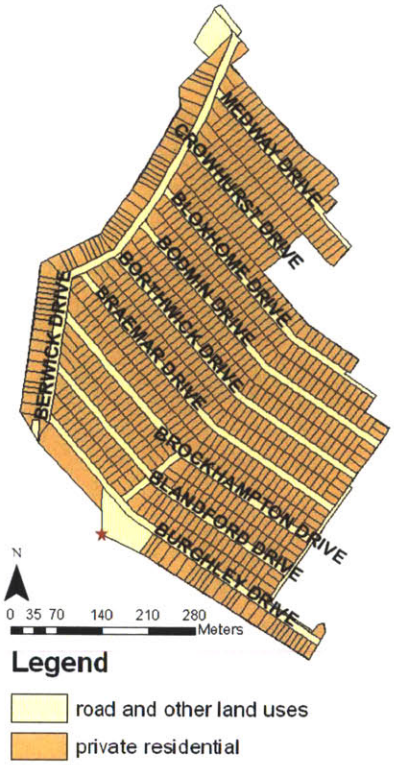


Figure 33 – Map of Serangoon sampling location and drainage area

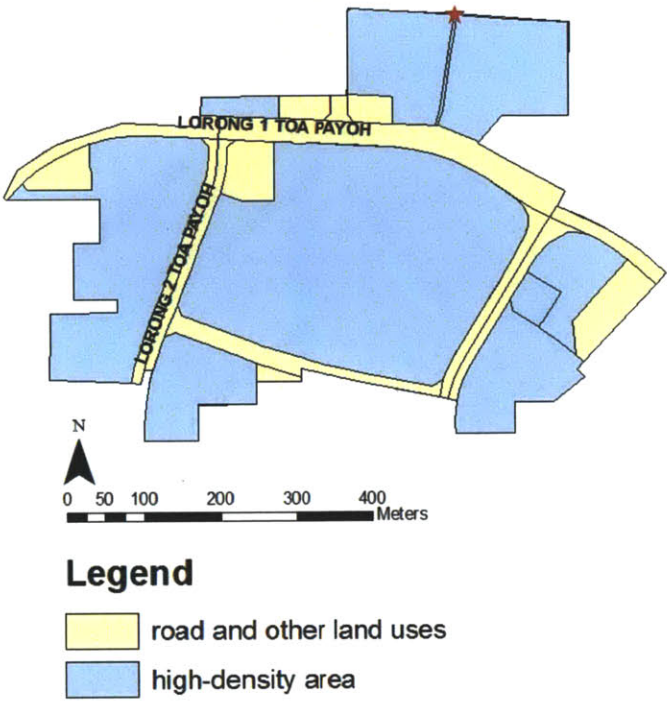


Figure 34 – Map of Toa Payoh sampling location and drainage area

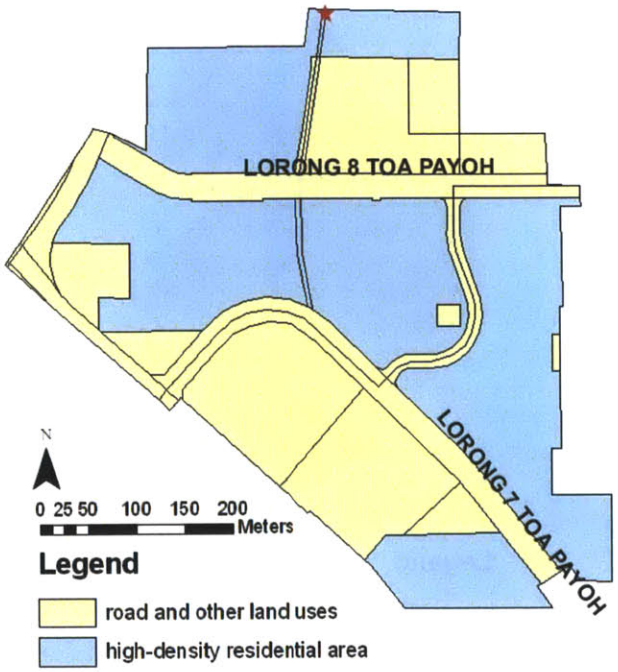


Figure 35 – Map of Lorong 8 sampling location and drainage area

Appendix III – Correlation Between Indicators

To see how well the distributions of each indicator resemble each other, I calculated Pearson correlation coefficients among the four indicators (Table 21). All correlations were generally low. The highest correlations were among the bacterial indicators total coliform, *E. coli*, and enterococci, which might be expected as they all belong to the same group of organisms. These correlations show how differently each indicator behaves and suggest limitations in using indicator bacteria to predict pathogen concentrations in Singapore.

Table 21 – Pearson correlation coefficients among four indicators

	Total coliform	***E. coli***	**Enterococci**	**Human factor**
Total coliform	-	0.46	0.42	0.28
E. coli	-	-	0.36	0.35
Enterococci	-		-	0.15
Human factor	-	-	-	-