



**IDENTIFICATION OF SUITABLE INDICATORS
FOR TRACING HUMAN FAECAL CONTAMINATION
AND SEWAGE IN URBAN CATCHMENTS**

**EVELINE EKKLESIA
SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING
2014**

**IDENTIFICATION OF SUITABLE INDICATORS
FOR TRACING HUMAN FAECAL CONTAMINATION
AND SEWAGE IN URBAN CATCHMENTS**

EVELINE EKKLESIA

SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

**A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy**

2014

Acknowledgements

First and foremost, I am very grateful to God who has given me provision in every area of my life throughout my PhD study. I would also like to extend my gratitude to the Singapore National Research Foundation (NRF) through the Singapore-MIT Alliance for Research and Technology (SMART) Center for Environmental Sensing and Modeling (CENSAM) for the funding support which has made this research possible. I would like to thank my supervisors, Asst. Prof. Lloyd Hock Chye Chua (NTU) and Dr. Peter Shanahan (MIT), who have patiently guided me and given me many valuable inputs along the way. I would also like to thank Asst. Prof. Janelle Thompson (MIT) for sharing her expertise on the microbiological work and supporting the research on the HF183 method and analysis.

Many thanks to Jean Pierre Nshimiyimana (MIT Master of Science 2010 graduate and presently NTU PhD student under Singapore-MIT Alliance scholarship) whose research has been a significant contribution to my study and also for working on the development of the HF183 quantitative method. Also, many thanks to Syed Alwi Bin Sheikh Bin Hussein Alkaff for sharing his experiences on the field work and giving assistance during the sampling campaign. I would like to express my gratitude to the Public Utilities Board (PUB) of Singapore for providing data and permission for sampling and to Vandhana Chathapuram Ravi for her assistance in the laboratory analysis of faecal indicator bacteria.

This study involved extensive (spatially) and intensive (in terms of resources such as time) environmental samples collection and lab analysis which were achieved through teamwork with many people. I, especially, would like to thank MIT Master of Engineering (MEng) graduates:

- Ryan Bossis, Genevieve Ho and Yangyue Zhang (graduated in 2011) for their help in January 2011 sampling which was the very first intensive sampling conducted,

- Suejung Shin, Janhvi Manoj Doshi, Laurie KelIndorfer and Kathyayani Shobhna Kondepudi (graduated in 2012) for their high cooperation in January 2012 sampling and
- MIT MEng 2013 graduates: Ndeye Awa Diagne for her work in the upstream and tracer study, Sophia Tsung Hwa Burkhart and Margaret Ann Hoff for the help during the field work in January 2013.

I would also like to thank NTU Bachelor of Engineering graduates:

- Wei Jing Ong, Wen Hao Aw, Qian Ying Wong, Chee Chong Chow, Yoke Yee Lai and Jun Jie Wong (graduated in 2011) for their support during June – July 2011 sampling period and
- Pin Yuan Tan, Edwin Weng Chin Wong, Yew Boon Teo and Hong Wei Poh (graduated in 2013) for their tireless effort during May – August 2012 intensive sampling campaign.

Also, I would like to thank Song Ha Le for her inputs especially during early years of my PhD study.

I am grateful to have my family, especially my mom and dad, who have always supported me through their endless prayers and love. I am also grateful for Christian who is always ready to support and consistently encourages me these last crucial and busy few years. I am thankful for Elis and my cell group friends for their steadfast faith in season and out of season which has inspired me to keep on believing. Last but not least, I would like to thank my fellow graduate students: Trieska, Namyo, Martin, Hazarki, Herry, Herlina, Kath, Bowo, Handarmin, Yudiand Wahyu for sharing experiences in this graduate students' life.

Table of Contents

Acknowledgements.....	i
Table of Contents.....	iii
Summary.....	vii
List of Figures.....	ix
List of Tables.....	xii
Chapter 1 Introduction.....	1
1.1. Background and motivation.....	1
1.2. Objectives.....	5
1.3. Scope of study.....	5
1.4. Outline.....	6
Chapter 2 Literature Review.....	8
2.1. Potential sources of surface water contamination.....	8
2.1.1. Sewer leakage and exfiltration possibility.....	8
2.1.2. Exfiltration and colmation layer.....	10
2.1.3. Sewer leakage detection and rehabilitation.....	11
2.1.4. Typical daily sewage flow pattern.....	12
2.2. The need for indicators of faecal contamination.....	13
2.3. Microbiological indicators.....	15
2.3.1. Conventional faecal indicators.....	15
2.3.2. Other faecal indicators.....	19
2.3.3. Conventional faecal indicators in tropical areas.....	22
2.3.4. Summary.....	24
2.4. Source tracking using microbiological indicators.....	24
2.4.1. Classical methods.....	25
2.4.2. Enzymatic methods.....	26
2.4.3. Library-dependent methods.....	30
2.4.4. Culture-independent, library-independent methods: host-specific marker.....	36
2.5. Physical, inorganic and organic indicators.....	44

2.5.1. Physical and inorganic indicators	44
2.5.2. Wastewater organic compounds	46
2.5.3. Chemical tracer studies	57
2.6. Temporal variation of indicator concentrations	60
2.6.1. Review of indicator temporal variation studies	60
2.6.2. Implication of temporal variation on sampling time for surface water quality monitoring	64
2.7. Influence of flow conditions.....	65
2.8. Implications of the literature review on this study.....	66
Chapter 3 Preliminary Data Analysis	70
3.1. Methodology	70
3.1.1. Data pre-processing.....	70
3.1.2. Cluster analysis	72
3.1.3. Principal component analysis.....	73
3.1.4. Summary.....	73
3.2. Data	74
3.2.1. Marina catchment one-day grab sample data	74
3.2.2. Toolbox Study data	78
3.3. Results and discussion.....	79
3.3.1. CA and PCA of Marina catchment one-day grab sample data	79
3.3.2. Diurnal variation of Toolbox Study data.....	82
3.3.3. Implications of the preliminary data analysis on this study.....	86
Chapter 4 Methodology.....	87
4.1. Parameter selection	87
4.2. Site selection	89
4.3. Sampling procedure	97
4.4. Laboratory analysis procedures	102
4.4.1. FIB analysis procedure.....	102
4.4.2. Chemical analysis procedure	105
4.4.3. HF183 quantitative method	108
4.5. Data analysis methods	113

4.5.1. Lognormal distribution	113
4.5.2. Geometric mean	114
4.5.3. Box-and-whisker plots	115
4.5.4. Representation of error	116
4.5.5. Correlation	120
4.5.6. Time series and autocorrelation	121
4.5.7. Normalised sets of samples	124
4.5.8. Artificial time series	124
4.5.9. Root-mean-square error	125
4.5.10. t-test	125
4.5.11. Wilcoxon-Mann-Whitney test	126
4.5.12. Analysis of variance	126
4.5.13. Kruskal-Wallis test	127
4.5.14. Tukey-Kramer test	127
Chapter 5 Diurnal Variation of Faecal Indicator Bacteria	128
5.1. Results	128
5.1.1. Log-normality of FIB concentration	132
5.1.2. Variation in FIB concentration	134
5.1.3. Field verification of sewer leakage	140
5.2. Data analysis and discussion	145
5.2.1. Diurnal variation of FIB concentration	145
5.2.2. Atypical trends in diurnal FIB concentration	152
5.2.3. Autocorrelation of FIB time series	154
5.2.4. Autocorrelation function for enterococci	156
5.2.5. Comparison of FIB concentration between daytime and nighttime	161
5.2.6. Significance of sampling time on FIB concentration	164
5.2.7. Non-conservative FIB behaviour	166
5.3. Conclusions	167
Chapter 6 Potential Chemical Tracers	170
6.1. Results	170
6.1.1. Log-normality of chemical data	172

6.1.2. Overview of chemical data	180
6.2. Data analysis and discussion	184
6.2.1. Overall correlation of chemical concentrations	184
6.2.2. Influence of land use on chemical concentrations.....	187
6.2.3. Chemical data segregation based on FIB	197
6.2.4. Associations among chemical parameters and FIB.....	206
6.2.5. Influence of age of sewers on chemical data	216
6.3. Conclusions.....	224
Chapter 7 Data Analysis based on HF183 Groups	227
7.1. Results	227
7.2. Data analysis and discussion	228
7.2.1. Analysis of all data based on HF group.....	228
7.2.2. Analysis of FIB data based on HF group and sampling time (daytime/nighttime)	233
7.2.3. Analysis of daytime data based on HF groups and land use	235
7.3. Conclusions.....	244
Chapter 8 Conclusions and Recommendations.....	245
8.1. Conclusions.....	245
8.2. Recommendations for future work.....	252
References	257
Appendix A. FIB and HF183 concentrations of raw sewage and non-urban water samples	277
Appendix B. Field blank samples	278
Appendix C. List of publications or communications	280

Summary

Faecal contamination, particularly from human sources, to surface waters poses a threat to public health especially when the surface water is used for water supply and/or for recreational purposes. Faecal indicator bacteria (FIB) have been used widely as surrogates for faecal contamination but were originally developed in temperate climates. FIB are known to experience growth and decay in the environment and therefore their use has been questioned, especially in tropical environments. This has motivated the present study in Singapore, a tropical and highly urbanised country, in which identification of human-related sources is important. This study aimed to suggest sampling protocols, identify suitable tracers as a function of land use and identify evidence of leaking sewers as sources of human faecal and sewage contamination in a highly urbanised catchment. Preliminary data analysis shows that sampling location influences surface water quality and surface water quality parameters have diurnal variations. Therefore, in this study, dry-weather samples were collected hourly over 12 and/or 24hr at 13 upstream areas with specific land uses: 6 high-density residential, 5 low-density residential, 1 commercial and 1 industrial. A suite of potential tracers which includes FIB, chemical parameters and HF183 is analysed.

Hourly sampling over 24hr conducted at eleven residential sites revealed diurnal variation of FIB which closely follows the pattern of sewer flow. Daytime FIB concentrations were significantly higher than nighttime FIB concentrations. Field tracer tests conducted also provide qualitative evidence linking sewage exfiltration and transport to surface drains via preferential flow paths. FIB concentrations in samples collected hourly from 04:00 – 07:00 were significantly lower than those collected from 12:00 – 15:00. Particularly, sampling at 12:00 and 14:00 results in significantly higher FIB concentrations while sampling at 05:00 and 04:00 or at 05:00 and 06:00 results in significantly lower FIB concentrations than sampling at other times throughout the day. Therefore, samples for water quality monitoring can be collected at these suggested sampling times.

Correlation analysis shows that land use is an important factor in drain water quality. There is stronger correlation among water-quality parameters when only samples with high normalised FIB concentrations are considered. Acetaminophen, orthophosphate, cholesterol, cholestanol, coprostanol, sucralose and saccharin are potential tracers for high-density residential areas. Chloride, cholestanol, coprostanol and caffeine are potential tracers for low-density residential areas.

In general, higher HF concentrations are accompanied by higher concentrations of FIB, faecal stanols, pharmaceuticals and sucralose. Correlation analysis and analysis of variance of FIB, chemical and HF data shows that high-density residential is more contaminated than low-density residential.

Overall, the diurnal pattern of FIB concentrations, agreement among tracers and more contaminated high-density than low-density residential support the hypothesis that leaking sewers are sources of human faecal and sewage contamination in urban storm drains in Singapore.

List of Figures

Figure 3.1. Map of Marina catchment	74
Figure 3.2. Sampling locations of Marina catchment one-day grab sample data (Symbol colours correspond to colour groupings in Figures 3.3 and 3.4.).....	77
Figure 3.3. Catchment map of the sampling location for the Toolbox Study	78
Figure 3.4. Dendrogram of CA for the Marina catchment one-day grab sample data (Numbers assigned to sampling locations are defined in Table 3.1.).....	80
Figure 3.5. Plot of PC1 vs. PC2 for the Marina catchment one-day grab sample data (Numbers in the plot are sampling locations which are defined in Table 3.1)	81
Figure 3.6. Hourly plot of fourteen Toolbox Study's parameters from 02 Feb 2009, 10:00 to 05 Feb 2009, 09:00	83
Figure 4.1. Sampling points of this study.....	93
Figure 4.2. Detailed catchment maps of all sampling locations.....	94
Figure 4.3. Refrigerated auto-sampler: (a) secured with metal enclosure and locked cage which was anchored to the ground (b) Avalanche® sampler (c) refrigeration compartment with 4-bottle (5-L each) configuration	98
Figure 4.4. Refrigeration compartment temperature vs. time for 14-bottle sampling event from 16 July 2012, 20:00 (first sample taken) to 17 July 2012, 10:25 (all samples retrieved from the auto-sampler).....	98
Figure 4.5. The extendable Nasco sampling pole used together with Whirl-Pak® bag	100
Figure 4.6. Typical box-and-whisker plots used in this study.....	116
Figure 4.7. Steps to create artificial time series	125
Figure 5.1. Histograms of raw (left) and log-transformed (right) 24-hr FIB concentrations: (a) TC (b) <i>E. coli</i> (c) enterococci	132
Figure 5.2. Probability plots of raw (left) and log-transformed (right) 24-hr FIB concentrations: (a) TC (b) <i>E. coli</i> (c) enterococci	133
Figure 5.3. Plots of 24-hr FIB concentrations and error bars	134
Figure 5.4. Field verification of sewer leakage, adapted from Diagne (2013) (GIS location data for sewer and main drain lines provided by PUB): (a) <i>E.</i>	

coliconcentrations on 14 January 2013 (b) <i>E. coli</i> concentrations on 17 January 2013	
(c) Location for the injection of tracer.....	142
Figure 5.5. Tracer test results at Lorong 8 Toa Payoh: (a) 24 January 2013, adapted from Diagne (2013). Monitoring ceased at 15:15 pm due to rain (b) 4 February 2013	144
Figure 5.6. Artificial FIB time series: (a) TC (b) <i>E. coli</i> (c) enterococci. Time sequence = 1 corresponds to 20:00	146
Figure 5.7. Box plots of normalised FIB concentrations grouped based on sampling time: (a) TC (b) <i>E. coli</i> (c) enterococci	148
Figure 5.8. Arithmetic means of normalised FIB concentrations with 95% confidence intervals: (a) TC (b) <i>E. coli</i> (c) enterococci.....	151
Figure 5.9. Correlograms of typical diurnal FIB concentrations: (a) TC (b) <i>E. coli</i> (c) enterococci	155
Figure 5.10. Correlograms of typical diurnal FIB concentrations excluding samples from Watten Drive and Verde: (a) TC (b) <i>E. coli</i> (c) enterococci	158
Figure 5.11. Arithmetic means of normalised FIB concentrations of samples obtained from sites exhibiting typical variation in FIB excluding samples from Watten Drive and Verde (Table 5.2) with 95% confidence intervals: (a) TC (b) <i>E. coli</i> (c) enterococci ...	159
Figure 5.12. Box plots of FIB concentrations grouped based on daytime and nighttime: (a) TC (b) <i>E. coli</i> (c) enterococci	161
Figure 6.1. Histograms of 13 chemical parameters.....	173
Figure 6.2. Probability plots of 13 chemical parameters.....	176
Figure 6.3. Box plots of 13 chemical parameters based on land uses (rs = raw sewage, H = high-density residential, L = low-density residential, C = commercial, I = industrial, • = geometric mean concentration)	182
Figure 6.4. Principle component analysis of 13 chemical parameters.....	188
Figure 6.5. PC1 vs. PC2 plot of 230 daytime samples (x = high-density residential, o = low-density residential, * = commercial, Δ = industrial)	189
Figure 6.6. Box plots of measured concentrations of 13 chemical parameters based on land use (H or L) and concentrations groups (1, 2 or 3) defined by the first segregation method.....	208

Figure 6.7. Box plots of 13 chemical parameters based on land uses, “high” dataset and age of sewers	218
Figure 7.1. Box plots of concentrations of FIB and specific conductance based on HF groups (x axis: 1 = high HF, 2 = medium HF and 3 = low HF concentration group)....	229
Figure 7.2. Box plots of chemical concentrations based on HF groups (x axis: 1 = high HF, 2 = medium HF and 3 = low HF concentration group)	230
Figure 7.3. Box plots of FIB concentrations of all sets of 24-hr samples based on HF group and daytime/nighttime (x axis: 1 = high HF, 2 = medium HF, 3 = low HF, D = daytime and N = nighttime).....	233
Figure 7.4. Box plots of FIB concentrations of sets of 24-hr samples that show typical FIB diurnal variation (Table 5.3) based on HF group and daytime/nighttime (x axis: 1 = high HF, 2 = medium HF, 3 = low HF, D = daytime and N = nighttime)	234
Figure 7.5. Box plots of the concentrations of FIB and specific conductance based on land uses and HF group (x axis: H = high-density residential, L = low-density residential, 1 = high HF, 2 = medium HF and 3 = low HF. For example, H1 means high-density residential data that have high HF concentrations.)	237
Figure 7.6. Box plots of the chemical concentrations based on land uses and HF group (x axis: H = high-density residential, L = low-density residential, 1 = high HF, 2 = medium HF and 3 = low HF. For example, H1 means high-density residential data that have high HF concentrations.)	238

List of Tables

Table 2.1. Faecal source related to FC/FS ratios (Feachem 1975)	18
Table 2.2. Correlation coefficients for mean indicator densities against swimming-associated gastroenteritis rates during summer, extracted from USEPA (1986).....	19
Table 2.3. Colilert, Colilert-18 and Enterolert (IDEXX 2010a).....	27
Table 2.4. Comparison of two enzymatic methods	29
Table 2.5. Primers for human- and cow-specific <i>Bacteroides-Prevotella</i> marker, extracted from Bernhard and Field (2000a)	37
Table 2.6. Differences of conventional PCR with QPCR.....	37
Table 2.7. <i>Bacteroides</i> persistence at different temperature	38
Table 2.8. Comparison of studies which analysed HF183 marker in tropical areas.....	41
Table 2.9. Sequences of primers used for human-specific <i>Bacteroides-Prevotella</i> marker.....	42
Table 2.10. PCR and its primers for the detection of human	42
Table 2.11. Pharmaceutical compounds and their persistence, extracted from Tixier et al. (2003)	51
Table 2.12. Comparison of several chemical tracer studies in tropical areas	59
Table 3.1. Measured indicator concentrations in Marina catchment one-day grab samples and Lilliefors test results.....	76
Table 4.1. Catchment characteristics: (a) high-density residential (b) low-density residential, commercial and industrial	91
Table 4.2. Sampling locations of raw sewage and non-urban water samples	96
Table 4.3. Auto-sampler cross-contamination test results	99
Table 4.4. Manual sampling's container cross-contamination test results.....	101
Table 4.5. List of parameters analysed by the appointed accredited laboratory with the test method, LOD and uncertainty	106
Table 4.6. Summary of melting temperatures and comparison with other studies ..	110
Table 4.7. Summary of LOD and samples below LOD.....	111
Table 4.8. Error of pipettes (provided by the manufacturer).....	117
Table 4.9. Absolute uncertainty of 100-mL flask or cylinder	118

Table 5.1. Summary of sampling program.....	129
Table 5.2. Comparison of FIB concentrations among different studies	131
Table 5.3. RMSE values for TC, <i>E. coli</i> and enterococci	153
Table 5.4. Kruskal-Wallis and Tukey-Kramer tests results for (a) six-hr and (b) four-hr groups	164
Table 6.1. Summary of non-detect samples and lab duplicate analysis.....	171
Table 6.2. Correlation matrix for all daytime data	186
Table 6.3. Correlation matrix for high-density residential areas.....	192
Table 6.4. Correlation matrix for low-density residential areas.....	193
Table 6.5. Correlation matrix for commercial area	194
Table 6.6. Correlation matrix for industrial area	195
Table 6.7. First segregation method for samples from high-density residential areas	199
Table 6.8. Second segregation method for samples from high-density residential areas	200
Table 6.9. Third segregation method for samples from high-density residential areas	201
Table 6.10. First segregation method for samples from low-density residential areas	202
Table 6.11. Second segregation method for samples from low-density residential areas.....	203
Table 6.12. Third segregation method for samples from low-density residential areas	204
Table 6.13. Correlation matrix of “high” dataset (58 samples) from high-density residential areas.....	210
Table 6.14. Correlation matrix of “moderate” dataset (20 samples) from high-density residential areas.....	211
Table 6.15. Correlation matrix of “low” dataset (29 samples) from high-density residential areas.....	212
Table 6.16. Correlation matrix of “high” dataset (54 samples) from low-density residential areas.....	213

Table 6.17. Correlation matrix of “moderate” dataset (23 samples) from low-density residential areas.....	214
Table 6.18. Correlation matrix of “low” dataset (13 samples) from low-density residential areas.....	215
Table 6.19. Correlation matrix of 32 samples from high-density residential areas, “high” dataset and old sewers.....	219
Table 6.20. Correlation matrix of 26 samples from high-density residential areas, “high” dataset and new sewers	220
Table 6.21. Correlation matrix of 35 samples from low-density residential areas, “high” dataset and old sewers.....	221
Table 6.22. Correlation matrix of 19 samples from low-density residential areas, “high” dataset and new sewers	222
Table 7.1. Tukey-Kramer test results of all data based on HF groups.....	232
Table 7.2. Two-way ANOVA test results of FIB data based on HF group and daytime/nighttime.....	235
Table 7.3. Tukey-Kramer test results of high-density residential data based on HF group.....	241
Table 7.4. Tukey-Kramer test results of low-density residential data based on HF group.....	242

CHAPTER 1 INTRODUCTION

1.1. Background and motivation

Urbanisation has resulted in the deterioration of surface water quality. This problem takes on a greater significance when surface water is used as a source of water supply and for recreational purposes. Singapore, an urbanised island city-state, is a unique test case of such a problem. It faces land scarcity and therefore, since the 1960s, it has recognised that acquiring a sustainable water supply is a vital matter (PUB 2010a). Singapore has four different sources of water supply, known as the Four National Taps. These are water from local catchment areas, imported water, reclaimed water (called NEWater) and desalinated water (PUB 2010a). For the local catchment areas, Singapore uses a network of concrete-lined storm drains, canals and rivers to convey rainwater to reservoirs. Subsequently, the water undergoes treatment for drinking water supply. This makes Singapore one of the few countries in the world to harvest urban stormwater on a large scale for its water supply (PUB 2010a). In addition, since 2006, Singapore has had plans to utilise these waterbodies for recreational purposes. This is part of the Public Utilities Board (PUB)'s Active, Beautiful and Clean (ABC) Waters Programme which is an effort to improve the water quality and the quality of living by harnessing the full potential of the waterbodies (PUB 2010c). However, urban stormwater runoff is often contaminated with pollutants including faecal pathogens (Arnone and Walling 2007) which can endanger public health. Pathogens are costly and difficult to analyse. Therefore faecal indicator bacteria (FIB) are used as surrogates for those pathogens and routinely monitored to evaluate the risk of pathogen contamination. The common bacterial indicators, which include total coliform, faecal coliform, *E. coli* and enterococci, were initially developed for use as indicators in temperate countries. However, different studies (Hazen 1988; Muñiz et al. 1989; Noble et al. 2004; Walters et al. 2011) have shown that FIB exhibit either net growth or die-off. The growth and survival of FIB are susceptible to environmental factors (Isobe et al. 2004). The most important environmental factors are temperature and sunlight (Noble 2004). These two factors make bacterial indicators potentially unreliable in tropical areas such as Singapore.

Nevertheless, Singapore has adopted the World Health Organization (WHO) bacterial indicators guideline based on enterococci. This guideline states that for 95% of the time, the counts of enterococci should be less than or equal to 200 per 100 mL in recreational waters (ocean beaches or freshwater bodies) (WHO 2003; NEA 2010).

The majority of large-scale waterborne disease outbreaks in the past, including waterborne disease upon contact with contaminated recreational water bodies, have been attributed to human faecal or sewage contamination (Arnone and Walling 2007). Sewage can enter stormwater systems due to on-site wastewater disposal via septic systems, abandoned sewer bypass locations, breaches in sanitary sewage infrastructure (Sauer et al. 2011), cross-connected pipes (Hyer 2007) and illicit dumping of septic waste or illicit discharge (Paul and Meyer 2008). However, in urban areas where septic systems and combined sewer overflows do not exist and where illegal connection or dumping is not prevalent, failing sanitary sewer infrastructure or sewer leakage may be the cause of human sewage contamination (Boehm et al. 2003; Wu et al. 2008; Sercu et al. 2009; Nguyen et al. 2011; Xu et al. 2011). PUB (2010d) stated that Singapore has separate sanitary sewer and storm drainage systems and 100% of urban areas in Singapore are served by the sewer system. Wastewater collected by the sewer system is directed to large deep tunnels leading to centralised wastewater treatment plants. The sewer system and its ancillaries are constructed to detailed and exacting specifications developed by PUB (2004; 2009). Moreover, PUB has also invested heavily in sewer rehabilitation. Despite this very well maintained system, sewer leakage may still occur. In a tropical setting with intense tropical rainstorms like Singapore, the drainage system needs to extend upstream to surface drains around nearly every building. Therefore, there is a potential for any sewage leak to enter the drain network. Tools to monitor and identify sources of human faecal contamination to the storm drainage system are critical for identifying and prioritising targets for urban water quality improvement. Unfortunately, FIB do not specifically indicate the presence of human faecal contamination. These concerns have motivated this study in Singapore where identification of human-related sources is important.

Besides FIB, a suite of tracers which include chemical parameters and human-specific HF183 *Bacteroides* 16S rRNA genetic marker is investigated in this study. Agreement among indicators and tracers is expected to provide different lines of evidence about bacterial contamination. Although FIB may grow, occur naturally or survive longer in tropical waters (Hazen 1988; Rivera et al. 1988; Muñiz et al. 1989), they are regulated in surface water by WHO guidelines and therefore commonly used worldwide and included as benchmarks, even in tropical areas (Kenzaka et al. 2001) such as Singapore. Wastewater organic components that are quantifiable by reasonably facile methods are explored as possible tracers. Studies have shown the applicability of the HF183 marker in tropical areas (Jenkins et al. 2009; Nshimiyimana 2010; Ahmed et al. 2010b; Boehm et al. 2011; Santiago-Rodriguez et al. 2012; Toledo-Hernandez et al. 2013). Particularly, Nshimiyimana (2010) confirmed the use of the HF183 assay as a viable technique for human faecal source tracking in Singapore by showing that HF183 genes obtained in Singapore were clustered within the same evolutionary group as genes obtained by Santoro and Boehm (2007) in temperate climates. The indicators and tracers can be ordered to provide information at multiple tiers from less sensitive methods, simple methods that are fast and cheap, to more sensitive and elaborate techniques, such as deoxyribonucleic acid (DNA) polymerase chain reaction (PCR)-based analysis that are more time consuming and costly. Simple methods could be used for screening purposes while the more involved procedures could be used when greater accuracy is required.

Studies by others report that FIB concentrations vary diurnally, from high at nighttime to low in daytime (Boehm et al. 2002; Whitman and Nevers 2004; Traister and Anisfeld 2006; Enns et al. 2012; Desai and Rifai 2013). Those studies were conducted in temperate areas except Enns et al. (2012) in tropical Miami and Desai and Rifai (2013) in subtropical Houston. Enns et al. and Desai and Rifai collected most of their samples from open areas which are exposed to sunlight. Besides temperature, sunlight is known as an environmental factor that strongly influences bacterial survival (Noble et al. 2004). Therefore, there is a question of whether FIB concentration varies similarly in a tropical urban area that is not exposed to

sunlight, such as in covered drains or shaded areas, and whether the variation informs the occurrence of human faecal and sewage contamination. This is important to suggest protocols for monitoring of human faecal and sewage contamination.

A study by Nanyang Technological University (NTU) in 2004-2007 (NTU 2008), in which dry-weather samples were collected in drainage channels, showed that mean *E. coli* and enterococci counts in Kranji catchment, Singapore, were high and did not comply with the United States Environmental Protection Agency (USEPA) guideline. As the NTU study was concerned with obtaining gross measures of bacterial contamination, sampling during the study was conducted at relatively downstream locations within the watershed and thereby integrated the dry-weather flow contributions from a variety of land uses. Therefore, land use impacts on bacterial contamination could not be ascertained. The study however revealed that the highest bacteria count was obtained from an area of which the land use comprised 70% high-density residential areas (NTU 2008). Desai and Rifai (2010) show that several urbanised sites, comprising residential areas with equal areas of grassland, exhibited higher overall *E. coli* concentrations than areas dominated by grassland. For a highly urbanised water catchment, a more rigorous categorisation into distinct land use types, such as high-density residential, low-density residential, commercial and industrial areas, within the urban context is desirable. In this study, samples were collected from sites with specific land uses and analysed for the suite of tracers indicated above. Hence, the characteristics of tracers in different types of urban land uses were explored. This investigation leads to the identification of suitable indicators according to these specific land use types. Agreement among tracers provides evidence of human faecal and sewage contamination.

Nshimiyimana (2010) reported that high *E. coli* concentrations were found in residential, horticultural and animal farming areas in Kranji catchment. In his study, PCR techniques (presence/absence tests) for identification of HF183 marker were also developed. The study shows that 94% of the 34 environmental water samples collected in residential, horticultural and animal farming areas were positive for

HF183 marker. The detection of bacterial contamination, especially in residential areas, is expected to be indicative of other urbanised areas in Singapore, where leakage from sewer pipes is a likely source of faecal contamination.

1.2. Objectives

The aims of this study are as follows.

- (i) To suggest sampling protocols that can be applied for the identification of human faecal and sewage contamination in “storm drains” in a highly urbanised catchment

This study collected samples from upstream storm drains. There is little use of the drains for recreation but they drain into canals and reservoirs that are used for recreation or will be used for recreation as part of the PUB’s ABC Waters Programme.

- (ii) To identify suitable tracers for human faecal and sewage contamination in a highly urbanised catchment as a function of land use

Since this study addresses concerns about human faecal and sewage contamination in urban areas, the tracers must be of human origin (from human activities, production or consumption). In addition, the tracers must be present in the waterways in sufficient amount that they are quantifiable with facile methods.

- (iii) To identify evidence of leaking sewers as sources of human faecal and sewage contamination in a highly urbanised catchment

1.3. Scope of study

The scope of this study is as follows.

- (i) To investigate how FIB concentrations vary with respect to time

Samples were collected hourly over 24 hr. Sampling was conducted at various sites in order to investigate if variation prevails and is consistent across sampling sites. Sampling times for identification of human faecal and sewage contamination and also for monitoring program are proposed.

- (ii) To investigate the use of wastewater chemical constituents as potential tracers
Certain manmade compounds indicate the presence of human faecal and sewage contamination. A non-parametric correlation analysis tool is used to analyse the relationships between tracer pairs with particular emphasis on relationships between tracers and biological indicators including total coliform, *E. coli* and enterococci. The relationships are expected to differ for different types of land use, hence this research seeks to identify indicator signatures that are unique to specific land use types.
- (iii) To investigate how HF183 varies as other indicators vary
HF183 concentrations were obtained using a laboratory analysis method developed by Nshimyimana et al.(2014). Since the method is still under further development, HF183 concentrations obtained were converted into categorical data. The relationship between HF183 and other parameters was illustrated with box plots and investigated with one-way ANOVA (followed by Tukey-Kramer test) and two-way ANOVA.

1.4. Outline

This thesis is organised as follows.

Chapter 1 Introduction describes the motivation, aims and objectives of this study.

Chapter 2 Literature Review provides a review of the literature related to this study. It starts with a discussion about the concern that motivates this study: potential sources of surface water contamination, particularly leaking sewers. Then, it discusses the need for indicators as surrogates for faecal contamination. The many options for indicators, microbiological and chemical, are also discussed. Source-tracking studies are also discussed in terms of their different approaches employed. Studies of temporal variation of FIB are reviewed. The influence of flow on surface water quality is also discussed. The chapter is wrapped up with a discussion of the implications of the literature review on this study.

Chapter 3 Preliminary Data Analysis describes and evaluates two datasets obtained from the Public Utilities Board, Singapore. The analysis results motivate the approach adopted in this study: hourly sampling at upstream locations.

Chapter 4 Methodology describes the parameters and sites selected for this study, the sampling and laboratory procedures employed in this study (FIB, chemical and HF183 analysis) and the data analysis methods used in this thesis.

Chapter 5 Diurnal Variation of Faecal Indicator Bacteria provides the FIB concentration results for the 24-hr sets of samples. The analysis shows the diurnal pattern of FIB concentrations in residential urban catchments. Sampling times are suggested for the identification of human faecal and sewage contamination and also for routine monitoring program.

Chapter 6 Potential Chemical Tracers provides the chemical tracer results. The analysis involves the evaluation of the use of chemical tracers for different land uses, the agreement of chemical tracers with FIB and the effect of the age of sewers on leaking sewers.

Chapter 7 Data Analysis based on HF183 Groups evaluates the agreement between HF183 and other parameters (FIB and chemical tracers).

Chapter 8 Conclusions and Recommendations conclude the findings of this study (including the suggestion can be made from this study) and recommend future works that are useful to complement this study.

CHAPTER 2 LITERATURE REVIEW

2.1. Potential sources of surface water contamination

Potential sources of surface water contamination include point and non-point sources. When pollutants are discharged from a single source at a discrete point, the source is categorised as point source pollution. This is usually associated with disposal of water for industrial, commercial or municipal purposes. Hence, it can feasibly be controlled through use of wastewater treatment technologies to remove pollutants before discharge and through regulatory permits, inspections, monitoring and compliance processes. When pollutants enter waterbodies at many locations from various sources, distributed diffusely over an area, the source is categorised as non-point source pollution. This can originate from the air (wet and dry deposition), land surface and subsurface. Examples of non-point source pollution are various land conversions including construction site erosion, traffic (Novotny 2003) and agricultural land (Pereira and Hostettler 1993) as well as animals (Hyer 2007) including domestic pets such as dogs and cats; wildlife such as raccoons, opossum, rats, squirrels and deer; and waterfowl such as geese, ducks and seagulls.

2.1.1. Sewer leakage and exfiltration possibility

In urban areas, most pollution is caused by human sewage contamination (Sercu et al. 2009; Sauer et al. 2011). Contaminated groundwater and surface water pose a risk for public health especially when they serve as drinking water supplies (Held et al. 2006). The risk of contaminated groundwater and surface water changes as wastewater composition constantly varies (Held et al. 2006). Besides, the mechanisms that introduce sewage into waterways are obscure. Sewage can enter stormwater systems due to on-site wastewater disposal via septic systems, abandoned sewer bypass locations, breaches in sanitary sewage infrastructure (Sauer et al. 2011), cross-connected pipes (Hyer 2007) and illicit dumping of septic waste or illicit discharge (Paul and Meyer 2008). In urban areas where septic systems and combined sewer overflows do not exist and where illegal connection or dumping is not prevalent,

failing sanitary sewer infrastructure or sewer leakage may be the cause of human sewage contamination.

Singapore, a highly urbanised environment, operates separate sanitary sewer (in short, sewer) and storm drainage (in short, drain) systems. Therefore, sewer leakage may be an important cause of faecal pollution of the drainage system. Sewer leakage may occur because of two factors: (i) ageing sewer lines (Amick and Burgess 2000; Held et al. 2006) and (ii) damaged sewer lines (for example, at the joints). The reticulation system in Singapore has been developed since early in the twentieth century and particularly since the 1930s when the first housing estates appeared, so there are numerous ageing sewers. Damage may be caused during backfilling at the end of construction or due to ground settling after backfilling.

Sewer lines in Singapore are generally laid deeper than storm drains and located below the groundwater table (where infiltration rather than exfiltration occurs). However, when the water table is low (during dry weather) or when sewer lines are at upstream locations (at shallow depth), sewage can exfiltrate from leaks in sewer pipes (Whitlock et al. 2002; Paul and Meyer 2008) through groundwater (Boehm et al. 2004) or unsaturated soil (Boehm et al. 2003; Sercu et al. 2009; Sauer et al. 2011) into surface water such as at the shoreline (Boehm et al. 2003; Boehm et al. 2004) or storm drains (Sercu et al. 2009; Sauer et al. 2011; Sercu et al. 2011b). Sewage exfiltration to nearby drainage systems is possible when the depth of drainage system is similar to the sanitary sewer lines (Sercu et al. 2009). For example, upstream sewer lines in Singapore, including sewers connecting buildings to the public sewers, are located at relatively high elevations and are generally above the groundwater table. Those that are near premises can be as shallow as only 1.4 m (Diagne 2013) and are prone to having joints displaced due to settling. Therefore, there is greater probability that these sewers will leak and contaminate surface water in nearby storm drains. This phenomenon had been mentioned by Hyer (2007) who described a contamination incident due to damaged laterals. Defects may also be caused by

development works that affect nearby sewers. Moreover, defects at the invert of sewer lines are prone to constant exfiltration (Wolf et al. 2006).

Sewage exfiltration from leaks in sewer pipes may travel to surface water such as storm drains via preferential pathways. One type of preferential pathway occurs due to the pore-size exclusion process. In the pore-size exclusion process, relatively larger particles are excluded from small pores and move only through the larger pores (Sinton et al. 1997). This may happen in porous strata such as fractured bedrock (Champ and Schroeter 1988). Therefore, transport of larger particles is dominated by convection and less by dispersion (Pang et al. 1998). This also implies that the transport of larger particles in groundwater can be faster than the average groundwater velocity (Sinton et al. 1997). Sewage contains bacteria which can be considered to be larger particles. Pang et al. (1998) tested whether preferential pathways of bacteria transport in groundwater existed. They reported that bacteria tended to travel faster than dye and attributed this to the pore-size exclusion process.

Each leak in a sewer line could be considered as a point source of pollution. However, there are likely to be many such leaks at unknown locations and in that sense they are diffuse sources of pollution. In other words, it is a point source on the micro scale (i.e. in a specific drain near a particular leak location) but a diffuse source on the macro scale (i.e. when considering the drains as a system). In this study, drains that service fairly good-sized areas were monitored; hence the pollution could be characterised as diffuse pollution.

2.1.2. Exfiltration and colmation layer

Exfiltration is an unsteady process (Held et al. 2006). It depends on sewer-related parameters, the geological environment and local climate conditions (Amick and Burgess 2000). Sewer-related parameters include the size and age of sewer lines, material, type and quality of construction and depth of flow within the sewer. The exfiltration rate is also subject to changes in the clogging layer (Held et al.

2006) including the colmation layer (a clogging layer typically develops from microorganisms and unsuspended solids, which leads to a certain self-sealing of the leak) immediately next to the pipe (Karpf et al. 2009). The exfiltration rate increases as the surrounding soil gets coarser and as the water level in the pipe increases (Karpf et al. 2009). The exfiltration rate is also higher in unsaturated conditions (Karpf et al. 2009).

The colmation process can be divided into two phases. The initial colmation phase is characterized by physical and chemical processes such as pores clogging (Karpf et al. 2009). The second phase is characterized by biological, chemical and physical processes (Karpf et al. 2009). Colmation is often described by the leakage factor which combines two characteristics of the colmation layer: its conductivity and inverse of the thickness. Pilot-scale experiments show that the first phase causes a fast decrease of the leakage factors (Vollertsen and Hvitved-Jacobsen 2003; Karpf et al. 2009), while the second phase exhibits low rates in the decrease of hydraulic conductivity compared to the first phase (Karpf et al. 2009). Klinger et al. (2007) investigated defects in real sewers and showed that the colmation layer seems to be vulnerable to variations of hydrostatic pressure and flow velocity, especially during extreme flow conditions (Held et al. 2006). Ellis et al. (2009) also mentioned that even for low flow conditions, there is little evidence that this sealing is stable. Klinger et al. (2007) reported that the colmation layer can form within days or hours in lab studies, but not in the field. Vollertsen and Hvitved-Jacobsen (2003) also recorded that the exfiltration rate increased rapidly for short periods now and then (short-term breakthrough of wastewater) in their pilot-scale experiment.

2.1.3. Sewer leakage detection and rehabilitation

An assessment of sewer pipes is usually accomplished with the help of closed-circuit television (CCTV) inspections. However, these CCTV records and the resulting sewer defect assessments do not reveal the potential for exfiltration as visually identified defects do not necessarily account for high exfiltration rates (Eiswirth et al. 1995; Held et al. 2006; Rutsch et al. 2006). Another approach is a direct measurement of

exfiltration using flow measurement. However, measurement of flow in real sewers to address small differences in flow due to exfiltration is challenging (Vollertsen and Hvitved-Jacobsen 2003).

Sewer rehabilitation can be carried out through the dig-and-replace method or trenchless methods (USEPA 1999). Generally trenchless methods cause fewer disturbances than the dig-and-replace method. Trenchless methods include slip lining, cured-in-place pipe and deformed and reformed liners. Slip lining is a new liner of smaller diameter placed inside an existing pipe. Cured-in-place pipe is a flexible fabric liner, coated with a thermosetting resin, inserted into the existing pipeline and cured to form a new liner. In Singapore, most leaks are fixed using a trenchless method; however, if the sewer pipe is in poor condition, the dig-and-replace method is carried out (PUB 2010e). In a country that has a comprehensive sewer network like Singapore, sewer rehabilitation will be a continuous programme. Sewer rehabilitation takes time and resources and sewers wear out continuously. Therefore, a good sewer leakage detection program is useful in prioritising areas for sewer rehabilitation.

2.1.4. Typical daily sewage flow pattern

Sewage flow is known to show a daily variation. This characteristic and the possible existence of a preferential pathway may be useful in indicating sewer leakage (discussed in Chapter 5). Therefore, the typical daily sewage flow pattern in sanitary sewer collection systems is described here. The typical daily sewage flow pattern is depicted by a wave form (Chesner and Pai 1981). The pattern generally has high flow during the daytime and low flow during the nighttime (especially late at night and early in the morning) and exhibit two peaks in discharge, resulting from morning and early evening water usage (Chesner and Pai 1981; Enfinger and Stevens 2006). The morning peak is generally higher than the early evening peak (Butler and Graham 1995). Variations in the wave form of any given collection system can take the form of time lags, advances in the wave form, or increases or decreases in observed peaks (Chesner and Pai 1981). These variations can be expected to occur on a day-to-day basis.

Sewer leakage may occur due to ageing or damaged sewer lines. However, sewage exfiltration only occurs when the water table is below the sewer elevation. Leaking sewage may travel through preferential pathways such as pore-size exclusion to storm drains. This study investigates the possibility of leaking sewage contaminating storm drains.

2.2. The need for indicators of faecal contamination

There are many potential sources of contamination to surface waters and the potential health hazards from these sources are varied. Untreated human faecal wastes present the most prevalent hazard to human health with regard to recreational water quality (Cabelli 1977). However, pathogenic microorganisms usually appear intermittently at low concentrations (Borrego et al. 1987) and they are difficult to cultivate and identify. Microscopic inspection and identification of specific pathogenic microorganisms is very costly. The routine examination of surface water thus needs a broader base than just pathogen detection. Indicators of surface water quality were introduced to address this need.

According to Cabelli (1977), indicators should ideally have the following properties: (i) They should be consistently and exclusively associated with the source of the pathogens. (ii) They must be present in sufficient numbers to provide an estimate of risk of illness. (iii) They also should approach the resistance to disinfectants and environmental stress of the pathogens potentially present at significant levels in the source. (iv) Finally, they should be quantifiable in recreational waters by reasonably facile and inexpensive methods and with considerable accuracy and specificity.

The best indicators have densities that correlate well with health hazards associated with a given type of pollution source. With regard to recreational waters, recreational water quality indicators are microorganisms or chemicals whose densities in the water can be quantitatively related to potential health hazards resulting from recreational use (Cabelli 1977). Since it is known that the discharge of human faecal

wastes represents a very common form of hazardous pollution with regard to recreational water use, the term “recreational water quality indicator” has been used synonymously with faecal indicator, i.e. microorganisms or chemicals that are specifically, consistently and exclusively associated with the faecal wastes of warm-blooded animals (Cabelli 1977).

The search for the best faecal indicators has evolved since the early 1970s (Cohen and Shuval 1973). Initially, the coliform group of bacteria was used. However, it was later shown to be associated with non-faecal sources, such as plants and soil (Cohen and Shuval 1973). Hence, the faecal coliform group of bacteria including faecal streptococci was suggested. Of these, *Escherichia coli* (*E. coli*) and enterococci levels have been found to be significantly correlated with gastrointestinal symptoms (Cabelli 1977). The fact that *E. coli* gave the best relationship to gastrointestinal illness is clearly understood because *E. coli* dominates the distribution of coliforms found in faecal wastes from humans and warm-blooded animals (Dufour 1977). Prüss (1998) stated that gastrointestinal symptoms are the most common health problem related to the count of indicator bacteria in recreational waters and there is a causal relationship between gastrointestinal symptoms and microbiological water quality. Although there has never been broad agreement on the bacterium or group of bacteria to be adopted as indicators of faecal contamination, organizations such as the World Health Organization (WHO) and United States Environmental Protection Agency (USEPA) have provided guidelines or maximum contaminant levels (MCLs). The USEPA (1986) recommendations for maximum allowable concentrations in freshwater are as follows: (i) geometric mean of no less than five 1-day samples within a 30-day period equal to 126 colony-forming unit (CFU)/100 mL for *E. coli* and 33 CFU/100 mL for enterococci and (ii) single-sample maximum-allowable-density equal to 235 CFU/100 mL for *E. coli* and 61 CFU/100 mL for enterococci.

Contamination in urban drainage systems by human faecal wastes, especially from leaking sewers, is of great concern. However, *E. coli* and enterococci may be unsuitable as indicators due to their ability to replicate in the natural environment

(discussed in Section 2.3.3). It is therefore worthwhile to evaluate the use of other microbiological and wastewater organic compounds for use as human faecal and sewage indicators or as components in combination with other indicators. A more detailed elaboration of the variety of indicator bacteria and organic compounds can be found in Section 2.3 and 2.5, while the development of microbiology techniques is discussed in Section 2.4.

2.3. Microbiological indicators

2.3.1. Conventional faecal indicators

a. Total coliform

Total coliform (TC) bacteria have the following characteristics: aerobic or facultative anaerobic, gram-negative (having thin peptidoglycan layer as part of the cell wall), non-spore-forming bacilli, and fermenting lactose when incubated at 35°C (Dufour 1977). They include the genera *Escherichia*, *Citrobacter*, *Klebsiella*, and *Enterobacter* which do not inhabit the intestinal tract of warm-blooded animals exclusively (Dufour 1977). Some are commonly found in the environment. Hence, the total coliform indicator is not faecal specific (USEPA 1986) and does not fulfil the first criterion that an indicator organism needs to be exclusively associated with the source of the pathogens (Cabelli 1977).

b. Faecal coliform

The faecal coliform (FC) group may be separated from the total coliform group by its ability to grow and ferment lactose at elevated temperatures, i.e. 44.5-45.5°C. Hence, it is also referred to as thermotolerant coliform (Feng et al. 2002). A faecal coliform index can be obtained from the total coliform index by using the ratio of faecal coliform to total coliform bacteria (USEPA 1986).

The faecal coliform group mainly grows in the intestines of warm-blooded animals. Evison and James (1973) reported that faecal coliform is present in human faeces at a

quantity of 4.3×10^{10} most probable number (MPN)/g faeces. In addition, scientists learned that coliforms' ecology, prevalence and resistance to stress differ greatly from those of pathogenic microorganisms (Scott et al. 2002). Thus, this limits the use of coliform as faecal pollution indicators. Therefore, specific microbes have been suggested as alternatives. The most common member is *Escherichia coli*, which constitutes 90-95% of the coliforms in faeces (Dufour 1977). Hence, it is often used as a surrogate measure for faecal coliform (Noble et al. 2004). However, this group still includes *Klebsiella* which can be found in carbohydrate materials without any indication of faecal pollution (Dufour 1977).

c. *Escherichia coli*

Escherichia coli can be easily detected by its ability to ferment glucose which is later changed to lactose. *E. coli* is used because of its abundance in human and animal faeces at concentrations much higher than the pathogens it predicts. Evison and James (1973) reported that *E. coli* in human faeces is 4.3×10^{10} MPN/g faeces. As *E. coli* can be found in a variety of warm-blooded animals, it is not possible to differentiate human from animal wastes based on this indicator only (Field et al. 2003).

With regard to climate, *E. coli* is a reliable indicator of faecal pollution in water sources in temperate climates (Evison and James 1973). However, in tropical and subtropical waters, there is evidence to suggest that *E. coli* is much less satisfactory due to their ability to grow or replicate in the environment or extraintestinal habitats (Scott et al. 2002). Hence, its abundance can change independently of source contributions.

d. Faecal streptococci and enterococci

Faecal streptococci (FS) are gram-positive (having a thick peptidoglycan layer as part of the cell wall), catalase-negative cocci that cleave esculin and are not inhibited by bile salts (Harwood et al. 2000). They occur in relatively high numbers in the excreta of humans and other warm-blooded animals (Sinton et al. 1993). Evison and James

(1973) reported that faecal streptococci in human faeces is 5×10^{10} MPN/g faeces. Since they can be found in other warm-blooded animals, they have the drawback of being non-specific to humans. They persist in the environment without multiplication outside the animal body. Faecal streptococci are also absent from pure waters, virgin soils and other environments that have no contact with human or animal life (Sinton et al. 1993).

Different species of faecal streptococci can be found in different sources (Sinton et al. 1993). *Streptococcus faecalis* and *Streptococcus faecium* are both found in humans and most other animals. *Streptococcus avium* is found in birds or poultry, but also occurs in humans and other animals. *Streptococcus gallinarum* occurs principally in birds, but has also been isolated from a human clinical specimen. *Streptococcus bovis* is found in the faeces of ruminants or cattle, often as the predominant faecal streptococcus, and in the faeces of other animals and humans. *Streptococcus equines* is the predominant streptococcus in the faeces of horses.

In 1984, there was a revision in nomenclature (Sinton et al. 1993). Hence, species such as *S. faecalis*, *S. faecium*, and *S. avium* became *Enterococcus faecalis*, *Enterococcus faecium*, and *Enterococcus avium*. Enterococci (Ent) are differentiated from other streptococci by their ability to grow in 6.5% NaCl at pH 9.6 and temperature 45 °C (Scott et al. 2002). Enterococci have been suggested as an alternative to *E. coli* for monitoring in freshwater due to their direct correlation to swimmer-associated gastroenteritis (Kinzelman et al. 2003). It has also been reported that, overall, the inactivation rates of enterococci are slower than those of total coliform (Noble et al. 2004).

The ratio of enterococci to other streptococci in faeces differs among vertebrate species (Sinton et al. 1993). This attribute could potentially provide a distinction between the different faecal sources. However, there is some ambiguity relating to the occurrence of different species of *E. faecalis* and *E. faecium* since both are found

in humans and animals. For *S. bovis* and *S. equines* however, it is less ambiguous as both are found only in animals.

Initially, the ratio of FC to FS was used to distinguish human and non-human sources (Geldreich and Kenner 1969). This was based on the assumption that FC and FS die off at identical rates. However, this assumption is inconsistent with the fact that the FC/FS ratio changes gradually. Enterococci survive better than FC, and FC survive better than *S. bovis* and *S. equines* (Feachem 1975), where *S. bovis* has 99-percent die-off within 24 hours (Geldreich and Kenner 1969).

Another attempt to differentiate faecal sources was made by obtaining a series of FC and FS concentrations through time. An improved estimation of the pollution source was expected to be achieved as shown in Table 2.1. The rationale behind this assertion is the fact that in human faeces, the better-surviving enterococci are the dominant FS group, hence the FC/FS ratio will tend to fall (Feachem 1975). Whereas, in faecal material from cattle or pigs, the more poorly surviving *S. bovis* and *S. equines* dominate, hence the ratio will tend to rise. However, the numerical value of this ratio changes rapidly and makes the results of the analysis questionable (Rutkowski and Sjogren 1987). Hence, it is no longer in use (Burnes 2003).

Table 2.1. Faecal source related to FC/FS ratios (Feachem 1975)

Initial FC/FS ratio	Change of FC/FS ratio through time	Probable faecal source
>4	Rise	Uncertain
	Fall	Human
<0.7	Rise	Non-human
	Fall	Uncertain

Correlation coefficients for mean indicator densities against swimming-associated gastroenteritis rates at marine and freshwater bathing beaches in the USA during summer suggest that enterococci is a better indicator in marine water, while both enterococci and *E. coli* are equally efficient for monitoring water quality in freshwater as shown in Table 2.2. The correlation coefficient for *E. coli* is slightly higher than enterococci in freshwater, but they are not significantly different.

Table 2.2. Correlation coefficients for mean indicator densities against swimming-associated gastroenteritis rates during summer, extracted from USEPA(1986)

Indicator	Correlation coefficients	
	Marine	Fresh
Enterococci	0.75	0.74
<i>E. coli</i>	0.52	0.80
Faecal coliform	-0.01	-0.08
Total coliform	0.19	NA

2.3.2. Other faecal indicators

a. *Bacteroides*

Bacteroides has some advantages over other indicators. Firstly, in human intestinal microflora, *Bacteroides* is the most common genus and it outnumbers the coliforms (Kenzaka et al. 2001). They occur at levels up to three orders of magnitude greater than faecal coliform bacteria in human sewage (Santoro and Boehm 2007), i.e. approximately 10^{12} *Bacteroides* CFU/g faeces (Seurinck et al. 2005). Secondly, *Bacteroides* is an obligate anaerobe. This limits its potential for re-growth. However, there is a drawback to this characteristic in that it is more difficult to grow this strict anaerobe with culture-based methods. To circumvent this problem, molecular-based approaches using specific 16S primers have been developed (Bernhard and Field 2000a). The specific 16S primers used are not able to distinguish *Bacteroides* from *Prevotella*, a different genus of bacteria under the same order *Bacteroidales*. *Bacteroides-Prevotella* molecular markers exhibit “host-specific” patterns. This means that faecal organisms from different animals can be distinguished on the basis of their genetic sequencing. An elaboration of the use of *Bacteroides-Prevotella* molecular marker as a faecal indicator can be found in Section 2.4.4.1.

b. *Bifidobacterium*

Similar to *Bacteroides*, *Bifidobacterium* has some advantages. Firstly, it is one of the major bacterial inhabitants of the human intestine. *Bifidobacteria* are present in high numbers in fresh human faeces (Evison and James 1973), up to 10^{10} CFU/g faeces

(Seurinck et al. 2005b). Secondly, *Bifidobacteria* are obligate anaerobic bacteria. Hence, they are unable to grow in most water environments after discharge. This implies that when *Bifidobacteria* are detected in a water sample, it indicates a recent faecal pollution event. Thirdly, *Bifidobacteria* are rarely found in animals. Certain species are specific to humans, e.g., *Bifidobacterium adolescentis* and *B. breve* (Seurinck et al. 2005).

c. *Clostridium perfringens*

Clostridium perfringens (*C. perfringens*) is an enteric, gram-positive, anaerobic, spore-forming bacterium found in human and animal faeces (Scott et al. 2002). The resistant nature of the spores allows enhanced survival in environmental waters, hence *C. perfringens* may be a better indicator of faecal pollution in tropical waters (Griffin et al. 2000). However, this indicator commonly needs time-consuming culture-based methods for identification (Savichtcheva and Okabe 2006).

d. *Rhodococcus coprophilus*

Rhodococcus coprophilus (*R. coprophilus*) is present in the faeces of domesticated herbivores, pasture runoff and associated receiving waters and sediments, but not in human faecal wastes (Seurinck et al. 2005). Hence, *R. coprophilus* could be used as an indicator of herbivorous faecal pollution in water.

e. Coliphages

Coliphages are viruses that infect *E. coli*. They are non-pathogenic to humans and are more similar to enteric viruses with respect to physical characteristics, environmental persistence and resistance to treatment processes than indicator bacteria (Cole et al. 2003). Hence, they have been proposed as animal virus water pollution indicators (Borrego et al. 1987). They are ubiquitous in warm-blooded animal excrement, in sewage and natural waters polluted by faecal material (Borrego et al. 1987) and hence are used as indicators of faecal contamination (USEPA 2001).

There are two main groups of coliphages: somatic coliphages and male-specific (F+) coliphages. Somatic coliphages attach directly to the lipopolysaccharide (outer membrane) of *E. coli*. F-factor is a plasmid that is a fertility factor in certain strains of *E. coli*. When it is present, it codes for the pilus formation that allows for transfer of nucleic acid from one bacterium to another. F+ coliphages are RNA or DNA viruses that infect via this F-pilus of male strains of *E. coli* (USEPA 2001). F+ RNA coliphages have been more fully characterized. It has also been reported that animal and human faeces contain specifically different serotypes of RNA coliphages (Scott et al. 2002). Hence, microbial source tracking research has largely focused on the F+ RNA coliphages.

F+ RNA coliphages are divided into four main subgroups (Scott et al. 2002). Group I coliphages are present in faeces and sewage from cattle, sheep, pigs and many zoo mammals (Griffin et al. 2000). Group II is highly associated with pig and human faecal contamination. Group III is specifically found in humans. Group IV coliphages have a higher incidence in wastes associated with birds and zoo mammals.

f. *Bacteroides fragilis* phage

Bacteroides fragilis (*B. fragilis*) phages are mainly found in human faecal samples (Savichtcheva and Okabe 2006). Therefore, they could serve as a specific index for human faecal pollution. They do not replicate in the environment. Their presence in the environment has been found to significantly correlate with the presence of human enteric viruses (Scott et al. 2002). However, the difficulty in performing the assay limits the usefulness of *B. fragilis* phages as an indicator.

g. Enteric viruses

Enteric viruses are excreted in high concentrations in human and animal faeces (Fong et al. 2005). They are also more resistant to extreme environmental conditions and conventional wastewater treatment than many other sewage-associated pathogens and bacterial indicators.

Adenoviruses are a subgroup of the enteric viruses and one of the most important human pathogens present in polluted water (Fong et al. 2005). They have been found to survive three to five times longer than poliovirus in seawater, wastewater and tap water. But, they are slow growing and often do not produce cytopathogenic effects in cells. Therefore, they are consistently underestimated when other fast-growing enteric viruses are present.

Human adenoviruses are the only human enteric viruses that contain double-stranded DNA instead of RNA (Fong et al. 2005). Hence, they are potentially more stable in various environments than other human enteric viruses. They are also more resistant to UV irradiation and other water purification treatments because they are able to use the host cell DNA repair mechanism to repair damage in their DNA caused by UV irradiation. However, adenoviruses are not excreted by every human, which can lead to inconsistent detection in sewage-impacted waters (McQuaig et al. 2012).

h. Enterophages

Enterophages are a newly identified group of phages that have been recently isolated from recreational waters (Santiago-Rodríguez et al. 2010). They have not been conclusively linked to human faecal pollution. Santiago-Rodríguez et al. (2010) reported that they are in the process of molecularly characterizing isolates from enterophages, including the sequencing of the viral genomes, to propose them as viral indicators and possible surrogates of enteric viruses in recreational waters. There are at least three groups of enterophages that can be present in water contaminated with faeces (Bonilla et al. 2010). Future studies on the survival of enterophages, such as effects of sunlight and the prevalence of enterophages and enteric viruses in recreational waters, are still needed.

2.3.3. Conventional faecal indicators in tropical areas

Factors such as pH, temperature, sunlight (solar, both ultraviolet/UV and visible irradiation), predation, osmotic stress, nutrient deficiencies, particulate levels, turbidity, oxygen concentrations and microbial community composition affect

bacteria inactivation once bacteria reach receiving waters (Noble et al. 2004). Noble et al. (2004) reported that among the above-mentioned factors, temperature is the most important environmental factor influencing the survival of enteric bacteria while the effects of total suspended solids (TSS) and nutrients on rates of inactivation were not found to be statistically significant.

An earlier study by Hazen (1988) showed that *E. coli* survived significantly longer in-situ in tropical waters as compared to temperate waters. This might be because the average nutrient concentrations as well as ambient air and water temperatures do not vary substantially during the year. Rivera et al. (1988) also showed that *E. coli* might naturally occur in some tropical areas. Hence, tropical waters might have high densities of *E. coli* in the absence of pathogens or faecal sources. Another study by Muñiz et al. (1989) showed that faecal coliform, *Streptococcus faecalis* and other faecal streptococci bacteria did not accurately indicate the presence of pathogens in tropical waters due to their rate of survival and possibly indigenous nature.

Noble et al. (2004) showed that total coliform, *E. coli* and enterococci degraded significantly more rapidly at 20°C than at 14°C. Sherr et al. (1988) reported that the digestion times of protozoa appeared to be significantly influenced by temperature. The digestion rate is faster at higher temperatures, which is consistent with higher predation rates of the indicator bacteria. Walters et al. (2011) concurred with this result.

Coliforms grow all the time in tropical drainage channels (Janelle R. Thompson, MIT, personal communication, 2011). However, other factors such as predation and sunlight may affect the net growth rate of the indicator bacteria. Foley et al. (2010) mentioned that during the day, sunlight deactivates bacteria at a rate faster than their growth rate, resulting in net die-off, while the converse may be true at night. The inconsistent findings with regards to the survival of indicator bacteria in tropical waters suggest that the maximum contaminant levels applicable for temperate environments, mentioned in Section 2.2, may not be relevant for tropical waters.

Although there are apparent problems with these conventional indicators, a consensus with regards to the choice of indicator bacteria for tropical environments seems unavailable. Temperate MCLs have, however, been used to assess contamination levels in tropical environments (Kenzaka et al. 2001).

2.3.4. Summary

Microbiological indicators can be divided into two categories: conventional faecal indicators and other faecal indicators. Conventional faecal indicators are bacterial indicators which are comprised of total coliform, faecal coliform, *E. coli*, faecal streptococci and enterococci, while other faecal indicators include bacteria (*Bacteroides*, *Bifidobacterium*, *Clostridium perfringens*) and viruses (coliphage, enterophage, *Bacteroides fragilis* phage, enteric viruses). Total coliform bacteria are not faecal-specific due to their heterogeneity. Similarly, faecal coliform is not faecal-specific due to the inclusion of *Klebsiella* as part of the indicator. *E. coli* are more faecal-specific than total and faecal coliform. However, when the need to differentiate human and other warm-blooded animal faecal contamination is necessary, *E. coli* does not have this capability. Besides, *E. coli* are also known to grow in the tropical environment, causing monitoring results to be inaccurate. Hence, other faecal indicators, such as the anaerobic bacteria *Bacteroides*, *Bifidobacterium*, and *Clostridium perfringens* have been suggested as alternative indicators because they grow in the environment to only a limited degree. However, this characteristic is disadvantageous since it is difficult to enumerate these bacteria by the culture-based method. Therefore, other tracers (discussed in Section 2.5) and molecular-based approaches are investigated (discussed in Section 2.4.4).

2.4. Source tracking using microbiological indicators

The monitoring of recreational water quality with faecal indicators is a long-established practice, from as soon as concern about the hazard of untreated human faecal wastes to human health first arose. However, the search for the best faecal indicator has not ended yet and the development of microbial source tracking methodology is still an emerging topic. Studies have developed different approaches

to source tracking. In this section, the development of the enumeration methodology of commonly used faecal indicator bacteria is discussed in Sections 2.4.1 and 2.4.2. Then, this section discusses studies about emerging source tracking using microbiological indicators in Sections 2.4.3 and 2.4.4. Those studies are discussed and compared in terms of their methodologies. The applicability of the emerging methodologies is also discussed. Commonly used methods for source tracking (classical and enzymatic methods which are based on viable bacteria and culture-independent, library-independent methods which identify both viable and nonviable bacteria) are described in more detail.

2.4.1. Classical methods

Classical methods are generally based on metabolic reactions (Rompré et al. 2002). There are two classical methods, namely multiple-tube fermentation and membrane filtration methods.

2.4.1.1. Multiple-tube fermentation technique

Multiple-tube fermentation (MTF) consists of inoculating a series of tubes with appropriate decimal dilutions of the water sample (Rompré et al. 2002). It is easy to implement, can be performed by a technician with basic microbiological training and is relatively inexpensive as it requires unsophisticated laboratory equipment (Rompré et al. 2002). However, the method is also labour-intensive and extremely time-consuming. This is because many dilutions have to be processed for each water sample and the process requires 48 hr for presumptive results and another 48 hr for confirmation results. The results of the multiple-tube fermentation are expressed as the most probable number (MPN) of microorganisms present.

2.4.1.2. Membrane filter technique/filtration method

The membrane filter (MF) method consists of filtering a water sample through a 0.45- μm filter which retains bacteria, incubating the filter on a medium and enumerating colonies on the filter (Rompré et al. 2002). However, a universal medium which allows for the optimal enumeration of various coliform species currently does not

exist. This is because bacteria species originate from different environments and are present in a wide variety of physiological states. In addition, this method is still not sufficiently specific and a confirmation stage requiring a further 24-hr analysis is needed.

A significant advantage of the MF technique over the MTF method is that with MF, the examination of larger volumes of water is feasible (Rompré et al. 2002). This leads to greater sensitivity and reliability. MF is also a relatively simple method to use since many samples can be processed in a day, with limited laboratory equipment, by a technician with basic microbiological training. Bacterial counts from membrane filtration are expressed as colony-forming units (CFU).

2.4.2. Enzymatic methods

Enzyme-substrate-based technology refers to the detection of microorganisms based on specific enzymatic activity. The addition of chromogenic and fluorogenic substrates to cultivation media (agar and liquid media) produce colour and fluorescence, respectively, upon cleavage by a specific enzyme. This has increased the sensitivity and rapidity of the methods compared to classical methods (Rompré et al. 2002). Tests based on the defined substrate technology are easy to use and give a more realistic estimate of indicators of bacteriological contamination in water. This method might be more expensive in terms of consumables than the classical methods, but the classical methods require additional confirmation steps which cause the costs incurred by both methods to be equivalent. In all cases, enzymatic methods require less manpower. Therefore, the commercial cost is lower (Rompré et al. 2002).

2.4.2.1. Multiple-well enzyme substrate formats

This substrate technology is based on the principle that only target microbes (TC including *E. coli* or enterococci) are fed and no substrates are provided for other bacteria. No additional confirmatory tests are needed. Several commercial test kits have been developed based on substrate technology such as Colilert (IDEXX Laboratories, Portland, ME, USA), Colisure (IDEXX Laboratories, Portland, ME, USA)

and ColiQuick (Hach, Loveland, CO, USA). The most widely used reagent among them is Colilert® for TC and *E. coli* enumeration (Rompré et al. 2002). IDEXX also produces Enterolert® for enterococci enumeration. The functioning of Colilert or Colilert-18 and Enterolert is compared in Table 2.3. The reagents are used together with the IDEXX Quanti-Tray® or Quanti-Tray®/2000 product to determine a statistical most probable number (MPN) enumeration. The MPN values were obtained using the Poisson law of normal distribution curve (Siang Lyn, Arachem (M) Sdn Bhd, personal communication, 2013). With Quanti-Tray/2000, up to 2,420 MPN/100 mL of sample can be enumerated without dilution. The 95% confidence limits for the Quanti-Tray/2000 can also be obtained (IDEXX 2010b).

Table 2.3. Colilert, Colilert-18 and Enterolert (IDEXX 2010a)

	Colilert and Colilert-18		Enterolert
	Total coliform	<i>E. coli</i>	Enterococci
Substrate (that is cleaved)	<i>ortho</i> -nitrophenyl- β -D-galactopyranoside (ONPG)	4-methyl-umbelliferyl- β -D-glucuronide (MUG)	4-methyl-umbelliferyl- β -D-glucoside
Enzyme	β -galactosidase	β -glucuronidase	β -glucosidase
Product	<i>o</i> -nitrophenol	4-methylumbelliferone	4-methylumbelliferone
Characteristics of product	Yellow-coloured	Yellow-coloured and blue fluorescent (when it is exposed to long-wave UV light, 365 nm)	Blue fluorescent (when it is exposed to long-wave UV light, 365 nm)

The assay of β -glucuronidase activity for *E. coli* identification appears to be less prone to false positives than the multiple tube fermentation technique that is not totally specific for *E. coli* (Cabral and Marques 2006). Yakub et al. (2002) also showed that defined substrate methodology such as with Colilert and Enterolert correlated well to membrane filtration for surface water and treated wastewater effluent samples. Sercu et al. (2011a) quantified the degree of false-positive results of Colilert and Enterolert in Quanti-Tray/2000 for samples from an urban creek by molecular methods and found that the false-positive rates under dry weather conditions were 4 to 8% for total coliform, 14 to 23% for *E. coli*, and 6 to 20% for Enterolert. The latter suggests that Colilert and Enterolert can be improved, for example by increasing selective inhibition of non-target bacteria to improve the assays accuracy.

2.4.2.2. Membrane filter using enzymatic methods

The classical agar medium used for total coliform bacteria and *E. coli* enumeration is modified with specific chromogenic and/or fluorogenic substrates for the detection of β -galactosidase and/or β -glucuronidase. Different commercial agar media are now available for the detection of total coliform and *E. coli*. One such product is the m-ColiBlue24 broth (Hach, Loveland, CO, USA). According to Hach(1999), the m-ColiBlue24 broth may allow a few varieties of non-coliform bacteria (*Pseudomonas*, *Vibrio*, and *Aeromonas spp.*) to grow and form red colonies. They can be distinguished from total coliform by the oxidase test. *E. coli* O157:H7 is also not produced as a blue colony, but only a red colony (total coliform counts) (Hach 2002). In addition, there are only two strains of *E. coli* that have tested positive (producing blue colony) with m-ColiBlue24 Broth. A comparison of the two enzymatic methods can be found in Table 2.4.

As mentioned above, the results of the multiple-tube fermentation and the multiple-well enzyme substrate formats are expressed as the most probable number (MPN) of microorganisms present, a statistical estimate of the mean number of microorganisms in the sample (Rompré et al. 2002). For MPN-based methods, samples must be adequately shaken to break up any clumps and provide an even distribution of bacteria. If the sample is not gently shaken, the MPN value may underestimate the actual bacterial density (USEPA 2005). On the other hand, the formula to derive the MPN value is rather biased and may overestimate the value(USEPA 2005), thus offsetting underestimates for samples in which clumps are not broken up. In contrast to MPN tests, the MF method enumerates the results directly in CFU and hence seems more accurate and precise. However, the MF method may also underestimate the number of viable bacteria (USEPA 2005) because the filter suppresses the bacteria and because stressed bacteria (common condition of faecal bacteria when they are in the environment) may need more time than the incubation time to recover and grow to produce a colony. The differences in recovery between the CFU-based method and MPN-based method are generally statistically insignificant(USEPA 2005). Bolton et al. (1982) and Fricker et al. (1997) reported that

the results obtained by the MPN method were in close agreement with the MF method. But, Gronewold and Wolpert (2008) observed that the MTF-derived MPN procedure is more variable than the MF-derived CFU procedure. Besides, MTF-derived MPN estimates are somewhat higher on average than CFU estimates on split samples from the same water bodies. This MPN and CFU intra-sample variability does not come from extrinsic sources of variability (such as departure from sampling protocol), but from intrinsic variability (such as natural dispersion of bacteria within sample containers and randomness).

Table 2.4. Comparison of two enzymatic methods

Multiple-well enzyme substrate formats (MPN) e.g. Colilert	MF using enzymatic methods (CFU) e.g. m-ColiBlue24 Broth	References
Advantages	Disadvantages	
This method can analyse turbid samples. Colilert is suitable for the quantification of coliforms and <i>E. coli</i> in a wide range of sample types.	This method is not suitable for samples with high turbidity as particulate matter will block the membrane.	Bolton et al. (1982) and Fricker et al. (1997)
Procedure is less complex.	Procedure is more complex.	Bolton et al. (1982)
This method does not deal with enumeration of colonies.	This method needs enumeration of colonies which may be difficult when colonies developing on a membrane coalesce.	Bolton et al. (1982)
If the sample is not diluted, enumeration can go to 2×10^3 with Quanti-Tray/2000 (but only 2×10^2 with Quanti-Tray).	If the sample is not diluted, enumeration can only go to 2×10^2 . Upper limit of the MF count is approximately 80.	Fricker et al. (1997) and Hach (1999)
Disadvantages	Advantages	
More expensive—cost of test with Colilert is approximately USD 6.70 per test.	Less expensive—cost of test with m-ColiBlue24 is approximately USD 2.65 per test.	Hach (2006)
A fluorescent lamp is needed.	No fluorescent lamp is needed.	Hach (2008)
This method offers a semi-quantitative enumeration (MPN).	This method offers a quantitative enumeration (CFU).	Rompré et al. (2002)

Colilert and m-ColiBlue24 are included in USEPA's list of approved microbiological methods for the detection of *E. coli* in ambient waters (USEPA 2005). Similarly, Enterolert is an approved microbiological method for the detection of enterococci in ambient waters (USEPA 2005).

2.4.3. Library-dependent methods

A library-dependent method is a method whereby a library of genotypic or phenotypic characteristics is constructed first and then used as a comparison standard for the identification of faecal sources. Genotypic characteristics are related to the genetic makeup of the target organism, while phenotypic characteristics are those that appear as the result of the interaction between genotypic characteristics and the environment. One strength of phenotypic over genotypic methods is that they are relatively inexpensive, hence many isolates can be characterized (Harwood et al. 2003). The inherent limitations of library-dependent methods are as follows: Firstly, the constructed library must be representative. This is costly and time- and labour-intensive. Secondly, results vary depending on geographical and temporal variability. This variability must be assessed because it affects the reproducibility of these methods.

2.4.3.1. Phenotypic methods

a. Antibiotic resistance

This method is based on the hypothesis that source animals treated with antibiotics will host bacteria in their gut which have developed resistance to those antibiotics (Farmer 2006). The process of this method involves sample collection, target organism isolation, and application of antibiotics to the isolated target organisms. These steps alone, excluding sample collection, require days of incubation.

The multiple-antibiotic-resistance (MAR) method relies upon the use of many antibiotics at only one concentration each. Parveen et al. (1997) used ten antibiotics on *E. coli* isolates from point and non-point sources in Apalachicola Bay, Florida, USA. Waters from urban areas harboured higher percentages of resistant *E. coli* strains than rural waters. This was probably due to the widespread use of chemotherapeutic drugs by residents in urban areas. Hence, human faeces were likely the source of faecal pollution.

Antibiotic resistance analysis (ARA) was coined to describe a method of discriminating faecal sources in which more than one concentration of each antibiotic were used to develop the antibiotic resistance pattern. Target microorganisms (Seurinck et al. 2005) in ARA studies have been faecal streptococci (Wiggins 1996; Hagedorn et al. 1999; Wiggins et al. 1999; Harwood et al. 2000), faecal coliforms (Whitlock et al. 2002; Burnes 2003), enterococci(Wiggins et al. 2003) and *E. coli*(Carroll et al. 2009).

Hagedorn et al. (1999) applied ARA with faecal streptococci to samples collected from a rural area with an agriculture-based economy and identified cattle to be the source of faecal pollution. Their results were confirmed when much lower faecal coliform numbers were found in the following year, following the installation of fences which resulted in limiting cattle access to nearby streams.

Wiggins et al. (1999) used ARA with faecal streptococci to classify isolates into a different number of possible sources. They found that the correct classification rate in samples that were collected from many different sources over a 4-year period were consistently more than 60%. Similarly correct classification rates of ARA with faecal streptococci were obtained by Harwood et al.(2000). According to Harwood et al. (2000), correct classification rates of 60 to 70% are suitable for regulatory agencies and ARA of faecal streptococci was better in terms of source prediction than ARA with other microbiological species. However, during the 4-year period of sample collection (five collection events from 1993 to 1997), Wiggins et al. (1999) used different sets of antibiotics in order to find a set of antibiotics that provided the best classification. This might indicate that the resistance patterns were only temporarily stable, i.e. over the span of one year (Wiggins et al. 2003), likely due to changes in the pattern of antibiotics use.

Whitlock et al. (2002) applied ARA with faecal coliforms to samples collected from Stevenson Creek, a subtropical urban watershed in Florida, USA. When the faecal coliform levels were elevated, wild animals were found to be the dominant contributors to faecal contamination, but when faecal coliform levels were low, other

sources of faecal coliforms such as humans and dogs became apparent. ARA with faecal coliforms have also been used for samples collected from Big Creek, a mixed-use watershed in Alabama, USA, where faecal coliform quantification was an integral part of an existing monitoring plan (Burnes 2003). During the baseline event, chicken and livestock were found to be responsible for the majority of faecal coliforms isolates in upstream reaches, while human sources appeared to contribute a substantial portion downstream near the confluence with the Chattahoochee River.

Wiggins et al. (2003) applied ARA to enterococci for six different watersheds in Virginia, USA. The six watersheds included four rural watersheds and two watersheds that had mixed urban and rural land uses. Libraries from several other watersheds were merged to provide a comprehensive database. This merged library was found to classify the samples better than the smaller library; however, it did not necessarily give the correct classification all the time.

ARA with *E. coli* was applied to Ningi Creek, an urbanizing catchment, in Queensland, Australia (Carroll et al. 2009). This study found that the nature of possible faecal pollution in a waterway can be directly correlated to the surrounding land use. In the upstream segment of the catchment, where the predominant land use was natural bushland and agriculture, the main source was found to be non-human. In the downstream reach, which was urbanised and included areas where onsite systems were used for the treatment and dispersal of wastewater, increasing human source isolates were identified.

Generally, antibiotic resistance method is not a good alternative method of source tracking because: (i) it requires days of incubation before the results are obtained (Farmer 2006) and (ii) it needs to be updated over a short period of time (one year) because resistance patterns depend on the pattern of antibiotics use and are therefore only temporarily stable (Wiggins et al. 2003). The antibiotic resistance method works well in differentiating contamination from urban vs. rural area (Parveen et al. 1997; Wiggins et al. 2003) or human vs. non-human sources (Whitlock

et al. 2002; Carroll et al. 2009). In contrast, when the catchment is all urbanised or when contamination is mainly human faecal and sewage contamination (such as Singapore especially Marina catchment), the antibiotic resistance method does not effectively answer the question of whether contamination occurs.

b. Carbon source utilization

Carbon source utilization (CSU) methods (also called the Carbon Utilization Profiles (CUP) method) is a method that is based on the differences among bacteria in their use of carbon and nitrogen sources for energy and growth (Seurinck et al. 2005). CSU analysis was applied to enterococci by Hagedorn et al. (2003) and Wallis and Taylor (2003) using different techniques. Hagedorn et al. (2003) used the commercial Biolog system, while Wallis and Taylor (2003) used the PhenePlate system. The former is based on the exchange of electrons generated during respiration (Miller and Rhoden 1991) and the latter is based on the fermentation kinetics of eleven carbohydrates by the bacterial isolates. The discriminatory ability of the PhenePlate system was compared with pulsed-field gel electrophoresis (PFGE) and the result was found to be satisfactory.

2.4.3.2. Genotypic methods

Most of the research on genotypic methods has concentrated on ribotyping, pulsed-field gel electrophoresis (PFGE), and repetitive element PCR methods. All of these methods are based on the assumption that specific markers or strains of bacteria are associated with specific animal species (Hartel et al. 2004).

a. Ribotyping

Ribotyping is a method of DNA fingerprinting used to identify conserved rRNA genes using oligonucleotide probes (Scott et al. 2002). A major advantage of ribotyping is that rRNA sequences are highly conservative and are present as multiple copies in bacterial genomes. Ribotyping has also proved to be a useful epidemiological technique for various bacteria, including *E. coli*, *Salmonella enteric*, *Vibrio cholera* O1 and *Vibrio vulnificus* (Parveen et al. 1999). However, this method is also known to

require labour-intensive procedures including cell lysis and DNA extraction, DNA cleavage by restriction enzyme, agarose gel electrophoresis separating the resulting fragments, Southern blot transfer of DNA fragments onto a membrane, hybridization with the DNA probes and detection of the hybridization signal (Severino and Brisse 2005). Another major drawback of ribotyping is that it shows limited discriminatory power in some bacterial groups. It is discriminatory only for those species like the *Enterobacteriaceae*, which possess a high number (approximately seven) of rRNA operons (Severino and Brisse 2005). Ribotyping with *E. coli* has been used by Parveen et al. (1999) and Carson et al. (2001). Both studies reported that this method was useful for differentiating human and non-human sources of faecal pollution.

b. Pulsed-field gel electrophoresis

Pulsed-field gel electrophoresis (PFGE) detects variations in nucleotide sequences of chromosomal DNA (Parveen et al. 2001) or polymorphisms using restriction enzymes, i.e. rare-cutting restriction endonuclease (Scott et al. 2002). The restriction endonuclease is an enzyme that recognizes a DNA sequence that occurs infrequently in the genome. The restriction digest is separated by gel electrophoresis of alternately pulsed, perpendicularly oriented electrical fields and stained, resulting in a fingerprint (Maier et al. 2000). PFGE is used to detect fragments of higher molecular weight, often greater than 40 kilobase pairs. According to Severino and Brisse(2005), PFGE is slightly more discriminatory than ribotyping. The PFGE technique was applied to *E. coli* for samples collected from a watershed with mixed land use (Simmons et al. 1995) and an urban watershed (Simmons et al. 2000), both in Virginia, USA. In both watersheds, non-human sources appeared to be a contributor of faecal pollution. However, Parveen et al. (2001) were not able to differentiate human-source and non-human-source isolates using PFGE profiles of *E. coli*. Parveen et al. (2001) also found that human-source PFGE profiles were less diverse than non-human-source isolates.

c. Repetitive element PCR

The repetitive element PCR method is based on the use of primers which are analogous to repetitive DNA elements present in various locations at certain interval

within the prokaryotic genome to produce highly specific genomic fingerprints (Scott et al. 2002). This method uses three methods of repetitive sequence analysis where each repetitive sequence targets a specific family of repetitive elements. The three methods are repetitive extragenic palindromic sequence PCR (REP-PCR) which targets REP sequences, enterobacterial repetitive intergenic consensus sequence PCR (ERIC-PCR), and PCR with extragenic repeating elements (BOX-PCR). REP, ERIC and BOX sequences are unrelated to each other (Maier et al. 2000), but because of their high copy number and variable spacing in the genome, they have a mutual use as targets for PCR fingerprinting.

Carson et al. (2003) compared repetitive element PCR and ribotyping. Repetitive element PCR was considered superior to the ribotyping method. This is because repetitive element PCR is simple, accurate, fast and can differentiate between closely related strains of bacteria (Dombek et al. 2000). It is simple because DNA fingerprints are generated by using whole cell suspensions, which eliminated the need for DNA purification. Dombek et al. (2000) applied repetitive element PCR with REP and BOX primers to *E. coli*. The results showed that the DNA fingerprints obtained with the BOX primer were more effective for grouping *E. coli* strains than the DNA fingerprints obtained with REP primer. Hence, repetitive PCR with the BOX primer may be useful for determining the source groups of *E. coli* isolates and the sources of closely related *E. coli* strains obtained from environmental samples.

In general, the implementation of genotypic methods (such as ribotyping, PFGE and repetitive element PCR) is not an effective choice in a very urbanised area or when contamination is mainly human faecal and sewage contamination (such as Singapore especially Marina catchment) because genotypic methods require the development of a library of target markers from different sources beforehand. Even after the library is constructed, genotypic methods (such as ribotyping) are usually labour-intensive procedures. Some genotypic methods also require certain criteria of the target markers which make them less flexible. For example, ribotyping is only discriminatory for species which possess a high number of rRNA operons. PFGE is only

used to detect fragments of higher molecular weight (often greater than 40 kilo base pairs).

2.4.4. Culture-independent, library-independent methods: host-specific marker

2.4.4.1. Host-specific bacterial markers: *Bifidobacterium*, *Bacteroides-Prevotella*

Bernhard and Field (2000a) identified seven potential host-specific 16S rDNA genetic markers: two in human and five in cow faecal DNA. The two human-specific markers were found in *Bacteroides-Prevotella* and *Bifidobacterium*, and the five cow-specific markers were found in *Bacteroides-Prevotella* (three) and *Bifidobacterium* (two). According to Bernhard and Field (2000a), the *Bacteroides-Prevotella* marker is better than *Bifidobacterium* because of its ease of detection and longer survival rate in water. The number of *Bacteroides* that can be found in human faecal matter also fluctuates less over time than *Bifidobacterium*. According to taxonomic classification, *Bacteroides* and *Prevotella* refer to different genera from different families (*Bacteroidaceae* and *Prevotellaceae*, respectively) but from the same order, *Bacteroidales*.

a. Specificity and sensitivity

As shown in Table 2.5, Bernhard and Field (2000a) used a common primer for both the human- and cow-specific *Bacteroides-Prevotella* marker. Bernhard and Field (2000b) then refined their research by designing specific PCR primers which were able to discriminate between human and ruminant sources of faecal contamination. The human-factor primers were HF134F, HF183F and HF654R. HF134F and HF654R targeted HF10 genes and HF183F targeted HF8 genes. HF8 genes are more specific, in terms of source identification, and more widely distributed among humans; therefore HF183F is more effective and favourable because it allows wider spatial suitability. Ahmed et al. (2008) also showed that HF183 marker was more reliable than HF134 marker because the HF134 marker can also be found in dogs.

Table 2.5. Primers for human- and cow-specific *Bacteroides-Prevotella* marker, extracted from Bernhard and Field (2000a)

Host specificity	Forward primer	Reverse primer	Enzyme used
Human	Bac32F	Bac708R	HaeIII
Cow	Bac32F	Bac708R	HaeIII
Cow	Bac32F	Bac708R	Acil
Cow	Bac32F	Bac303R	None

Previously, the host-specific marker assay was designed to determine the presence or absence of a particular source—a qualitative measure (Bernhard and Field 2000a). Recently a quantitative assay for human-specific HF183 *Bacteroides* 16S rRNA genetic marker (in short, HF183 marker) has been developed using real-time quantitative PCR (RT-PCR or QPCR) methods (Seurinck et al. 2005a). This real-time PCR method is more sensitive than conventional PCR. The different sensitivity can be attributed to different reverse primers, number of cycles, polymerase and dye used (Seurinck et al. 2005a), as summarised in Table 2.6. Both conventional PCR and QPCR use the same forward primer, HF183F by Bernhard and Field (2000b). However, the two methods use different reverse primers: Bac708R by Bernhard and Field (2000a) for conventional PCR and Bac242R by Seurinck et al. (2005a) for QPCR. The use of Bac242R allows shorter target base pairs (there are only 82 base pairs between HF183F and Bac242R). In addition, the higher number of cycles, the use of Hot Start DNA polymerase and SYBR Green I dye result in higher sensitivity than a lower number of cycles, use of the common *Taq* polymerase and traditional dye ethidium bromide.

Table 2.6. Differences of conventional PCR with QPCR

	Conventional PCR	QPCR
Reverse primer	Bac708R by Bernhard and Field (2000a)	Bac242R by Seurinck et al. (2005a)
Number of cycles	35 cycles	40 cycles
Polymerase	<i>Taq</i> polymerase	Hot Start DNA polymerase
Dye	Traditional dye ethidium bromide	SYBR Green I

b. Persistence

Kreader(1998) reported that persistence of PCR-detectable DNA from an anaerobic bacterium, *Bacteroidesdistasonis*, is influenced by temperature and predation, similar to the survival of aerobic bacteria detected by culture. At low incubation or storage temperatures, at which predators and degradative processes are less active, the PCR target is preserved for an extended time(Seurinck et al. 2005a; Schulz and Childers 2011) regardless of the source water temperature(Kreader 1998). A summary of *Bacteroides* persistence at different temperatures is tabulated in Table 2.7. Walters and Field (2009) reported that natural sunlight did not significantly affect the persistence of either total (both live and dead cells) HF *Bacteroidales* DNA markers or cultivable *E. coli*. According to Bernhard and Field (2000a), the seawater matrix affects *Bacteroidales*(the order of *Bacteroides-Prevotella*) less than *E. coli*. This is also supported by Schulz and Childers (2011) who indicate that *Bacteroidales* markers had slower decay rates at higher levels of salinity.

Table 2.7. *Bacteroides* persistence at different temperature

	Kreader(1998)	Seurinck et al. (2005a)
<i>Bacteroides</i>	<i>B. distasonis</i>	Human-specific HF183 <i>Bacteroides</i> 16S rRNA
Method	Conventional PCR	QPCR
28-30°C	30°C: Detected up to 1 day	28°C: Decreased two log-units after 6 days
24°C	Detected up to 2 days	NA
12-14°C	14°C: detected up to 5 days	12°C: decreased one log-unit after 10 days
4°C	Detected for at least 2 weeks	Did not decrease significantly throughout 24 days

c. Application of HF183 marker

Qualitative studies of the host-specific *Bacteroides-Prevotella* marker have recently gained in popularity and have been widely used(Shanks et al. 2006; Gawler et al. 2007; Santoro and Boehm 2007; Jenkins et al. 2009; Nshimiyimana 2010). The host-specific *Bacteroides-Prevotella* marker has been used for samples obtained from a forested watershed in Oregon, USA in which six different forward primers, including HF183F were used(Shanks et al. 2006). Santoro and Boehm (2007) showed in a

qualitative study that there was no relationship between the occurrence of the HF marker and faecal indicator bacteria (FIB) abundance for individual samples. However, when the occurrence of the HF marker and FIB abundance were considered together by year, a higher incidence of the HF marker was correlated with higher FIB abundance.

In a quantitative study, Sercu et al. (2009) reported that the human-specific *Bacteroides* marker was not consistently correlated with FIB, although there is exclusive occurrence of high human-specific *Bacteroides* marker concentrations with high FIB concentrations. Sauer et al. (2011) also did not find correlation between human *Bacteroides* genetic marker and *E. coli* or enterococci culture results.

Quantitative risk assessment studies by Ashbolt et al. (2010) have estimated that levels of $>8.6 \times 10^2$ copies/100mL of human *Bacteroides* genetic marker may pose a health risk in recreation waters. This level of *Bacteroides* genetic marker corresponds to 1.5×10^3 HF genome equivalent/100mL according to the method described by Nshimyimana et al. (2014). However, the limitation of the QPCR method is that it detects DNA and thus cannot differentiate between viable and nonviable bacteria (Kaushik et al. 2012). The use of QPCR data in health risk assessment could give rise to false positives, because some of the cells detected may be dead cells that are not capable of infecting humans (Kaushik et al. 2012).

Despite the inability to differentiate viable and nonviable bacteria, the performance of the HF183 marker has gained positive reviews. Santoro and Boehm (2007) reported that there was greater than 99% sequence similarity between sequences obtained from multiple states within the USA (Oregon, Michigan, California) and multiple countries (Belgium, Canada, Australia) leading to the conclusion that there appears to be little genetic variability in the sequence type targeted by the HF183F and Bac708R primers. Gawler et al. (2007) also used HF183 marker in four countries, namely France, Ireland, Portugal and the United Kingdom, and reported that the HF183 marker was generally applicable in all the countries.

Besides temperate areas, HF183 marker also has good specificity in tropical areas such as Bangladesh (Ahmed et al. 2010b), Puerto Rico (Santiago-Rodriguez et al. 2012; Toledo-Hernandez et al. 2013), Hawaii (Boehm et al. 2011) and Kenya (Jenkins et al. 2009), as summarised in Table 2.8. Jenkins et al. (2009) used HF183 as one of the five quantitative assays and reported that it demonstrated strong potential to be used in diverse geographic regions in Kenya and other African countries. Toledo-Hernandez et al. (2013) also reported that HF183 marker had the lowest detection frequencies at sites in Puerto Rico with low-level human development and density and the highest detection frequencies at sites impacted by sewage treatment plants. Boehm et al. (2011) reported that HF183 marker was detected in a drain sample in Hawaii, suggesting exfiltration of septage or cesspool discharge to the drain.

Nshimyimana (2010) detected HF183 marker in drain water samples in an urban residential area in Singapore that is completely served with sewer lines and without obvious illicit discharge to the drain. Therefore, there is a question whether exfiltration of sewage to the drain can be the possible cause of contamination. This study further explores this question.

Table 2.8. Comparison of studies which analysed HF183 marker in tropical areas

Reference	Sampling location	Type of sampling site	Type of analysis	Outcome related to this study
Ahmed et al. (2010b)	Dhaka, Bangladesh	urban (but with slum area not served with proper sanitation system - developing country); lake	Quantitative	HF 183 marker was detected in 13 of 15 human faecal samples tested and was not detected in 28 of 30 animal faecal samples. HF183 marker was detected in 1 dog and 1 cat sample. HF183 concentrations in lake water samples were 3.9×10^4 to 6.3×10^7 genomic units/100mL which is 1–2 orders of magnitude lower than that found in human faeces or sewage.
Santiago-Rodriguez et al. (2012)	Puerto Rico	urban and rural; river	Qualitative	HF183 marker was detected mostly in samples collected downstream of wastewater treatment plants, suggesting that its presence indicates human faecal matter input.
Toledo-Hernandez et al. (2013)	Puerto Rico	urban and rural; river	Qualitative	HF183 marker was highly specific as it did not react with non-target samples. HF183 marker showed results consistent with predictable sources: sites near low-level human development and density yielded the lowest detection frequencies and sites impacted by sewage treatment plants yielded the highest detection frequencies among the study sites.
Boehm et al. (2011)	Hawaii, USA	tropical rain forests with small farms and urban areas; mostly river	Qualitative	HF183 marker was detected in 2 cesspool samples and 6 of 52 river samples collected over 2 years. HF183 marker was detected in 1 of 6 drain samples, suggesting exfiltrated septage or cesspool discharge can make its way to the drain.
Jenkins et al. (2009)	Kenya	rural (developing country); river	Quantitative	HF183 marker was detected in 7 of 12 human faecal samples and 4 of 5 sewage samples.
Nshimiyimana (2010)	Kranji Catchment, Singapore	urban and rural; drain	Qualitative	HF183 marker of environmental surface water samples collected in Singapore was clustered within the same evolutionary group as HF183 marker of surface water samples collected in the USA.

Nshimiyimana (2010) used the HF183 assay and in a phylogenetic tree compared the genes obtained in Singapore, atropical climate, with the genes obtained by Santoro and Boehm (2007) in temperate climates. Nshimiyimana (2010) concluded that they were clustered within the same evolutionary group. This confirmed the use of the

HF183 assay as a viable technique for human faecal source tracking in Singapore. For qualitative monitoring, Nshimiyimana (2010) demonstrated that high sensitivity in detecting human faecal matter in water could be obtained using nested PCR analysis with the HF183 assay. Nshimiyimana et al. (2014) also have optimised QPCR methods for environmental surface water samples from Singapore. The sequences for the primers and the pair-up of the PCR and primers can be found in Table 2.9 and Table 2.10, respectively.

Table 2.9. Sequences of primers used for human-specific *Bacteroides-Prevotella* marker

Primer	Sequence (5'-3')	Reference
Bac32F	AACGCTAGCTACAGGCTT	Bernhard and Field (2000a)
Bac708R	CAATCGGAGTTCTTCGTG	Bernhard and Field (2000a)
HF183F	ATCATGAGTTCACATGTCCG	Bernhard and Field (2000b)
Bac242R	TACCCCGCCTACTATCTAATG	Seurinck et al. (2005a)

Table 2.10. PCR and its primers for the detection of human

PCR	Forward primer	Reverse primer	Reference
Nested PCR (step 1)	Bac32F	Bac708R	Nshimiyimana (2010)
Nested PCR (step 2)	HF183F	Bac708R	Nshimiyimana (2010)
Real-time PCR	HF183F	Bac242R	Seurinck et al. (2005a) Sercu et al. (2009)

The study by Sercu et al. (2009) reported that there is exclusive occurrence of high HF183 marker concentrations with high FIB concentrations was conducted in a temperate area. Nshimiyimana (2010) has shown the applicability of qualitative HF183 in a tropical area (Singapore). Therefore, it is worth investigating if high HF183 marker concentrations (quantitative HF183) also occur with high FIB concentrations in tropical areas.

d. Summary

The discussion of host-specific bacterial markers above highlights *Bacteroides-Prevotella* as a potential marker because of its ease of detection, longer survival rate

in water and relatively stable abundance in the human faecal matter (Bernhard and Field 2000a). Specifically, it highlights HF183 marker because it targets a more specific gene that is likely found in humans only (Bernhard and Field 2000b) and is applicable in many areas: North America (the USA, Canada and Puerto Rico), many countries in Europe (Belgium, France, Ireland, Portugal and the United Kingdom), Australia, tropical Africa (Kenya) and Asia (Singapore and Bangladesh) (Gawler et al. 2007; Santoro and Boehm 2007; Jenkins et al. 2009; Nshimiyimana 2010; Ahmed et al. 2010b; Santiago-Rodriguez et al. 2012). The methodology to detect HF183 marker has undergone development from a qualitative to a quantitative methodology. The quantitative methodology (QPCR) has higher sensitivity because it uses Bac242R as the reverse primer, has 40 cycles, uses Hot Start DNA polymerase and SYBR Green I as the dye (Seurinck et al. 2005a).

2.4.4.2. Host-specific enteric viruses

Viral markers can be used to predict the occurrence of human viruses (Fong et al. 2005) which may be part of human faeces and therefore may indicate human faecal contamination. Nonetheless, a limitation of using viral markers is that their concentrations appear to be low in sewage compared to bacterial markers such as *Bacteroides* (Ahmed et al. 2010a). Therefore, a large sample volume needs to be analysed for these markers to be detectable. Viral markers can be detected with PCR-based methods. PCR viral detection is less laborious and time-consuming than cell culture methods (Fong et al. 2005) and is also more specific and sensitive. PCR is capable of detecting viruses that are either difficult to grow in cultured cells or replicate without producing cytopathogenic (destructive) effects in cells. However, as PCR detects nucleic acids from both infectious and non-infectious viruses, the data derived from direct PCR is most useful when the objective is solely to track the source of water pollution or when the infectivity of these viruses is not of public health concern.

Qualitative studies using viral markers have been conducted by Boehm et al. (2003), Fong et al. (2005) and Ahmed et al. (2010a). Fong et al. (2005) detected human

enteroviruses, bovine enteroviruses and human adenoviruses using reverse transcription PCR and nested PCR. Nested PCR is generally more specific and exhibits a higher level of sensitivity in detecting enteric viruses from environmental samples. These qualitative studies do not provide information regarding the magnitude of faecal pollution in the streams under investigation. Hence, a quantitative method, real-time PCR, is being developed (Ahmed et al. 2010a). Despite the ability to detect host-specific enteric viruses, Ahmed et al.(2010a) recommend that multiple markers (viral and bacterial) should be used to identify human faecal pollution accurately.

2.5. Physical, inorganic and organic indicators

2.5.1. Physical and inorganic indicators

Common physical and inorganic indicators include the following: water temperature, specific conductance, turbidity, boron, chloride, chloride/bromide ratio (Cl/Br ratio) and orthophosphate.

Direct discharge from point sources may be relatively cooler or warmer than the water temperature in the drainage system (Hyer 2007). Therefore, water temperature can be useful to identify those point sources. However, when the discharge is not direct, temperature may not show substantial differences. In Singapore, direct discharge from point sources is prohibited, therefore it is supposedly non-existent. If there is human sewage contamination, it likely originates from leaking sewers since Singapore has an extensive sewerage reticulation system. Discharge from leaking sewers, if any, is indirect and most probably in the form of seepage with limited flow; therefore water temperature is not likely to be an effective indicator in this case.

The ionic strength of sewage is usually far greater than that of the natural environment (Hyer 2007). Hence, the overall electrical conductivity in water in the drainage system may be elevated if sewage is present. Vandenberg et al. (2005) also reported that electrical conductivity is a good tracer because it has a close

relationship to chloride, despite the fact that it is affected by many other parameters and processes. In addition, electrical conductivity can be measured easily in real-time without the addition of synthetic tracers. Specific conductance is temperature-compensated conductivity; it reports conductivity as if the sample had been at 25 °C (YSI 2007).

Direct discharge of sewage into a stream can also increase turbidity levels (Hyer 2007). However, Hyer (2007) did not use turbidity in his study since the particulates that contributed to elevated turbidity levels settled out rapidly. In the case of indirect discharge, turbidity is, moreover, an ineffective indicator. Indirect discharge in Singapore is investigated in this study; therefore turbidity is not an advisable option as an indicator.

Boron concentrations are relatively low in natural waters (Hyer 2007). Boron, in the form of sodium borate (borax) is used in cleaning products. Thus, the presence of boron indicates likely contamination by sewage containing household cleaning residues. Boron present in animals and humans is maintained by urinary excretion. Sutherland et al. (1999) reported that daily urinary and faecal boron excretion averaged 85% and 8% of dietary boron, respectively. Together with faecal coliform bacteria and surfactants, boron was declared by Hyer (2007) as the most sensitive tracer in the identification of minor sewage sources. Held et al. (2006) also mentioned that boron can be related to wastewater origin with high plausibility.

Elevated chloride concentrations can indicate the presence of sewage (Hyer 2007). Chloride is an ideal tracer because it is conservative, easily measured, present at much higher levels near the source and innocuous (Vandenberg et al. 2005). Chloride analysis is relatively inexpensive. Chloride is a poor indicator in temperate climates because it is a component of road salt which is widely applied during the winter. It may also be nonspecific in coastal regions, where it can be introduced to freshwater by intermixed salt water and sea spray. Singapore has an area of 714 km², of which 134 km² has been added by land reclamation (Tortajada et al. 2013). Both catchments

near the coast and reclaimed land may be affected by seawater intrusion; therefore chloride must be interpreted cautiously. The Cl/Br ratio is a good indicator of sewage as bromide concentrations in sewage are not elevated relative to those in the natural environment (Hyer 2007). However, in areas where road salting is performed widely during the winter months, it is difficult to identify whether the elevated Cl/Br ratio is the result of the presence of either sewage or halite in streams.

In natural waters and wastewaters, phosphorus exists mostly as phosphates (Eaton et al. 2005). Investigation of phosphate has special importance in environmental samples (Aydin et al. 2010) because it can contribute to freshwater eutrophication (Berkheiser et al. 1980). Inorganic phosphate is adsorbed to small particles in the soil (Busman et al. 1997). The amount of phosphate adsorbed by soil increases as the amount of phosphate in solution increases and vice versa. Soil particles can act either as a source or a sink of phosphate to the surrounding water depending on conditions. Orthophosphate, PO_4^{3-} , is the simplest form of phosphate. Orthophosphate is often called simply "phosphate" by chemists and also non-technical people. A study of karst waters showed that samples with positive results for *E. coli* and/or enterococci tended to have higher concentrations of orthophosphate (Zheng et al. 2013).

2.5.2. Wastewater organic compounds

A series of wastewater organic compounds have been suggested for use as tracers. They include natural organics (such as faecal sterols and stanols, which are reproductive hormones) and artificial organics (such as caffeine, pharmaceutical compounds, artificial sweeteners, chemicals in laundry detergents, antimicrobials and disinfectants, cotinine, antioxidants, insect repellents, insecticides, flame/fire retardants, fragrances and plasticisers).

According to Hyer (2007), wastewater organic compound analysis is useful because it is a whole-sample extraction, unlike bacterial analysis which is mostly performed on a subset of the bacteria. However, the occurrence of trace organic chemicals in the water is not static (Dickenson et al. 2011). It will depend on usage patterns (for

example, pharmaceutical prescription practices) and environmental factors (although to a lesser degree than indicator bacteria). Wastewater organic compounds should ideally have the following characteristics (Hyer 2007): (i) Their environmental fate and transport characteristics should be similar to the fate and transport of sewage; and, (ii) They should occur in sewage far above the analytical detection limit. In order for the tracers to exist in concentrations far above the detection limit, the tracers should be present/used regularly and constantly. This implies that consumer habits do not change or that the compound is not phased out within a few years. In addition, the quantities discharged with wastewaters should be sufficient to permit analytical quantification after dilution and dissipation in the environment (Buerge et al. 2003).

An elaboration of each of potential organic compound tracers is provided in the following.

a. Faecal sterols and stanols

Cholesterol is a plant and animal steroid (Wilkison et al. 2002). Higher water temperature in tropical zones may facilitate in-situ conversion of cholesterol to stanols. Cholesterol undergoes hydrogenation and is microbially reduced into coprostanol (5β -cholestan- 3β -ol) and related isomers within the intestines of humans and other higher animals (Seurinck et al. 2005). Férézou et al. (1978) showed that of the total neutral sterols in human faecal matter, the human body eliminates 9.5% cholesterol and the rest are bacterial transformation products which are mainly coprostanol (65%), phytosterols (18%), coprostanone (<6%), cholestanol (<2%), and Δ^7 -cholestenol (<2%). So, the major human faecal sterol was coprostanol. There was on average 10 times more coprostanol in human faeces than in the faeces of any other animal (including cat and pig) on a dry-weight basis (Leeming et al. 1996). On the other hand, the metabolism of cholesterol by microorganisms within the environment usually generates cholestanol (Hagedorn and Weisberg 2009).

A study in Canada showed that the relationship of cholesterol and coprostanol was highly significant (Dutka and El-Shaarawi 1975). Cholesterol and coprostanol were

two most frequently detected wastewater organic compounds in streams in the USA (Kolpin et al. 2002). Coprostanol degrades rapidly under aerobic conditions; the reported half-lives are generally less than 10 days at 20°C (Isobe et al. 2002). Thus, the presence of coprostanol in an aerobic environment can be considered to be an indication of recent faecal input to the waters (Isobe et al. 2004). Coprostanol is also hydrophobic and thus it is strongly sorbed onto sediments. It is stable in anaerobic environments. The extent of contamination could be deduced by comparing coprostanol concentrations with those in raw sewage. For example, if coprostanol concentrations are two orders of magnitude lower than that in raw sewage, it would imply that the waters contain raw sewage with a dilution ratio of 100 (Isobe et al. 2002).

Leeming and Nichols (1996) reported coprostanol is better correlated with enterococci than faecal coliform in Derwent estuary, Tasmania, Australia. Enterococci are generally regarded as hardier, more resistant to environmental stress and better indicators of sewage pollution. In contrast, Isobe et al. (2002) reported coprostanol is better correlated with *E. coli* than total coliform and faecal streptococci in Mekong Delta, Vietnam. They also reported coprostanol is better correlated with total coliform than faecal streptococci despite the fact that total coliform bacteria are known to be unreliable faecal indicators because soil microbes or other non-faecal bacteria can be mistakenly counted as total coliform. In addition, Isobe et al. (2002) also reported that coprostanol is correlated with *E. coli* in West Malaysia as strongly as in Mekong Delta, Vietnam. Coprostanol is found to be correlated with *E. coli* in groundwater and in canal waters in both West Malaysia and Mekong Delta, Vietnam. The contrasting behaviour of coprostanol correlation with enterococci/faecal streptococci and *E. coli*/faecal coliform may be caused by the different climate in which the sampling locations are located: temperate for Tasmania, Australia and tropical for West Malaysia and Mekong Delta, Vietnam. However, Shah et al. (2007) reported that high coprostanol levels were not necessarily associated with high faecal coliform counts and high levels of coprostanol can be indicative of human contamination but are not exclusively so.

b. Caffeine

Caffeine is contained in beverages (coffee, tea and caffeinated soft drinks), foods (chocolate products) and drug products (Wu et al. 2008). Coffee and tea are reported to be two of the most widely consumed beverages in the world (Odegaard et al. 2008). A wide variety of both prescription and over-the-counter (OTC) or non-prescription drugs contain caffeine at levels similar to caffeine-containing food and beverage products (Barone and Roberts 1996). It is known that caffeine enhances the effect of certain analgesics in cough, cold, and headache medication (Buerge et al. 2003).

Odegaard et al. (2008) reported that 71% of a cohort (aged 45-74 years old) studied in Singapore consumed more than one cup of coffee per day. This is equal to more than 115.9 mg caffeine per day. According to Wong et al. (2001), the average weight of people in Singapore aged between 40 and 79 years is 59.9 kg. Hence, the caffeine consumption in Singapore is equal to 1.9 mg/kg, which is relatively low compared to adult caffeine consumption in the USA, UK and Denmark which is typically in the range of 4 to 7 (Barone and Roberts 1996).

Caffeine is excreted in the urine of individuals who have ingested the substance. Caffeine levels of between 20 and 300 µg/L have been measured in domestic wastewater (Scott et al. 2002) and between 7 and 73 µg/L in influents of wastewater treatment plants (Buerge et al. 2003). Caffeine levels in receiving waters are much lower. Buerge et al. (2003) reported that caffeine concentrations in effluents of wastewater treatment plants ranged from 0.03 to 9.5 µg/L. Buerge et al. (2003) also reported that the caffeine concentrations in lakes varied from 6 ng/L (in a sparsely populated area) to 164 ng/L (in a more densely populated area), while caffeine concentrations in rivers were found to have concentrations of up to 250 ng/L.

According to Buerge et al. (2003), caffeine is considered a suitable tracer because caffeine concentrations in lakes are linearly correlated with the anthropogenic burden by domestic wastewaters where anthropogenic burden is the ratio of

population per water throughflow. In addition, caffeine consumption at comparable amounts is assured to continue into the future (Buerge et al. 2003). The Public Utilities Board (PUB), Singapore has found that caffeine in the drainage system can be attributed to the direct discharge of beverages from food courts into open drains, compromising the use of caffeine as an indicator in such situations (Hans S. Eikaas, PUB, personal communication, 2010). However, another study in Singapore (Wu et al. 2008) has attributed elevated caffeine levels to sewer leakage.

c. Pharmaceutical compounds

Pharmaceuticals typically have properties favouring their use as wastewater tracers (Benotti and Brownawell 2007) and can be related to wastewater origin with high plausibility (Held et al. 2006). They are relatively water-soluble and non-volatile (Benotti and Brownawell 2007). Pharmaceutical compounds can reach detectable concentrations in surface water if the use is sufficiently large and the compounds show some mobility and persistence in the aquatic environment (Buser et al. 1999). Pharmaceutical compounds include over-the-counter (OTC) drugs (such as acetaminophen or paracetamol and ibuprofen) and prescription drugs (such as sulfamethoxazole and trimethoprim).

Ibuprofen has an estimated annual global production of several kilotons (Buser et al. 1999). It is the third-most popular drug in the world. It is also excreted to a significant degree (70-80% of the therapeutic dose) as the parent compound or in the form of metabolites. Its mobility in the aquatic environment is also high. It undergoes significant degradation in the human body and wastewater treatment plants. It is also degraded under non-sterile conditions in daylight as well as in the dark. In spite of this, ibuprofen remains detectable in surface waters: up to 8 ng/L in lakes and rivers in Switzerland and in the North Sea (Buser et al. 1999). A search of the literature failed to reveal data on the usage of ibuprofen in Singapore. Ibuprofen has been listed as one of the four most often-used anti-inflammatory pharmaceuticals in China (Peng et al. 2008). The other three are acetaminophen, aspirin and diclofenac. The annual production of ibuprofen in China is greater than 1,000 tons. In contrast, the

annual production of naproxen, another anti-inflammatory drug, is only about 300 tons per year. The production and consumption patterns of these pharmaceuticals correspond well with the high occurrence of ibuprofen and less frequently detected naproxen in water samples (Peng et al. 2008). Table 2.11 compares the persistence of various pharmaceutical compounds. It can be observed that ibuprofen is one of the few pharmaceuticals that is not eliminated by phototransformation. Data for the Marina Reservoir catchment in Singapore (Wu et al. 2008) show concentrations of ibuprofen from 37 ng/L in a downstream location (reservoir) up to 195 ng/L in an upstream location (storm drain).

Diclofenac is used worldwide and has a production volume estimated to be in the hundreds of tons annually (Buser et al. 1998). It is the 12th-most used pharmaceutical in Malaysia (Al-Odaini et al. 2010). It can be used as an analgesic. Diclofenac has negligible adsorption onto sediment particles and minimal chemical and biological degradation (Buser et al. 1998). However, it photodegrades rapidly. Buser et al. (1998) reported that diclofenac concentrations in a major tributary to a lake were significantly higher (up to 370 ng/L) than in the outflow of the lake (up to 12 ng/L). The latter concentration is comparable to diclofenac concentrations reported by Al-Odaini et al. (2010) in Malaysia (17 ng/L in river water) and Wu et al. (2008) in Singapore (4 – 26 ng/L in river and reservoir water).

Table 2.11. Pharmaceutical compounds and their persistence, extracted from Tixier et al.(2003)

Pharmaceutical compounds	Remarks
Ibuprofen	Eliminated by sedimentation, biodegradation
Diclofenac	Eliminated by phototransformation
Naproxen	Eliminated by biodegradation, phototransformation
Ketoprofen	Eliminated by biodegradation, phototransformation
Carbamazepine	Persistent
Clofibrac acid	Persistent

Although less used than over-the-counter medications, some prescription drugs may also constitute useful tracers. Lim and Yap (1999) reported that the ten most

prescribed drugs by polyclinic (a clinic that provides subsidised outpatient medical care, health screening and pharmacy services) doctors in Singapore for all ages, in descending order, are antihistamines, acetaminophen, throat medications, nifedipine, beta-blockers, antacids, mist Benadryl expectorant (MBE), Procodin™ cough syrup, amoxicillin and vitamin B. In the same study, amoxicillin is listed as the most prescribed antibiotic in Singapore (accounting for 70% of all antibiotic prescriptions) and acetaminophen as one of the most prescribed drugs for younger (below 40 years old) age groups. Similarly, acetaminophen is one of the most commonly used drugs in the world, with nearly 28 billion doses sold in the USA in 2008 (Hutson 2010). It is also the most sold over-the-counter pharmaceutical in pharmacies in Malaysia (Al-Odaini et al. 2010). Critchley et al. (2005) showed that roughly 5% of consumed acetaminophen is excreted in urine. According to Wilkison et al. (2002), acetaminophen was one of the most frequently detected over-the-counter pain medications in stream samples. It was detected at a concentration of 10 ng/L in the Langat River, Malaysia (Al-Odaini et al. 2010).

Nifedipine is one of the most-prescribed drugs by polyclinic doctors in Singapore for the 40 years and above age group (Lim and Yap 1999) and the most common cardiac drug in use based on a study in Singapore by Yap and Chan (1998). This drug appears to be a popular choice among doctors probably because of its potency, non-interaction with serum lipids, and relatively few side effects (Yap and Chan 1998). Henry (1980) reported that 80% of the consumed nifedipine is excreted through the renal system (70% on the first day) and less than 15% of nifedipine is excreted in faeces.

d. Artificial sweeteners

Artificial sweeteners are used in beverages, food and other consumer products such as toothpaste. Most of the artificial sweeteners consumed are excreted and no industrial or agricultural use of artificial sweeteners is known. Artificial sweeteners are source-specific and are easily quantified (Buerge et al. 2009). Artificial sweeteners include aspartame, acesulfame-k, sucralose, saccharin and cyclamate. Aspartame is

the artificial sweetener that is permitted as a general purpose food additive in Singapore (AVA 2006). Other artificial sweeteners that are allowed with permission in Singapore are acesulfame-k, sucralose and saccharin(AVA 2006). Cyclamate was banned in United States in 1970, but is registered in Europe and many other countries (Buerge et al. 2009).

Aspartame can be broken down under elevated temperature or at high pH. Buerge et al. (2009) assumed that aspartame was quickly biodegraded in wastewater treatment plants. Both acesulfame-k and sucralose are persistent: they are not removed by the treatment process in wastewater treatment plants. Buerge et al. (2009) detected sucralose in domestic wastewaters and natural waters, while saccharin was found at lower levels in treated wastewater. In the same study, Buerge et al. reported that acesulfame-k was consistently detected in untreated and treated wastewater and in most surface waters and that the concentrations of acesulfame-k in rivers were high. Acesulfame-k was also detected in groundwater and its concentrations increased with population in the catchment area and decreased with water throughflow. Acesulfame-k, saccharin and cyclamate are anionic compounds at pH values typical of natural waters. Therefore, they are expected to be quite mobile in the subsurface. In addition, hydrolysis, photolysis and biodegradation of acesulfame-k and saccharin play only a minor role in its environmental fate, properties that are consistent with the requirements for a good tracer.

e. Chemicals in laundry detergents

Chemicals in laundry detergents include long-chain alkyl benzenes, several metabolites of alkylphenolpolyethoxylate surfactants, optical brighteners/fluorescent whitening agents, and sodium tripolyphosphate. The use of laundry detergent chemicals can only be attributed to general human or industrial sources, but not faecal pollution (Scott et al. 2002). Surfactants are the substances that provide the cleaning action of detergents. There are ionic and non-ionic surfactants. However, in the USA, two-thirds of the total surfactants used is ionic surfactants, in which anionic

surfactants amount to more than nine tenths of the ionic surfactants used (Eaton et al. 2005).

Linear alkylbenzene sulfonate (LAS) is the most widely used anionic surfactant in commercial detergents and has been used in the USA since 1965 (Rapaport and Eckhoff 1990). Long-chain alkyl benzenes are used in the production of LAS. LAS is a highly biodegradable surfactant, unlike branched alkyl benzene sulfonate (the predecessor). It is mostly removed during sewage treatment by biodegradation and sorption (Rapaport and Eckhoff 1990). LAS is used to standardise the methylene blue active substance (MBAS) method (Eaton et al. 2005). LAS analysis using this nonspecific anionic surfactant method (MBAS method) is highly correlated with the analysis methods that are specific to LAS (Rapaport and Eckhoff 1990).

Alkylphenolpolyethoxylates (APEOs) are non-anionic surfactants used in laundry detergents and textile manufacturing (Wilkison et al. 2002). They are also components in many paints, herbicides, pesticides and cosmetics. Nonylphenolpolyethoxylates (NPEOs) constitute a large portion of non-anionic surfactants. Studies have shown that the biodegradation products of NPEOs are very stable and toxic to marine and freshwater species (Ying 2004). Biodegradation products of NPEOs are nonylphenol (NP) and nonylphenoethoxylates (NPnEOs or NPEOns, where n indicates the number of ethoxy units). NP, nonylphenolmonoethoxylate (NPEO1), and nonylphenoldiethoxylate (NPEO2) have been classified as endocrine disrupting compounds (EDCs) because they have been reported to cause estrogenic responses in aquatic organisms (Ying 2004).

Measurements for NPEOs were conducted by Mendez-Sagel (2010) in Kranji catchment, Singapore in January 2010. Mendez-Sagel used polyethylene devices (PEDs) to accumulate hydrophobic organic contaminants (HOCs) including NPEOs, greatly reducing the need to sample large volumes of water and perform complicated extraction steps. NP and NPEO1 were not detected in residential areas, but were surprisingly found in non-residential areas which included horticulture, animal farming and military land uses. This is counter-intuitive since one would expect NP

and NPEO1 to be detected more readily in residential areas where detergents are in use, however there are numerous small on-site wastewater treatment plants in the non-residential areas and these are likely a source of domestic wastewater, including laundry wastewater.

Fluorescent whitening agents (FWAs) are highly fluorescent, moderately water-soluble organic compounds with a high affinity for cellulosic material. They are minor components in detergents and on the average contribute only 0.15% of the total mass of laundry detergents (Poiger et al. 1996). The purpose of the FWAs is to adsorb to laundry and to improve or restore its whiteness during washing. However, FWAs are only partially bound to fabrics; a considerable fraction (5-80%) remains in the washing water and is discharged to the sewers (Poiger et al. 1996). FWAs are not easily biodegradable, but readily degraded photochemically (Poiger et al. 1996). Poiger et al. (1996) reported that the concentrations of the two most frequently used FWAs (FWA 1 and FWA 2) ranged in river water from 0.04 to 0.57 µg/L. Dixon et al. (2009) attempted to test for FWAs in the field using cotton pads which were placed in drains in Kranji catchment, Singapore. The FWAs should be adsorbed to the cotton pads and fluoresce when held under a black light. However, the attempt failed because the cotton pads attracted a substantial amount of particulate matter which circumvented the test.

f. Antimicrobials and disinfectants

Disinfectants are antimicrobials commonly found in deodorants, bar soaps, body washes, liquid hand soaps, toothpastes, dishwasher powders, fabrics and plastics (Wilkison et al. 2002). Triclosan is one of the most common disinfectants in use. The presence of triclosan in environmental samples is a concern because it may cause the development of antiseptic-resistant bacteria strains. According to Singer et al. (2002), concentrations of triclosan in treated wastewater effluent are typically in the 42-213 ng/L range and concentrations in receiving rivers range from 11 to 98 ng/L. Similar results were obtained by Lindström et al. (2002) who reported triclosan concentrations in receiving waters of up to 74 ng/L. Triclosan in surface water may be

eliminated by photodegradation (Singer et al. 2002). Triclosan is quite stable against hydrolysis and fairly persistent in sediment. A potential biotransformation product of triclosan is methyl triclosan. However, environmental concentrations of methyl triclosan can be considered to be negligible. Although triclosan can react with free chlorine to produce chloroform, a known carcinogen, the chemical reactions require an excessive chlorine concentration far beyond chlorine concentrations found in household tap water (Fiss et al. 2007). Therefore, products containing triclosan have been approved for sale by many authorities around the world including the authorities within the ASEAN countries (Khaiat 2005) which includes Singapore .

g. Plasticisers

Plasticisers include DBP (di-n-butyl phthalate), DEP (diethyl phthalate), DEHP (diethylhexyl phthalate or bis(2-ethylhexyl) phthalate) (Hyer 2007) and BPA (bisphenol A) (Kolpin et al. 2002). DEP and DEHP are plasticisers for polymers and resins (Hyer 2007). Approximately 95% of DEHP is used as a plasticiser in polyvinylchloride (PVC) (Wams 1987) because of its low cost. DEHP is emitted to the environment during the production of plastics and plastic products, during their use and after disposal. DEHP is also readily adsorbed by organic soil particles.

h. Other potential organic tracers

Other potential anthropogenic organic tracers for wastewater, including cotinine, BHA (butylated hydroxyanisole/2,6-di-tert-butylphenol), BHT (butylated hydroxytoluene/2,6-di-tert-butylbenzoquinone), DEET (N,N-diethyl-meta-toluamide) and flame retardants, have been measured by Wilkison et al. (2002) in a stream in Kansas City, Missouri, USA influenced by wastewater and combined sewer overflows. Cotinine is the result of the body's rapid assimilation and metabolism of nicotine (the active ingredient in tobacco and smoking-cessation products) (Wilkison et al. 2002). BHA and BHT are antioxidants which are added as preservatives to foods and to prevent brittleness in food packages and plastics(Wilkison et al. 2002). DEET is the active ingredient in most commercial insect repellents and is removed during laundry

and bathing (Wilkison et al. 2002). Flame retardants are materials that inhibit the spread of fire. They are used in fabrics, carpets and plastics (Wilkison et al. 2002).

Perfluorochemicals (PFCs) have also been investigated as potential tracers (Nguyen et al. 2011). Perfluorochemicals are a group of manmade chemicals that have been used extensively as surface coating and protectant formulations for paper and cardboard packaging products, carpets, leather products and textiles that repel water, grease and soil (USEPA 2011). However, PFCs have been detected in rainwater collected in Singapore (although at low concentrations) (Nguyen et al. 2011). Dreyer et al. (2010) reported that air-borne (particle phase) PFCs are deposited from the atmosphere by precipitation: the relationship between PFC wet deposition and PFC air concentration may be established via precipitation amounts. Therefore, the existence of PFC in surface water may originate from a natural process (precipitation) and not particularly indicate the occurrence of human faecal and sewage contamination.

2.5.3. Chemical tracer studies

Studies have been done using several wastewater organic compounds as tracers of contamination. For example, in the USA, Kolpin et al. (2002) reported that detergent metabolites, plasticisers and steroids contributed most of the total measured concentration of organic wastewater contaminants. In China, Peng et al. (2008) found that the total concentrations of acid pharmaceuticals and phenolic compounds were moderately correlated with coprostanol. Peng et al. also found that the maximum concentrations of most pharmaceutical and personal care products and coprostanol were found in urban streams. However, many of the studies were conducted in temperate areas. This study is designed to investigate potential tracers in a tropical urban area. Therefore, some studies that were conducted in tropical areas are summarised in Table 2.12 and discussed below.

Different studies implement different approaches (such as potential tracers selected to be analysed and the lab methodology employed). The approaches adopted are influenced by the objectives that the studies aim to achieve such as development of

lab analysis methodology (Al-Odaini et al. 2010), assessing the occurrence of the analysed tracers in the environment (Mendez-Sagel 2010; Nguyen et al. 2011; Spongberg et al. 2011; Xu et al. 2011) or finding the correlation with conventional tracers (Isobe et al. 2002; Wu et al. 2008). This study aims to identify tracers that are practical in terms of sample collection and lab analysis. In terms of sample collection, a surface water sample itself is collected instead of accumulating targeted compounds by other mean such as cotton pads (Dixon et al. 2009). Selected parameters must be those that exist at high-enough concentrations for measurement in water samples. In terms of lab analysis, selected parameters must be quantifiable in the sample by reasonably facile methods and with considerable specificity.

Isobe et al. (2002) have explored the correlation of faecal sterols and stanols with faecal indicator bacteria (FIB). Wu et al. (2008) showed good correlation between caffeine and faecal coliform but with only three samples. This study is designed to explore the correlation of FIB with more types of wastewater organic compounds (such as pharmaceuticals and artificial sweeteners) with a good number of samples to give a more representative correlation.

Nguyen et al. (2011) and Xu et al. (2011), who studied PFCs and different kinds of emerging organic contaminants, respectively, discuss the possibility that those compounds originate from leaking sanitary sewers, however, they did not show any further evidence other than the occurrence of those compounds in tropical urban catchment. This study aims to test this hypothesis by seeking more evidence through collecting 24-hr samples, analysing a suite of practical parameters and analysing the results obtained.

Table 2.12. Comparison of several chemical tracer studies in tropical areas

Reference	Sampling location	Type of sampling site	Potential tracers investigated	Outcome related to this study
Spongberg et al. (2011)	Costa Rica	urban and rural; river	34 pharmaceutical and personal care products or PPCPs (including caffeine, ibuprofen, diclofenac, acetaminophen and triclosan)	Eighty six samples were analysed. The frequencies of triclosan and caffeine detection were 34% and 29%, respectively (they are two out of five most frequently detected compounds in the study).
Isobe et al. (2002)	West Malaysia and Mekong Delta, Vietnam	urban and rural; river	faecal sterols and stanols, FIB (total coliform, <i>E. coli</i> , faecal streptococci)	Coprostanol is better correlated with <i>E. coli</i> than faecal streptococci.
Al-Odaini et al. (2010)	West Malaysia	urban and rural, river	18 pharmaceuticals (including diclofenac, acetaminophen and nifedipine), 4 synthetic hormones	The study developed a lab analysis methodology that suits Malaysian surface water samples. The study also reported detection of the parameters in river water.
Dixon et al. (2009)	Kranji Catchment, Singapore	urban and rural; drain	fluorescent whitening agent (FWA), FIB (total coliform, <i>E. coli</i>)	The attempt of FWA analysis failed because the cotton pads used to adsorb FWA attracted a substantial amount of particulate matter which circumvented the test.
Mendez-Sagel (2010)	Kranji Catchment, Singapore	urban and rural; drain	nonylphenol ethoxylates (NPEO) and nonylphenol (NP)	NP and NPEO were not detected in residential areas, but were found in non-residential areas (horticulture, animal farming and military land uses) probably due to numerous small on-site wastewater treatment plants which are likely a source of domestic wastewater including laundry wastewater.
Wu et al. (2008)	Marina catchment, Singapore	urban drainage; reservoir	6 pharmaceuticals (including caffeine, ibuprofen and diclofenac), faecal coliform, chemical oxygen demand (COD)	Caffeine was well correlated with faecal coliform in Rochor Canal. This correlation was derived from three points (upstream/drain, mid-stream/river and downstream (river close to reservoir) samples).
Nguyen et al. (2011)	Marina catchment, Singapore	urban drainage;reservoir	19 perfluorochemicals (PFCs)	Thirteen PFCs were detected. PCA plots show samples from more upstream locations are not clustered together with samples from downstream locations. Nguyen et al. also mention but do not pursue a concern that contaminants from leaky sewers seep into surface waters.
Xu et al. (2011)	Marina catchment, Singapore	urban drainage; reservoir	13 emerging organic contaminants (EOCs) which include alkylphenol ethoxylate metabolites, hormones, pharmaceuticals (such as caffeine and ibuprofen), plasticiser (bisphenol A) and a pesticide	Xu et al. reported that it appears likely that EOCs enter surface water by the same route: non-point sources diffusing from the subsurface into canals and rivers from leaking sewer lines or uncontrolled discharge points.

2.6. Temporal variation of indicator concentrations

2.6.1. Review of indicator temporal variation studies

Faecal indicator bacteria (FIB) have been sampled and monitored at different sampling frequencies, from low temporal resolution such as seasonal to high resolution such as hourly, as summarised in Table 2.14. Kacar (2011) sampled five non-tidal rivers in an area with high population density and intensive agricultural and industrial activities in Turkey seasonally (four seasons) and found significant variations in faecal bacteria counts and seasonal effects. Cabral and Marques (2006) carried out weekly sampling at the Febros River, Portugal. They found considerable variation of total coliform and faecal coliform from week to week. Rodgers et al. (2003) also studied temporal variability of faecal coliform. The most intensive sampling done was twice weekly at eight sampling sites in Aberdeenshire, UK. They found bacterial concentrations at all sampling sites were extremely variable as indicated by the standard deviations. Rodgers et al. (2003) attributed this variability to peaks associated with rainfall events and variations during normal baseflow conditions. Koirala et al. (2008) performed time-series analysis on daily measurements of total coliform (TC) obtained from Little River, East Tennessee, USA. They observed that TC was likely to be influenced by precipitation over the short term (6 weeks and less) and by discharge (baseflow) over the long term (up to 55 weeks).

Some studies have observed a diurnal cycle of FIB (TC, FC, *E. coli* and enterococci) at beaches (Boehm et al. 2002; Enns et al. 2012) and streams (Traister and Anisfeld 2006; Desai and Rifai 2013) caused by sunlight-induced die-off, with higher concentrations in the morning and lower in the afternoon. Traister and Anisfeld (2006) assessed diurnal variation of *E. coli* levels by sampling once every two hours over the course of a 24-hr period under baseflow conditions at two rural sites in the Upper Hoosic River, Massachusetts, USA. They sampled five days at each site. Both sites have more than 70% forest area, but one site has more proportion of urban and residential areas (9.4%) than the other (2.8%). They found that the more urbanised stream exhibited greater diurnal variability but lower baseflow-to-stormflow

variability compared to the less urbanised stream. However, it is important to note that the two watersheds which they sampled for 24 hr have a very low degree of urbanisation compared to 100% urbanised areas in Singapore. They also found that bacterial levels were generally higher in the night and early morning than in the afternoon. Statistical analysis showed that the difference between daytime and nighttime samples was significant, which was attributed to sunlight-induced die-off during the daytime. Boehm et al. (2002) also found that FIB (TC, FC and enterococci) concentrations in Huntington Beach, California, USA were highest in the middle of the night and fell to levels near or below the detection limit at midday when solar radiation peaks. Mesocosm experiments by Boehm et al. (2002) showed that FIB concentrations declined when exposed to sunlight but did not rebound after sunset. This result does not support the idea that bacterial reactivation and/or growth is the cause of high FIB concentrations at night. Enns et al. (2012) also observed that enterococci at Hobe Cat Beach in Miami, Florida, USA peaks when solar radiation levels are low. Beach visitors (humans and dogs) are sources of enterococci. They usually visit the beach during daytime (when there is sunlight); therefore high enterococci concentrations were expected to occur at daytime. However, Enns et al. found that solar inactivation at daytime likely affects the enterococci concentrations more than the contributions from beach visitors.

Desai and Rifai (2013) reported an *E. coli* diurnal sag due to daytime sunlight die-off and nighttime regeneration and which is not due to contamination. In contrast, Davies-Colley et al. (2004) observed a sharp spike of *E. coli* concentration due to contamination from a dairy cow herd crossing a stream in New Zealand. *E. coli* concentrations temporarily elevated to more than 100x background levels, peaking at 50,000 MPN/100mL. McDaniel et al. (2013) did laboratory controlled experiments with manure and observed similar behaviour. They reported that samples collected 0.5 min after manure deposition had *E. coli* concentrations that were one to two orders of magnitude greater than the samples collected at later times after deposition. They also observed that *E. coli* concentrations downstream of the cowpie decreased quickly, within the first 5 min.

Table 2.14. Summary of studies related to indicators temporal variation (in sequence of sampling resolution: from low to high)

Reference	Sampling location	Type of sampling site	Highest sampling resolution applied	Indicators analysed	Outcome related to this study
Kacar (2011)	Turkey	Urban; river; temperate	Seasonally for 3 years	FC, FS	FIB counts and seasonal changes showed a significant variation.
Cabral and Marques (2006)	Portugal	Rural; river; Mediterranean	Weekly for 35 successive weeks	TC, FC, FS, Ent	FIB measurements showed a considerable variation from week to week.
Rodgers et al. (2003)	Aberdeenshire, UK	Rural; stream; temperate	Twice weekly for about 2.5 months	FC	FIB concentrations at all sites were variable in terms of peaks associated with rainfall events and variations during normal baseflow conditions.
Koirala et al. (2008)	Tennessee, USA	Urban and rural; river; temperate	Daily for 5 years and 3 months	TC	This study only involved data analysis and not collection. TC was likely to be influenced by precipitation over the short term and by discharge (baseflow) over the long term.
Whitman and Nevers (2004)	Chicago, USA	Urban; shoreline; temperate	Twice a day	<i>E. coli</i>	Due to solar disinfection, FIB concentrations in the afternoon were lower than in the morning.
Sercu et al. (2009)	California, USA	Urban; drain; temperate	Three times in one day at two sampling points (6 samples collected)	<i>E. coli</i> , Ent, HF183	Temporal variation of HF183 marker indicated sewage exfiltrated from sewer pipes into storm drains.
Traister and Anisfeld (2006)	Massachusetts, USA	Forests with small proportion of urban areas; stream; temperate	Two-hourly for 24 hr (10 d in total) at two sampling sites	<i>E. coli</i>	The more urbanised stream exhibited greater diurnal variability but lower baseflow-to-stormflow variability compared to the less urbanised stream. Generally bacterial levels were higher in the night and early morning than in the afternoon due to sunlight-induced die-off during the daytime.
Enns et al. (2012)	Florida, USA	Urban; shoreline; tropical (Miami)	Hourly for 10 d	Ent	Enterococci peaks when solar radiation levels are low.
Boehm et al. (2002)	California, USA	Urban; shoreline; temperate	Hourly for 2 weeks	TC, FC, Ent	FIB concentrations were highest in the middle of the night and fell to levels near or below the detection limit at midday when solar radiation peaks.
Desai and Rifai (2013)	Texas, USA	Urban; stream (2 out of 3 are concrete-lined); subtropical (Houston)	10-min interval for 12 hr (7 sampling events), 30-min interval for 24 hr (6 sampling events)	<i>E. coli</i>	Observed a bacteria diurnal sag or BDS (marked with daytime exponential decay, which is strongly influenced by water temperatures and solar radiation intensities, followed by an exponential nighttime regeneration) and reported that 4 to 6 samples taken between 06:00 and 18:00 may be sufficient to define the BDS depending on stream conditions. They reported that background indicator bacteria in urban catchments are relatively elevated even without influence from diurnal variations and stormflow inputs.

Desai and Rifai (2013) also reported that background indicator bacteria in urban catchments are relatively elevated even without influence from diurnal variations and stormflow inputs. Sercu et al. (2009) analysed HF183 *Bacteroides* marker in addition to FIB from samples collected in urban catchments in Santa Barbara, California, USA. They found that temporal variation of HF183 *Bacteroides* marker indicated a repeated external source instead of a continuous input from in situ growth. The repeated external source was not direct contamination such as cross-connections because there were no visible sewage solids during sampling, but it might be sewage that exfiltrated from buried sewer pipes and flowed through unsaturated soil into storm drains. However, these inferences were made from a very limited data set of only three storm-drain water samples that were compared to three creek water samples.

Generally, the studies summarised here have shown that temporal variation exists regardless of the sampling resolution used. Some have attributed temporal variation to the difference between stormflow and baseflow surface water quality (Rodgers et al. 2003; Koirala et al. 2008) and diurnal variation in die-off due to sunlight (Boehm et al. 2002; Traister and Anisfeld 2006; Enns et al. 2012).

To the best of the author's knowledge, there is no study on the temporal variation of indicators in tropical urban areas other than Enns et al. (2012) which was conducted at a tropical urban shoreline and Desai and Rifai (2013) at subtropical urban streams. Those studies and other temperate studies (Boehm et al. 2002; Traister and Anisfeld 2006) showed the same characteristic: lower bacterial levels in the daytime than the nighttime due to sunlight induced die-off. Although literature sources reported that daytime FIB concentrations are lower than nighttime, Desai and Rifai (2010) point out the complexity of this phenomenon and its dependence on flow, turbidity, total suspended solids, temperature and the location/land use of the monitoring point (upstream or downstream and rural/urban). Therefore, there is a question whether the same characteristic prevails in surface water from tropical urban catchments

where there is only slight temperature variation between daytime and nighttime and where there is no effect of sunlight in mostly-covered or shaded drains.

2.6.2. Implication of temporal variation on sampling time for surface water quality monitoring

Assessing whether a water body is meeting water quality criteria is made difficult by the high variability of indicator bacteria concentrations over time. Many water quality programs focus primarily on low-frequency dry-weather sampling (Traister and Anisfeld 2006). The frequency of monitoring or sampling affects the number of FIB samples that exceeds the standard. Leecaster and Weisberg (2001) showed that sampling frequency affects the number of high faecal indicator bacteria events that are detected. According to their study, sampling five days per week resulted in missing 20% of the total or faecal coliform exceedances observed when sampling seven days per week. Sampling three days per week resulted in missing approximately 45% of the threshold exceedances. Sampling once per week resulted in missing approximately 75% of the exceedances. Sampling once per month missed approximately 95% of the exceedances. Therefore, Traister and Anisfeld (2006) suggested that it is important to support sampling programs that capture a fuller range of stream conditions.

Besides sampling frequency, sample collection time can dramatically influence the concentration of indicator bacteria detected (Boehm et al. 2002) and hence different sample collection times could result in different management decisions (Enns et al. 2012). Whitman and Nevers (2004) indicated that samples should be collected either at a standard time of day (same time every day) that corresponds to maximum public exposure or in the early morning to obtain the most conservative estimate of indicator bacteria counts. Boehm et al. (2002) and Kwasi et al. (1999) also recommended that ideally samples should be collected in the early morning before sunlight has had a chance to reduce FIB concentrations. However, it is important to note that those studies sampled surface waters at beaches that are exposed to sunlight. Therefore, when the surface water is not exposed to sunlight (such as

surface waters in covered drain or shaded area), this early morning sampling suggestion may not be applicable.

In addition to sample collection time, there is concern regarding the subsequent time needed for sample analysis, which is commonly about 24 hours. However, Boehm et al. (2002) reasoned that developing more rapid methods of detecting FIB cannot resolve this problem because FIB concentrations are highly variable. For example, two samples collected in 10-min interval may result in different FIB concentrations: one concentration below the regulatory guideline and one concentration above the regulatory guideline. So, even if FIB could be analysed real-time, high-frequency FIB measurements could not be used in making and informing decisions such as beach closure—some sort of time average would still be needed to smooth out the highly variable FIB concentrations.

2.7. Influence of flow conditions

Flow conditions have been shown to be a significant factor in the influence of faecal sources in a watershed (Hartel et al. 2004). As shown by Ahmed et al. (2010a), during stormflow periods, different transport pathways can potentially change the relative contributions from different sources of faecal contamination, allowing for example more remote influx of contaminants and more widespread loading of contaminants. Ahmed et al. reported that the concentrations of faecal indicator bacteria in water samples collected after wet weather conditions were generally higher compared to dry conditions. They also reported that, overall, viral markers and zoonotic pathogens were more frequently detected in wet than dry weather conditions. Observations by Burnes (2003), Hyer and Moyer (2003), Hartel et al. (2004), Fong et al. (2005), Meays (2005) and Schilling et al. (2009) however were inconsistent with the findings of Ahmed et al. (2010a). According to Burnes (2003), who implemented ARA with faecal coliform methodology, runoff from rainfall appeared to dilute baseline readings and at the same time introduced more sources. Hartel et al. (2004) applied ribotyping with *E. coli* and found that in general, a higher percentage of presumptive *E. coli* colonies were confirmed to be *E. coli* from water samples collected during baseflow

conditions compared with stormflow conditions. Hyer and Moyer (2003), who also applied ribotyping with *E. coli*, showed that the relative contributions from each major source were about the same during both baseflow and stormflow periods. In addition, Fong et al. (2005), who used human and bovine enteric viruses as markers, reported that there was no immediate response in viral detection to rainfall events. Meays(2005) reported that the amount of precipitation in a year may only have a minor effect on faecal coliform counts. Schilling et al. (2009) found that although *E. coli* concentrations were clearly higher during periods of higher discharge, the linear relation of daily *E. coli* to daily discharge was not particularly strong. Hence, it appears that runoff still transports bacteria to drains, independent of the discharge.

Some factors are believed to contribute to the above observations. The first factor is disturbance of sediment (Hyer and Moyer 2003; Meays 2005). A “sloughing off” of bacteria from the bed-sediment surface may contribute to the concentration of microorganisms during low flow (dry weather). The second factor is direct discharge from bacteria point sources to the stream or drain (Meays 2005; Schilling et al. 2009). The third factor is subsurface flows to the stream (Meays 2005; Schilling et al. 2009). A rainfall event may not produce much surface flow, but runoff in the form of subsurface flows generated from the event would still transport microorganisms to the stream or drain network. Hence, high microorganism counts from a rainstorm could persist even after the discharge rate drops. This factor is most often cited in the context of on-site disposal and subsurface treatment of domestic wastewater which are more common in rural areas. In urban environments, however, it is applicable to subsurface contamination caused by leaky sewers, especially sewers at shallow elevation such as connections to the buildings.

2.8. Implications of the literature review on this study

There are many sources of surface water contamination. The sources of contamination that pose the most prevalent hazard for human health are human faecal sources of contamination (Cabelli 1977). Thus, methods are needed to evaluate the potential presence of human faecal contamination and particularly associated

pathogenic organisms in surface water. However, the cultivation and identification of pathogenic microorganisms are difficult and costly. Hence, indicators are utilised. Different kinds of indicators have been introduced but the most widely used are bacterial indicators, non-pathogenic organisms that are believed to be associated with pathogenic organisms but which are more easily measured. However, faecal indicator bacteria (FIB) are not specific to the presence of human faecal contamination. Using FIB as an indicator could lead to costly and inconvenient false-positive-based beach closures (Sercu et al. 2009). On the other hand, outbreaks of gastroenteritis have been recorded despite coliform counts at a seemingly safe level (Scott et al. 2002). In addition, the bacterial indicators which are currently used in tropical countries were developed for temperate countries. A number of studies have raised questions about the suitability of these indicators in tropical countries because the survivability of FIB such as *E. coli* and enterococci is influenced by many factors, the most significant of which is temperature (Noble et al. 2004) and which is markedly different between tropical and temperate climates. These concerns have motivated this study of Singapore which is a tropical country and is mostly urbanised, and in which identification of human-related sources is important.

In this study, a suite of tracers, including FIB, chemical parameters and HF183, is analysed. FIB may grow, occur naturally or survive longer in tropical waters (Hazen 1988; Rivera et al. 1988; Muñiz et al. 1989), nevertheless they are analysed in this study because they are widely used as indicators in regulatory guidelines for tropical areas (Kenzaka et al. 2001) including Singapore. There are many different types of possible chemical tracers with different laboratory methodologies which result in different detection limits. This study aims to find potential tracers that are practical to be used on a routine basis. Therefore, selected possible chemical tracers (Section 4.1) are those that are quantifiable by reasonably facile methods. Many studies have shown the applicability of the HF183 marker in temperate and also tropical areas (Gawler et al. 2007; Santoro and Boehm 2007; Jenkins et al. 2009; Nshimiyimana 2010; Ahmed et al. 2010b; Boehm et al. 2011; Santiago-Rodriguez et al. 2012; Toledo-Hernandez et al. 2013). Particularly, Nshimiyimana (2010) compared HF183 genes

obtained in Singapore, atropical climate, with the genes obtained by Santoro and Boehm (2007)in temperate climates and concluded that they were clustered within the same evolutionary group. This confirmed the use of the HF183 assay as a viable technique for human faecal source tracking in Singapore.

Studies reported that FIB had temporal variation over seasonal, weekly and biweekly time scales (Rodgers et al. 2003; Cabral and Marques 2006;Kacar 2011). Other studies reported that FIB concentrations vary diurnally, from high at nighttime to low in daytime (Boehm et al. 2002; Whitman and Nevers 2004; Traister and Anisfeld 2006; Enns et al. 2012; Desai and Rifai 2013). Those studies were conducted in temperate areas except Enns et al. (2012) in tropical Miami and Desai and Rifai (2013) in subtropical Houston. However, Enns et al. and Desai and Rifai collected most of their samples from open areas which are exposed to sunlight. Desai and Rifai (2010) reported that there is complexity in FIB diurnal variation since the variation depends on such factors as temperature and location/land use of the monitoring point (upstream/downstream and urban/rural). Therefore, there is a question of whether FIB concentration varies similarly in a tropical urban area that is not exposed to sunlight such as covered drains or shaded areas.

Other studies reported the association between certain chemical tracers (such as coprostanol and caffeine) and FIB (Isobe et al. 2002; Wu et al. 2008). This study aims to investigate the correlation of a suite of possible tracers (which includes faecal sterols and stanols, pharmaceuticals and artificial sweeteners) and FIB by collecting a representative sample (in terms of number of samples and sampling locations). Since the chemical tracers can be associated with wastewater, the agreement among parameters provides evidence of human faecal and sewage contamination.

Santoro and Boehm (2007) reported that a higher incidence of HF183 marker was correlated with higher FIB abundance. As well, Sercu et al. (2009) reported that high HF183 marker concentrations occurred exclusively when there was also a high FIB concentration. However, those studies were conducted in temperate areas.

Therefore, there is a question of whether the same incidence (occurrence of high HF183 with high FIB concentrations) also occurs in tropical urban areas such as Singapore.

Some studies suggest but do not substantiate that surface water contamination may originate from leaking sewers (Sercu et al. 2009; Nguyen et al. 2011; Xu et al. 2011). Sercu et al. (2009) mentioned that sewage probably exfiltrated from buried sewer pipes and flowed through unsaturated soil into storm drains in urban areas. This study aims to pursue more definitive evidence of leaking sewers which may be obtained from (i) the variation of FIB concentrations, (ii) agreement between chemical tracers and (iii) HF183. FIB variation has been attributed to exponential growth during nighttime and sunlight-induced die-off during daytime when the sampling site is exposed to sunlight. Therefore, there is a question of whether the variation of FIB concentrations can be related to sewage leaking to urban storm drains which are not exposed or are only partially exposed to sunlight. In addition, there is a question of whether the chemical tracers (particularly, their association with FIB) may provide another line of evidence of leaking sewers. Toledo-Hernandez et al. (2013) showed that HF183 detection frequency was low at tropical sites with low-level human development and density and high at tropical sites impacted by sewage treatment plants. Boehm et al. (2011) found HF183 marker in a tropical drain and attributed this to the exfiltration of septage or cesspool discharge to the drain. Nshimiyimana (2010) reported the presence of HF183 marker in the storm drains in Singapore however the area where he sampled is not served by septic tanks or cesspools but sewer lines. Therefore, there is a question of whether the presence of HF183 marker in the storm drains originates from leaking sewer lines.

CHAPTER 3 PRELIMINARY DATA ANALYSIS

Two datasets provided by the Public Utilities Board (PUB) of Singapore were analysed for this study. The datasets were analysed as part of an initial effort to characterise the spatial and temporal distributions of human faecal and sewage contamination. The two datasets consist of (i) Marina catchment one-day grab sample data and (ii) Toolbox Study data (Hans Eikaas, PUB, personal communication, 2010). Section 3.1 provides details of the methodology used to analyse the datasets. Results of the data analysis are discussed and the implications of the results of this study are elaborated in Section 3.3.

3.1. Methodology

3.1.1. Data pre-processing

A large number of the available statistical tests depend upon the assumption that the data being analysed are normally distributed. However, it is commonly known that environmental data are typically highly skewed and do not follow a normal distribution. The data tend to be positively skewed or right-skewed since environmental data often have a lower bound of zero (negative values are impossible) and outliers are often present (Helsel and Hirsch 2002). The non-normal distribution of data will affect the choice of the statistical analysis used.

Data for multivariate data analysis were tested to determine the degree to which measured concentrations were normally distributed. Three normality tests, available in Matlab (The MathWorks, Inc., Natick, MA, USA), were considered: Lilliefors test (`lillietest`), Jarque-Bera test (`jbttest`) and Kolmogorov-Smirnov test (`kstest`). The Kolmogorov-Smirnov test compares the data to a standard normal distribution, while the Lilliefors test and Jarque-Bera test compare the data to a normal distribution. Both the Lilliefors test and Jarque-Bera test are suitable when a fully-specified null distribution is unknown and hence its parameters must be estimated. The test statistic of the Jarque-Bera test depends on the sample skewness

and kurtosis for a small sample size. The Lilliefors test is a modification of the Kolmogorov-Smirnov test which provides a reasonably powerful two-sided goodness-of-fit test (Taylor et al. 1986). Estimates of the Lilliefors test's significance levels are available for $N \geq 3$, where N is the number of samples, and thus may be readily applied to small sample sizes. For this reason, the Lilliefors test was used to test for data normality. The hypotheses for the normality test consist of H_0 = the data are normal (this null hypothesis may not be rejected due to small sample size) and H_1 = the data are not normally distributed. These hypotheses were tested at the 95% confidence level (or 5% significance level).

When the raw data sets did not all conform to a normal distribution, the data were transformed so as to create a dataset that was normally distributed. For example, logarithmically transforming a log-normally distributed dataset creates a normally distributed dataset. The Box-Cox transformation was also attempted to transform the data. The Box-Cox transformation is one of the most frequently used power transformation methods (McGrath et al. 2004; Templ et al. 2008). It is designed to determine an optimal transformation while fitting a linear regression model (Kannel et al. 2007). However, not all the data could be normalised through log or Box-Cox transformation. For those data for which normalisation was not possible, the condition for data normality in the application of multivariate data analysis was relaxed.

In comparing between different water quality parameters, standardisation of the data is needed to ensure that each variable has the same influence in the analysis (Kannel et al. 2007), especially for multivariate techniques that are based on distance coefficients (Templ et al. 2008). The normal or normalised data were thus standardised by shifting the mean to 0 and scaling the variance to 1 (converting into Z-score) as shown in Equation 3.1.

$$E(X) = \mu, Var(X) = \sigma^2$$
$$\text{If } X \sim N(\mu, \sigma^2) \text{ then } Z = \frac{X - \mu}{\sigma} \sim N(0, 1) \quad (3.1)$$

3.1.2. Cluster analysis

Cluster analysis (CA) (or clustering) divides objects into groups such that the objects in a group will be similar to one another and different from the objects in other groups. CA results depend on the choice of the data transformation, the clustering method, and the distance measure (Templ et al. 2008). The choice of data transformation, including standardisation, is discussed in Section 3.1.1. The choice of clustering method and the distance measure used are discussed below.

There are two main types of hierarchical clustering: agglomerative and divisive cluster analyses. Agglomerative clustering first treats the data as individual clusters. Then, it merges the closest pair of clusters one at a time until only a certain number of clusters is left. In contrast, divisive clustering first treats all data as one cluster. Then, it splits off the farthest point one at a time until there is a certain number of clusters formed. Hierarchical agglomerative CA was performed in this study on the standardised data by means of Ward's method and Euclidean distance. Ward's method is the best performing hierarchical clustering method (Templ et al. 2008); it attempts to minimise the sum of squared errors of any two clusters (Panda et al. 2006). Ward's method is less susceptible to noise and outliers. Euclidean distance is one of the most frequent measures of distance—that is the degree of separation between different objects or clusters. In general, the Euclidean distance is calculated as in Equation 3.2.

$$d(p, q) = \sqrt{\sum_{k=1}^n (p_k - q_k)^2} \quad (3.2)$$

where d is Euclidean distance, p and q are data objects or data points, n is the number of dimensions, attributes or parameters, p_k and q_k mean the k^{th} attributes of data object p or q , respectively. Ward's hierarchical clustering method in conjunction with the Euclidean distance measure, available in Matlab, was used. This method has also been used by other authors (Panda et al. 2006; Astel et al. 2007).

Clusters produced by hierarchical clustering are often illustrated with tree diagrams or dendrograms. The cophenetic correlation coefficient is a measure of how faithfully the tree represents the dissimilarities among observations. This measure can be used as a coarse indicator for the potential performance of different clustering methods (Trakhtenbrot and Kadmon 2006). The magnitude of this value should be very close to 1 for a high-quality solution (MathWorks 2013).

3.1.3. Principal component analysis

Principal component analysis (PCA) yields new (virtual) variables (principal components) that are artificially constructed from an existing set of (real) variables. The principal components (PC) are constructed so as to explain as much of the variation within the real variables as possible. Each principal component is completely independent and uncorrelated with the other principal components. Generally, principal components with eigenvalues greater than 1.0 will be used to simplify multivariate data because eigenvalues of this magnitude show there are strong correlations among variables (McKillup and Dyar 2010). Every principal component is an eigenvector. Each absolute value of the eigenvector's coefficients represents the degree of influence of the corresponding parameter, attribute or dimension (Manly 2005).

3.1.4. Summary

In sum, data pre-processing involved a normality test, data transformation, and data standardisation to Z-score. Raw data were first tested for normality using the Lilliefors test at the 5% significance level. If the data were found to be non-normal, log transformation or Box-Cox transformation was done. The data were then standardised to be used in the multivariate data analysis. Then, multivariate data analyses such as CA and PCA were done to classify the data. The CA results were presented in a dendrogram, while the PCA results were usually presented as a plot between the first and second principal components.

3.2. Data

3.2.1. Marina catchment one-day grab sample data

Marina catchment has an area of 10,000 ha, one-sixth the area of Singapore (PUB 2010b) and is 100% urbanised. Because Marina catchment is the largest and most urbanised catchment in Singapore, monitoring surface water quality in this catchment is especially important. Marina catchment is located in the heart of the city, and comprises the central business district. Figure 3.1 shows the land use characteristics of this catchment.

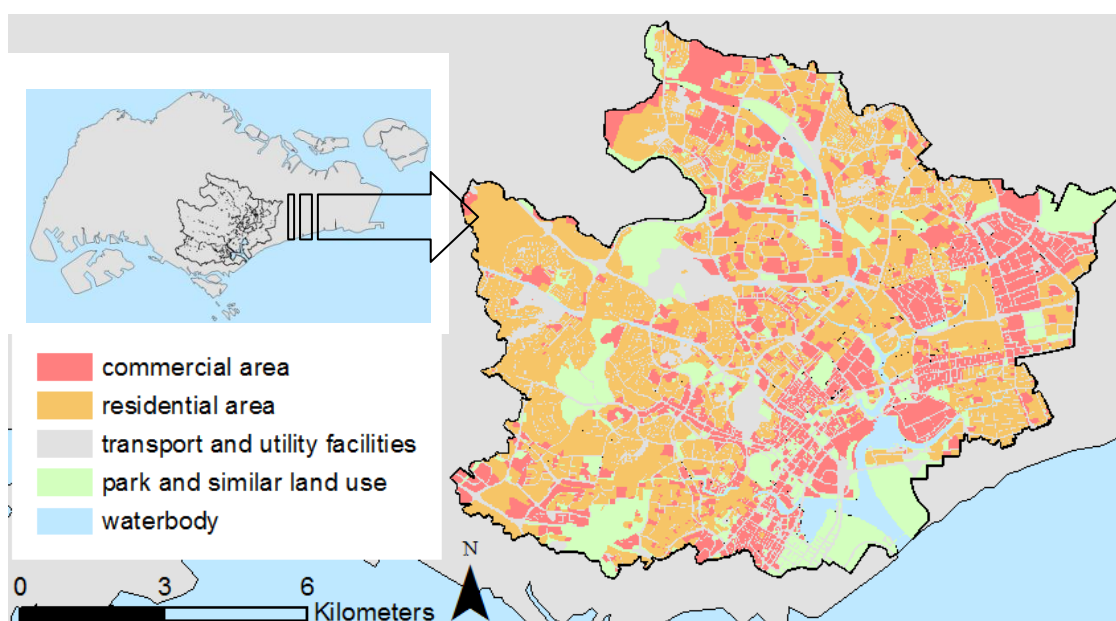


Figure 3.1. Map of Marina catchment

The one-day grab sample data from the Marina catchment contains measured concentrations of 16 water quality variables for samples collected at 21 locations in the catchment on 18 January 2010. The sampling locations are mapped in Figure 3.2. Out of 16 water quality parameters, only twelve were used for analysis: faecal coliform, *E. coli*, pH, turbidity, DSBP (distyryl biphenyl) which is a fluorescent whitening agent, cholesterol, caffeine, ibuprofen, naproxen, gemfibrozil, diclofenac and ketoprofen. Conductivity and salinity were removed from the suite of parameters for data analysis because they are highly influenced by distance from the sea and could cause spurious results in the analysis. Coprostanol and cholestanol were removed because they contained a significant number of non-detect values which

would have hampered the multivariate data analysis. In addition, data for 19 sampling locations out of the 21 sampling locations were considered as the other 2 locations contained non-detect faecal coliform and *E. coli* which were implausible because they were environmental surface water samples from a highly urbanised catchment. The water quality parameters and locations considered in the analysis are provided in Table 3.1. There are five non-detect pharmaceutical concentrations (one naproxen, one gemfibrozil and three ketoprofen concentrations) which were replaced with 0.1 ng/L and one non-detect cholesterol which was replaced with 5 ng/L. Those replaced values are printed in boldface in Table 3.1. The value of 0.1 ng/L was chosen because it is an extremely low concentration which is close to the lowest reading of 0.093 ng/L (ketoprofen). The value of 5 ng/L was chosen because it is close to the lowest reading of cholestanol (5.04 ng/L, not shown), which is measured the same way with cholesterol.

The Lilliefors test results showed that the hypothesis of normal distribution for ibuprofen, naproxen and diclofenac cannot be rejected. A further Lilliefors test on \log_{10} -transformed data (not shown) revealed that the hypothesis of normal distribution for the logarithm of caffeine and ketoprofen also cannot be rejected. A Lilliefors test on Box-Cox transformed data (not shown) also revealed that the hypothesis of normal distribution for Box-Cox transformed pH, turbidity and cholesterol cannot be rejected. In contrast, this hypothesis can be rejected for faecal coliform, *E. coli*, DSBP and gemfibrozil whether or not transformed. The condition for data normality for these parameters was relaxed: Box-Cox transformations were, however, used. When the appropriate transformation for each parameter had been done, the data were standardised.

Table 3.1. Measured indicator concentrations in Marina catchment one-day grab samples and Lilliefors test results

No	Sampling location	Faecal Coliform	<i>E. coli</i>	pH	Turbidity	DSBP (Fluorescence)	Cholesterol	Caffeine	Ibuprofen	Naproxen	Gemfibrozil	Diclofenac	Ketoprofen
		CFU/100mL	CFU/100mL		NTU	µg/L	µg/L	µg/L	ng/L	ng/L	ng/L	ng/L	ng/L
1	Rocher River Estuary	6,800	3,700	7.36	7.02	1.94	5.223	0.24	60.9	3.6	6.0	1.5	0.1
2	Kallang Riverside Park	1,500	800	7.57	6.44	1.89	2.031	0.03	48.6	2.7	8.2	2.1	0.7
3	Kallang Basin Aeration Area	1,700	800	7.74	5.63	1.77	1.567	0.04	44.9	3.5	5.7	2.6	0.1
4	Bridge near Tanjong Rhu Road	14	7	7.84	5.55	1.81	1.950	0.02	52.4	2.5	5.2	2.8	3.6
5	Marina Channel Aeration Area	12	6	8.02	5.81	1.66	1.955	0.02	37.1	1.8	3.7	3.7	3.9
6	Near Marina Barrage	7	4	8.06	6.23	1.65	0.659	0.02	39.4	0.1	3.5	1.9	0.7
7	Marina Bay Aeration Area	4	3	8.05	5.63	1.65	2.019	0.05	39.2	4.0	3.7	2.5	0.1
8	Asian Civilization Museum	440	220	8.02	6.09	1.70	0.728	0.08	43.9	3.5	3.9	2.9	0.7
9	Read Bridge	3,600	1,800	7.96	7.94	1.82	76.440	0.03	44.9	3.3	3.8	1.9	0.3
10	Jiak Kim Bridge	1,100	500	7.69	4.88	1.88	0.251	0.02	48.5	0.1	3.5	3.2	0.4
11	Little India MRT	120,000	80,000	7.32	1.82	0.75	3.380	0.47	22.4	1.6	0.1	0.9	1.1
12	Newton Church	36,000	19,000	7.18	4.91	1.49	8.841	0.54	57.2	2.0	0.9	1.6	1.9
13	Drainage Near Mayne Road	6,300	3,100	7.08	1.31	2.52	6.581	0.36	63.8	0.2	0.1	1.0	0.1
14	Kg Kayu Bridge	3,300	1,800	7.82	5.54	2.41	4.047	0.09	73.1	4.4	9.4	3.0	2.1
15	Upper Boon Keng RD	1,900	1,000	7.89	7.92	1.93	4.836	0.39	70.4	3.4	9.8	2.5	0.5
16	Marina Channel Aeration Area (1m)	7	4	8.01	6.72	1.66	3.198	0.01	41.2	1.3	3.5	1.6	0.7
17	Marina Channel Aeration Area (2m)	5	2	8.03	5.87	1.64	0.014	0.01	40.0	1.4	3.6	3.4	1.5
18	Marina Channel Aeration Area (3m)	18	11	7.95	5.69	1.66	54.131	0.01	41.0	2.5	4.0	4.0	3.4
19	Marina Channel Aeration Area (4m)	14	9	7.92	6.29	1.62	0.005	0.01	42.1	1.5	3.3	1.4	1.7
Lilliefors test results		1	1	1	1	1	1	1	0	0	1	0	1
p values		0.001	0.001	0.037	0.001	0.006	0.001	0.001	0.118	0.438	0.009	0.500	0.007

For the Lilliefors test results, 0 means the hypothesis of normal distribution cannot be rejected and 1 means the hypothesis of normal distribution can be rejected.

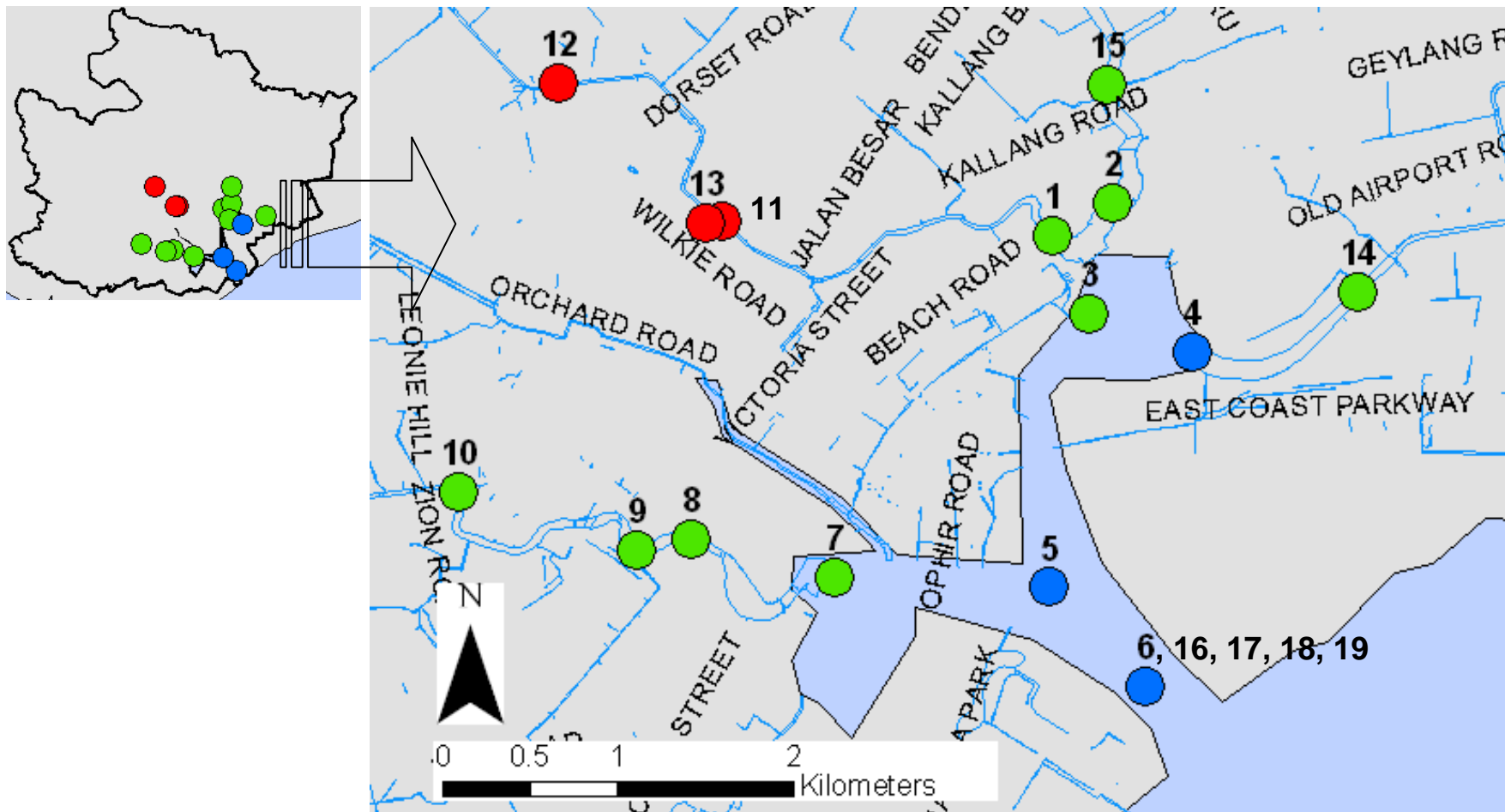


Figure 3.2. Sampling locations of Marina catchment one-day grab sample data (Symbol colours correspond to colour groupings in Figures 3.3 and 3.4.)

3.2.2. Toolbox Study data

The Toolbox Study collected analytical results for 23 water quality variables analysed for samples collected at hourly intervals over three consecutive days (02 February 2009, 10:00 – 05 February 2009, 09:00) (Hans Eikaas, PUB, personal communication, 2010). The samples were collected during dry weather, at a low density residential area comprised of single-family housing. The location and catchment boundary of the sampling site are shown in Figure 3.3. The sampling location is a box culvert located at the intersection of Dunearn Road and Vanda Road. The catchment is steep, and the elevation difference between the highest point in the catchment and sampling location is about 15 m.

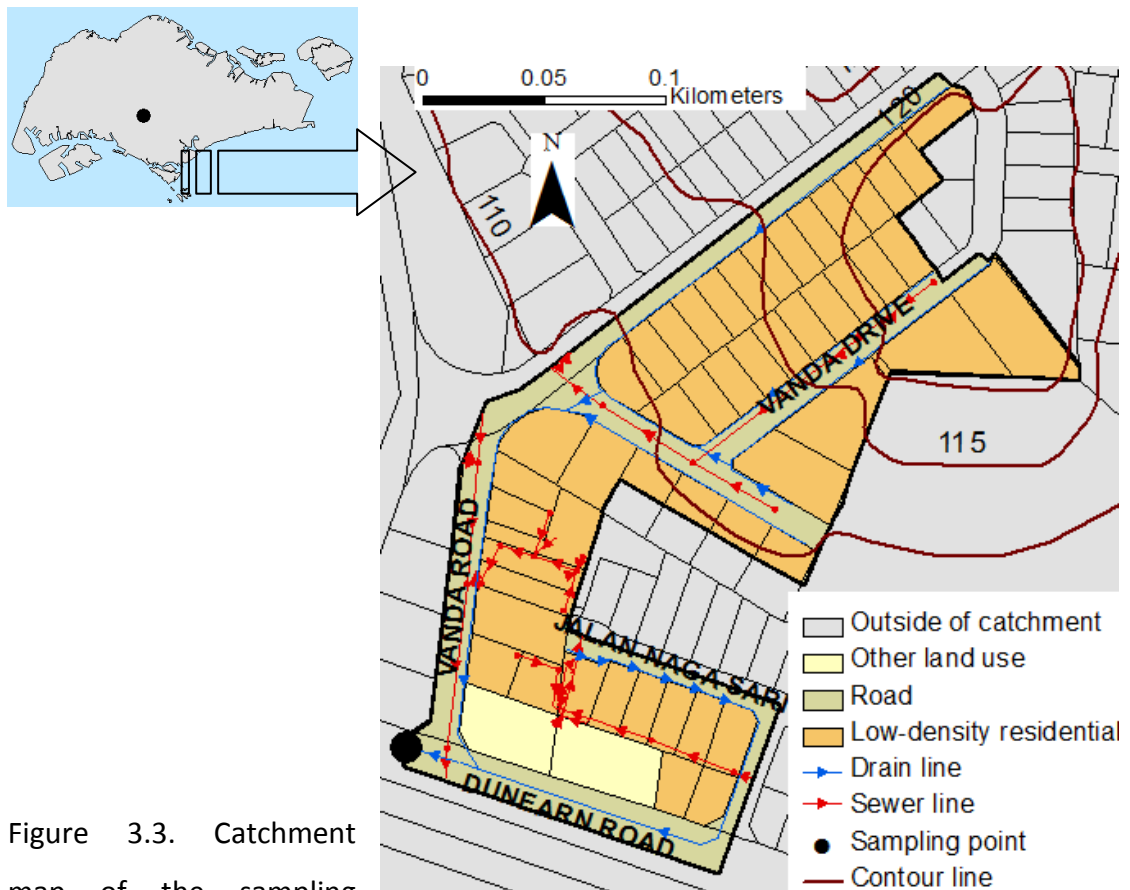


Figure 3.3. Catchment map of the sampling location for the Toolbox Study

3.3. Results and discussion

3.3.1. CA and PCA of Marina catchment one-day grab sample data

Cluster analysis results are presented in Figure 3.4. The cophenetic correlation coefficient of the dendrogram is 0.71 which implies that there is reliable dissimilarity (distance) among the observations. The high cophenetic correlation coefficient also implies that the cluster analysis method employed in this chapter (as elaborated in Section 3.1.2) has performed well in clustering the Marina catchment data. Out of nineteen sampling locations (where the sampling location in Figure 3.4 matches those in Table 3.1), seven (locations 6, 16, 19, 4, 5, 17, 18) had close Euclidean distances. This means that the water quality, as defined by the parameters listed in Table 3.1, at the seven sampling locations is similar. It can be observed in Figure 3.2 that these sampling locations are located in the downstream part of the catchment. Therefore, the seven sampling locations are grouped as the “reservoir” cluster. Similarly, the next nine sampling points (locations 1, 15, 14, 2, 3, 10, 7, 8, and 9) are located at the mid-stream portion of the catchment. Therefore, they are classified as the “river” cluster. The other three sampling points (locations 11, 12, 13) are located at the upstream portions of the catchment and therefore they are classified as the “drain” cluster. The fuzzy c-means clustering method was also applied (not shown) and the samples were clustered similarly with the clusters obtained from the hierarchical agglomerative cluster analysis.

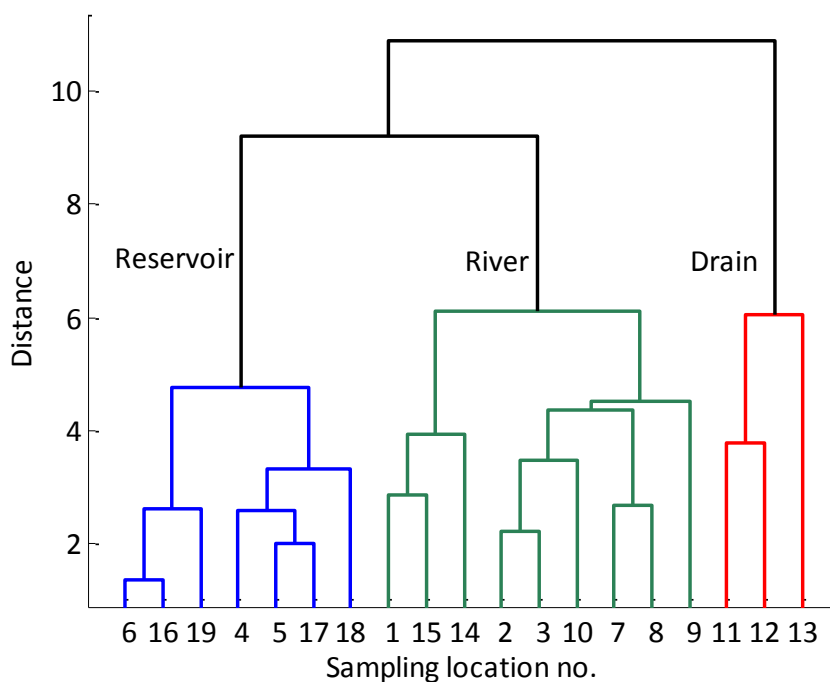


Figure 3.4. Dendrogram of CA for the Marina catchment one-day grab sample data (Numbers assigned to sampling locations are defined in Table 3.1.)

Principal component analysis (PCA) was carried out next. The first two principal components explain 61% of the observed variation. A plot containing these two principal components (PC1 and PC2) is shown in Figure 3.5. PC1 was mostly influenced by *E. coli* and faecal coliform, whereas PC2 was mostly influenced by gemfibrozil, naproxen and ibuprofen. Similarly to the cluster analysis results, the data can be grouped into three clusters (the “reservoir”, “river” and “drain” clusters) as symbolised in Figure 3.5.

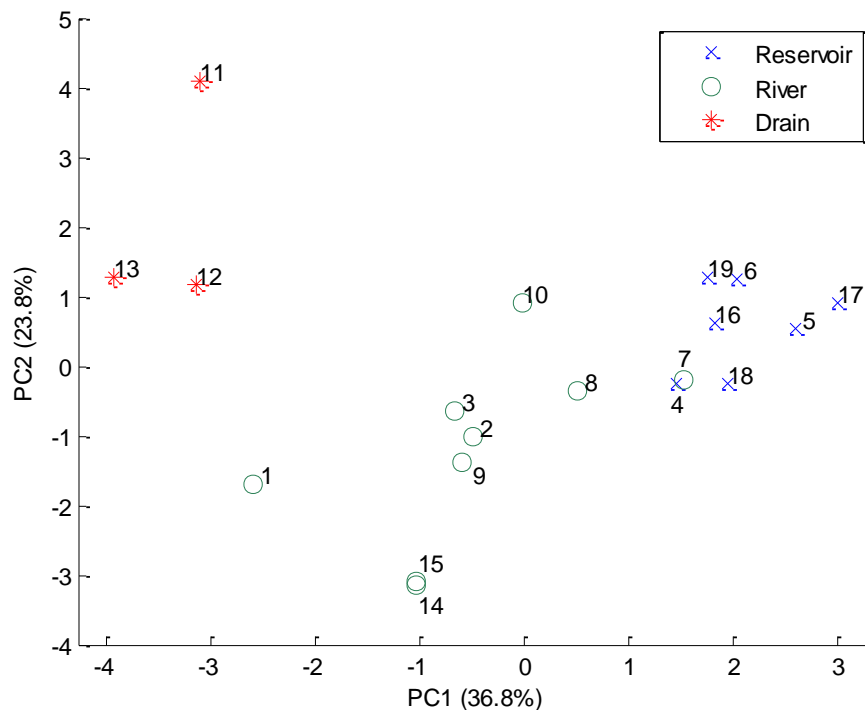


Figure 3.5. Plot of PC1 vs. PC2 for the Marina catchment one-day grab sample data (Numbers in the plot are sampling locations which are defined in Table 3.1)

It is hypothesised that land use may be an important factor in drain water quality because the sub-catchments associated with the “drain” sampling points consist of different type of land uses. On the other hand, land use may not be such an important factor for “river” and “reservoir” water quality because the catchments draining to “river” and “reservoir” sampling points typically comprise a mix of land uses. The surface water quality characteristics of the “river” and “reservoir” samples are not specific to a single land-use type and therefore not as distinct from one another. The cluster and principal component analysis involve chemical parameters (such as pharmaceutical compounds) which can be associated with humans. Therefore, the surface water quality contamination may have originated from sewage. Figure 3.2 shows the colour-coded map of the CA and PCA results. Circle symbols filled with red colour in Figure 3.2 represent samples from the upstream portions of the catchment (“drain”) which are nearest to potential sewage sources such as leaking sewers. Green represents samples from the mid-stream (intermediate) portion of the catchment (“river”). Blue represents samples from downstream portions of the catchment (“reservoir”) which is located the farthest

from the source. There is overall better surface water quality in the downstream portion of the catchment than in the intermediate or upstream portions of the catchment because the distance from the source provides opportunity for dilution and for the contaminants to settle or degrade.

3.3.2. Diurnal variation of Toolbox Study data

Visual inspection of the plots of each parameter versus time revealed that 14 parameters peak at similar times: faecal coliform, enterococci, somatic coliphage, pH, temperature, turbidity, total phosphorus, orthophosphate, detergent as MBAS, cholesterol, cholestanol, coprostanol, epicholestanol and epicoprostanol. This observation suggests that the parameters studied may originate from the same source and may indicate the suitability of those parameters as tracers for human faecal and sewage contamination. Plots of the hourly variation of all 14 parameters are presented in Figure 3.6 in which nighttime samples (defined as samples collected from 20:00 to 07:00) are indicated by the shaded time intervals. Faecal coliform and enterococci are widely used indicators of faecal contamination. Coprostanol is the most abundant sterol found in human faecal matter (Férézou et al. 1978; Leeming et al. 1996) and hence it is indicative of human faecal pollution.

A closer inspection of Figure 3.6 (a) and (b) reveals that there are generally two types of orthophosphate and detergent as MBAS peaks: peaks that coincide with faecal coliform and enterococci peaks as described above and peaks that do not coincide with faecal coliform and enterococci peaks. This suggests that there are two possible sources of orthophosphate and detergent as MBAS in drain water. The first is leaking sewers, which in turn suggests there may be insufficient soil adsorption to attenuate orthophosphate during subsurface transport from the underground sewers into the surface drains or minimal adsorption if orthophosphate from leaking sewers travels to the surface water through macropore pathways such as subsurface voids. Insufficient soil adsorption is plausible if soapy water has seeped through the soils for years and therefore the available sorption sites have been exhausted. Walter et al. (1996) reported this phenomenon in Cape Cod, MA, USA. Continuous orthophosphate

loading from the disposal of secondarily treated sewage had filled available sorption sites and therefore orthophosphate was found mobile and detected in a downstream waterbody. The second possible source is direct discharge of household laundry wastewater into drain lines instead of sewer lines. Many private residential home owners carry out laundry washing in the backyard and discharge laundry wastewater directly into the surface water drains. The orthophosphate and detergents from this source would not necessarily be expected to peak at the same time as bacterial indicators whereas leakage from sewers clearly would.

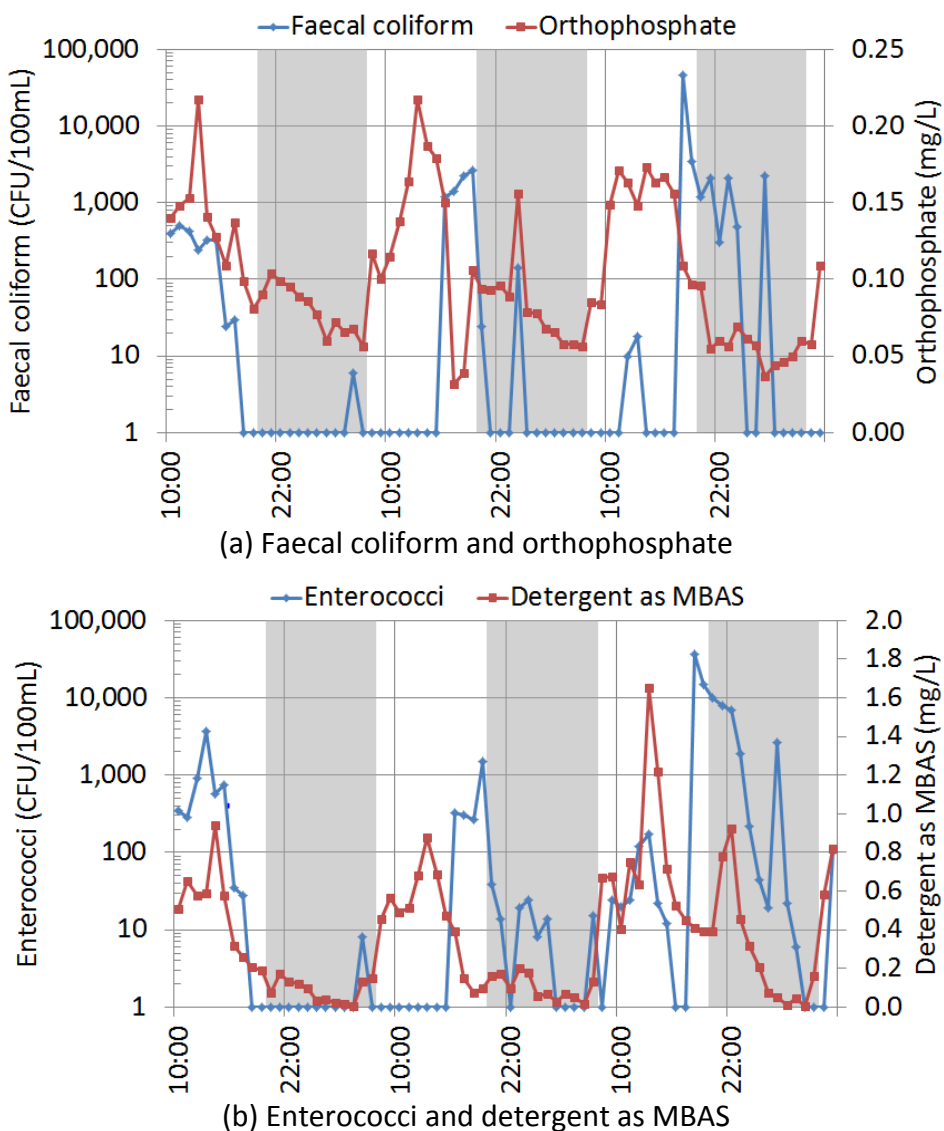
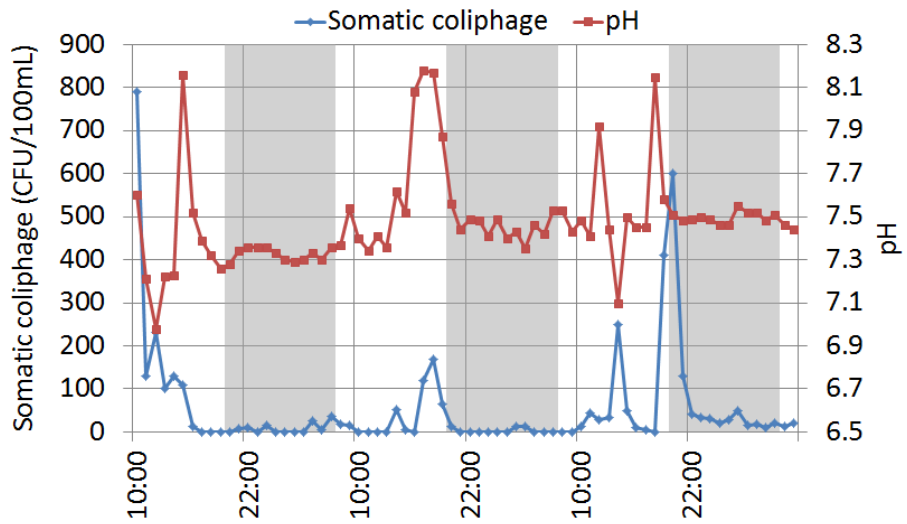
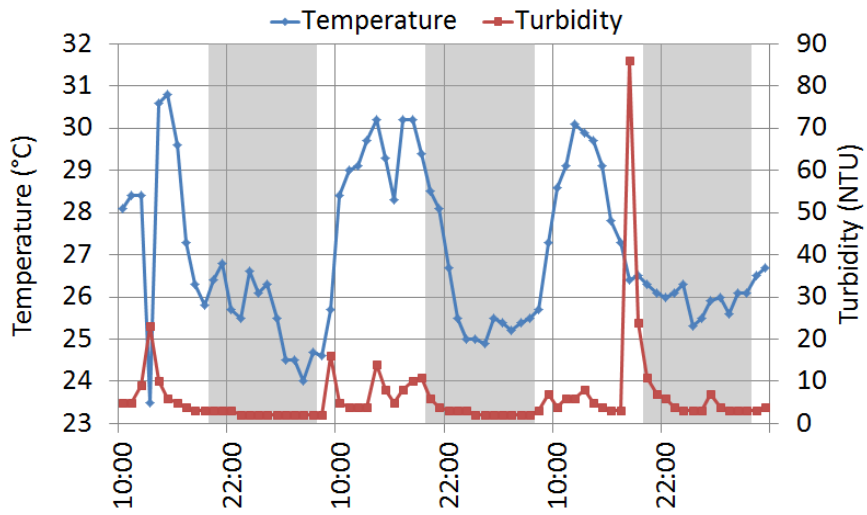


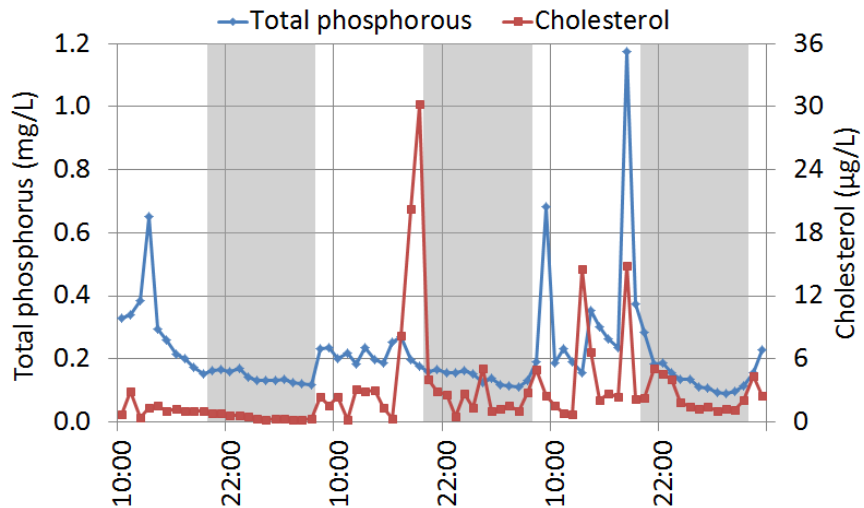
Figure 3.6. Hourly plot of fourteen Toolbox Study's parameters from 02 Feb 2009, 10:00 to 05 Feb 2009, 09:00



(c) Somatic coliphage and pH



(d) Temperature and turbidity



(e) Total phosphorus and cholesterol

Figure 3.6.(continued) Hourly plot of fourteen Toolbox Study's parameters from 02 Feb 2009, 10:00 to 05 Feb 2009, 09:00

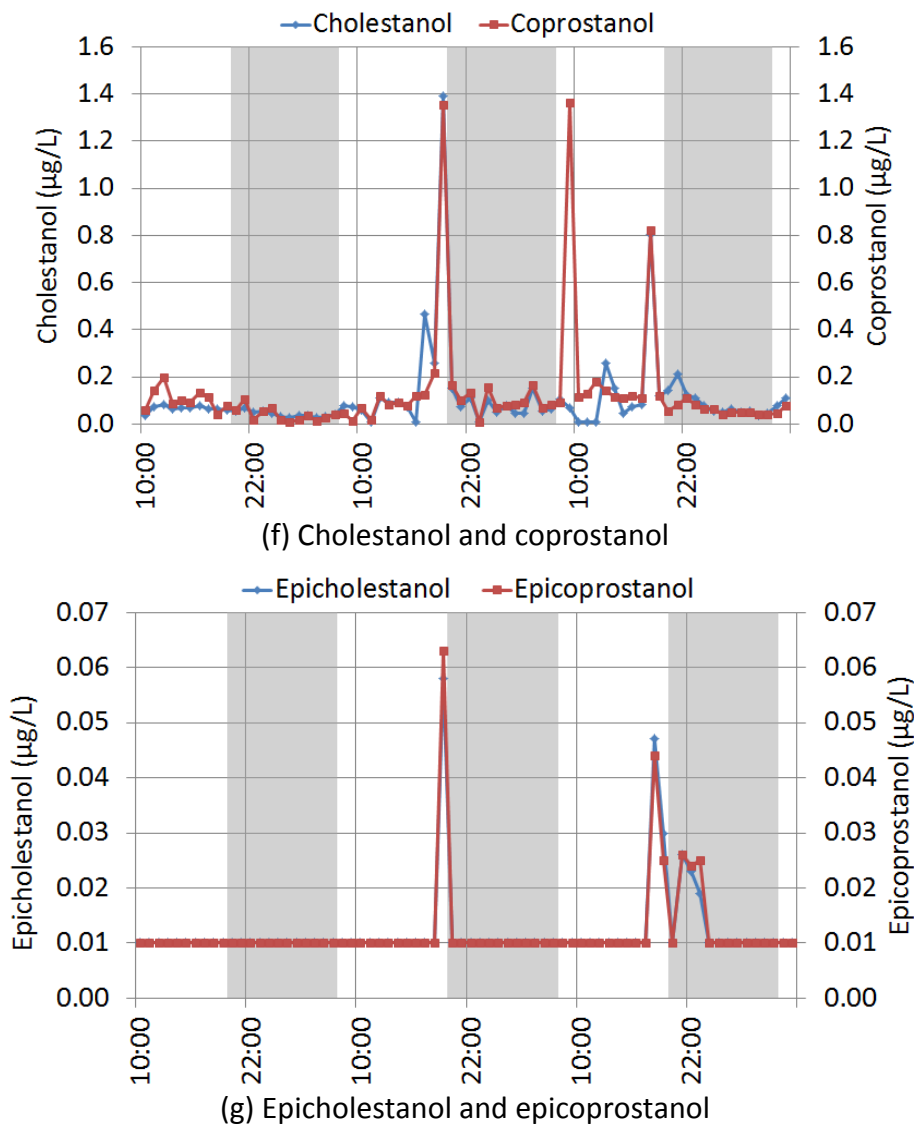


Figure 3.6.(continued) Hourly plot of fourteen Toolbox Study’s parameters from 02 Feb 2009, 10:00 to 05 Feb 2009, 09:00

Figure 3.6 also provides an indication of a diurnal pattern. The concentrations are higher during the daytime (unshaded time interval from 08:00 to 19:00 in Figure 3.6) and lower during the nighttime (shaded time interval from 20:00 to 07:00 in Figure 3.6). This behaviour appears to be repeated for the three days that sampling was conducted, although large variations in peak concentrations can be observed during the three days. This observed diurnal pattern strongly suggests the exfiltration of sewage from the reticulation system. This hypothesis is supported by two considerations. First, the concentration pattern mimics the typical daily sewage flow pattern in which high flow is observed during the daytime and low flow is observed

during the nighttime (especially late at night and early in the morning), as discussed in Section 2.1.4. Secondly, the catchment has a steep hilly topography. This makes it more likely the sewer lines lie above the ground water table allowing exfiltration from the sewer pipes.

3.3.3. Implications of the preliminary data analysis on this study

Multivariate data analysis of the Marina catchment data reveals that the samples can be grouped into a “drain” cluster for upstream samples, a “river” cluster for mid-stream samples and a “reservoir” cluster for downstream samples. This implies that the sampling location influences the surface water quality. In addition, the type of land uses in the upstream areas may impart distinct surface water quality characteristics. The finding of the Marina catchment data analysis and the hypothesis that land use affects drain water quality provide guidance for site selection in this study: sampling should be conducted in upstream areas with specific land uses in order to distinguish the effects of land use on water quality.

The plots of hourly Toolbox Study data (Figure 3.6) suggest diurnal variation of surface water quality parameters. Besides, the plots show that different water quality parameters associated with sewage peak at about the same time. The diurnal variation follows typical daily sewage flow patterns. This and the similar behaviour among parameters imply that the signal may originate from sewage which leaks and exfiltrates to surface water drains. In addition, the agreement among parameters also implies the potential suitability of the parameters (other than the FIB) as tracers. The findings from the analysis of the Toolbox Study data leads to the following questions: (i) do similar diurnal patterns exist at other locations, (ii) is sewage exfiltration the cause of the observed diurnal patterns and (iii) if exfiltration is shown to be the cause, what are the potential tracers.

CHAPTER 4 METHODOLOGY

The findings in Chapter 3 influenced the approaches adopted in parameter selection and the field sampling campaign for this study. Field sampling consisted of continuous sampling at field sites chosen primarily based on land use. The continuous sampling technique adopted is consistent with the PUB Toolbox Study (Hans Eikaas, PUB, personal communication, 2010). This is a departure from the commonly used grab-sampling method, which is opportunistic, and was adopted because time variation in FIB concentrations was expected to be a significant feature of the investigation. The adopted methodology is further elaborated in this chapter.

4.1. Parameter selection

Possible parameters and methodologies that could be explored for use as potential indicators of faecal contamination are elaborated in Sections 2.3 through 2.5. After considering the resources available, available analytical methods, and practicality, 25 parameters were chosen to be analysed in the field and at three laboratories. In the NTU Environment Laboratory, the parameters analysed were faecal indicator bacteria (FIB) which were comprised of total coliform (TC), *Escherichia coli* (*E. coli*) and enterococci. Physical parameters which included water temperature, specific conductance and salinity were measured in the field. Inorganic parameters (chloride, bromide, boron and orthophosphate), detergents (as methylene blue active substances or MBAS), an antimicrobial (triclosan), faecal sterols (cholesterol, cholestanol and coprostanol), plasticizers (DBP and DEHP), a stimulant (caffeine), pharmaceutical compounds (acetaminophen, ibuprofen and diclofenac) and artificial sweeteners (acesulfame-k, sucralose and saccharin) were analysed in an accredited laboratory. A genetic marker (HF183 which is a human-specific *Bacteroides* marker) was analysed in the MIT Parsons Laboratory.

E. coli and enterococci were analysed because they are recommended by USEPA (1986) and are widely used as FIB. Total coliform was also analysed because the methodology that is explained in Section 4.4.1 enables us to obtain this parameter at

the same time as *E. coli*. Water temperature, specific conductance and salinity are physical parameters that can be measured using a typical handheld water-quality meter. Inorganic parameters such as chloride, bromide and boron were analysed as suggested by Hyer (2007). Orthophosphate was analysed because Zheng et al. (2013) showed that samples with positive results for *E. coli* and/or enterococci tended to have higher orthophosphate concentrations. Surfactants, predominantly linear alkylbenzene sulfonate (LAS), were measured using the MBAS test because they are a prevalent part of detergents and hence of wastewater, especially municipal wastewater. Triclosan was included because Hyer (2007) categorised triclosan as one of the organic compounds that was identified as a useful tracer for sewage. Cholesterol and coprostanol were analysed because they were detected frequently in streams in USA (Kolpin et al. 2002). Coprostanol and cholestanol are two biotransformation products of cholesterol in human body (Férézou et al. 1978). DEHP was analysed because it is a widely used PVC component (Wams 1987) and PVC is a commonly used material in sewer pipes. Cholestanol and DBP were also analysed because they were analysed by the same laboratory procedure as cholesterol and DEHP.

A variety of pharmaceuticals and caffeine were also investigated. Caffeine is a widely used chemical compound in consumed products such as beverages, foods and drugs (Wu et al. 2008). Hence, it is a good indicator of human faecal contamination. However, as direct discharge into the drainage system is possible, the presence of caffeine must be interpreted carefully. Acetaminophen is a drug commonly consumed in Singapore. Diclofenac and ibuprofen are used worldwide (Buser et al. 1998; Buser et al. 1999). Nifedipine, one of the most commonly prescribed drugs by doctors in Singapore (Lim and Yap 1999), was also considered. However, the standard is extremely expensive, approximately SGD 3,000 per 1 g of dehydronifedipine (electronic communication with an accredited laboratory, 25 November 2010). Nifedipine is not suitable as an indicator and was ruled out since indicators should be quantifiable in recreational waters by reasonably inexpensive methods (Cabelli 1977). Acesulfame-k, sucralose and saccharin were analysed because Buerge et al.

(2009) showed that they can be present in measurable quantities in surface waters. Aspartame, a class of artificial sweetener, is not included because it is known to be easily degraded. Cyclamate, another artificial sweetener, is also not used since it is not permitted to be used in Singapore (AVA 2006). Other food additives like monosodium glutamate (MSG) were considered but not chosen because it is known to be easily degraded in the environment (Philip M. Gschwend, MIT, personal communication, 2010). HF183 marker was analysed because Nshimiyimana (2010) showed the suitability of this tracer in Singapore through its phylogenetic analysis.

The list of parameters was reduced during the course of this research since some of the parameters proved to be mostly non-detect. As a result, bromide and triclosan were only analysed for samples in January 2011. Diclofenac and acesulfame-k were only analysed for samples in January 2011, June-July 2011 and January-April 2012.

4.2. Site selection

Analysis of the Marina Catchment data in Section 3.2.1 revealed that surface water samples could be grouped according to whether samples were collected from downstream, midstream or upstream locations in the drainage system. This suggests that land use may have an influence on the water quality in the drainages. This is consistent with results obtained by Desai and Rifai (2010) who sampled in urban areas in Houston and found that the spatial variability of *E. coli* concentrations is highly correlated with land use. Therefore, sampling points were identified based on delineation of sub-catchments according to land use. Each of these catchments was mapped and analysed using ArcGIS Desktop 10 (Esri, Redlands, CA, USA). Sampling locations were selected based on technical and practical reasons which include percentage of the major land use represented (the higher the percentage, the better), dry-weather flow rate (in order to obtain sufficient sample volume), area at the site available for equipment installation and access to the sampling point. About 34 sites were considered, visited and delineated. Out of these, 13 sites were selected. They represent the drainage outlets from six high-density residential areas consisting of high-rise apartment buildings, five low-density residential areas consisting of single-family housing units, one commercial and one industrial area. All of these 13

sites are concrete-lined drains. Catchment characteristics are shown in Table 4.1, sampling points in Figure 4.1 and detailed catchment maps in Figure 4.2. Sewer lines are also shown in Figure 4.2 to show that the sanitation practice in all sampling sites is to discharge to a sanitary sewer line. However, the sewer lines drawn in Figure 4.2 only indicate the main sewer network and do not indicate the shallow (upstream) sewer lines such as a sewer line serving a building and its connection to the main sewer network.

Table 4.1. Catchment characteristics: (a) high-density residential (b) low-density residential, commercial and industrial

(a)

Major land use		High-density residential				
Sampling site	Lorong 8 Toa Payoh	Toa Payoh North	Lorong 6 Toa Payoh	Ang Mo Kio Avenue 10	Choa Chu Kang Crescent	Punggol Central
Coordinates	103° 51' 19.2" E, 1° 20' 29.1" N	103° 50' 59.6" E, 1° 20' 32.2" N	103° 50' 48.4" E, 1° 19' 52.7" N	103° 51' 16.0" E, 1° 22' 31.9" N	103° 45' 7.6" E, 1° 24' 4.8" N	103° 54' 45.9" E, 1° 23' 53.1" N
Total area (ha)	24	30	30	30	37	91
% by land use type:						
High-density residential	54	77	53	69	84	64
Low-density residential	0	0	0	0	0	0
Commercial (including hotel, civic & community institution, educational institution and place of worship)	23	5	30	18	4	8
Industrial (business)	0	0	0	0	0	0
Agricultural	0	0	0	0	0	0
Transportation (road, light rapid transit), utility	15	18	14	12	11	17
Sports and recreation	0	0	0	0	0	0
Other (including park, reserve site, open space in urban and waterbody)	8	0	2	1	2	11

Table 4.1. (continued) Catchment characteristics: (a) high-density residential (b) low-density residential, commercial and industrial

(b)

Major land use	Low-density residential					Commercial	Industrial
Sampling site	Serangoon Garden	Jalan Siantan	MacPherson Lane	Watten Drive	Verde	Bras Basah (Waterloo St)	Sims Place
Coordinates	103° 51' 33.8" E, 1° 21' 38.5" N	103° 47' 54.1" E, 1° 19' 22.7" N	103° 53' 4.9" E, 1° 19' 50.9" N	103° 48' 22.4" E, 1° 19' 39.0" N	103° 45' 9.7" E, 1° 23' 29.9" N	103° 51' 12.0" E, 1° 18' 3.0" N	103° 52' 48.0" E, 1° 18' 58.4" N
Total area (ha)	29	21	8	7	7	15	7
% by land use type:							
High-density residential	0	0	0	12	0	0	2
Low-density residential	78	80	86	65	76	0	0
Commercial (including hotel, civic & community institution, educational institution and place of worship)	0	1	0	0	0	65	10
Industrial (business)	0	0	0	0	0	0	63
Agricultural	0	0	0	0	0	0	0
Transportation (road, light rapid transit), utility	20	18	14	22	24	31	21
Sports and recreation	1	0	0	0	0	0	0
Others (including park, reserve site, open space in urban and waterbody)	1	1	0	2	0	4	5

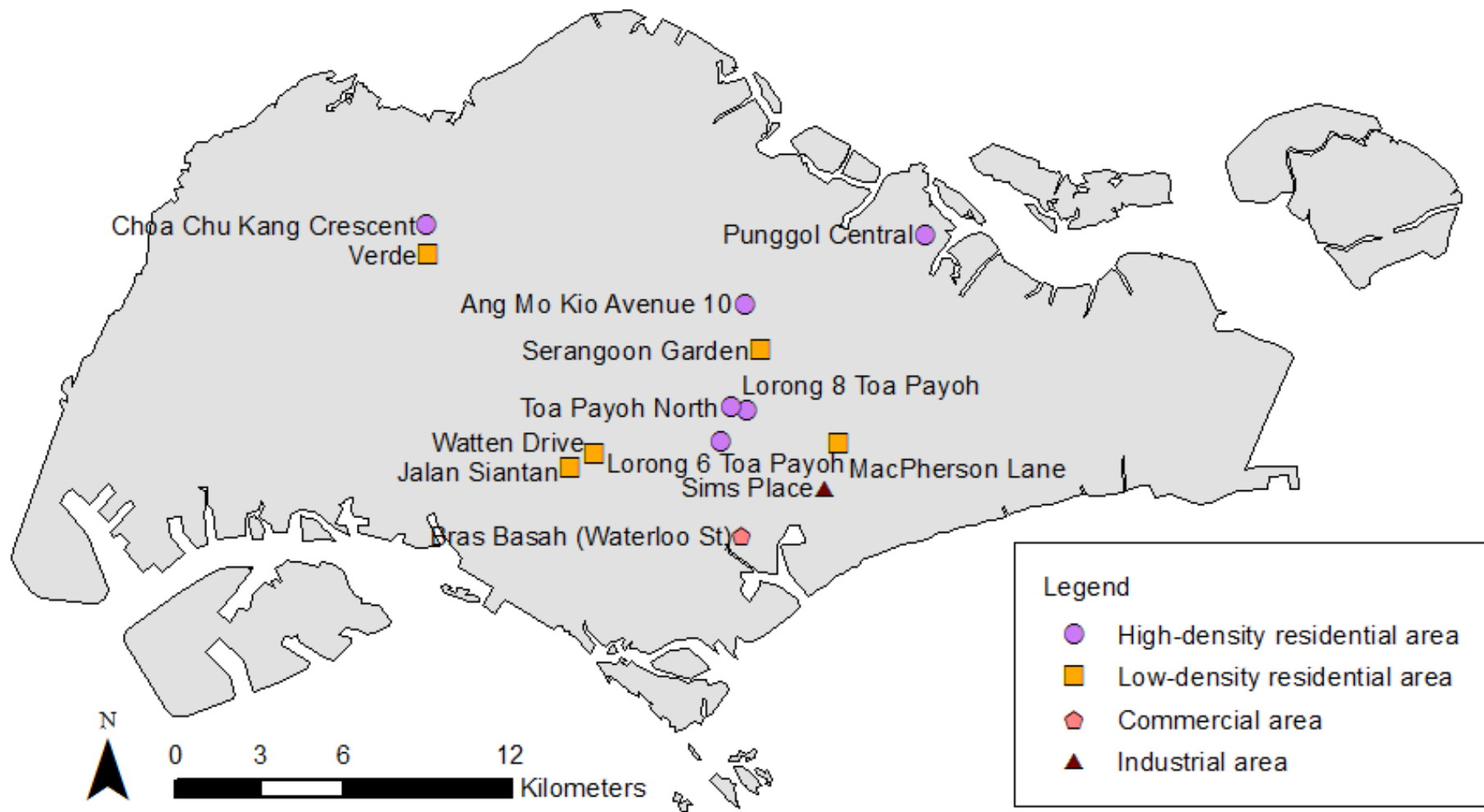
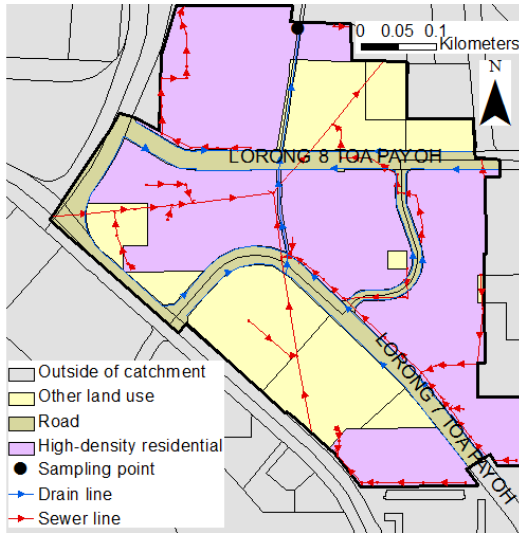
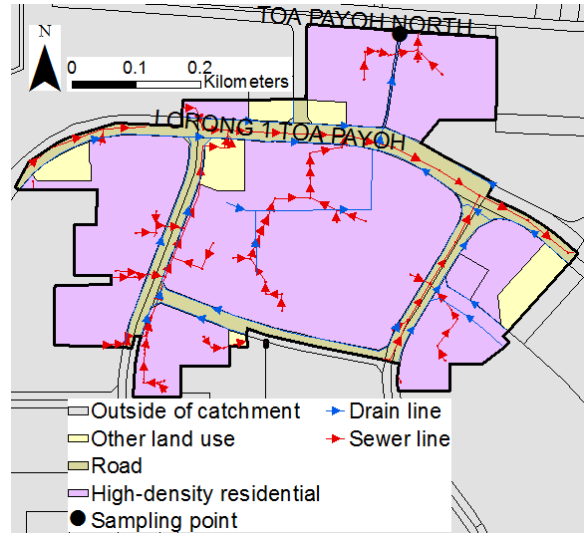


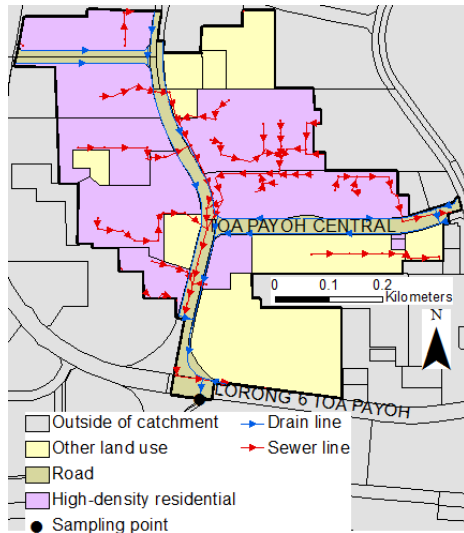
Figure 4.1. Sampling points of this study



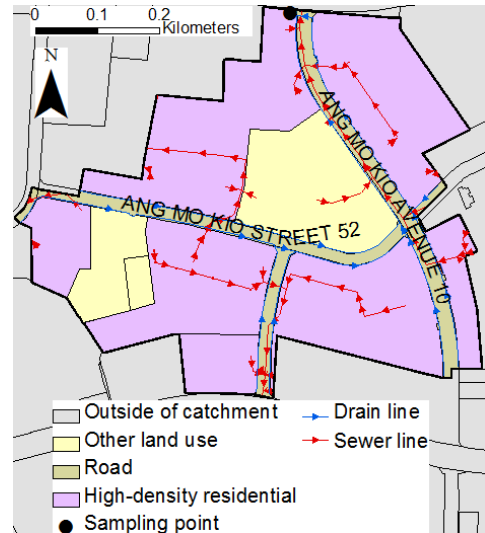
(a) Lorong 8 Toa Payoh



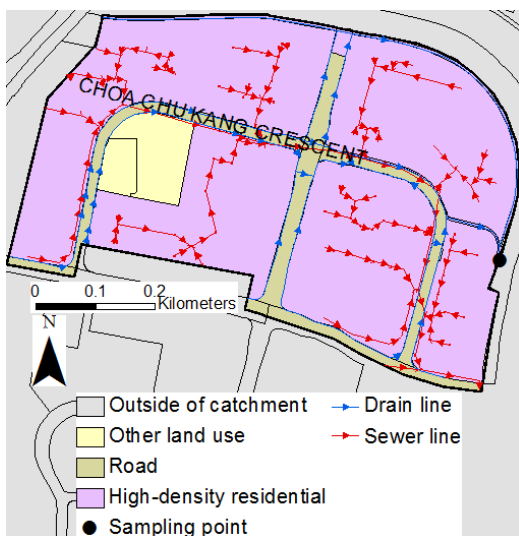
(b) Toa Payoh North



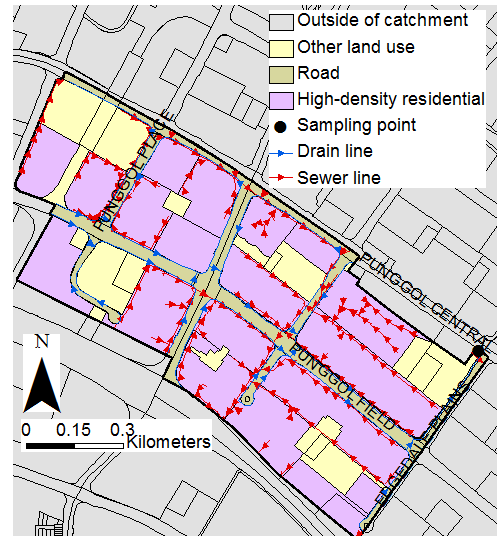
(c) Lorong 6 Toa Payoh



(d) Ang Mo Kio Avenue 10

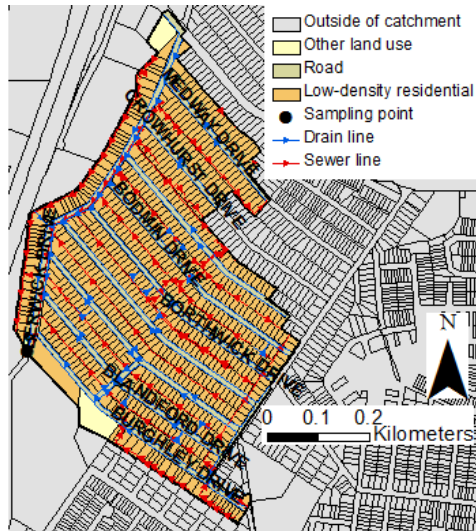


(e) Choa Chu Kang (CCK) Crescent

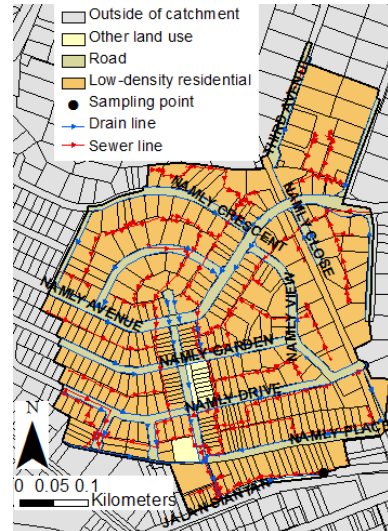


(f) Punggol Central

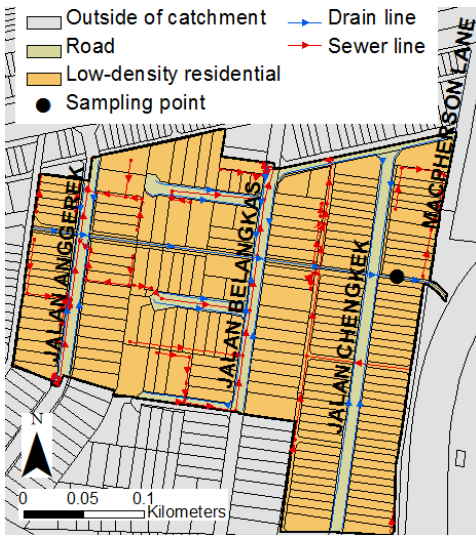
Figure 4.2. Detailed catchment maps of all sampling locations



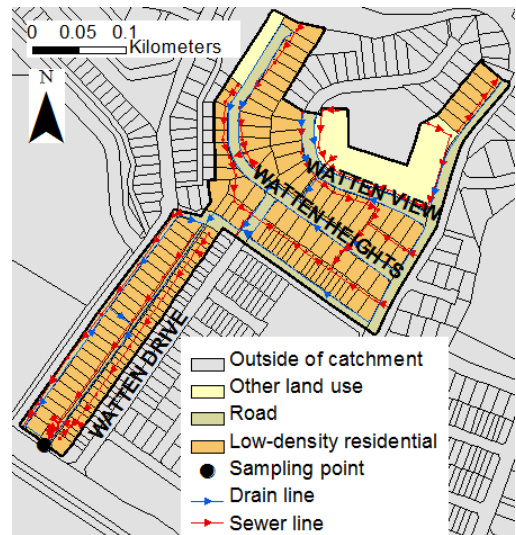
(g) Serangoon Garden



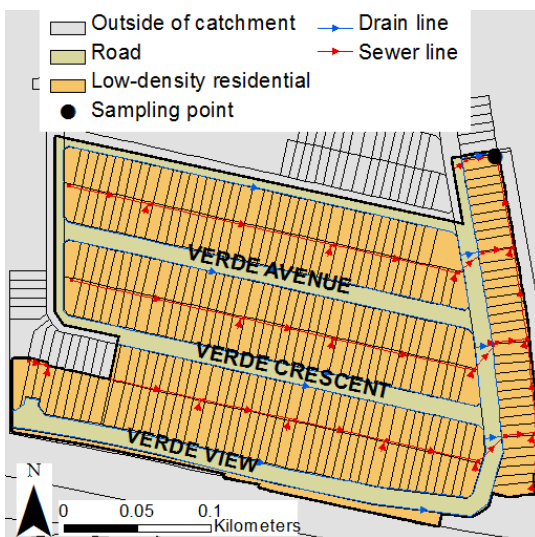
(h) Jalan Siantan



(i) MacPherson Lane



(j) Watten Drive



(k) Verde

Figure 4.2. (continued) Detailed catchment maps of all sampling locations

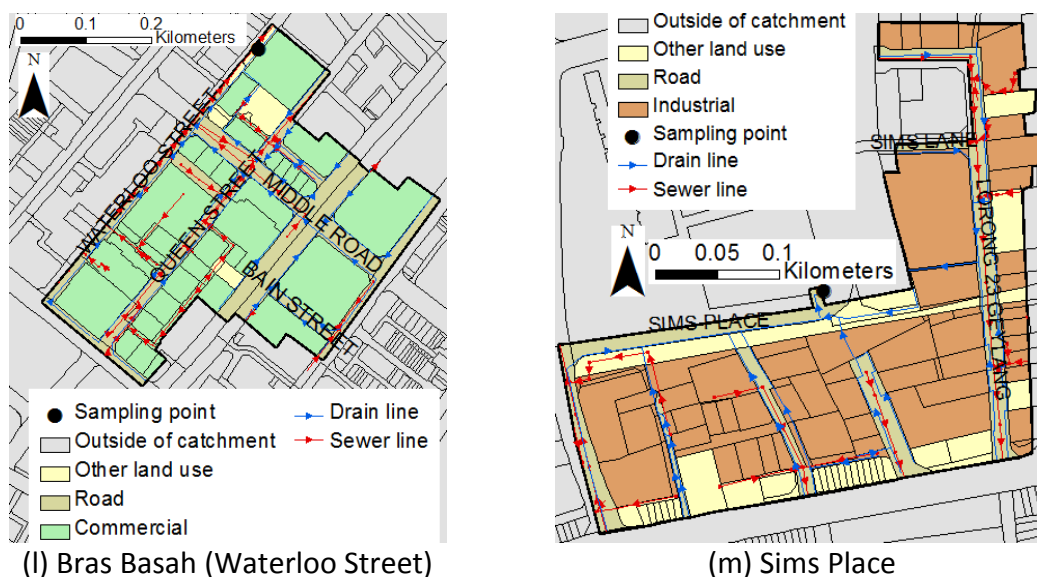


Figure 4.2. (continued) Detailed catchment maps of all sampling locations

In addition, five raw sewage samples were collected from manholes in residential areas as listed in Table 4.2. Four non-urban water samples were also collected from two streams in MacRitchie catchment as listed in the same table. The results of FIB and HF183 analyses of the raw sewage and non-urban water samples are tabulated in Appendix A.

Table 4.2. Sampling locations of raw sewage and non-urban water samples

Sample name	Type of samples	Location address/description	Type of land use	Coordinates
RS1	raw sewage	manhole in front of Bedok Garden no. 22	low-density residential	103° 57' 8.7" E, 1° 19' 21.4" N
RS2	raw sewage	manhole (ID 272878) at Choa Chu Kang Avenue 4 Block (Blk) 441	high-density residential	NA
RS3	raw sewage	manhole at Choa Chu Kang Crescent Blk 692A	high-density residential	NA
RS4	raw sewage	manhole at Choa Chu Kang St. 54 Blk 790	high-density residential	NA
RS5	raw sewage	manhole at Choa Chu Kang Avenue 3 Blk 413	high-density residential	NA
PS1, PS3	non-urban water	canalised stream at MacRitchie catchment	non-urbanised and preserved	103° 48' 25.5" E, 1° 21' 13.9" N
PS2, PS4	non-urban water	stream that can be accessed from boardwalk near Petaling Hut at MacRitchie catchment	non-urbanised and preserved	103° 48' 16.7" E, 1° 21' 11.3" N

NA denotes data are not available.

4.3. Sampling procedure

In this study, sampling was conducted only under dry weather (baseflow) conditions in order to exclude rainfall runoff as a potential source of contamination. Dry-weather flow contributes a significant volume of the water discharged from a storm drainage system. A study in Sacramento, CA reported that slightly less than half the volume of water discharged from a storm drainage system was not directly attributable to runoff (Field et al. 1994). However, dry-weather flow is often contaminated by other entries such as sanitary wastewater, industrial and commercial pollutants, failing septic-tank systems and vehicle-maintenance activities. Therefore, field monitoring during dry weather that pin-points periods with greater potential for pollutant entries are likely to reveal contamination. In most urban areas, drainage system flows due to stormwater runoff will cease within a few to several hours following the storm event.

The samples were taken at hourly intervals following the Toolbox Study approach (see Section 3.2.2). This sampling method will reveal time variations of FIB concentrations and possible associations with sewage flows. Samples were collected over 24 hr (from 08:00 to 07:00 the next day or from 20:00 to 19:00 the next day) or over 12 hours (from 08:00 to 19:00 the same day). Samples were collected using two Avalanche[®] refrigerated auto-samplers (Teledyne Isco, Lincoln, NE, USA) (Figure 4.3). The auto-samplers had two types of bottle configurations: 4-bottle (5-L each) and 14-bottle (950-mL each). The samplers were powered by deep-cycle-marine batteries and designed to stay refrigerated at 4°C for 48 hr. This provided ample time for sample storage prior to collection. A HOBO[®] U10 temperature data logger (Onset Computer Corporation, Cape Cod, MA, USA) was installed in the refrigeration compartment of the auto-sampler to record the temperature every 5 minutes. An example plot of refrigeration compartment temperature vs. time during a 14-hr-and-25-min storage time is shown in Figure 4.4. This method of sample storage is considered adequate especially as McCarthy et al. (2008) reported that *E. coli* samples could remain in non-refrigerated auto-samplers for up to 24 hr.

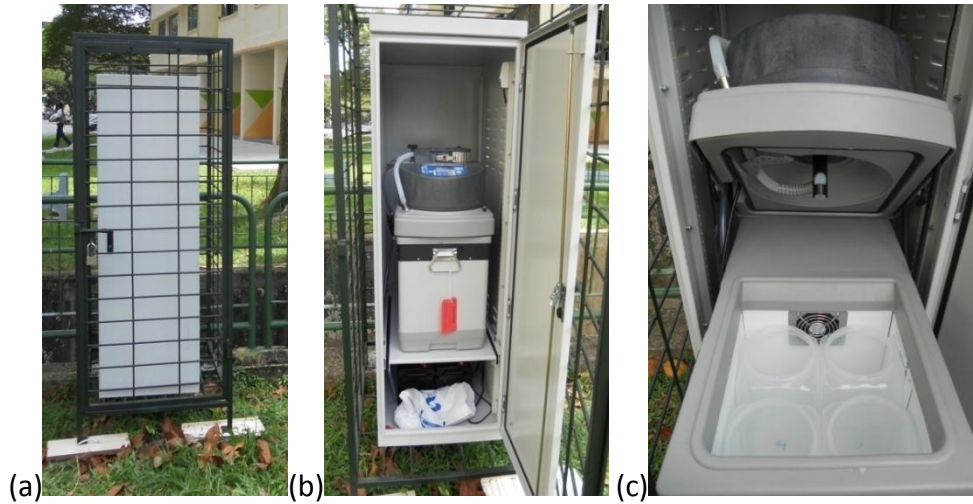


Figure 4.3. Refrigerated auto-sampler: (a) secured with metal enclosure and locked cage which was anchored to the ground (b) Avalanche® sampler (c) refrigeration compartment with 4-bottle (5-L each) configuration

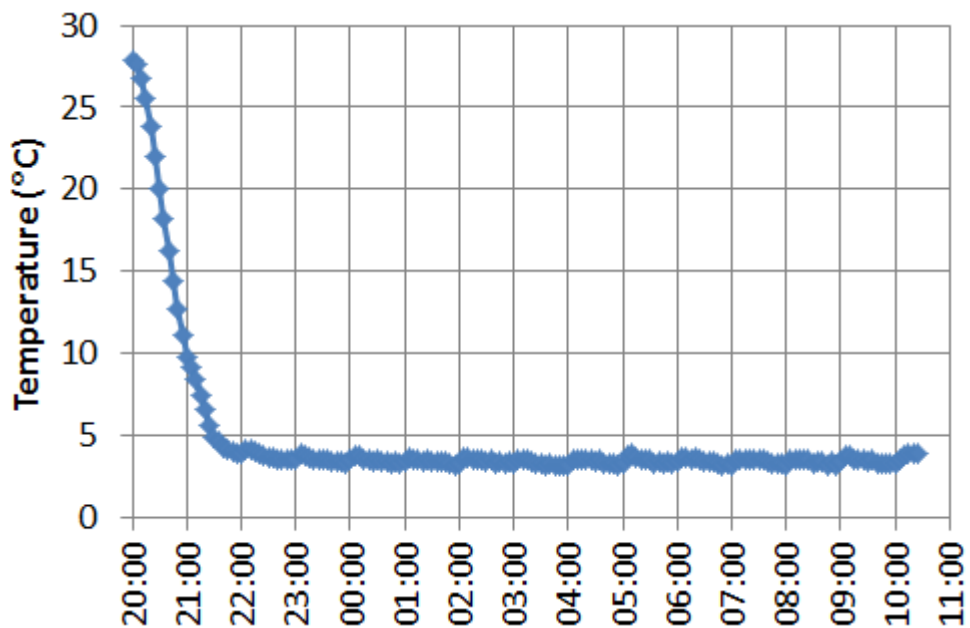


Figure 4.4. Refrigeration compartment temperature vs. time for 14-bottle sampling event from 16 July 2012, 20:00 (first sample taken) to 17 July 2012, 10:25 (all samples retrieved from the auto-sampler)

One potential disadvantage of the automated systems is that the tubing and sampling system (e.g. strainers) are not sterilised between samples, so there is a possibility that sample-to-sample cross-contamination might occur. However, according to Grant et

al. (2001), the purge step in between sampling events is able to prevent cross-contamination in automated sampling. In addition, carryover from sampler lines was tested in a laboratory by Solo-Gabriele et al. (2000) who showed that carryover between sequential river and sterile water samples was only 0.5% (4 MPN/100mL *E. coli* carryover from 800 MPN/100mL average *E. coli* in the samples) with two rinses of the sampler line. This carryover quantity is small relative to the variability of the analytical procedure, and thus carryover from one sample to the other was considered insignificant (Solo-Gabriele et al. 2000). In this study, the samplers were programmed to perform either one automated rinse cycle (purge – rinse – purge – take sample – purge) or two automated rinse cycles (purge – rinse – purge – rinse – purge – take sample – purge) between sample collections. Cross-contamination was tested in the field by first going through the rinse cycle and then sucking environmental water followed by going through the rinse cycle again and then sucking sterile water (either ultrapure or autoclaved tap water) through the strainer and sampler hose on three occasions for one rinse cycle and three occasions for two rinse cycles. For both types of samples (environmental water and sterile water) collected, sterile water was used for rinsing. Although the *E. coli* carryover percentages of 1.1% and 1.9% for the one rinse and two rinse configurations are higher than the 0.5% *E. coli* carryover obtained by Solo-Gabriele et al. (2000), this result is considered acceptable since it remains within the expected variability of the test method.

Table 4.3. Auto-sampler cross-contamination test results

Sampling date and time		% carryover		
Environmental water	Sterile water	TC	<i>E. coli</i>	Enterococci
One rinse cycle				
28 Feb 2012 12:40	28 Feb 2012 12:45	0.8	0.3	1.3
28 Feb 2012 12:50	28 Feb 2012 12:55	1.6	2.8	0.3
28 Feb 2012 13:00	28 Feb 2012 13:05	0.3	0.3	1.4
Average		0.9	1.1	1.0
Two rinse cycles				
17 Apr 2012 11:00	17 Apr 2012 11:10	0.0	0.1	0.1
16 Jul 2012 10:50	16 Jul 2012 10:55	0.9	1.4	9.5
24 Sep 2012 11:20	24 Sep 2012 11:30	9.6	4.3	2.9
Average		3.5	1.9	4.2

Twelve-hr samples (from 08:00 to 19:00) were also collected manually either by sampling directly in the drains, when the sampling point was an open drain, or by using a Nasco Model B01367WA extendable sampling pole (Nasco, Fort Atkinson, WI, USA) when the sampling point was a closed drain, in which entry was prohibited by regulation. The sampling pole was fitted with sterile 532-mL Whirl-Pak® bags (Nasco, Fort Atkinson, WI, USA) (Figure 4.5) following the manufacturer's instructions (Nasco 2010). Sampling was conducted with care so as not to disturb the bottom sediment. All manually collected samples were kept in ice coolers from the time of collection until delivery to the laboratories.



Figure 4.5. The extendable Nasco sampling pole used together with Whirl-Pak® bag

The procedure for manual sampling varied slightly during the course of this study. During most of the study, the total volume of each collected sample was composited in a container before the sample was separated for the different laboratory analyses. The container used to composite the samples was first rinsed with drain water immediately before sample collection. Container cross-contamination was checked on six occasions in the field. The results are tabulated in Table 4.4 and show that FIB carryover percentages are very low. For the manual sampling campaign before May 2012, the samples were not composited but were instead collected in separate

containers straightaway. These separate samples were all collected within a few minutes. For auto-sampling, different samples used different auto-sampler bottles and therefore each sample was composited in respective auto-sampler bottle. The auto-sampler bottles were cleaned with Decon®90 (Decon Laboratories Limited, East Sussex, UK) which is phosphate-free and able to reduce gram negative (*Pseudomonas aeruginosa*) and positive bacteria (*Staphylococcus aureus*) by greater than five orders of magnitude (Decon undated).

Table 4.4. Manual sampling’s container cross-contamination test results

Sampling date and time		% carryover		
Environmental water	Sterile water	TC	<i>E. coli</i>	Enterococci
10 May 2012 8:00	10 May 2012 8:50	0.0	0.0	0.0
14 May 2012 8:00	14 May 2012 8:30	0.1	0.0	0.0
21 May 2012 14:00	21 May 2012 14:27	0.0	0.0	0.0
22 May 2012 9:00	22 May 2012 9:30	0.0	0.0	0.0
2 Jul 2012 14:00	2 Jul 2012 14:40	0.2	0.4	0.0
16 Jul 2012 16:00	16 Jul 2012 17:00	0.1	0.0	0.0
Average		0.1	0.1	0.0

A total of 5 L of water sample was collected every hour for samples collected from 08:00 to 19:00. One-quarter liter of the sample was used for the analysis of chloride, bromide and boron. Two liters of the samples were used for the analysis of triclosan, faecal sterols, plasticizers, caffeine, pharmaceutical compounds and artificial sweeteners. One-half liter of the sample was analysed for orthophosphate and detergent as MBAS. One liter of the sample was kept for duplicate analysis of chemical parameters. Lastly, about 0.25 L of the sample were kept in three sterile Whirl-Pak® bags for FIB analysis in the NTU Environment Lab and 1 L was filtered through Millipore Sterivex filter cartridges. Section 4.4.3.1 describes the filtration process in greater detail. The filtered samples were used for subsequent HF183 quantitation. A total of 950 mL of water was collected every hour for samples collected overnight, from 20:00 to 07:00. These samples were analysed only for FIB and HF183.

Physical parameters including temperature, specific conductance and salinity were measured in the field using a conductivity meter, YSI Model 30 digital meter (YSI,

Inc., Yellow Springs, OH, USA), following manufacturer's instructions (YSI 2007) and recorded on field sheets, where other field observations are also noted. The conductivity meter was calibrated according to manufacturer's instructions (YSI2007) using a 1413 μS @ 25 °C conductivity standard (Eutech Instruments Pte Ltd, Singapore) to calibrate to freshwater conductivity range. Blank samples were poured into sterile 532-mL Whirl-Pak® bags and transported together with the bacteria samples in order to monitor if there was any significant cross-contamination during sample handling or transportation in the ice cooler. Field blanks were mostly negative for total coliform, *E. coli* and enterococci (Appendix B). Field blanks that were positive had extremely low FIB concentrations compared to minimum FIB concentrations of drain water samples presented in Table 5.1. This indicated that there was only an insignificant degree of contamination during sample handling or transportation in the ice cooler. All samples were kept in ice coolers from right after collection and during transportation to the laboratories. In the laboratories, the samples were refrigerated at 4°C and preserved according to the test methods' requirements.

4.4. Laboratory analysis procedures

4.4.1. FIB analysis procedure

Samples were analysed upon receipt in the NTU Environment Lab not more than 38 hr after sampling time. Although regulatory guidelines often specify shorter holding times (Eaton et al. 2005), the maximum 38-hr holding time is not believed to have compromised measurement reliability based on the experience of previous researchers. According to Pope et al. (2003), there was no significant difference in *E. coli* densities measured with the MPN method between time 0 and 48 hr. Other studies required long holding times when water samples were required to be shipped overnight on ice to the laboratory (Duris et al. 2013) or stored overnight in refrigerators and air-transported the following day to the laboratory (Davies-Colley et al. 2004) without any apparent compromise in accuracy. FIB are mesophilic bacteria and with proper preservation (refrigerated at 4°C), it is expected that FIB will be below their minimum growth temperature (8°C for *E. coli*), at which internal transport

processes are so slow that growth cannot occur (Madigan et al. 2011). In this study, the first samples of the 24-hr FIB sampling results typically had the longest holding time (time between sample collection and the start of FIB analysis) compared to the other 23 samples in the series. From the FIB results observation, there is no tendency that these first samples were consistently higher or lower compared to later samples. In addition, the overall 24-hr FIB sampling results did not show any decreasing or increasing trend.

Indicator bacteria samples were analysed using products from IDEXX (IDEXX Laboratories, Westbrook, ME, USA) as discussed in Chapter 2. They were Colilert® (reagent for total coliform and *E. coli* enumeration) which complies with Standard Methods (SM) 9223B (Eaton et al. 2005), Enterolert® (reagent for enterococci enumeration) which is an approved method (USEPA 2005) and Quanti-Tray®/2000. The use of Quanti-Tray®/2000 produced MPN results which have high precision because 49 large and 48 small wells are used in every tray (Davies-Colley et al. 2004). The Quanti-Tray®/2000 enumerates from one to 2,420 MPN/100 mL without dilution. However, based on prior experience with bacterial samples in Singapore (Dixon et al. 2009), concentrations were expected to exceed 2,420 MPN/100 mL and dilutions were necessary.

The planning of the dilution series involves the following considerations (Cochran 1950). The first is the range of reading values to be covered. The second is the volume of samples to be taken during the sampling activity in order to meet the sample volume requirement for the dilutions. In this study, the samples were diluted by three different dilution ratios (ratios between the final volume and the aliquot samples). The samples from the first day of sampling at each location are analysed with dilution ratios of 1, 100, and 10,000. If the FIB results can be predicted to be in a certain range from previous results, dilution ratios were adjusted. For example: higher dilution ratios (10, 1,000, and 100,000) were used to capture higher FIB ranges; two (1 and 100) instead of three dilution ratios were used for non-urban water samples since they were not expected to be in the range of dilution ratio 10,000.

All samples were vigorously shaken before analyses. A 100-mL sample of each dilution was prepared in a washed and autoclaved glass bottle. For dilution ratio 1, each 100-mL sample was measured in a sterile 100-mL graduated cylinder. For dilution ratio 100, 1 mL of sample and 99 mL of ultrapure or Milli-Q water (Millipore, Billerica, MA, USA) were mixed. The sample was measured using an Eppendorf Research® pipette (Eppendorf AG, Hamburg, Germany). A new sterile pipette tip was used for each sample and for each dilution. The ultrapure water was measured using a sterile 100-mL graduated cylinder. For dilution ratio 10,000, 1 mL of the 1:100 dilution and 99 mL of ultrapure water were mixed.

After the different sample dilutions were prepared, the samples were prepared for bacterial incubation. For each 100-mL dilution sample, one aliquot of either Colilert or Enterolert reagent was added. The bottle was shaken until the Colilert or Enterolert reagent dissolved. The mixture was then poured into a labeled Quanti-Tray®/2000. The tray was then sealed with an IDEXX Quanti-Tray® Sealer in a Quanti-Tray®/2000 Rubber Insert. The trays were incubated at 35°C ± 0.5°C (for the total coliform/*E. coli* trays) or at 41°C ± 0.5°C (for enterococci trays) for 24-28 hr following the manufacturer's instructions (IDEXX 2007; IDEXX 2008; IDEXX undated).

After 24 hr, the trays were read for positive large and small wells (IDEXX 2007; IDEXX 2008). For the samples with Colilert, wells that are yellow in colour are positive for total coliform and wells yellow in colour that also fluoresce under 365-nm UV light are positive for *E. coli*. For the samples with Enterolert, wells that fluoresce under 365-nm UV light are positive for enterococci. After counting the cells, the results and confidence limits were interpreted using the MPN tables provided by the manufacturer (IDEXX undated). For diluted samples, the MPN was scaled up by the dilution ratios (IDEXX 2007; IDEXX 2008).

There may be different values from different dilution ratios for the same sample because the distribution of organisms within the sample may be not uniform even though the sample was vigorously shaken. In fact, the distribution of organisms may

be far from homogeneous (Cochran 1950). The analysis determines the density of that part of the sample from which the dilution is made, which may be very different from the average density over the whole sample. This source of error can be more important than any error in the dilution method itself. Hence, when more than one dilution ratio gives readings that are not all fertile or all sterile, the reading with lower uncertainty (discussed in Section 4.5.4) was selected. Sixteen out of 895 samples (1.8%) exceeded the upper range of total coliform ($>24,196,000$ MPN/100mL after 10,000-fold dilution) and were replaced with the highest reading (24,196,000 MPN/100mL). For TC, the TC concentration for one sample was replaced with 241,960 MPN/100mL because the 100-fold dilution gave $>241,960$ MPN/100mL but 10,000-fold dilution gave $<10,000$ MPN/100mL.

4.4.2. Chemical analysis procedure

The analyses for the chemical parameters were contracted out to an appointed accredited laboratory. The laboratory was in charge of processing the samples. Most of the chemical parameters were analysed using methods that are available in the Standard Methods (Eaton et al. 2005). Caffeine, pharmaceuticals and triclosan were analysed using the liquid chromatography-tandem mass spectrometry (LCMSMS) method in accordance with EPA Method 1694 (USEPA 2007). Artificial sweeteners were also analysed using the LCMSMS method as per the appointed laboratory's in-house method which was developed following the method described in Scheurer et al. (2009). The laboratory analytical methods used, limit of detection (LOD) and uncertainty for each chemical compound are summarised in Table 4.5.

Table 4.5. List of parameters analysed by the appointed accredited laboratory with the test method, LOD and uncertainty

No	Parameter	Complete chemical name	Standard	Test method	Technique	LOD	Uncertainty (Jan 2011)	Uncertainty (Jun 2011 onwards)
1	Chloride, Cl	Chloride	NaCl	SM 4110B	IC	0.1 mg/L	± 6%	5.7%
2	Bromide, Br	Bromide	NaBr	SM 4500-Br (D)	FIA	0.5 mg/L	± 6%	NA
3	Boron, B	Boron	H ₃ BO ₃	SM 3120B	ICP-OES	1.5 µg/L	± 4%	3.5%
4	Orthophosphate, PO ₄ -P	Phosphate	KH ₂ PO ₄	SM 4500-P (G)	FIA	3 µg/L	± 15%	16.3%
5	Detergent as MBAS	Methylene blue active substances	Sodium lauryl sulfate	SM 5540C	UV	50 µg/L	± 5%	5.0%
6	Triclosan	2,4,4'-trichloro-2'-hydroxydiphenyl ether	Triclosan	EPA 1694	LCMSMS	0.1 µg/L	± 10-20%	NA
7	Cholesterol	Cholest-5-en-3β-ol	Cholesterol	Der-GCMS (USEPA 8270C)	GCMS	0.1 µg/L	± 10-20%	26.6%
8	Cholestanol	5α-cholestan-3β-ol	Cholestanol	Der-GCMS (USEPA 8270C)	GCMS	0.1 µg/L	± 10-20%	25.2%
9	Coprostanol	5β-cholestan-3β-ol	Coprostanol	Der-GCMS (USEPA 8270C)	GCMS	0.05 µg/L	± 10-20%	21.7%
10	DBP	Di-n-butyl phthalate	DBP	USEPA 8270C	GCMS	2 µg/L	± 10-20%	34.9%
11	DEHP	Bis(2-ethylhexyl)phthalate	DEPH	USEPA 8270C	GCMS	2 µg/L	± 10-20%	34.9%
12	Caffeine	3,7-dihydro-1,3,7-trimethyl-1H-purine-2,6-dione	Caffeine	EPA 1694	LCMSMS	0.01 µg/L	± 10-20%	7.2%
13	Acetaminophen or paracetamol	4-acetamidophenol	4-acetamidophenol	EPA 1694	LCMSMS	0.01 µg/L	± 10-20%	8.4%
14	Ibuprofen	2-(4-isobutylphenyl)propanoic acid	Ibuprofen	EPA 1694	LCMSMS	0.01 µg/L	± 10-20%	19.6%
15	Diclofenac	2-(2,6-dichloranilino)phenyl acetic acid	Diclofenac sodium salt	EPA 1694	LCMSMS	0.01 µg/L	± 10-20%	13.4%
16	Acesulfame-k	1,2,3-oxathiazin-4(3H)-one,6-methyl-2,2-dioxide, potassium salt	Acesulfame-k	Developed from Scheurer et al. (2009)	LCMSMS	0.01 µg/L	± 10-20%	14.2%
17	Sucralose	1,6-dichloro-1,6-dideoxy-β-D-fructofuranosyl-4-chloro-4-deoxy-α-D-galactopyranoxide	Sucralose	Developed from Scheurer et al. (2009)	LCMSMS	0.01 µg/L	± 10-20%	25.1%
18	Saccharin	1,2-benzisothiazol-3(2H)-one,1,1-dioxide	Saccharin	Developed from Scheurer et al. (2009)	LCMSMS	0.01 µg/L	± 10-20%	12.2%

1 mg/L = 1 ppm (part per million). 1 µg/L = 1 ppb (part per billion). NA denotes the parameters were not analysed.

Chloride was measured with the ion chromatography or IC (SM 4110B) method that is applicable to surface and wastewaters (Eaton et al. 2005). Bromide was measured with flow injection analysis (FIA) using the conditions set in SM 4500-Br (D) instead of the IC (SM 4110B) method because the water samples had high concentrations of chloride. If the IC method was used, the water samples would have had to be diluted which would have resulted in less sensitive detection (about twice of the detection limit of the FIA method) (Pik H. Wong, Setsco Services Pte Ltd, personal communication, 2011). Orthophosphate was measured with FIA using the conditions set in SM 4500-P (G) (Eaton et al. 2005). Boron was analysed with inductively coupled plasma (ICP) according to SM3120B which utilises very high temperature to completely dissociate molecules and hence results in significant reduction of chemical interferences and a low detection limit (Eaton et al. 2005). The analysis also involved the use of optical emission spectroscopy (OES) as quality control through examination of the spectral background surrounding the emission lines used for metal (i.e. boron) determination (Eaton et al. 2005). Detergent was measured with the methylene-blue active substances (MBAS) method. This is a non-specific method that measures anionic surfactants using LAS as a standard (Eaton et al. 2005). LAS is presumed to be the dominant surfactant in domestic wastewater.

Cholesterol, cholestanol, and coprostanol are semivolatile organic compounds that were derivatized and then run by gas chromatography-mass spectrometry (GCMS) following USEPA Method 8270C. DBP and DEHP were also analysed with GCMS using the USEPA Method 8270C. In GCMS, the sample is injected into a gas chromatograph (GC) column, which is temperature-programmed to separate the analytes, before detection by a mass spectrometer (MS) connected to the GC. Cholesterol, cholestanol, coprostanol, DBP and DEHP were analysed as total concentrations. Caffeine, acetaminophen, ibuprofen, diclofenac, acesulfame-k, sucralose, saccharin and triclosan were extracted using solid-phase extraction (SPE) under acidic conditions. The extracts were concentrated and analysed by LCMSMS method. Caffeine and acetaminophen were run in the positive electrospray ionisation (ESI+) mode, while the rest were run in the negative electrospray ionisation (ESI-) mode.

Caffeine, acetaminophen, ibuprofen, diclofenac, acesulfame-k, sucralose, saccharin and triclosan were analysed as dissolved concentrations.

For these chemical parameters, due to finite resources, not all samples were analysed. The samples to be analysed were chosen based on the bacteria counting results, i.e. to be roughly coincident with the high and low bacteria counts. Duplicates were analysed to assess the precision of the analytical process. The number of duplicates and the coefficient of variation are reported in Chapter 6. Lab blanks were analysed for every run of chemical analysis to ensure no contamination had occurred. All blanks were below detection. Samples with non-detect values were replaced with half of respective LOD for data analysis.

4.4.3. HF183 quantitative method

4.4.3.1. Environmental water sample filtration

The 1-L water sample collected for HF183 quantitation was filtered through an EMD Millipore Sterivex™-GS 0.22 µm filter unit (EMD Millipore Corporation, Billerica, MA, USA) by one of the following two methods. The first method was using a sterile syringe immediately after sample collection. The second method was pumping through sterilised silicon tubing (sterilised with 10% bleach and rinsed with tap water) with a peristaltic pump (Cole-Parmer, Vernon Hills, IL, USA), within at most 62 hr from sample collection and generally within 48 hr. The first method was employed on-site when manual sampling was conducted, while the second method was employed in the NTU Hydraulic Lab when sampling was done with the auto-sampler. The second method was conducted in the lab because auto-sampling resulted in the collection of many samples at once and therefore it was more practical to filter the samples in the lab using the peristaltic pump which enabled filtration of four samples at once. There is a holding time of up to 62 hr with the second method because the time after sample collection was used for FIB analysis, in which case samples were preserved at 4 °C prior to analysis. Seurinck et al. (2005a) indicated that human-specific *Bacteroides* marker did not decrease significantly throughout 24 d when the sample

was kept at 4 °C in their study. The volume filtered was between 100-1000 mL depending on how much water could pass until the filter clogged. The volume filtered was recorded, after which the filters were stored at -80 °C until DNA extraction.

4.4.3.2. DNA extraction

DNA was extracted from the filters using PowerPlant® Pro DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA) according to the manufacturer's protocols. Environmental DNA samples were kept on ice during the extraction procedures and the extracted DNA samples were stored at -20 °C (Nshimiyimana et al. 2014).

4.4.3.3. QPCR

Analysis by quantitative polymerase chain reaction (QPCR) followed a methodology developed by Nshimiyimana et al. (2014) for DNA water samples from Singapore. The HF183 marker was amplified using the human-specific HF183F primer and Bac242R primer (Seurinck et al. 2005a) and quantified by QPCR using LightCycler® 480 Real-Time PCR system and software v. 1.5.0 (Roche Applied Sciences, Indianapolis, IN, USA). QPCR reaction mixtures consisted of 10 µL of KAPA SYBR® FAST2X Master Mix (KAPABIOSYSTEMS, Woburn, MA, USA), 10 µM of each primer, and 1 µL of DNA template. Each sample was analysed in duplicate or triplicate.

Amplification followed the manufacturer's instructions: reactions were subjected to a pre-incubation (denaturation) step of 95 °C for 3 min, followed by 45 - 50 thermal cycles consisting of 95 °C for 10 s, 53 °C for 20 s and 72 °C for 1 s. QPCR crossing point (Cp) values were examined to verify consistency across duplicate or triplicate analyses. All QPCR experiments contained positive controls (standards) and at least a duplicate of one negative control. Tenfold serial dilutions of plasmid DNA containing the HF183 marker were used to generate a standard curve of Cp values versus target DNA concentration using a least-square fit for each QPCR run (Nshimiyimana et al. 2014). Quantification of the HF183 marker by QPCR was linear over the range of 10^1 to 10^8 copies/QPCR ($R^2 \geq 0.98$).

a. Melting temperature

Melting temperature analysis was done to ensure the specificity of the QPCR assay (Seurinck et al. 2005a), especially for samples that resulted in high copies/QPCR values. Table 4.6 summarises the melting temperatures (T_m) of this study and compares them with other studies. Melting temperatures of sample amplicons were confirmed to be within 2 standard deviations (s.d.) of the mean T_m associated with QPCR standards at concentrations of 10^1 - 10^6 copies/QPCR ($79.7\text{ }^\circ\text{C} \pm \text{s.d. } 1.0$) while late-stage PCR artefacts were associated with standards at concentrations $>10^6$ copies/QPCR ($T_m = 80.8\text{ }^\circ\text{C} \pm \text{s.d. } 2.2$). The melting temperatures were close to those found by Nshimyimana et al. (2014) and also Seurinck et al. (2005a) for concentrations of 10^1 - 10^6 copies/QPCR. The melting temperatures of QPCR analyses in February 2012 were consistently in the range of $74.1\text{ }^\circ\text{C} \pm \text{s.d. } 0.3$ for concentrations of 10^1 - 10^6 copies/QPCR and $74.8\text{ }^\circ\text{C} \pm \text{s.d. } 1.7$ for concentration $>10^6$ copies/QPCR. It is possible that the heating system of the QPCR did not reach the melting temperature; however, this is not expected to adversely affect the results. All HF values are used in the data analysis in Chapter 7.

Table 4.6. Summary of melting temperatures and comparison with other studies

Standard curves	10^1 - 10^6 copies/QPCR		$>10^6$ copies/QPCR	
	Mean T_m ($^\circ\text{C}$)	s.d.	Mean T_m ($^\circ\text{C}$)	s.d.
This study: 48 standard curves	79.7	1.0	80.8	2.2
Nshimyimana et al.(2014)	78.9	0.2	80.0	1.9
Seurinck et al. (2005a)	78.4	0.2	NA	NA
This study: 5 standard curves in February 2012	74.1	0.3	74.8	1.7

NA denotes not available.

b. Limit of detection

The limit of detection (LOD) is defined as the lowest concentration of the HF183 marker detectable with real-time PCR in freshwater compared with background fluorescence. The LOD was defined because the amplification and detection of lower levels of HF183 marker were not reproducible (Seurinck et al. 2005a). The LOD for each 96-well plate was calculated as the 99% confidence interval of the C_p values of

sample blanks (i.e. negative controls) or as the Cp value at 45 cycles if no signal was apparent.

Table 4.7. Summary of LOD and samples below LOD

LOD range in copies/QPCR	Number of plates	Number of samples < plate LOD	Number of samples < study-wide LOD (150 GE/100mL)
0 - 10	24	5	5
>10 - 20	3	0	0
>20 - 50	15	13	1
>50 - 100	5	6	0
>100 - 500	6	3	0
Total	53	27	6

Table 4.7 summarises the LODs of QPCR analyses in this study and the number of samples below LOD. Twenty-seven out of 644 samples (4%) were below each plate’s LOD. More than half of the QPCR analyses (24 out of 53 plates) had detection limits ranging from 0 to 10 copies/QPCR. After adjusting for sample volumes, these correspond to a conservative detection limit of 150 genome equivalents per 100 mL (GE/100mL). Therefore, 150 GE/100mL was set as the study-wide LOD for subsequent statistical analyses, in agreement with Nshimiyimana et al. (2014). QPCR results that were below the study-wide LOD were treated as non-detected (ND) values for plotting and replaced with the study-wide LOD for quantitative data analysis. One non-urban water sample which was below 150 GE/100mL and had plate LODs > 20-50 copies/QPCR was a reference sample and not used in the quantitative data analysis; therefore its value was not replaced with 150 GE/100mL. QPCR results that were below each plate’s LOD but above the study-wide LOD were kept as the calculated values.

c. Inhibition analysis

The impact of inhibition on QPCR was investigated by spiking 10^5 copies of the positive control plasmid bearing the HF marker (pHF183) into an aliquot from each sample before QPCR amplification and comparing the results measured by QPCR with and without addition of the spike (Nshimiyimana et al. 2014). A sample was

considered to be significantly inhibited if the spiked sample was quantified as having less than 65% of the added amount of HF marker. This corresponds to both the 95% confidence interval for quantification of the QPCR calibration curve and observed variability between technical replicates. Samples with inhibition were diluted tenfold and re-quantified by QPCR. About 11% of the samples (70 out of 644 samples) required additional dilution to remove inhibition.

For these HF183 data, the author completed all environmental water sample collection and filtration in Singapore, about one-fifth of the DNA extraction, QPCR and inhibition laboratory analyses during her one-semester research residency in MIT and all data pre-processing and analysis described in this thesis. The remaining DNA extraction, QPCR and inhibition laboratory analyses were conducted by Jean Pierre Nshimiyimana when he was engaged as a research associate for this project and based at MIT.

d. Estimation of HF183 marker genome equivalents

To convert from QPCR-detected HF183 marker copies to genome equivalents (GE), the efficiency of DNA extraction was determined. Then, genome equivalents of the HF183 marker in environmental samples could be calculated using Equation 4.1 (Nshimiyimana et al. 2014).

$$\frac{GE}{mL} = \frac{C_{HF} \times V_{Elute}}{V_{Template} \times V_{Sample} \times F_{Elute} \times E} \quad (4.1)$$

where C_{HF} is the number of HF183 marker copies detected by the QPCR (copies/QPCR), V_{Elute} (μ l) is the volume of buffer in which DNA extracts are suspended following purification, $V_{Template}$ (μ l) is the volume of DNA extract added to the QPCR, V_{Sample} (μ l) is the volume of environmental sample subjected to DNA extraction, F_{Elute} (= 0.75) is the fraction of sample DNA that is eluted, accounting for known volumetric losses in the extraction protocol and E is the efficiency of HF183 marker recovery from DNA extraction procedure.

4.5. Data analysis methods

Graphical and statistical analyses were carried out on the data obtained. Bacterial indicator data were plotted using Microsoft Excel (Microsoft, Redmond, WA, USA) to examine qualitatively the diurnal variation of the bacterial indicators. Statistical analyses were performed using Microsoft Excel, Matlab (The MathWorks, Inc., Natick, MA, USA) and JMP Pro v. 10 (SAS Institute, Inc., Cary, NC, USA). A variety of graphical and statistical techniques were used to analyse the data, and these techniques are described in this section.

4.5.1. Lognormal distribution

The distribution of the concentrations of particles, chemicals, and organisms in the environment generally appears to be right-skewed (Wymer and Wade 2007). This includes microbiological densities whether measured as CFU from plate counts, MPN estimates, and QPCR GE. Right-skewed distributions exhibit a long right-side tail and is common when a variable is bounded on the left (for example no value less than zero) but unbounded on the right. Logarithmic transformation of right-skewed data tends to result in a more nearly symmetric or normal distribution which may justify statistical analysis of the log-transformed data using the assumption of normality (Wymer and Wade 2007). Such a dataset is said to be lognormally (\log_{10} transformation) distributed.

The lognormal distribution has been used for environmental data in many studies. For example, Cheung et al. (1991) show that the frequency distributions of *E. coli* and staphylococci counts in 667 samples collected from nine beaches in Hong Kong were lognormal. Gannon and Busse (1989) found that a lognormal distribution best describes the variation of bacterial indicator organisms (faecal coliforms, *E. coli*, faecal streptococci, enterococci), and hence used log-transformed values in all tests for statistical significance. Pipes et al. (1977) found that a series of samples collected from a flow of water at 5-min intervals over a 5-hr period showed the same type of variability as most of the data sets collected over one-month periods. They fit lognormal better than other types of distributions.

The suitability of the lognormal distribution for the data in this study is shown with histogram and probability plots created with Matlab. Both are visualisation methods to identify the underlying distribution of the data. In this study, the histogram was created by grouping the data into ten equally spaced containers or bins. The number (frequency) of data in each bin can be obtained from the histogram. Normal probability plots were created to graphically assess whether the data come from a normal distribution. If the data are normal the plot would be linear. Histograms and probability plots of original and log-transformed values were compared. The data is light-tailed if the right, upper end bends below the straight line and the left, lower end bends above the line. In light-tailed distributions, the extreme parts of the distribution spread out less relative to the width of the center than in normal distributions. This means the probability of observing a value far from the median is less than in normal distributions. In contrast, the samples are heavy-tailed if the right, upper end bends above the straight line and the left, lower end bends below the line. In heavy-tailed distributions, the extreme parts of the distribution spread out more than in normal distributions. In other words, the probability of observing a value far from the median is more than in normal distributions.

4.5.2. Geometric mean

Log-transformation leads to the use of geometric means for environmental data. The use of the geometric mean has been criticized because (i) the geometric mean may be biased towards a lower value and (ii) places less emphasis on large values (Parkhurst 1998). However, Wymer and Wade (2007) argue that not only does the expected value of the geometric mean depend on the sample size, N , but also the probability of exceedance (likelihood). Hence, although the expected value of the geometric mean declines with increasing sample size, the likelihood that the sample geometric mean will exceed the compliance limit increases.

Parkhurst (1998) supports the use of the arithmetic mean over the geometric mean as a better predictor of health effects. However, the assertion that the arithmetic mean is a better predictor was made in the context of mass balance. Nevertheless, health effects are predicted in the context of human exposure to contaminated

recreational waters, but there is no clear or defined relationship between health and the arithmetic mean (Wymer and Wade 2007). Instead, exposure-response models for freshwater (Dufour 1984) and marine (Cabelli 1983) beaches indicate a linear relationship between risk of illness and the geometric mean of indicator densities (i.e. *E. coli*, enterococci). In addition, Wade et al. (2006) showed a linear relationship between risk of illness and the geometric mean of *Enterococcus* QPCR cell equivalents (CE). Hence, the geometric mean is the more protective of public health (Wymer and Wade 2007).

The log of the geometric mean is actually the arithmetic mean of logarithmic values or mean log (Equation 4.3). It is often used instead of the geometric mean itself. Mean log can be translated into geometric means without loss in meaning because of its direct correspondence (i.e. antilog of mean log is the geometric mean) (Wymer and Wade 2007).

$$\text{geometric mean} = \sqrt[N]{x_1 x_2 \cdots x_N} \quad (4.2)$$

$$\log(\text{geometric mean}) = \frac{1}{N}(\log x_1 + \log x_2 + \cdots + \log x_N) \quad (4.3)$$

where N is the number of sample and x is the value.

4.5.3. Box-and-whisker plots

Box-and-whisker plots were constructed with Matlab. The box (Figure 4.6) indicates the median, 25th percentile and 75th percentile values. The maximum whisker length is 1.5 times the interquartile range or IQR (the difference between 75th and 25th percentiles). Values beyond the whiskers are considered as outliers. The arithmetic mean or geometric mean is superimposed with a black circle sign (●). The arithmetic mean or geometric mean can be compared with the median in the box plots to assess normality. When the arithmetic mean or geometric mean falls far from the median, the data are skewed. If the arithmetic mean or geometric mean is above the median, the data are right-skewed. If the arithmetic mean or geometric mean is below the

median, the data are left-skewed. The number of samples analysed is indicated by the number above the box plot.

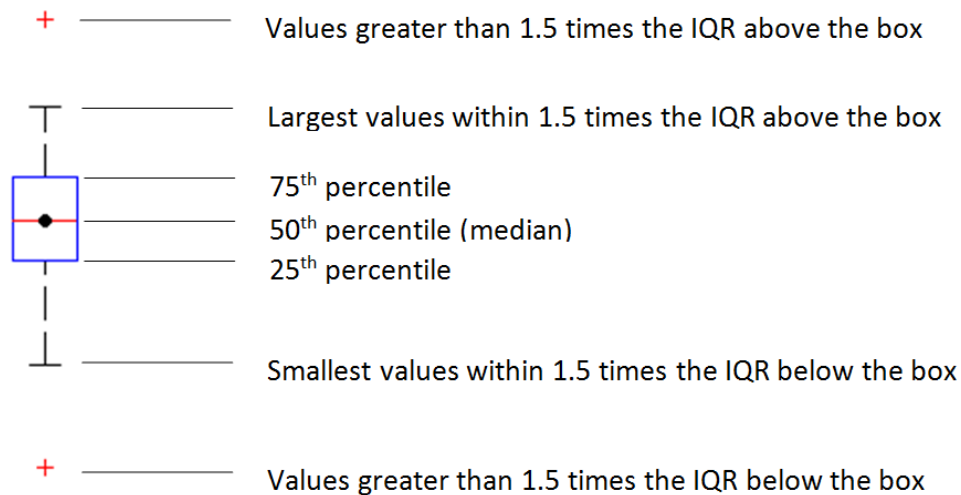


Figure 4.6. Typical box-and-whisker plots used in this study

4.5.4. Representation of error

Error bars are often used in graphical depictions of measured data to convey a sense of the range of variability or error within the measurements (Cumming et al. 2007). Error bars may give information describing the data, i.e. descriptive statistics like range and standard deviation, or information about inferences, i.e. inferential statistics such as standard error and confidence interval. Since error bars are of two distinct types, the type of error bar used should be stated explicitly in the graph or supporting text. Moreover, the value of N (sample size or number of independently performed experiments—and not number of replicates) must be stated regardless of the type of error bar chosen.

Range and standard deviation (SD) show how the data are spread. SD is the degree to which individuals within the sample differ from the sample mean (Cumming et al. 2007). Unlike range, SD will be unaffected by sample size. About two thirds of the data points will lie within $\pm 1 SD$ about the mean and about 95% of the data points will be within $\pm 2 SD$ about the mean. The standard error ($SE = SD/\sqrt{N}$) varies inversely with the square root of N (Cumming et al. 2007). The SE bar is shorter when

the experiment is repeated more often or when more samples are measured. The confidence interval (CI) is the $mean \pm t_{(N-1)} \times SE$, where $t_{(N-1)}$ is a critical value from tables of the t statistic. The critical value varies with N . For $N \geq 10$, the 95% CI (95% confident that the true population mean, μ , lies within the 95% CI) can be approximated as $\mu \pm 2 SE$. Hence, CI can be thought of as an SE bar that has been adjusted so as to be interpreted the same way regardless of N .

Errors of FIB concentrations based on the MPN method were calculated based on propagation of errors (Peters et al. 1974; Harris 2007) due to the dilution (pipette and measuring cylinder) and the 95% confidence interval of the MPN value. Information regarding the error of pipettes was provided by the manufacturer (Eppendorf AG, Hamburg, Germany) and is summarised in Table 4.8. Two types of pipettes were used to measure 1 mL or 10 mL depending on the dilution made. A conservative estimate of $\sigma_{pipette}$ that can be applied for all FIB results is 0.8%.

Table 4.8. Error of pipettes (provided by the manufacturer)

	Eppendorf Model Research® variable		
	100-1,000 µL	500-5,000 µL	
Volume setting (mL)	1	0.5	5
Systematic error (inaccuracy) (%)	-0.41	0.2	0.31
Random error max. (imprecision) (%)	0.12	0.6	0.15

While pipettes were used to measure the water sample to be diluted, 100-mL volumetric flasks or graduated cylinders were used to measure the sample for a dilution ratio of 1 and the diluent (ultrapure water) for dilution ratios greater than 1. The absolute uncertainty of the flask or cylinder is shown on the flask or cylinder surface and is tabulated in Table 4.9. A conservative estimate of $\sigma_{cylinder}$ is the highest relative uncertainty, i.e. $1/90 \times 100\% = 1.1\%$.

Table 4.9. Absolute uncertainty of 100-mL flask or cylinder

Type of 100-mL flask or cylinder	Absolute uncertainty (\pm mL)	Volume of sample or diluent measured (mL)		
Volumetric flask	0.1	100	-	-
Blue cylinder	0.5	100	99	90
Brown cylinder	1	100	99	90
White cylinder	1	100	99	90

Representation of error of the MPN value was taken from the 95% confidence limit given by the manufacturer (IDEXX Laboratories, Westbrook, ME, USA). The relative uncertainty of the confidence limit was the difference between the MPN value and its lower or upper confidence limit divided by the MPN value. This relative uncertainty was taken as σ_{conc} .

Analyses with a dilution ratio of 1 did not require the use of a pipette and hence only $\sigma_{cylinder}$ and σ_{conc} are considered in computing errors (Equation 4.4). Analyses of samples with dilution ratios of 10 and 100 entailed all sources of errors mentioned above (Equation 4.5) while those with dilution ratios of 1,000 and 10,000 include all sources of errors with double pipette and cylinder error owing to the need for a second serial dilution (Equation 4.6). Finally, analyses with a dilution ratio of 100,000 include triple pipette and cylinder error due to three serial dilutions (Equation 4.7).

$$\sigma_{error} = \sqrt{(\sigma_{cylinder})^2 + (\sigma_{conc})^2} \quad (4.4)$$

$$\sigma_{error} = \sqrt{(\sigma_{pipette})^2 + (\sigma_{cylinder})^2 + (\sigma_{conc})^2} \quad (4.5)$$

$$\sigma_{error} = \sqrt{2(\sigma_{pipette})^2 + 2(\sigma_{cylinder})^2 + (\sigma_{conc})^2} \quad (4.6)$$

$$\sigma_{error} = \sqrt{3(\sigma_{pipette})^2 + 3(\sigma_{cylinder})^2 + (\sigma_{conc})^2} \quad (4.7)$$

Errors of predicted target HF concentrations based on measured Cp values were calculated based on propagation of errors throughout the HF183 analysis process from sample filtration to the QPCR analysis. There are five sources of uncertainty

from sample filtration to QPCR analysis preparation. The first uncertainty is uncertainty of filtered sample volume or $\sigma_{V_{sample}}$ which is estimated to be 1%. The second uncertainty is uncertainty related to the use of half of the Sterivex filter in the DNA extraction process or $\sigma_{F_{DNA}}$ which is estimated to be 1%. The third uncertainty is uncertainty in pipetting a volume of DNA template in the DNA extraction process (50 μL) or $\sigma_{V_{DNA}}$ which is also estimated to be 1%. The fourth uncertainty is uncertainty in the efficiency of the DNA recovery from the Sterivex filter or $\sigma_{E_{DNA}}$ which was measured in the lab (24.4%). The last uncertainty is uncertainty in pipetting the volume of template per QPCR (1 μL) or $\sigma_{V_{QPCR}}$ which is estimated to be 1%. In addition, there is uncertainty in the standard (calibration) curve or σ_{SD} (%) which is derived from the uncertainty of the \log_{10} computed copies/ μL (abscissa of the standard curve). This uncertainty is calculated using the following formula (Harris 2007).

$$S_x = \frac{s_y}{|m|} \sqrt{\frac{1}{k} + \frac{1}{N} + \frac{(y-\bar{y})^2}{m^2 \sum(x_i - \bar{x})^2}} \quad (4.8)$$

where x is \log_{10} of computed copies/ μL , y is Cp, s_x is uncertainty in x , s_y is standard deviation of y , m is the value of the slope, k is the number of replicate measurements of the unknown, N is the number of data points for the calibration line, \bar{y} is the mean value of y for the points on the calibration line, x_i is the individual value of x for the points on the calibration line and \bar{x} is the mean value of x for the points on the calibration line.

The propagation of HF error is calculated using following formula.

$$\sigma_{error} = \sqrt{(\sigma_{V_{sample}})^2 + (\sigma_{F_{DNA}})^2 + (\sigma_{V_{DNA}})^2 + (\sigma_{E_{DNA}})^2 + (\sigma_{V_{QPCR}})^2 + \sigma_{SD}^2} \quad (4.9)$$

4.5.5. Correlation

Correlation reveals association but does not prove causation (Berthouex and Brown 2002). That is, observing that y increases when x increases does not mean that an increase in x causes the increase in y . Evidence of causation comes from knowledge of the underlying mechanistic behavior of the system. A causal relationship between two variables is found with regression. In correlation, the two variables stand equal (neither is the predictor nor the predicted) while in regression, the two variables of interest are assigned particular roles. One (x) is treated as the independent (predictor) variable and the other (y) is the dependent (predicted) variable. However, the coefficient of determination (R^2) of regression is equal to the square of the correlation coefficient between the two variables.

There are two types of correlations: parametric and nonparametric correlations. Pearson correlation is a parametric correlation which measures the strength of linear relationship between two variables (USEPA 2006). A linear association implies that as one variable increases, the other increases or decreases linearly. Pearson correlation is sensitive to the presence of one or two extreme values, especially when sample sizes are small (USEPA 2006). On the other hand, Spearman rank correlation and Kendall correlation are nonparametric correlation coefficients which measure all monotonic (nonlinear) relationships (Helsel and Hirsch 2002). Equation 4.10 gives the Pearson correlation coefficient (r), while equation 4.11 gives the Spearman rank correlation coefficient (r_s).

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (4.10)$$

$$r_s = \frac{\sum x_i^2 + \sum y_i^2 + \sum d_i^2}{2\sqrt{\sum x_i^2 \sum y_i^2}} \quad (4.11)$$

where r is the Pearson correlation coefficient, x_i and y_i are the values or ranks of the pair of samples, \bar{x} and \bar{y} are the sample means of x_i and y_i , r_s is the Spearman rank correlation coefficient, and d is the squared difference between x_i and y_i . As the name suggests, the Spearman rank correlation coefficient requires the variables to be

expressed in terms of rank order rather than in quantitative units. Hence, if the variables are numeric, they will be converted to ranks. Spearman rank correlation is similar to the signed-rank test in which data values ranked further apart are given more weight, whereas Kendall correlation is similar to the sign test in which all positive differences between data pairs are assigned a +1 without regard to the magnitude of the differences (Helsel and Hirsch 2002).

In this study, correlation analyses are carried out in order to establish trends between FIB and tracers. Matlab's `corr` function enables obtaining correlation coefficients for many pairs of parameters simultaneously. The Spearman rank correlation coefficient is used. The correlation analysis is carried out at the 5% level of significance. The null hypothesis is that the data are not correlated and this is a two-sided hypothesis test. Hence, when the p-values for the two-sided test are less than 10%, the null hypothesis that there is no correlation is rejected.

4.5.6. Time series and autocorrelation

Shumway and Stoffer (2011) define time series as a collection of random variables indexed according to the order in which they are obtained in time. Such variables could be measured continuously in time; however, because of restrictions inherent in the method of collection, continuous time series are approximated by discrete series at equally spaced points in time. Hence, adopting a sufficient sampling rate is important; otherwise important trends and processes over time could be missed or misrepresented.

The values of the random variables in the time-series are usually plotted on the vertical axis, or ordinate, with the time scale as the abscissa to reveal trend or periodicity, if any. The values at adjacent time periods are usually connected by lines to reconstruct visually some original hypothetical continuous time series. The obvious correlation introduced by sampling of adjacent points in time can severely restrict the applicability of many conventional statistical methods traditionally dependent on the assumption that these adjacent observations are independent and identically

distributed (Shumway and Stoffer 2011). For example, the t-test is based on an assumption that the observations are normally distributed, random, and independent. Lack of independence in a time series will bias the estimate of the variance and invalidate the t-test (Berthouex and Brown 2002).

Time-series analysis is the systematic approach to answer mathematical and statistical questions posed by these time correlations. An example of time-series analysis is autocorrelation or serial correlation. Autocorrelation is correlation of adjacent or nearby values within the time series (Berthouex and Brown 2002). It is a measure of the strength of relationship between successive observations (USEPA 2006). The general formula for the sample autocorrelation coefficient at lag k is:

$$r_k = \frac{\sum_{t=k+1}^N (z_t - \bar{z})(z_{t-k} - \bar{z})}{\sum_{t=1}^N (z_t - \bar{z})^2} \quad (4.12)$$

where N is the total number of observations in the time series, k is the lag expressed as a number of time intervals, z_t and z_{t-k} are the values at sequence t and $t - k$ respectively, \bar{z} is the mean of z_t and r_k is the sample autocorrelation coefficient which estimates the population autocorrelation coefficient (ρ_k). This set of coefficients (r_k) is called the autocorrelation function (ACF) (Berthouex and Brown 2002). It is common to graph r_k as a function of k in a correlogram. If there is no overall trend (decreasing or increasing) but there is regular repetition (fluctuation) of similar or dissimilar sections within a sequence, the correlogram will show autocorrelation at regular lag intervals (McKillup and Dyar 2010).

For autocorrelation, r_k at lag zero will always be 1, indicating perfect correlation. A positive correlation for lag k shows that when one observation is high, the observation at time lag k ahead or behind is also high. Conversely, if the observation is low, the observation at time lag k distant away is also low. A negative correlation indicates that observations taken at time lag k apart tend to be opposite in magnitude, one being high and the other being low. If measurements were made hourly, strong positive autocorrelation between z_t and z_{t-24} might indicate serial

dependence in the form of a daily cycle. If measurements were made daily, strong positive autocorrelation between z_t and z_{t-7} might indicate a weekly cycle (Berthouex and Brown 2002). Similarly, if z represents monthly averages, strong positive autocorrelation between z_t and z_{t-12} might indicate an annual cycle (Berthouex and Brown 2002).

Any test of the significance of r_k will only give a realistic result when the following are fulfilled. (i) Usually the autocorrelation coefficients are calculated for the lag up to but not larger than a quarter of the number of observations or $N/4$ (Berthouex and Brown 2002; McKillup and Dyar 2010). (ii) A series of $N \geq 50$ is needed for the ACF to be reliable (Berthouex and Brown 2002). The former is because as the lag interval increases, the number of overlapping values decreases so the actual number of values being correlated is fewer (McKillup and Dyar 2010). If a statistical computer program is used, it must be used cautiously because they often give autocorrelations for every possible lag, including even short sequences (McKillup and Dyar 2010). The latter relates to a 95% confidence interval for r_k that is $r_k \pm 1.96/\sqrt{N - k + 3}$. For $r_k = 0$, a series of $N \geq 50$ is needed in order that the confidence interval will be ± 0.28 or less (Berthouex and Brown 2002). Any autocorrelation coefficient that lies outside the confidence interval of $r_k = 0$ in a correlogram is significant. It is an evidence of serial correlation, therefore, the time points are dependent.

Some cases of r_k are common. (i) The series is purely random noise, so all r_k will be small. (ii) Beyond some value of k the correlation “dies out” (Berthouex and Brown 2002). Large values of r_k give information about the structure of the time-series process; small ones can be ignored. The lag at which r_k becomes negligible identifies the time between samples at which observations become independent. This system behavior makes physical sense because many factors (for example, weather, daily work patterns) change from day to day, thus gradually reducing the strength of the system memory. (iii) The series has a purely regular cycle. For example, an hourly series has diurnal cycle if it has strong positive correlations at multiples of 24 hr.

4.5.7. Normalised sets of samples

The sampling campaign in this study was conducted over 12- and 24-hr periods due to practical constraints, but represent relatively short series. The 24-hr sets of samples are used to create artificial time series (discussed in Section 4.5.8), while the 12-hr sets of samples are used in data segregation to investigate if better correlation coefficients could be obtained (Chapter 6). Since the sets of samples were collected from different sampling locations and at different dates, they have different FIB concentration ranges due to differences in background concentrations. Therefore, each sample set is normalised using the following formula before they are used for construction of the artificial time series or data segregation.

$$\text{normalised conc} = \frac{(\log_{10} \text{conc} - \log_{10} \text{min conc})}{(\log_{10} \text{max conc} - \log_{10} \text{min conc})} \quad (4.13)$$

where *conc* is concentration, *min conc* is minimum concentration of each 12-hr or 24-hr sample set and *max conc* is the maximum concentration of each 12-hr or 24-hr sample set.

4.5.8. Artificial time series

Artificial time series are constructed in order to assess diurnal patterns from measurements of FIB concentration spanning over a number of days. The steps to create the artificial time series are illustrated in Figure 4.7. Firstly, the FIB concentrations are normalised to each 24-hr minimum and maximum values. Secondly, since most 24-hr sampling events started at 20:00, the events that started at 08:00 are rearranged while the sequence within the 12-hr time frame is preserved. In other words, the last 12 samples of a particular 24-hr sample set are treated as the first 12 samples and the first 12 samples becoming the last 12 samples. In this way, all 24-hr samples sets are made to consistently start at 20:00 and end at 19:00. The artificial time series are then constructed by chaining the normalised concentrations one after another in chronological order. As the FIB time series is assumed to be a stationary stochastic process, lining up the different time series (sampled at different times and locations) is valid. After this, autocorrelation coefficients and correlograms

(Section 4.5.6) were determined for the artificial time series data using Microsoft Excel.

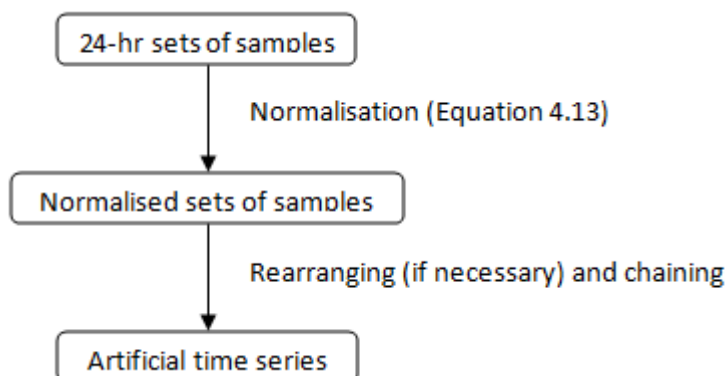


Figure 4.7. Steps to create artificial time series

4.5.9. Root-mean-square error

Root-mean-square error (*RMSE*) is used to measure the deviation of a time series from its average value. This analysis was conducted on the normalised sample set. *RMSE* is calculated using following formula.

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (x_t - \bar{x}_t)^2}{N}} \quad (4.14)$$

where N is the number of samples in one set of samples, t is the sequence of the sample in one set of samples, x_t is the normalized value at sequence t and \bar{x}_t is the average (arithmetic mean) normalized value at sequence t of all sets of samples.

4.5.10. *t*-test

The *t*-test is used to compare two groups of data, to determine if one group tends to contain values that are larger than the other (Helsel and Hirsch 2002). The *t*-test is used when both population variances are unknown, either equally or unequally (Tamhane and Dunlop 2000). It is based on an assumption that the observations are independent, random, and normally distributed (Berthouex and Brown 2002). The data are independent in the sense that there are no pairings of data between

observation 1 of group 1 and observation 1 of group 2 (Helsel and Hirsch 2002). The use of the t -test is acceptable since the data in this study have random characteristics: collected from different sampling locations and from different (non-consecutive) sampling dates. Further, the number of samples for each parameter was large ($N > 30$) and data normality was confirmed by histogram and probability plots. In cases where it is known ahead of time that a group is expected to be larger, a one-sided test is used. Otherwise, the test is two-sided (Helsel and Hirsch 2002). In this study, two-sided t -tests for unequal variances were performed using Microsoft Excel on the log-transformed FIB concentrations.

4.5.11. Wilcoxon-Mann-Whitney test

The Wilcoxon-Mann-Whitney test is a nonparametric test to compare two independent groups of data. The data are converted to ranks. The null hypothesis states that there is no significant difference between the two groups of data. The alternative hypothesis states that there is significant difference between the two groups of data. The Matlab function for Wilcoxon-Mann-Whitney test, `ranksum`, was used. It is a two-sided test at 95% confidence interval and for large samples, the normal approximation is used (the z stat is calculated and the p value is obtained from the z stat) (Tamhane and Dunlop 2000).

4.5.12. Analysis of variance

Analysis of variance (ANOVA) is a method for testing two or more groups to determine whether their sample means could have been obtained from populations with the same true mean (Berthouex and Brown 2002). One-way ANOVA is used when there is only one factor to be compared, for example, HF group. Two-way ANOVA is used when there are two factors to be compared, for example, HF group and sampling time. One-way and two-way ANOVA tests are conducted on the data using JMP Pro v. 10.

4.5.13. Kruskal-Wallis test

The Kruskal-Wallis test is a nonparametric version of the classical one-way ANOVA. The Kruskal-Wallis test is similar to ANOVA but uses the ranks of the data rather than their numeric values. Unlike ANOVA, the Kruskal-Wallis test is not affected by outliers. The Matlab function for the Kruskal-Wallis test, `kruskalwallis`, is used in this study. The significance was tested at 5% level.

4.5.14. Tukey-Kramer test

The Tukey-Kramer test is a multiple comparison test which is similar to a *t*-test. When there are multiple comparisons being made, the probability of making a type I error (incorrect rejection of null hypothesis) increases. Therefore, the Tukey-Kramer test shrinks the error rate of individual comparison to keep the simultaneous family error rate at the desired level (5%) in order to achieve valid statistical comparisons (Berthouex and Brown 2002). One-way ANOVA followed by Tukey-Kramer tests are conducted on the data using JMP Pro v. 10. Kruskal-Wallis followed by Tukey-Kramer tests are conducted on the data using Matlab function, `multcompare (using tukey-kramer as ctype)` when Kruskal-Wallis test result shows $p \text{ value} \leq 0.05$.

CHAPTER 5 DIURNAL VARIATION OF FAECAL INDICATOR BACTERIA

Preliminary data analysis (presented in Section 3.3) has revealed that sampling at upstream locations may enable better water quality characterisation for this study. In addition, analysis of the Toolbox Study data (Hans Eikaas, PUB, personal communication, 2010) in Section 3.3.2 showed that hourly sampling reveals diurnal variation of faecal indicator bacteria (FIB) and other potential tracers. This diurnal variation suggests the possibility that the contamination comes from sewage exfiltration as discussed in Section 3.3.2. This hypothesis was tested through a series of field experiments in which dry weather samples were collected at hourly intervals, over 24-hr periods, at eleven sampling locations and tested for FIB.

5.1. Results

Fourteen sets of 24-hr FIB samples were collected hourly at 11 sampling sites as shown in Table 5.1. Of the 11 sites, six are classified as high-density residential areas and five as low-density residential areas. Sampling was conducted three times at Lorong 8 Toa Payoh and twice at Jalan Siantan on different dates to test repeatability of results. Sampling mostly started at 20:00 with the exception of sampling at Lorong 8 Toa Payoh, MacPherson Lane and Ang Mo Kio Avenue 10, which started at 08:00 instead.

Out of the 336 (14 x 24) samples to be obtained, 3 samples were missed due to power failure of the auto-samplers and 11 samples could not be considered to be dry weather samples due to rainfall. These 14 samples were left out from the analysis presented in this chapter. In addition, twelve samples have total coliform (TC) values greater than the maximum of 24,196,000 MPN/100mL. For these samples, the value of 24,196,000 MPN/100mL has been assigned.

Table 5.1. Summary of sampling program

Sampling site	Type of residential land use	Sampling date	Number of dry-weather samples collected	Total Coliform (MPN/100mL)		<i>E. coli</i> (MPN/100mL)		Enterococci (MPN/100mL)	
				Min	Max	Min	Max	Min	Max
Lorong 8 Toa Payoh	High-density	9-10 January 2012*	19	706,000	24,196,000	29,090	6,867,000	7,940	241,960
		17-18 January 2012*	21	77,010	24,196,000	410	250,000	10,120	141,360
		18-19 January 2012*	20	677,000	24,196,000	1,733	77,010	6,570	198,630
Jalan Siantan	Low-density	6-7 March 2012	23	54,750	2,046,000	2,720	198,630	2,690	41,060
		19-20 March 2012	24	41,060	2,098,000	6,830	195,000	613	13,960
Punggol Central	High-density	21-22 May 2012	24	38,730	1,732,900	3,360	579,400	960	141,360
Lorong 6 Toa Payoh	High-density	12-13 June 2012	24	836,000	24,196,000	9,590	1,421,000	520	51,720
MacPherson Lane	Low-density	20-21 June 2012*	24	8,600	820,000	52	187,000	270	12,460
Ang Mo Kio Avenue 10	High-density	27-28 June 2012*	24	2,142,000	24,196,000	13,140	2,603,000	5,040	36,540
Serangoon Garden	Low-density	9-10 July 2012	24	12,110	771,000	816	384,000	18	22,470
Toa Payoh North	High-density	16-17 July 2012	23	141,360	6,488,000	4,880	141,360	860	602,000
Watten Drive	Low-density	23-24 July 2012	24	81,640	743,000	750	173,000	1,414	4,960
CCK Crescent	High-density	29-30 July 2012	24	34,480	155,310	190	21,870	411	10,500
Verde	Low-density	30-31 July 2012	24	496,000	10,462,000	10,100	1,850,000	1,090	43,520
Total number of dry-weather samples collected			322						

* denotes sampling started at 08:00. Maximum TC printed in boldface indicates actual concentration is greater than 24,196,000 but was replaced with 24,196,000 for plotting and analysis.

A comparison of the mean and range of FIB concentrations of this study with other studies is summarised in Table 5.2. Values underlined in the table indicate that they are faecal coliform or faecal streptococci instead of *E. coli* or enterococci. *E. coli* is the dominant bacterium in the faecal coliform group (Dufour 1977). Concentration values boldfaced in the table indicate they are median instead of geometric mean. For Isobe et al. (2002), data were obtained from Table S2 in the supporting information published on the journal website, except for min TC which was estimated from Figure 6 in their paper. For Sercu et al. (2009), data were estimated from Figures S1 and S2 in the supporting information published on the journal website. For Chua et al. (2010), Isobe et al. (2004) and Traister and Anisfeld (2006), data were estimated from figures in the publication. For Desai and Rifai (2013), the values presented are listed in Table 3 of their publication. For USEPA (1983), the data are obtained from Table 6-18 of the report. For Kacar (2011), FC data were taken from Bakircay River and FS data were taken from K. Menderes River in Table 2 of their publication. Geometric means of *E. coli* and enterococci concentrations of the 24-hr sets of samples of this study are comparable with other studies in urban areas both in tropical and temperate climates (USEPA 1983; Wu et al. 2008; Sercu et al. 2009). In general, studies in urban areas (this study, Wu et al. 2008, Sercu et al. 2009 and USEPA 1983) have higher geometric mean of FIB than other studies with a mix of urban and/or peri-urban area (Isobe et al. 2002; Isobe et al. 2004; Traister and Anisfeld 2006). Based on the limited study survey listed in Table 5.2, studies in tropical areas (this study, Isobe et al. 2002 and Isobe et al. 2004) have wider ranges of FIB concentrations than other studies in temperate areas (USEPA 1983; Traister and Anisfeld 2006; Sercu et al. 2009; Kacar 2011).

Table 5.2. Comparison of FIB concentrations among different studies

Study	Sampling location	Type of sampling site and sample	Unit	Total coliform			<i>E. coli</i> /faecal coliform			Enterococci/faecal streptococci		
				Geomean/ median	Min	Max	Geomean/ median	Min	Max	Geomean/ median	Min	Max
24-hr sets of samples of this study	Singapore	urban drainage; tropical; dry-weather; separate sewers	MPN/100mL	647,578	8,600	24,196,000	20,509	52	6,867,000	5,532	18	602,000
Wu et al. (2008)	Singapore	urban drainage, reservoir; tropical; probably dry-weather; separate sewers	CFU/100mL	NA	NA	NA	<u>21,209</u>	<u><500</u>	<u>304,000</u>	NA	NA	NA
Hans S. Eikaas, PUB, personal communication (2010) as listed in Table 3.1	Singapore	urban drainage, reservoir; tropical; probably dry-weather; separate sewers	CFU/100mL	NA	NA	NA	166/ <u>310</u>	2/ <u>4</u>	80,000/ <u>120,000</u>	NA	NA	NA
Chua et al. (2010) - Phase 2	Singapore	urban drainage (high-density residential); tropical; dry-weather; separate sewers	MPN/100mL	NA	NA	NA	5,000	100	2,000,000	NA	NA	NA
Isobe et al. (2002)	Malaysia	urban and rural; river; tropical; probably dry-weather; combined sewers	cell/100mL	20,271	500	960,000	5,370	200	270,000	NA	NA	NA
Isobe et al. (2002)	Mekong Delta, Vietnam	urban and rural; river; tropical; probably dry-weather; combined sewers	cell/100mL	22,311	10	8,600,000	3,279	1	1,100,000	<u>1,025</u>	<u>1</u>	<u>100,000</u>
Isobe et al. (2004)	Mekong Delta, Vietnam	urban and rural; river; tropical; dry-weather; combined sewers	CFU/100mL	19,953	3	15,848,932	3,162	1	2,511,886	1,585	<u>32</u>	<u>2,511,886</u>
Desai and Rifai (2013)	Texas, USA	urban; 3 streams (2 concrete-lined); subtropical; dry-weather; separate and combined sewers	MPN/100mL	NA	NA	NA	4,636	265	80,567	NA	NA	NA
Sercu et al. (2009)	Santa Barbara, CA, USA	urban drainage; temperate; dry-weather; separate sewers	MPN/100mL	NA	NA	NA	19,953	<1	158,489	3,162	<1	>31,623
USEPA (1983)	USA	urban, suburban, peri-urban; river; temperate; wet-weather; combined and separate sewers	CFU/100mL	NA	NA	NA	<u>21,000</u>	<u>300</u>	<u>281,000</u>	NA	NA	NA
Kacar (2011)	Turkey	urban; river; temperate; probably dry-weather; combined sewers	CFU/100mL	NA	NA	NA	<u>7,800</u>	<u>85</u>	<u>50,000</u>	<u>19,000</u>	<u>280</u>	<u>63,000</u>
Traister and Anisfeld (2006) - site 5	Massachusetts, USA	peri-urban; stream; temperate; dry- and wet-weather; separate sewers	MPN/100mL	NA	NA	NA	316	23	2,276	NA	NA	NA

NA means not analysed.

5.1.1. Log-normality of FIB concentration

Raw FIB concentrations are plotted as histograms in Figure 5.1 (left) and in probability plots in Figure 5.2 (left). The figures indicate that the raw FIB concentrations are heavily right-skewed. Therefore, FIB concentrations were log-transformed prior to analysis. Figure 5.1 (right) and Figure 5.2 (right) show that log-transformed FIB concentrations better resemble a normal distribution after transformation. The normal probability plots of TC in Figure 5.2 show a plateau at the upper detection limit of 24,196,000 MPN/100mL. This upper limit corresponds to the dilution ratio of 10,000 which was the highest dilution ratio adopted in this study.

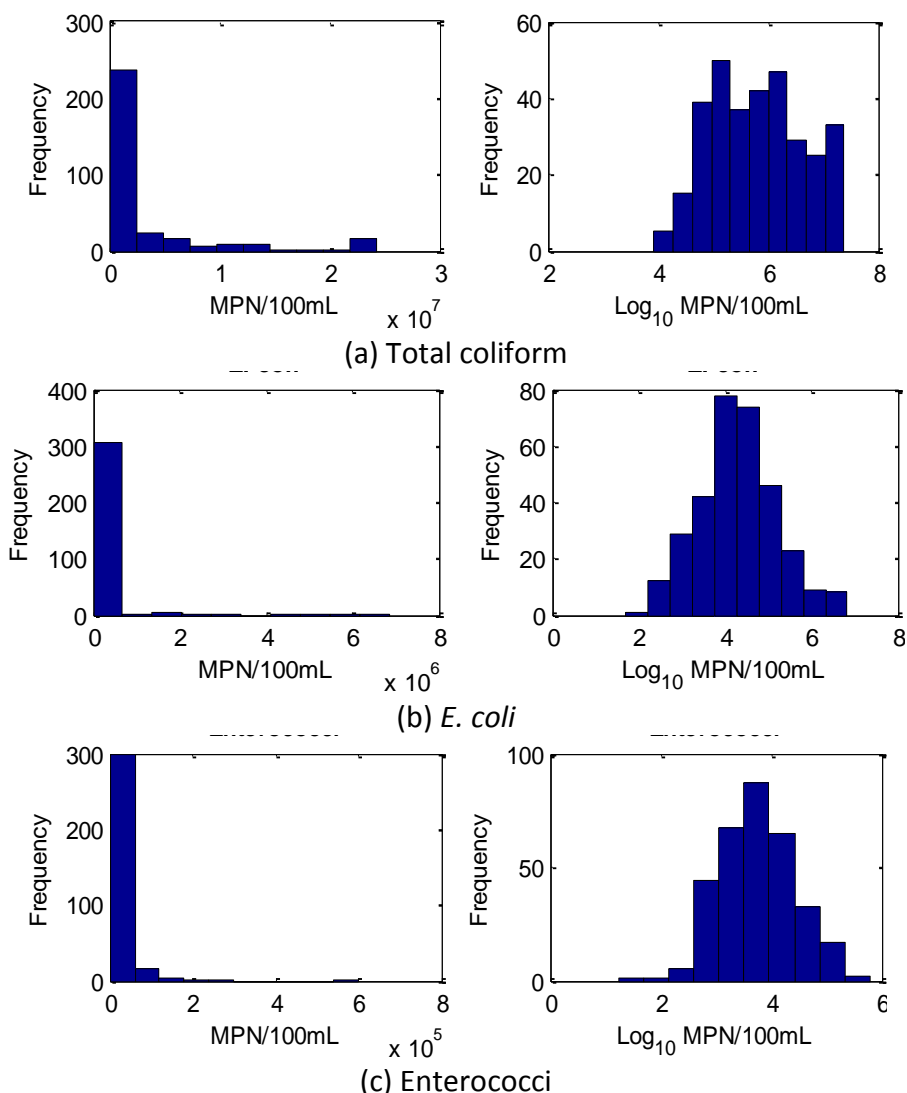
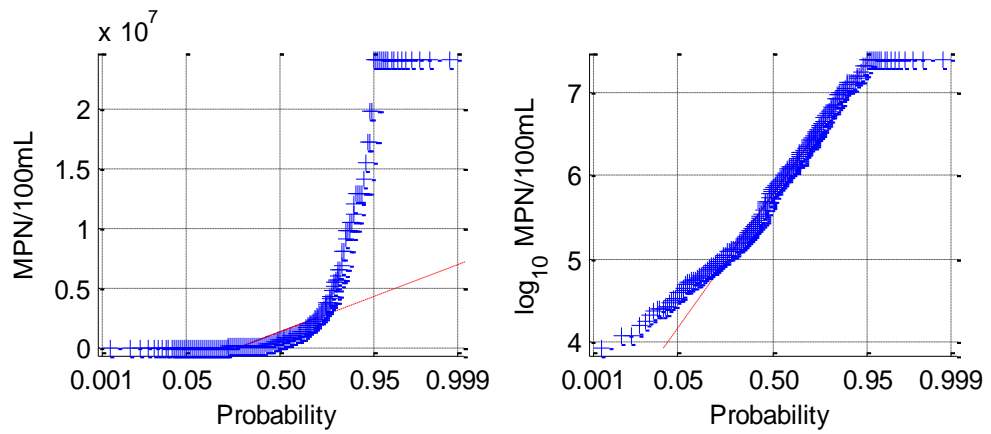
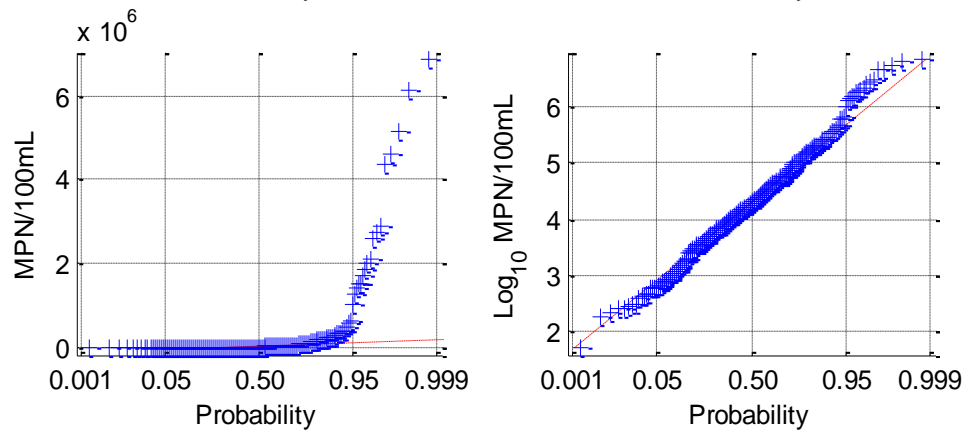


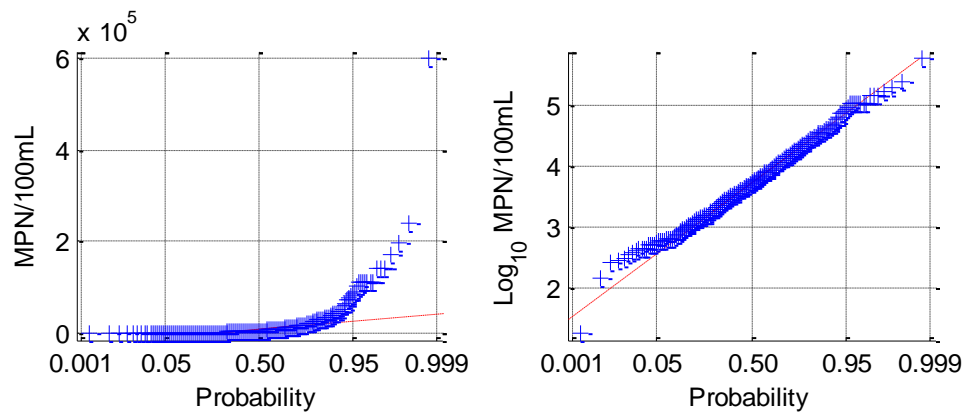
Figure 5.1. Histograms of raw (left) and log-transformed (right) 24-hr FIB concentrations: (a) TC (b) *E. coli* (c) enterococci



(a) Total coliform



(b) *E. coli*

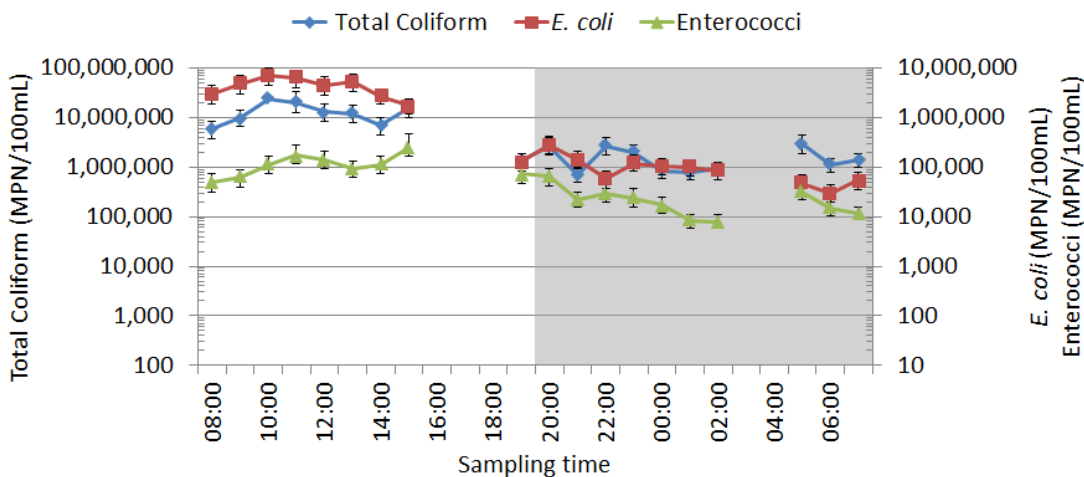


(c) Enterococci

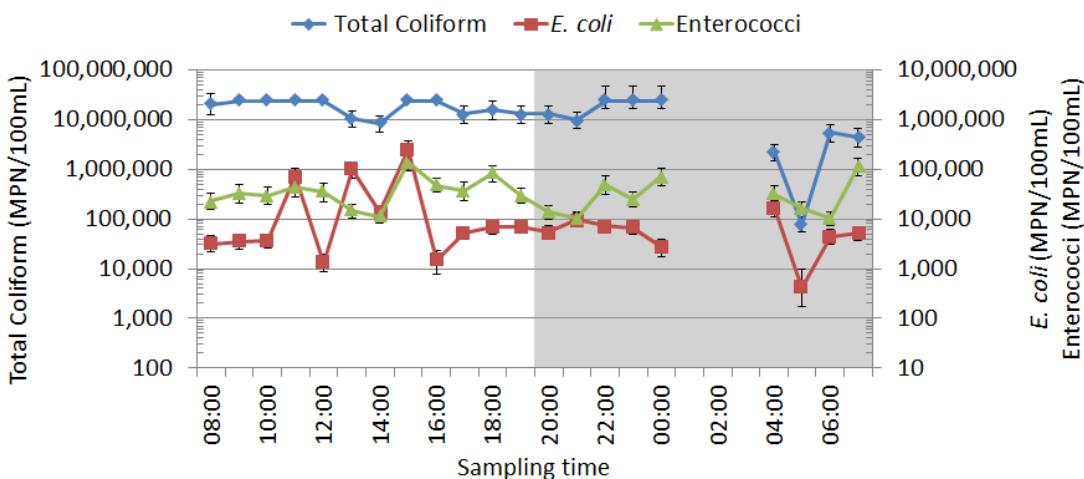
Figure 5.2. Probability plots of raw (left) and log-transformed (right) 24-hr FIB concentrations: (a) TC (b) *E. coli* (c) enterococci

5.1.2. Variation in FIB concentration

The 24-hr sets of measured FIB concentrations are plotted in Figure 5.3. The error bars were calculated based on Equations 4.4 – 4.7 which represent the errors associated with the estimation of the MPN (at the 95% confidence interval) and sample dilution (use of pipette and graduated cylinder) as discussed in Section 4.5.4. As mentioned above, the gaps in data are because samples were missed due to power failure or deviations from dry-weather conditions. Twenty-four-hour sampling was repeated at Lorong 8 Toa Payoh and Jalan Siantan, but not at the other sites. Repeated samplings were conducted to confirm that the results obtained are representative of general conditions.

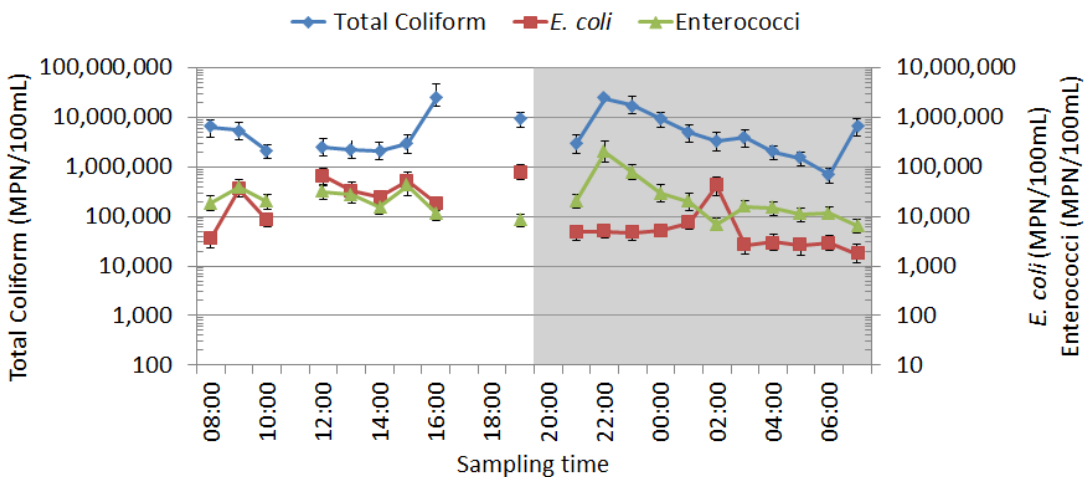


(a) Lorong 8 Toa Payoh, 9-10 January 2012

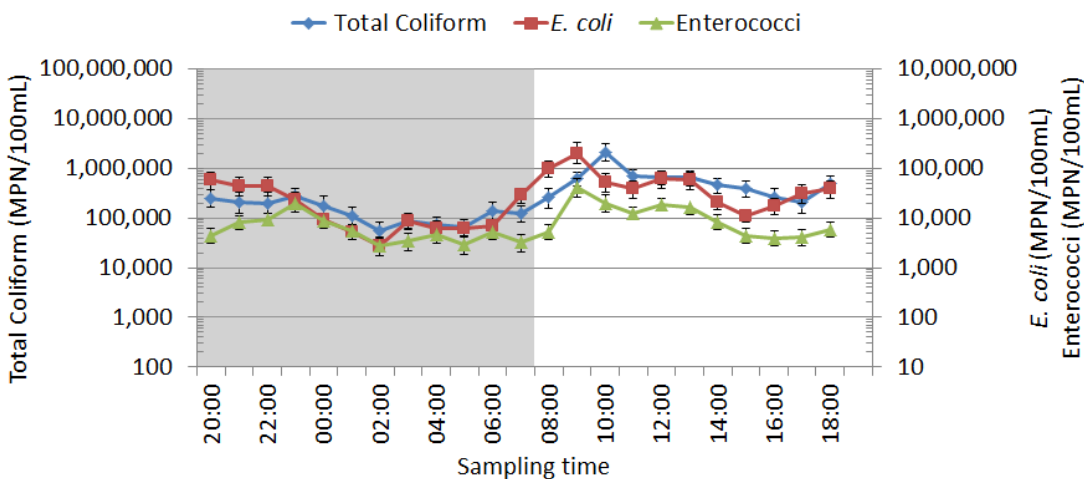


(b) Lorong 8 Toa Payoh, 17-18 January 2012

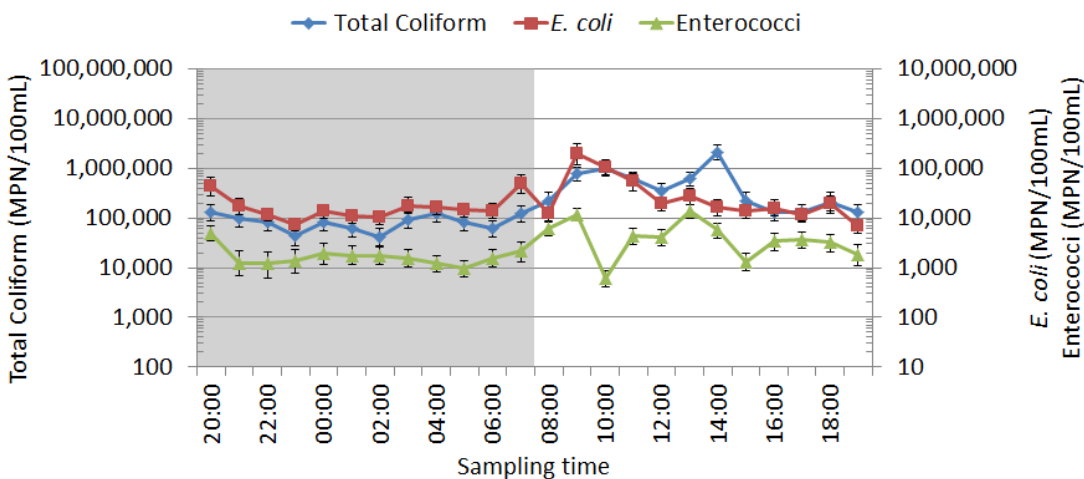
Figure 5.3. Plots of 24-hr FIB concentrations and error bars



(c) Lorong 8 Toa Payoh, 18-19 January 2012

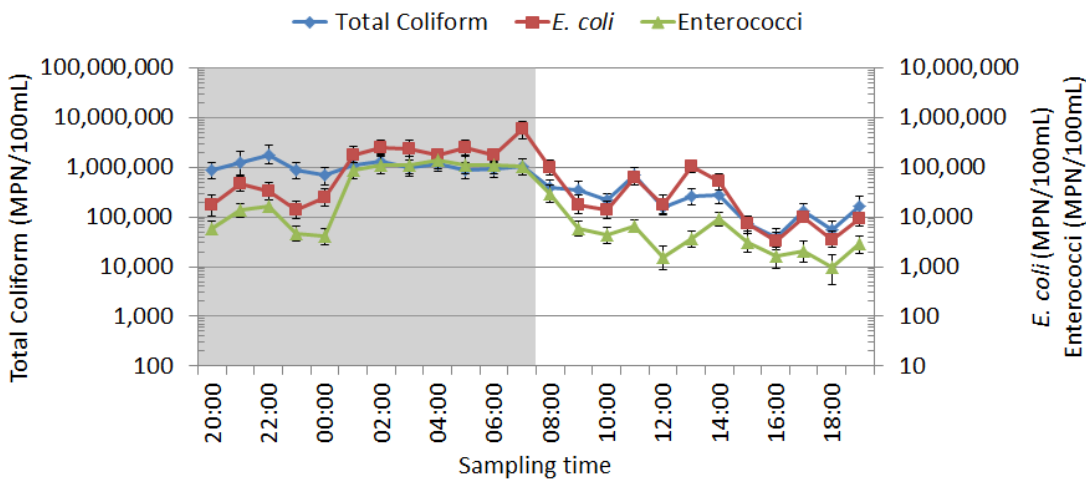


(d) Jalan Siantan, 6-7 March 2012

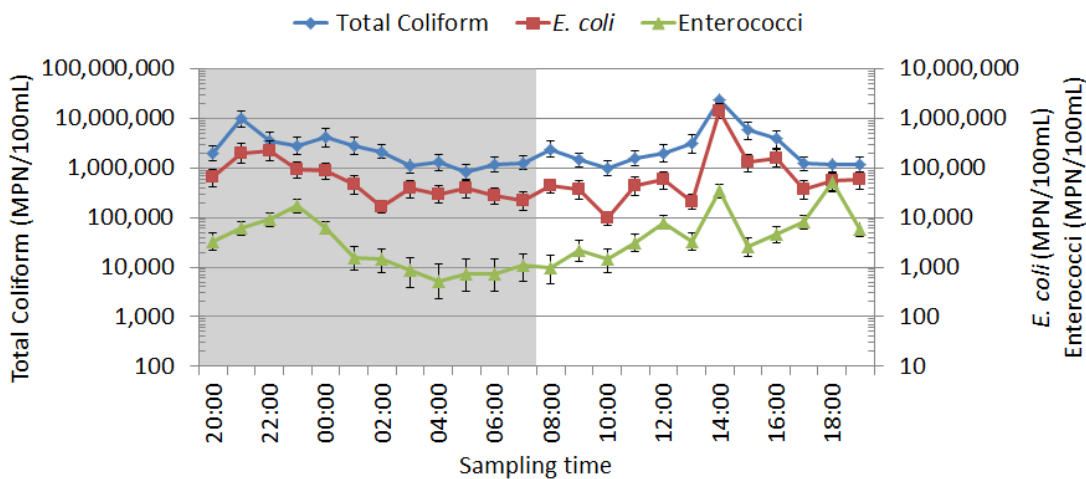


(e) Jalan Siantan, 19-20 March 2012

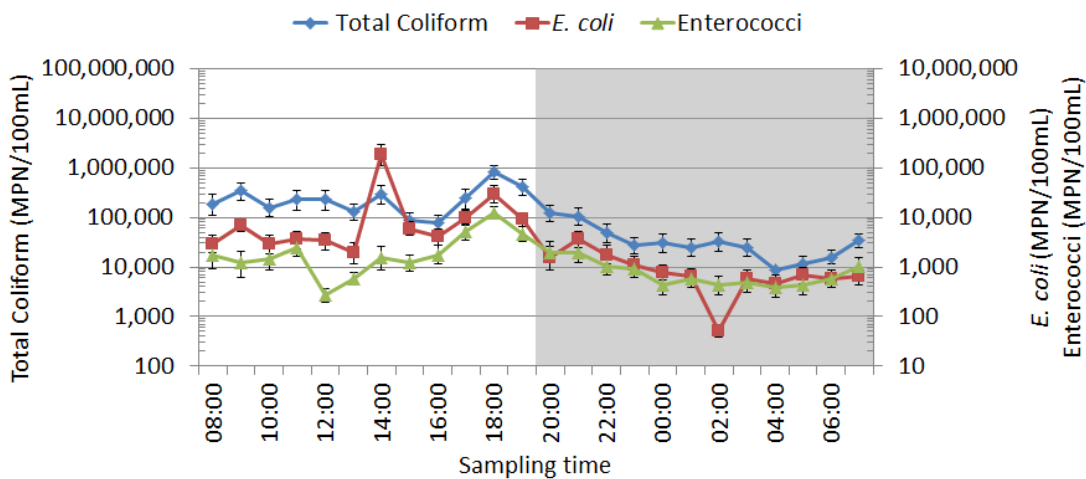
Figure 5.3. (continued) Plots of 24-hr FIB concentrations and error bars



(f) Punggol Central, 21-22 May 2012

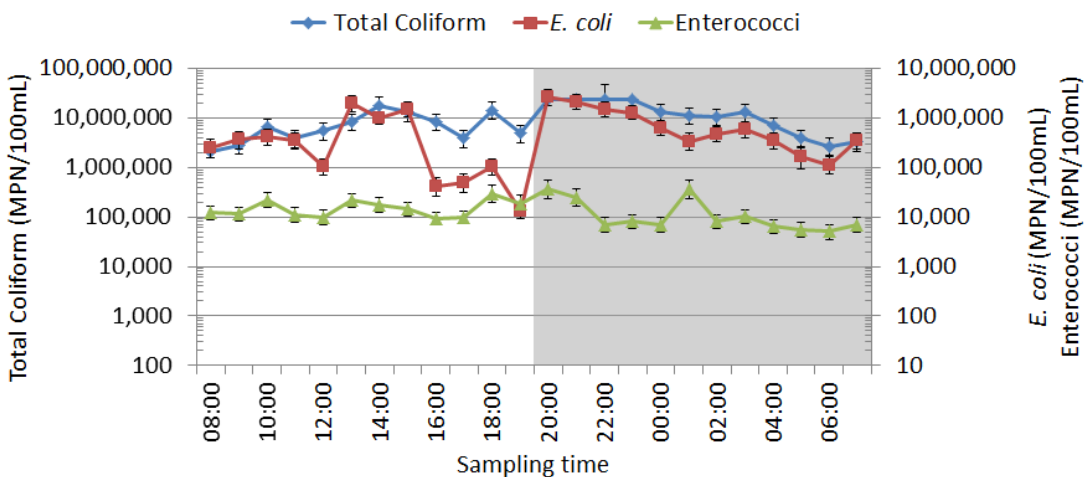


(g) Lorong 6 Toa Payoh, 12-13 June 2012

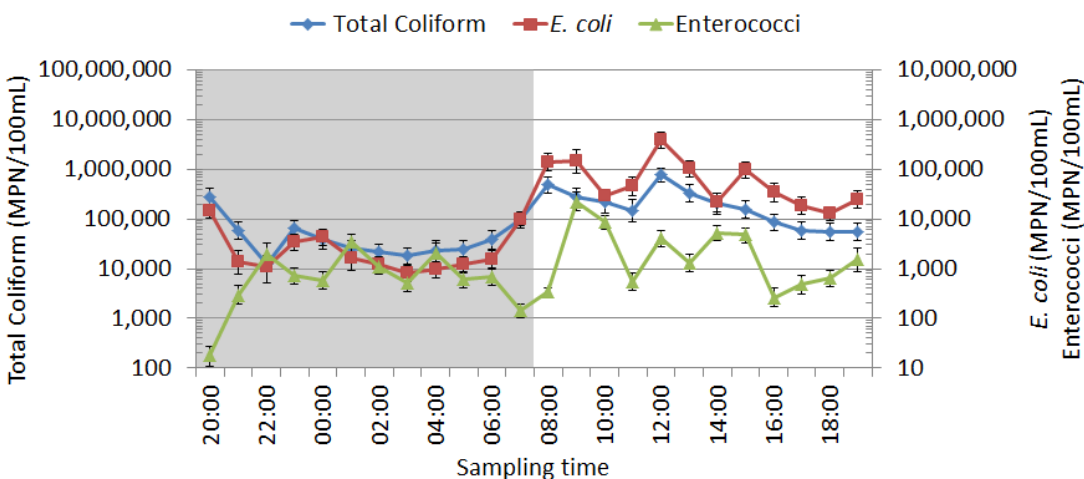


(h) MacPherson Lane, 20-21 June 2012

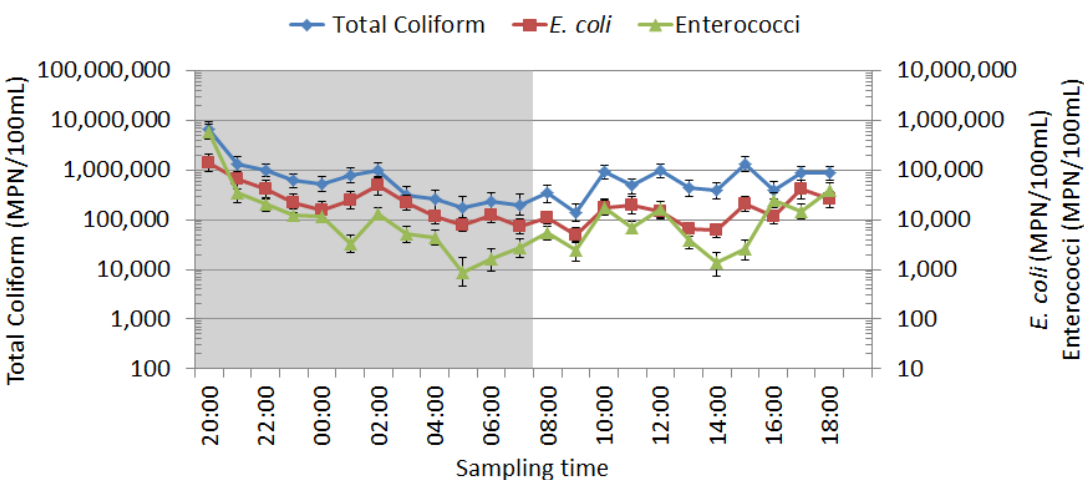
Figure 5.3. (continued) Plots of 24-hr FIB concentrations and error bars



(i) Ang Mo Kio Avenue 10, 27-28 June 2012

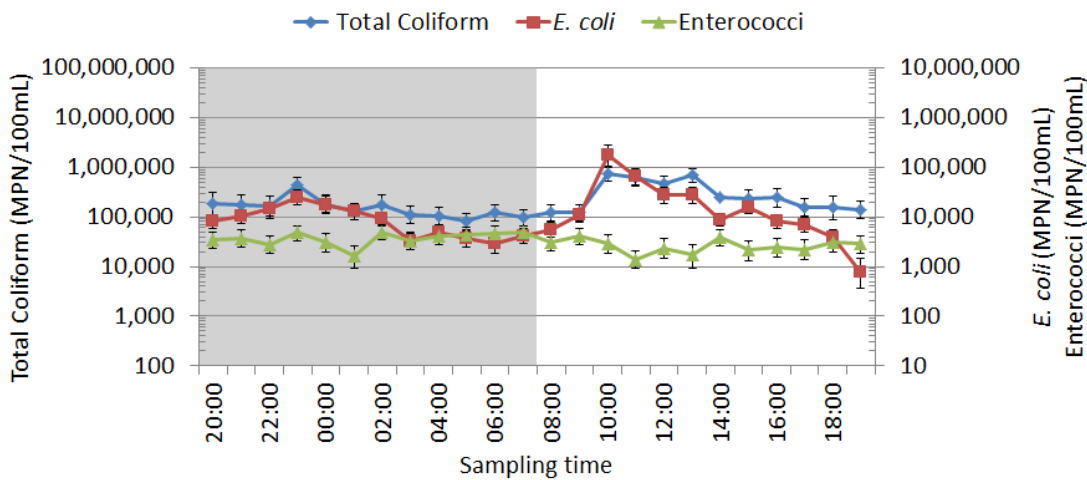


(j) Serangoon Garden, 9-10 July 2012

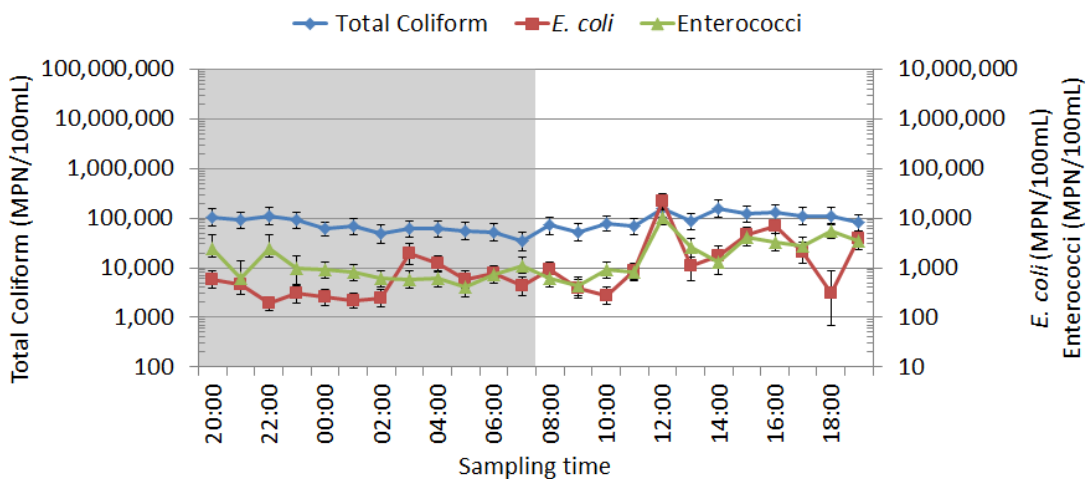


(k) Toa Payoh North, 16-17 July 2012

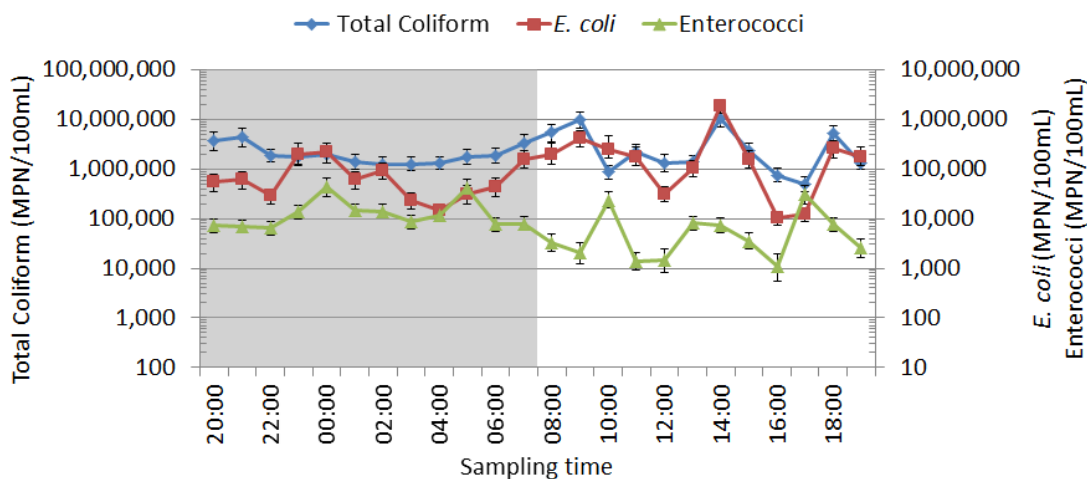
Figure 5.3. (continued) Plots of 24-hr FIB concentrations and error bars



(l) Watten Drive, 23-24 July 2012



(m) Choa Chu Kang Crescent, 29-30 July 2012



(n) Verde, 30-31 July 2012

Figure 5.3. (continued) Plots of 24-hr FIB concentrations and error bars

A majority of the plots show FIB concentrations that are high in the daytime (defined as samples collected from 08:00 to 19:00) and low during nighttime (defined as samples collected from 20:00 to 07:00, indicated by the shaded time intervals in Figure 5.3). Some of the plots show high nighttime FIB concentrations in addition to high daytime FIB concentrations: the 24-hr sampling set at Lorong 8 Toa Payoh on 17-18 January 2012, at Ang Mo Kio Avenue 10 and at Verde. Some of the plots show less variation of FIB concentrations throughout the 24-hr period, for example, the 24-hr sampling set at Watten Drive. The Punggol Central 24-hr sampling set shows roughly consistent concentrations of the different indicator bacteria but all with very high nighttime concentrations compared to the daytime concentrations. All of the 24-hr sampling sets, excluding Punggol Central, have high daytime FIB concentrations regardless of nighttime concentrations. In other words, the high daytime FIB concentrations occur across 10 out of 11 sampling sites. These high daytime FIB concentrations also occur on different sampling dates at the same sampling sites, for example, at Jalan Siantan on both 6-7 March 2012 and 19-20 March 2012.

Traister and Anisfeld (2006) carried out bi-hourly sampling over a total of ten 24-hr periods at two rural sites in the Upper Hoosic River, Massachusetts, USA. They observed that in general, *E. coli* was high (approximately 40-1,000 MPN/100mL) during the nighttime period from 20:00 to 11:00 and low (approximately 25-400 MPN/100mL) during the daytime period from 11:00 to 20:00. Traister and Anisfeld attributed this observation to the diurnal pattern of sunlight-induced die-off. Similarly, Desai and Rifai (2013) also reported diurnal variations in *E. coli* that are marked by a daytime exponential decay strongly influenced by water temperatures and solar radiation intensities, followed by an exponential nighttime regeneration. However, Traister and Anisfeld also noted that the diurnal die-off pattern was not consistent at the more shaded of their two sites; two out of five sets of samples from the more shaded site had insignificant difference between daytime and nighttime samples. Similarly, Desai and Rifai reported that the pattern for samples collected from a site partially underneath a bridge was not as apparent as the pattern for

samples collected from sites exposed to sunlight. In the present study, sampling was done in urban storm drains that are mostly covered. In cases where the drains are not covered or only partially covered (Jalan Siantan, MacPherson Lane and Watten Drive), the sampling sites are well shaded by roadside trees and/or adjacent buildings. Hence, sunlight-induced die-off is unlikely to be a significant factor in this study. It is therefore postulated that the observed diurnal variations in FIB concentration, which are consistent with sewage flows as mentioned in Section 3.3.2, are associated with wastewater discharge from leaking sewers.

5.1.3. Field verification of sewer leakage

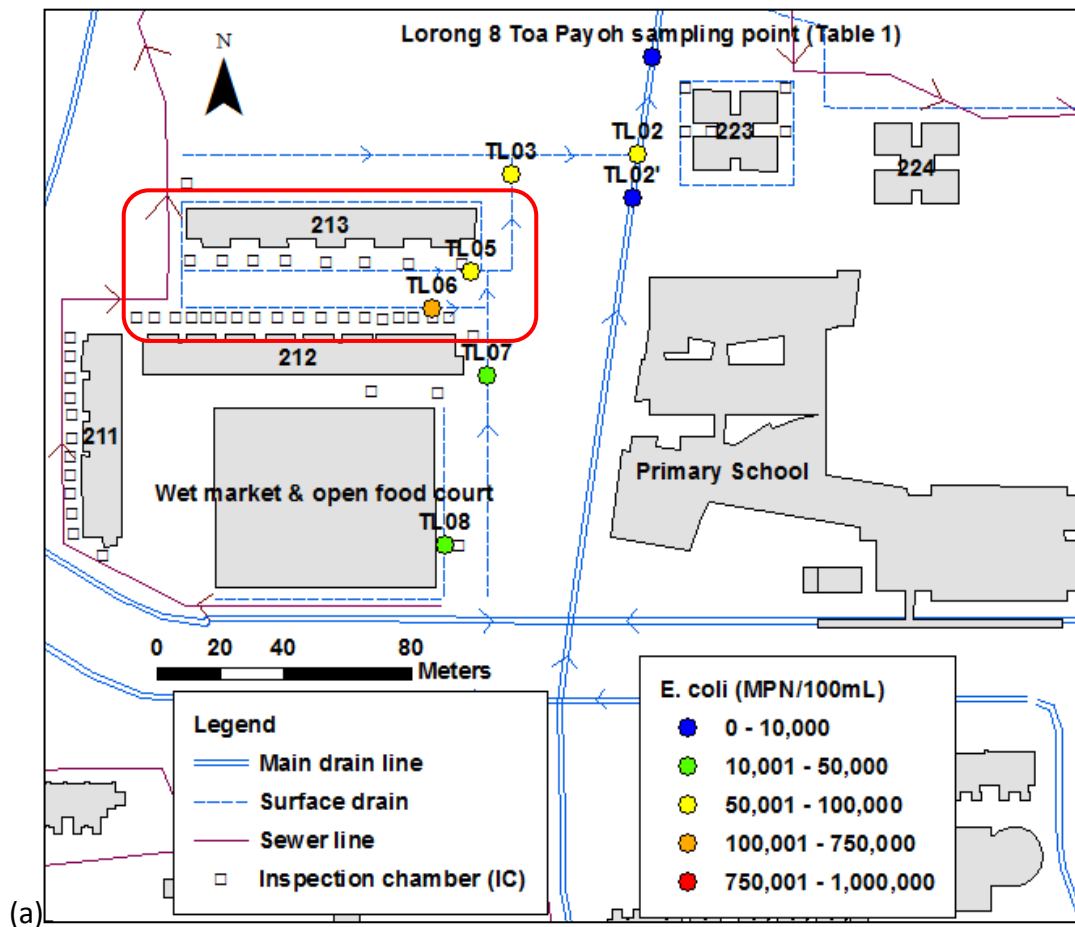
Sauer et al. (2011) indicated that tracer studies can be the most effective approach to assess the presence of hotspots of contamination (e.g. from leaking sewers to nearby drain lines). A tracer study with brine was therefore planned to demonstrate the possible bacterial contamination of surface drains via subsurface pathways as a result of sewage exfiltration from damaged pipes. In Singapore, the groundwater table is generally shallow (about 2 m below ground surface) and sewer pipes are usually laid at deeper levels; therefore infiltration into sewers, instead of exfiltration out of sewers, is generally expected. However, upstream sewer lines such as sewers connecting discharge pipes from buildings or properties to the public sewers are generally shallow and above the groundwater table. Therefore, it is hypothesised that sewer leakage may occur for example at joints that may have been damaged during backfilling or ground settling after backfilling operations at the end of construction. Tan et al. (2009) mention that Singapore suffers from a high incidence of subsidence, which can lead to breaks in sewers. The leaking sewage can travel through preferential pathways in the soil eventually discharging into surface drains through weep holes, cracks or joints. In addition, the bedding beneath the sewer line can carry sewage downstream from leaks. At points where the source of the leak is hydraulically connected via preferential pathways to a drain line, the leaking sewage can be intercepted and therefore flow into the drain line. Regulations (PUB 2004) require that sewer bedding be constructed from granular material or concrete and be

compacted, laid evenly, well rammed and worked. There is therefore a possibility that voids due to uneven placement of pipe bedding or the granular bedding layer itself may create preferential pathways (Reynolds and Barrett 2003; Shanahan 2009). Lundy and Gogel (1988) have discussed preferential pathways in sewer bedding as a means to transport gasoline downstream from leaks from an underground storage tank. The travel time through the subsurface can be short (Doshi 2012) in which case the FIB concentration in the surface drain in the vicinity of these sources would closely mimic sewage flow patterns.

The study site is shown in Figure 5.4. A wet market, an open food court, a school and several blocks of residential apartments (Blocks 211, 212 and 213) are located in the vicinity of the study site. Wet markets are outdoor markets selling fresh produce and are commonly found of many countries in the region. The location of the Lorong 8 Toa Payoh sampling point (Table 5.1) is shown in the figure. The Lorong 8 Toa Payoh sampling point receives flow from surface drains that intercept storm runoff from the apron areas surrounding Apartment Blocks 212 and 213. Inspection chambers (ICs) are shallow manholes that intercept sewage and grey water from properties and direct wastewater into public sewers. The exact age of the sewer in this area is not available, but this is an older section of Singapore and thus, based on Shin (2012), more likely to be affected by leakage.

Prior to the tracer tests, water samples were collected from the drains in the vicinity of Apartment Blocks 212 and 213 indicated in Figure 5.4 (a). The samples collected were analysed for TC, *E. coli* and enterococci. The results for *E. coli* sampled on 14 January 2013 are presented in Figure 5.4 (a). The results show that TL05 and TL06 were locations with high FIB concentrations. A repeat sampling at TL05 and TL06 was conducted on 17 January 2013, during which additional upstream sites, TL05-01 through -04 and TL06-01 through -04, were also tested (Figure 5.4 (b)). The data from these samples allowed the location for possible leakage to be narrowed to the row of

ICs immediately south of Apartment Block 213. Tracer tests were conducted by introducing brine as a tracer at these ICs.



(a) Figure 5.4. Field verification of sewer leakage, adapted from Diagne (2013) (GIS location data for sewer and main drain lines provided by PUB): (a) *E. coli* concentrations on 14 January 2013 (b) *E. coli* concentrations on 17 January 2013 (c) Location for the injection of tracer

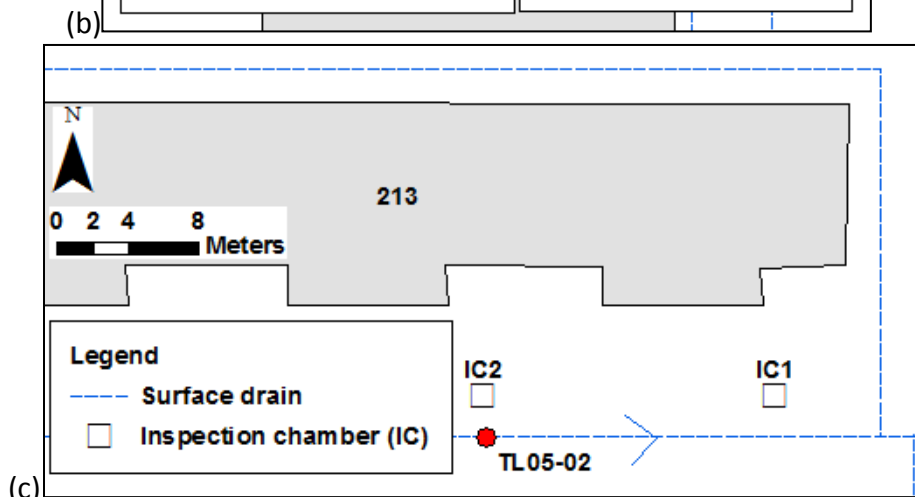
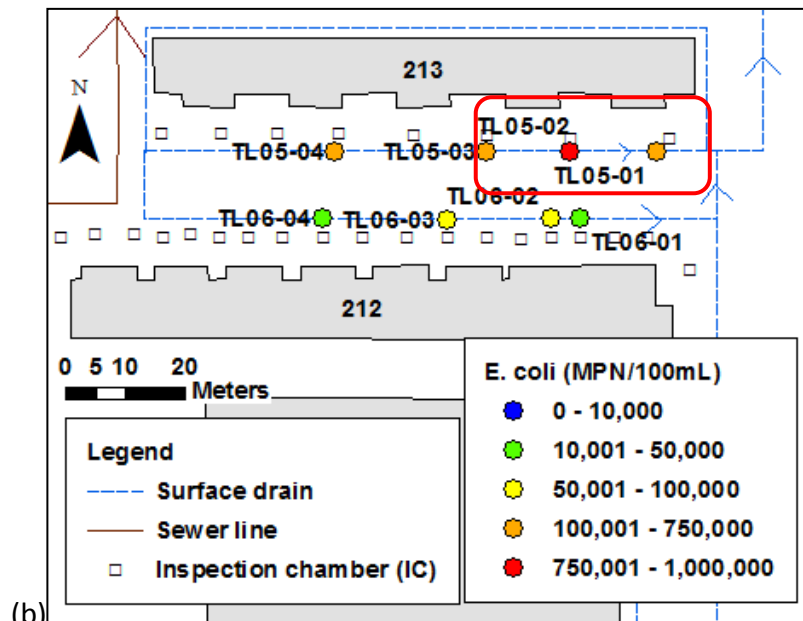


Figure 5.4. (continued) Field verification of sewer leakage, adapted from Diagne (2013) (GIS location data for sewer and main drain lines provided by PUB): (a) *E. coli* concentrations on 14 January 2013 (b) *E. coli* concentrations on 17 January 2013 (c) Location for the injection of tracer

For each day of the tracer test, a 200-ppt (part per thousands) brine solution was prepared by mixing 8 kg of table salt with 40 L of hot water. Brine was injected into inspection chambers IC1 and IC2 (Figure 5.4 (c)) on 24 January 2013 at 11:15 to simulate an instantaneous pulse of sewage. In addition, table salt mixed with ice was placed on the benching of the IC to simulate a continued release of sewage. The specific conductance was monitored at TL05-02 at 10 min intervals with a YSI 30 (YSI, Inc., Yellow Springs, OH, USA) conductivity meter. A spike in specific conductance was

observed at 12:35, about 1 hr 20 min after the start of injection (Figure 5.5 (a)) and a second spike was observed at 14:55, about 3 hr 40 min after the start of injection (Figure 5.5 (a)). The two spikes may indicate there are multiple preferential pathways or they could have originated from the two emplaced sources. The tracer experiment was repeated using only an instantaneous injection of brine on 4 February 2013 at 11:20. On this occasion, a spike was observed at 13:50, about 2 hr 30 min after the start of brine injection (Figure 5.5 (b)).

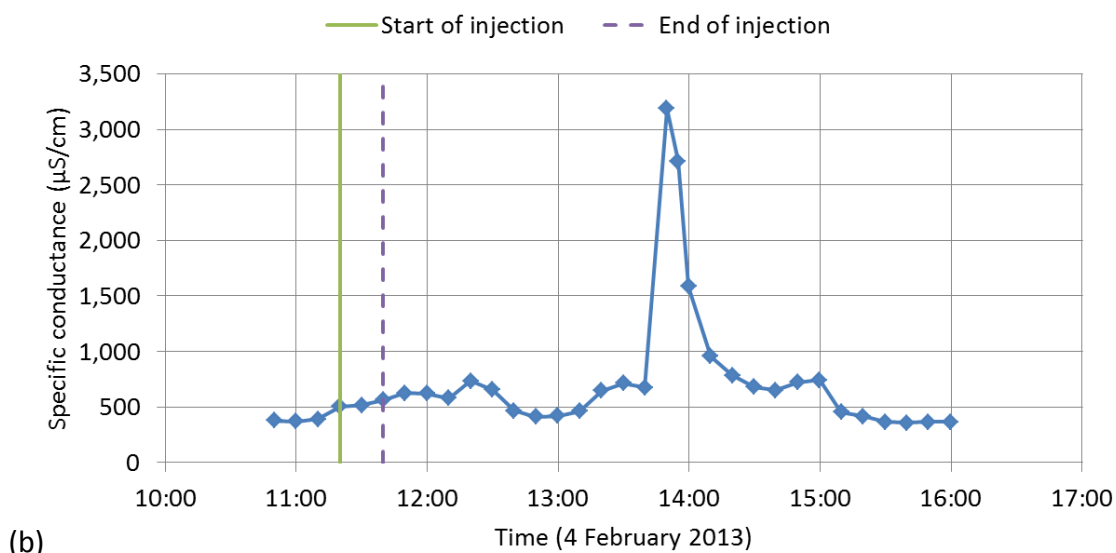
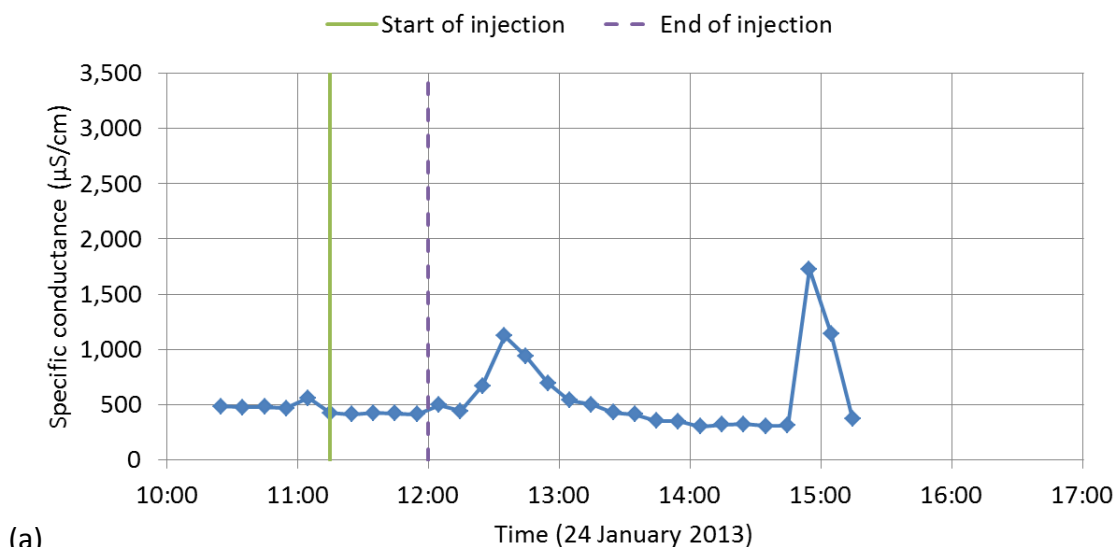


Figure 5.5. Tracer test results at Lorong 8 Toa Payoh: (a) 24 January 2013, adapted from Diagne (2013). Monitoring ceased at 15:15 pm due to rain (b) 4 February 2013

The spikes of specific conductance in Figure 5.5 occurred within a few hours from the start of brine injection indicate that one or more preferential flow paths, which allow for a direct connection from the leaking sewers to the storm drains, exist at this location. This preferential subsurface pathway or pathways allow the travel time from the sewer leakage to nearby drain to be about 2-3 hr. This suggests that the diurnal spikes in FIB concentrations result from sewage exfiltration and flow into surface drains through preferential flow paths.

5.2. Data analysis and discussion

5.2.1. Diurnal variation of FIB concentration

The Toolbox Study (Section 3.3.2) highlighted that FIB concentrations may reveal diurnal patterns and Section 5.1.3 above provides a possible explanation for this observation. An analysis of the data obtained from continuous 24-hr sampling at the six high-density residential areas and five low-density residential areas (Table 5.1) was performed in order to elucidate possible temporal patterns in the data. To do this, a normalisation methodology was developed in which log-transformed FIB concentrations were normalised based on the minimum (min) and maximum (max) log-transformed FIB concentrations for each sampling set (Section 4.5.7, Equation 4.13). This method of normalisation accentuates the “shape” of the 24-hr FIB concentration data, and allows the identification of a diurnal pattern, if any. It also allows for comparison across sites since the effect of different sites having different ranges of background FIB concentrations is removed.

All normalised FIB concentrations were chained into a continuous artificial time series in order to observe if a repeatable pattern across sampling sites exists. The methodology to create the artificial time series is described in Section 4.5.8. The normalised FIB concentrations of Lorong 8 Toa Payoh (9-10 January 2012, 17-18 January 2012 and 18-19 January 2012), MacPherson Lane (20-21 June 2012) and Ang Mo Kio Avenue 10 (27-28 June 2012) had to be rearranged because sampling started

at 08:00 at these sites, unlike the other sites where sampling started at 20:00. The normalised 24-hr sampling data were then chained in sequence in the order of the sampling dates shown in Table 5.1. The artificial time series are shown in Figure 5.6 where each 24-hr sequence represents a complete 24-hr set of samples starting at 20:00 at a particular day and location. It can be observed that generally the plots show high FIB concentrations during the daytime (second half of each 24-hr interval). This pattern appears repeatable despite the fact that the time series was formed by chaining data from different sites. This shows qualitatively that the diurnal pattern is prevalent across the different sites.

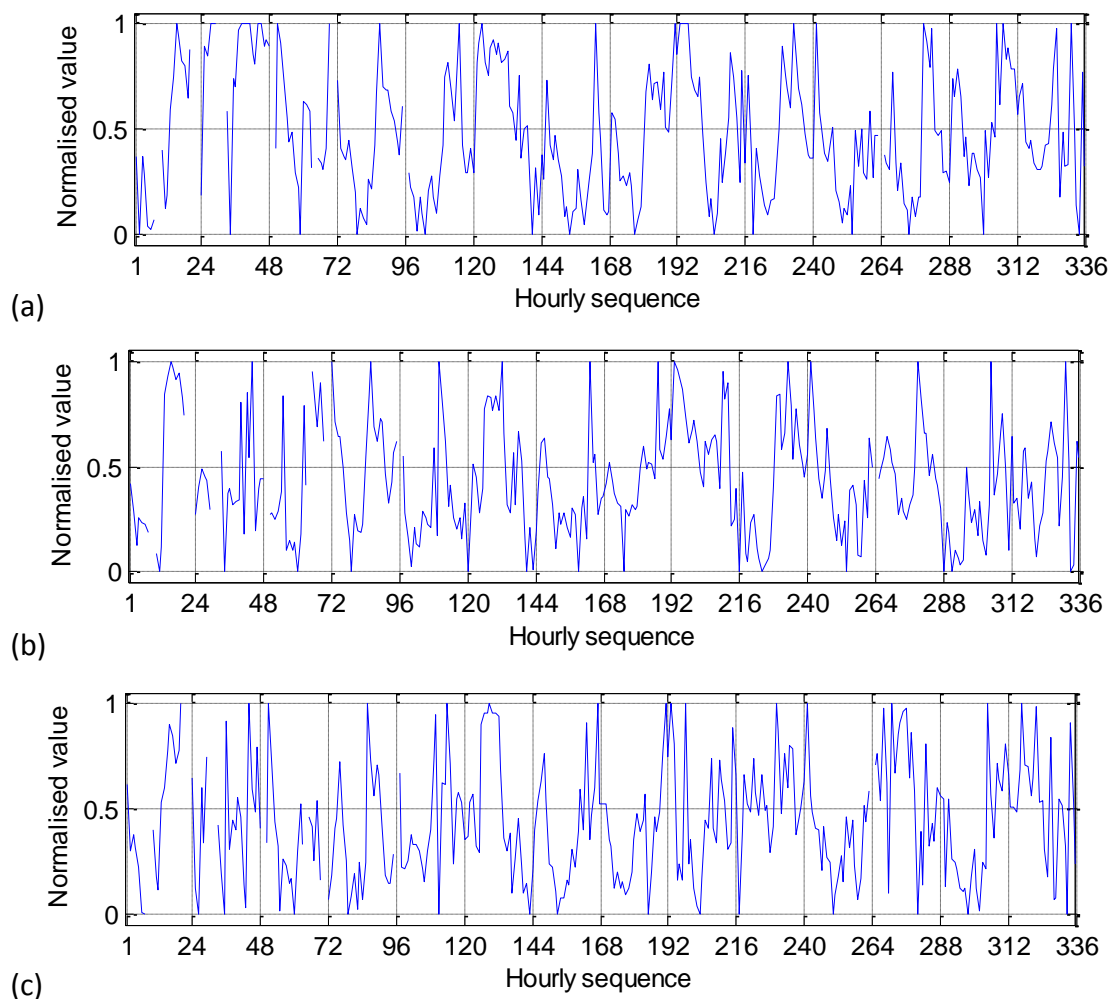
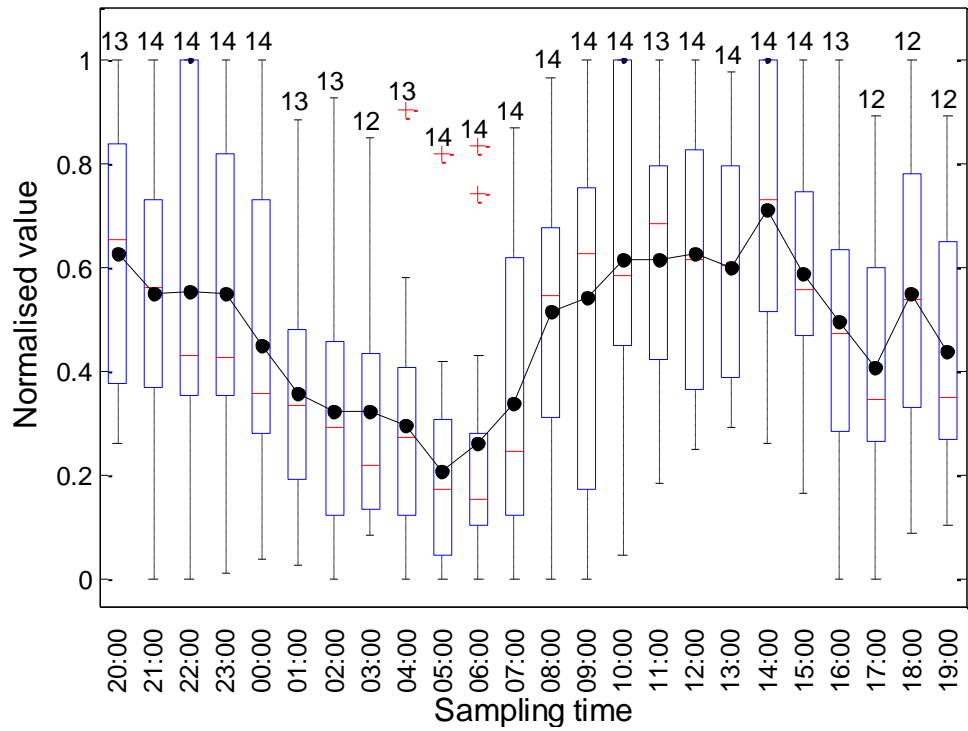
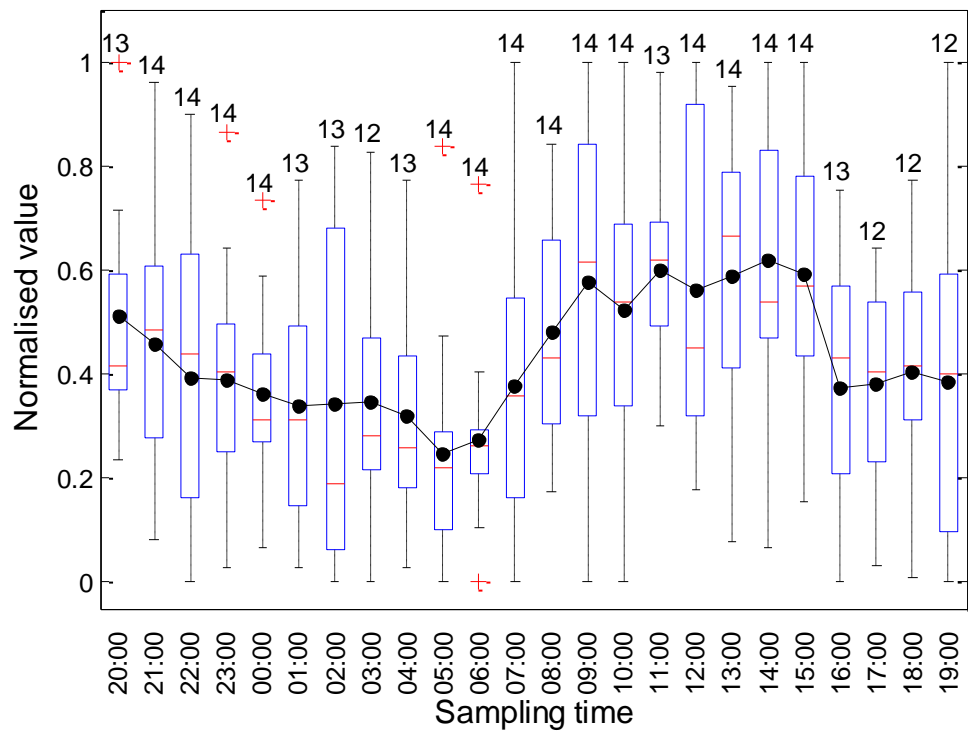


Figure 5.6. Artificial FIB time series: (a) TC (b) *E. coli* (c) enterococci. Time sequence = 1 corresponds to 20:00

The normalised FIB concentrations were grouped based on sampling time and shown as box-and-whisker plots in Figure 5.7 for TC, *E. coli* and enterococci. The arithmetic means of the normalised FIB concentrations are superimposed on the box plots with a black circle symbol (●) connected by trend lines. The number above each data point indicates the number of samples. The trends of the arithmetic mean and median confirm the results observed in the artificial time series that generally FIB concentrations are high in the daytime, peaking at around 10:00 – 14:00 and low in the nighttime, with a minimum at around 04:00 – 06:00. This trend is more significant for TC and *E. coli* and less so for enterococci. The diurnal pattern of FIB concentrations in the present study is different from other studies (Traister and Anisfeld 2006; Desai and Rifai 2013). In those studies, the diurnal pattern of FIB concentration is marked by high FIB concentrations in the nighttime and low concentrations in the daytime due to sunlight-induced die-off. The difference in diurnal variations between the present study and these published studies may be attributed to less influence of sunlight in this study since most major drains in Singapore are covered or shaded by buildings and trees. However, if the absence of sunlight is the only difference, no pattern would be observed in this study because other studies reported insignificant differences between daytime and nighttime FIB concentrations for samples collected from shaded areas (Traister and Anisfeld 2006; Desai and Rifai 2013). Therefore, the observed diurnal pattern in this study, which is marked by variations similar to sewage flow patterns, clearly points to leaking sewers as the source of contamination (Section 5.1.2 and verified in Section 5.1.3).

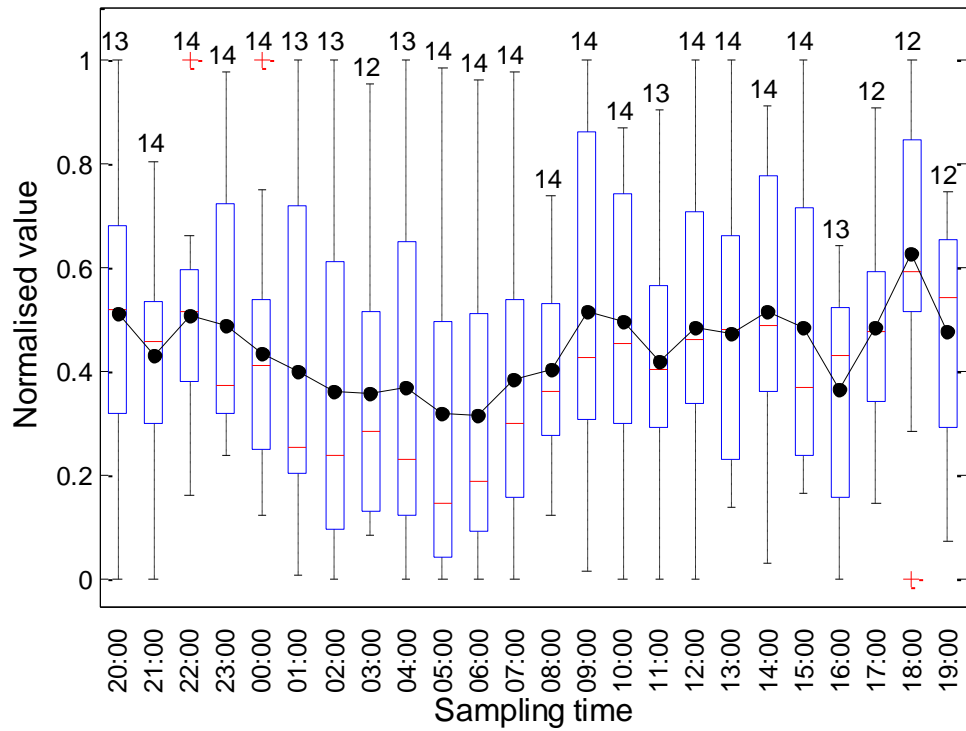


(a)



(b)

Figure 5.7. Box plots of normalised FIB concentrations grouped based on sampling time: (a) TC (b) *E. coli* (c) enterococci



(c)

Figure 5.7. (continued) Box plots of normalised FIB concentrations grouped based on sampling time: (a) TC (b) *E. coli* (c) enterococci

The arithmetic means of the normalised FIB concentrations shown in Figure 5.7 are re-plotted in Figure 5.8 showing the 95% confidence intervals (represented by dashed lines). The confidence interval was found with the test of two-sided *t*-interval (Tamhane and Dunlop 2000).

$$\bar{x} - t_{N-1, \alpha/2} \frac{s}{\sqrt{N}} \leq \mu \leq \bar{x} + t_{N-1, \alpha/2} \frac{s}{\sqrt{N}} \quad (5.1)$$

where \bar{x} is the observed average normalised value of each sampling hour, N is the number of normalised value for each sampling hour, $t_{N-1, \alpha/2}$ is the $\alpha/2$ critical point of the *t*-distribution with $N - 1$ degrees of freedom, s is the observed standard deviation of the normalised values for each sampling hour and μ is the true mean of normalised value of each sampling hour. The `ttest` function in Matlab (The MathWorks, Inc., Natick, MA, USA) was used for this analysis. The population means

(true means) of normalised FIB concentrations are within these error bars with 95% confidence.

Figures 5.7 and 5.8 show similar patterns for the different FIB. *E. coli* shows a more distinct diurnal pattern compared to enterococci and is likely to be more closely associated with human faecal and sewage contamination than enterococci. For example, Duris et al. (2013) reported that *E. coli* is better correlated with the human *Bacteroides* genetic marker than enterococci. They also found that the correlation of *E. coli* with pathogen genes was higher compared to enterococci. In the same study, Duris et al. (2013) also reported that small positive and significant correlation was found between *E. coli* and caffeine, a common indicator for sewage contamination. On the other hand, there was no significant correlation found between enterococci and caffeine. Sauer et al. (2011) showed that enterococci analysed with QPCR did not correlate with the human *Bacteroides* genetic marker, while *E. coli* exhibited better correlation. In general, the present findings and these other studies show that *E. coli* is more closely associated with human faecal and sewage contamination than enterococci, and is therefore a better indicator.

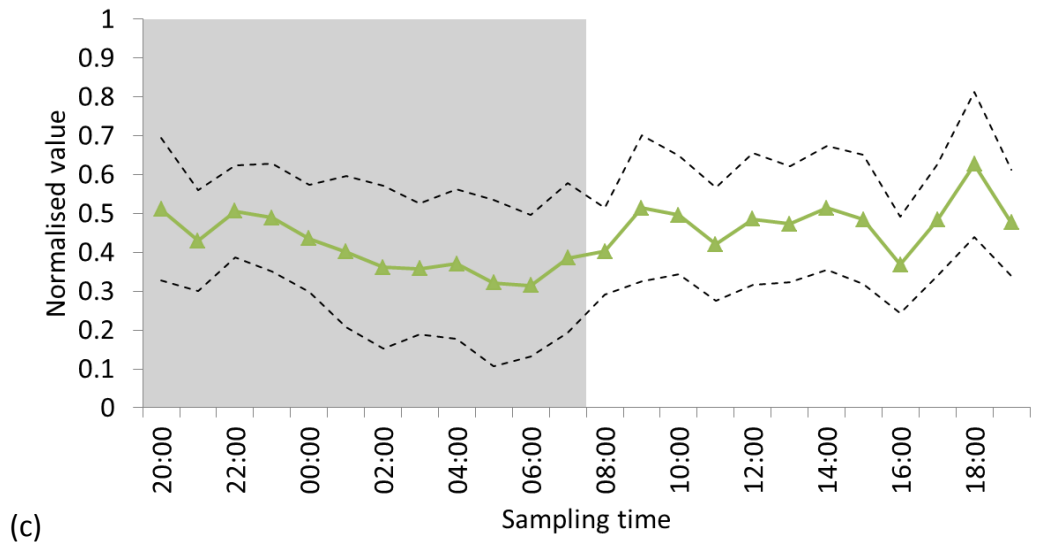
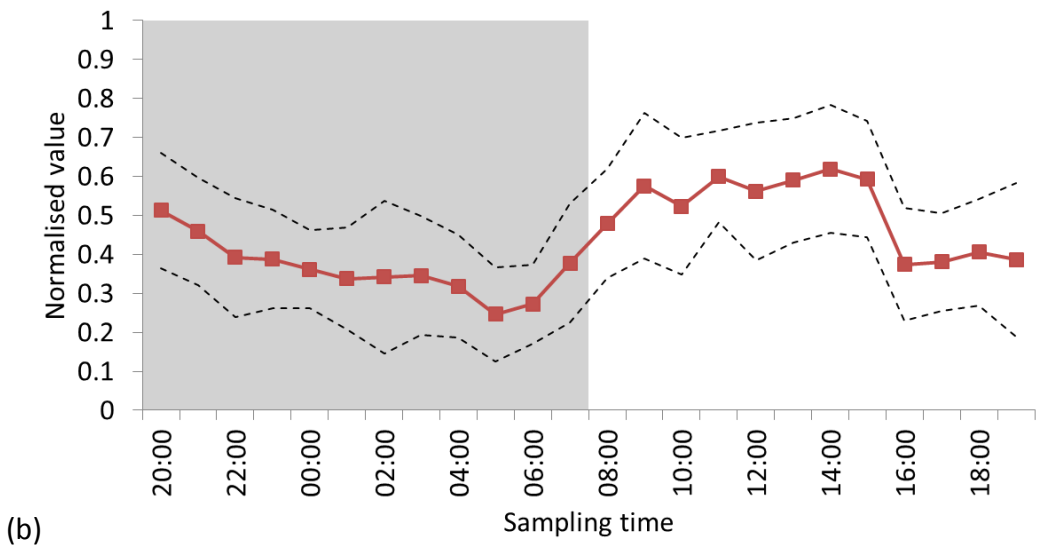
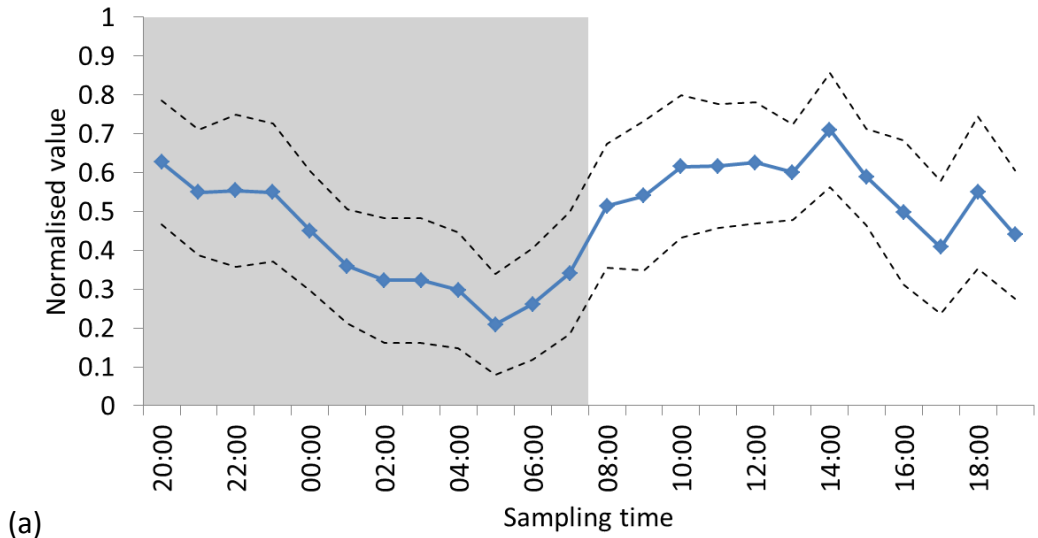


Figure 5.8. Arithmetic means of normalised FIB concentrations with 95% confidence intervals: (a) TC (b) *E. coli* (c) enterococci

5.2.2. Atypical trends in diurnal FIB concentration

In this section, each 24-hr set of samples is individually analysed in order to identify sites where the diurnal variation of FIB may deviate from the typical diurnal pattern characterised by high daytime FIB concentrations and low nighttime FIB concentrations (Section 5.2.1). Closer inspection of each 24-hr plot (Section 5.1.2) however, reveals that FIB concentrations do not follow the typical diurnal pattern at all sampling locations. An investigation was therefore undertaken to identify the likely causes.

The root-mean-square error (RMSE) statistic (Section 4.5.9, Equation 4.14) was employed to quantify deviations of FIB concentrations from a true diurnal pattern. RMSE values for TC, *E. coli* and enterococci of each 24-hr set of samples were summed and the results are shown in Table 5.3. The sampling sites with the highest sum of RMSE values are highlighted in the table. Observations of the FIB data for these sites (Figure 5.3 (b), Figure 5.3 (f) and Figure 5.3 (i)) reveal atypical behaviour, with high nighttime FIB concentrations in addition to high daytime FIB concentrations. In addition, the sites with the three highest RMSE values for enterococci are also highlighted in Table 5.3, the significance of which is discussed in Section 5.2.4.

Table 5.3 shows that the normalised FIB concentrations at Punggol Central have the highest sum of RMSE values. Punggol Central is a newly developed high-density residential estate with several construction works in progress. Punggol Central exhibits consistently elevated nighttime FIB concentrations in samples from 01:00 to 07:00 (Figure 5.3 (f)). The high nighttime FIB concentrations originated from unknown sources (presumably from the construction sites since these are the main land use differences compared to other high-density residential areas observed during the sampling periods) during the night. The normalised FIB concentrations at Lorong 8 Toa Payoh (17-18 January 2012) and Ang Mo Kio Avenue 10 have the second highest sums of RMSE. Both Ang Mo Kio and Toa Payoh are mature high-density housing

estates and the high RMSE values can be attributed to the existence of wet markets located within these catchments (wet markets are not present in other catchments). The activities at wet markets start in the early hours of the morning. Although the used water from wet markets is supposed to be channelled into the sewer system, it is likely that some of the used water overflows to surface drains. Depending on market operations, the impact of wet market activities on the measured FIB concentrations may not be consistent throughout the sampling campaign. This provides an added degree of noise in the data compared to sewer flows which peak and trough at relatively definite times consistent with sewer flow patterns. The variability introduced to the data due to the existence of wet markets at such sites can therefore lead to greater discrepancies in the FIB concentration time series. It is thus observed that two sets of samples from Lorong 8 Toa Payoh (9-10 January 2012 and 18-19 January 2012) can be classified as showing typical diurnal variation while one set of samples from Lorong 8 Toa Payoh (17-18 January 2012) shows atypical diurnal variation.

Table 5.3. RMSE values for TC, *E. coli* and enterococci

Sampling site	Sampling date	RMSE			Sum of RMSE	Type of diurnal variation
		TC	<i>E. coli</i>	Enterococci		
Lorong 8 Toa Payoh	9-10 January 2012	0.27	0.26	0.27	0.81	typical
	17-18 January 2012	0.39	0.19	0.27	0.86	atypical
	18-19 January 2012	0.27	0.27	0.22	0.76	typical
Jalan Siantan	6-7 March 2012	0.16	0.20	0.25	0.61	typical
	19-20 March 2012	0.23	0.24	0.23	0.70	typical
Punggol Central	21-22 May 2012	0.39	0.33	0.39*	1.11	atypical
Lorong 6 Toa Payoh	12-13 June 2012	0.27	0.22	0.21	0.70	typical
MacPherson Lane	20-21 June 2012	0.19	0.16	0.21	0.57	typical
Ang Mo Kio Avenue 10	27-28 June 2012	0.30	0.30	0.26	0.86	atypical
Serangoon Garden	9-10 July 2012	0.20	0.23	0.23	0.67	typical
Toa Payoh North	16-17 July 2012	0.23	0.28	0.22	0.72	typical
Watten Drive	23-24 July 2012	0.23	0.17	0.35*	0.75	mixed**
CCK Crescent	29-30 July 2012	0.20	0.27	0.25	0.72	typical
Verde	30-31 July 2012	0.23	0.20	0.30*	0.73	mixed**

Bold type indicates sites with atypical diurnal pattern. * denotes three highest RMSE values for enterococci. ** mixed refer to typical diurnal behaviour for *E. coli* and TC but atypical for enterococci.

5.2.3. Autocorrelation of FIB time series

Autocorrelation analysis (Section 4.5.6, Equation 4.12) was carried out to detect repeating patterns in the data. The analysis was conducted on the normalised FIB data from sites that are classified as showing typical or mixed diurnal variations (Table 5.2). The results are plotted as correlograms in Figure 5.9 where the dashed lines represent the 95% confidence limits. Since eleven of the 24-hr sample sets showed typical FIB variations, there are 264 (11 x 24) sequences used in this autocorrelation analysis. However, out of the 264 sequences, 11 data points were not considered in the analysis because the samples were either compromised by rain (8 samples) or were not collected due to equipment failure (3 samples). Sixty-six (264/4) lags were thus used for the autocorrelation analysis.

Figure 5.9 indicates that the correlograms peak at 24 and 48 hr, except for enterococci, which peaks only at 48 hr. The autocorrelation coefficients at 24- and 48-hr lag for TC and at 48-hr lag for enterococci exceed the 95% confidence limits indicating that these correlations are significant at $p < 0.05$. The autocorrelation coefficients at 24- and 48-hr lag for *E. coli* are only slightly below the 95% confidence limits. The significant positive correlation at 24-hr lag confirms the existence of a typical diurnal cycle. The presence of positive correlation at 24-hr lag indicates that TC and *E. coli* concentrations monitored 24 hr apart are associated. Since the autocorrelation was analysed on normalised FIB concentrations constructed from data obtained from different sampling sites, it means the concentrations monitored 24 hr apart possess similar trends, reaching high or low values at regular intervals within a 24-hr period.

Although the correlations fluctuate over the range of the lags with peaks repeated every 24 hr (Figure 5.9 (a) and (b)), the correlations at 48 hr are about the same magnitude as at 24 hr. The regular repetition of the autocorrelation function without the presence of an appreciable overall trend indicates that the FIB variation is

stationary. This implies that peaks in FIB concentrations occur regularly at approximately the same times across all sites and sampling days.

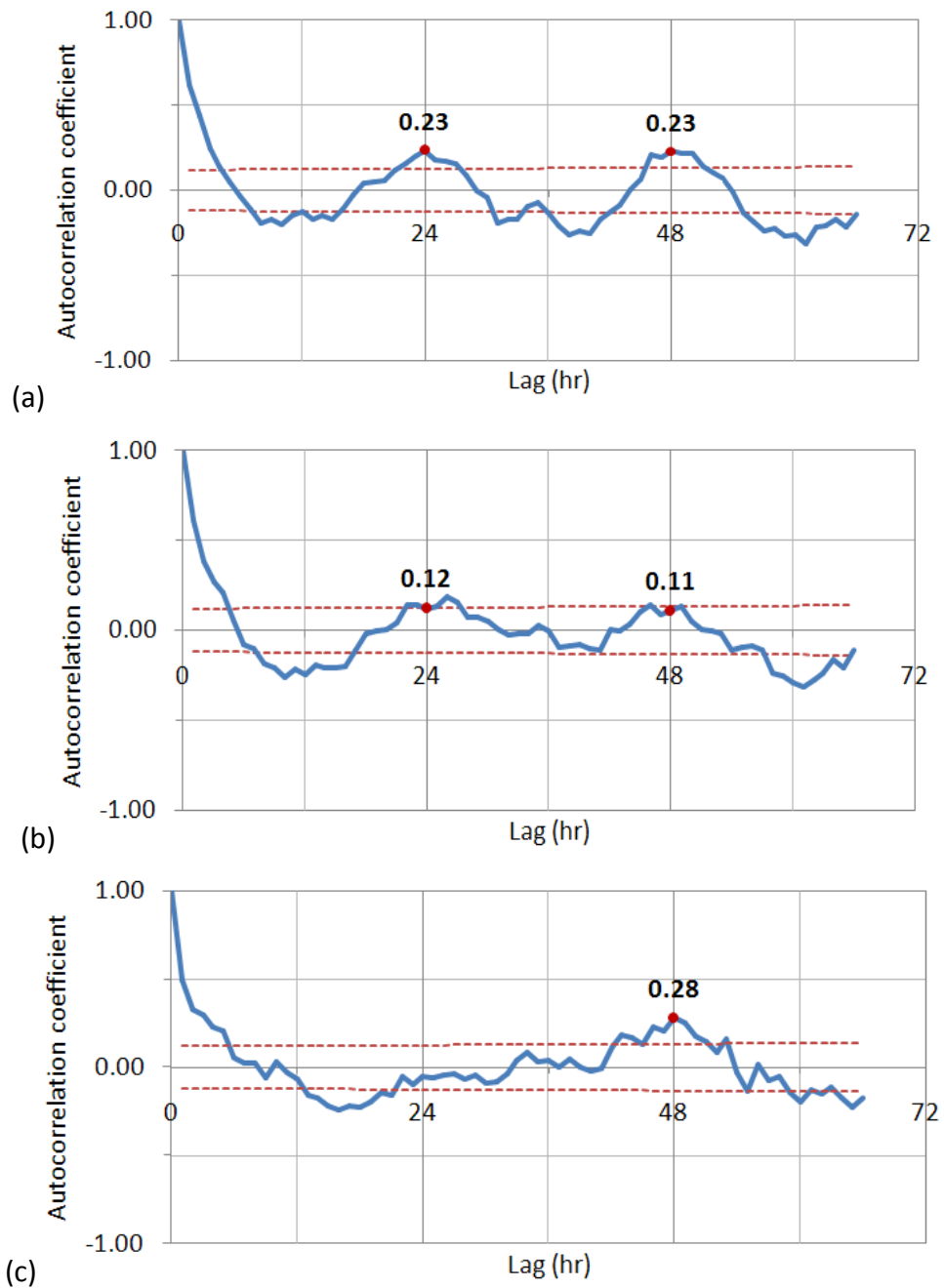


Figure 5.9. Correlograms of typical diurnal FIB concentrations: (a) TC (b) *E. coli* (c) enterococci

5.2.4. Autocorrelation function for enterococci

Unlike the autocorrelation functions for TC and *E. coli*, which peak at lags of 24 and 48 hr, enterococci only peaks at a lag of 48 hr (Section 5.2.3). Further analysis was carried out to investigate the lack of a significant peak of the autocorrelation function for enterococci at a lag of 24 hr. It is observed from Table 5.2 that the sets of 24-hr samples for Watten Drive and Verde have the highest RMSE values for enterococci, after Punggol Central (which is classified as site showing atypical diurnal variation due to the presence of construction activities). Despite the high RMSE values for enterococci, RMSE values for *E. coli* at Watten Drive and Verde are low (Table 5.2). This shows that *E. coli* is a better indicator than enterococci because *E. coli* is more closely associated with human faecal and sewage contamination as discussed in Section 5.2.1. It means that *E. coli* is insensitive to noise but shows greater response to contamination compared to enterococci. This is in agreement with the observation in Section 5.2.1 that *E. coli* shows stronger diurnal variation than enterococci. Other studies also reported the superiority of *E. coli* compared to enterococci in freshwater. Chua et al. (2010) reported that mean dry weather enterococci concentrations were not significantly correlated with percentage development of catchment land use while *E. coli* were significantly correlated. In addition, USEPA (1986) also reported that in freshwater the correlation of enterococci with swimming-associated gastroenteritis rates is slightly lower than *E. coli* while in marine water, enterococci are more strongly correlated than *E. coli*, as summarised in Table 2.2. Therefore, enterococci may be less suitable in freshwater (like drain water that was collected in this study) and instead more suitable in marine water (beach/shoreline).

Autocorrelation analysis for FIB data was repeated after excluding the data from Watten Drive and Verde. Hence, a total of nine 24-hr sets of samples were analysed, giving rise to 216 (9 x 24) data points used in the autocorrelation analysis. The number of lags considered was 54 (216 / 4) lags. Figure 5.10 shows correlograms with the 95% confidence limits represented by dashed lines. Figure 5.10 (c) indicates that with the exclusion of data from Watten Drive and Verde, the correlograms peak at 24

and 48 hr for enterococci. The autocorrelation coefficients at 24- and 48-hr lag are positive and exceed the 95% confidence limits (significant at $p < 0.05$). This is a significantly different result than Figure 5.9 (c) where enterococci did not peak at the 24-hr lag. Comparisons of Figure 5.10 (a) and (b) with Figure 5.9 (a) and (b) show that the autocorrelation coefficients at 24- and 48-hr lag for TC are slightly higher with this reduced dataset, while for *E. coli*, the autocorrelation coefficients are now significant at the 5% level (exceed 95% confidence limits).

Because the enterococci diurnal variations for Watten Drive and Verde are atypical, the sets of samples from Watten Drive and Verde are excluded from the further analysis in this chapter. Diurnal pattern of enterococci in Watten Drive and Verde may be masked by background concentrations because of weaker leaking-sewer signal. Watten Drive and Verde may have weak leaking-sewer signals because their catchment sizes are smaller and the percentages of low-density residential land use are also smaller than the other low-density residential areas sampled in this study (Table 4.1). The smaller catchment sizes and percentages of low-density residential indicate less number of building or property connections in these areas. It is suspected in Section 5.1.3 that sewer leakage to storm drains occurs at sewers connecting discharge pipes from buildings or properties to the public sewers which are generally shallow and above the groundwater table. Therefore, the lesser number of building or property connections in Watten Drive and Verde may have caused the weaker leaking sewer signal.

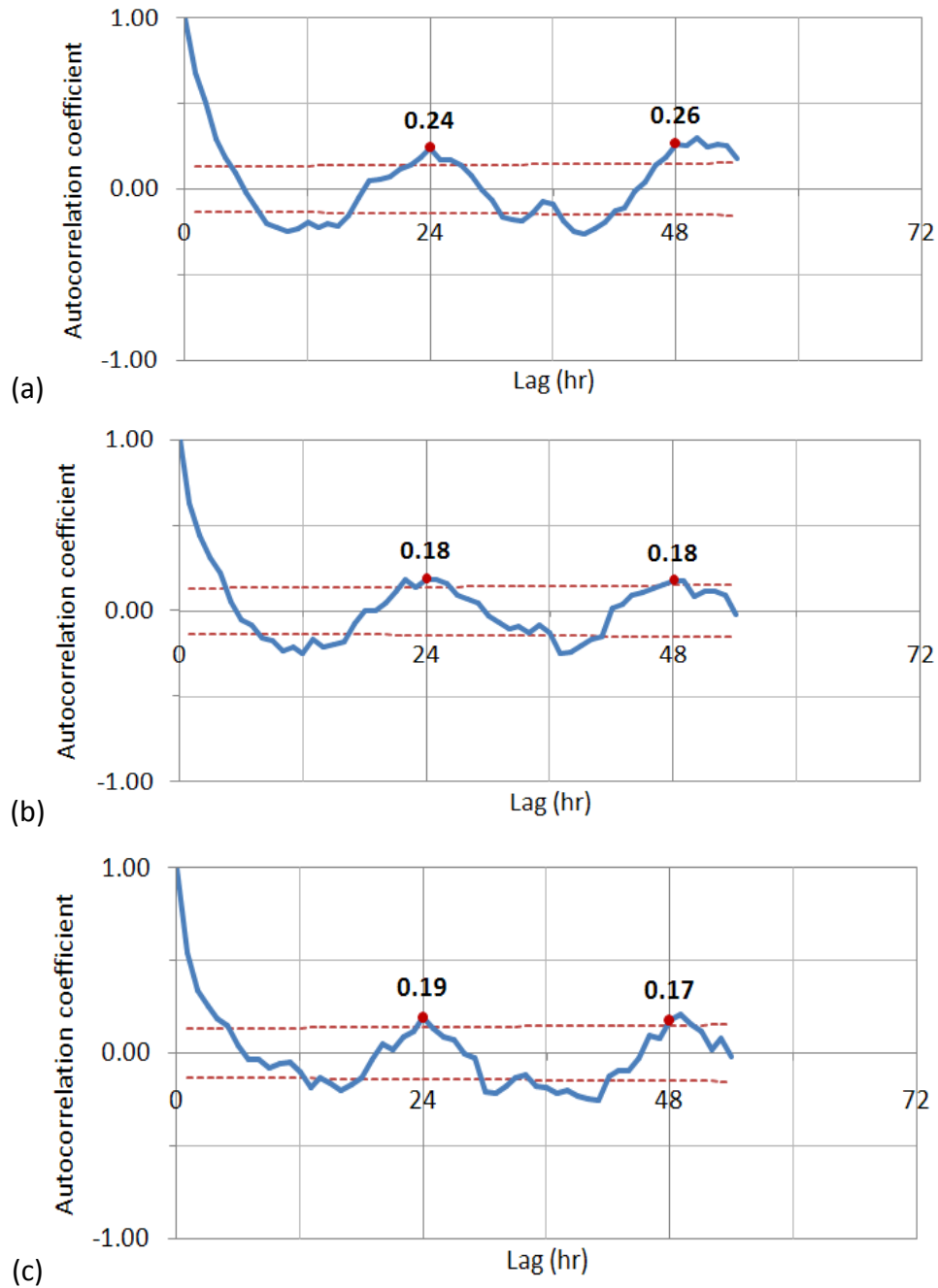


Figure 5.10. Correlograms of typical diurnal FIB concentrations excluding samples from Watten Drive and Verde: (a) TC (b) *E. coli* (c) enterococci

The arithmetic means of normalised FIB concentrations with typical diurnal variation, plotted above in Figure 5.8, are again plotted in Figure 5.11, but on a 36-hour time scale where the readings at 20:00 to 07:00 are repeated at the end (19:00) of the time series. Figure 5.11 thus allows the patterns during the daytime-nighttime

transition to be visualised more easily. Figure 5.11 also differs from Figure 5.8 in that FIB concentrations from Watten Drive and Verde are excluded. In contrast to Figure 5.8, which includes FIB concentrations at all stations showing typical diurnal variation of at least TC and *E. coli*, in Figure 5.11 all three FIB, including enterococci, show similarly high daytime FIB concentrations and low nighttime FIB concentrations. In addition, Figure 5.11 shows that FIB spike strongly in the morning and again, but more weakly, in the evening. The morning and evening spikes mimic the peaks in typical daily sewage flow patterns (Enfinger and Stevens 2006).

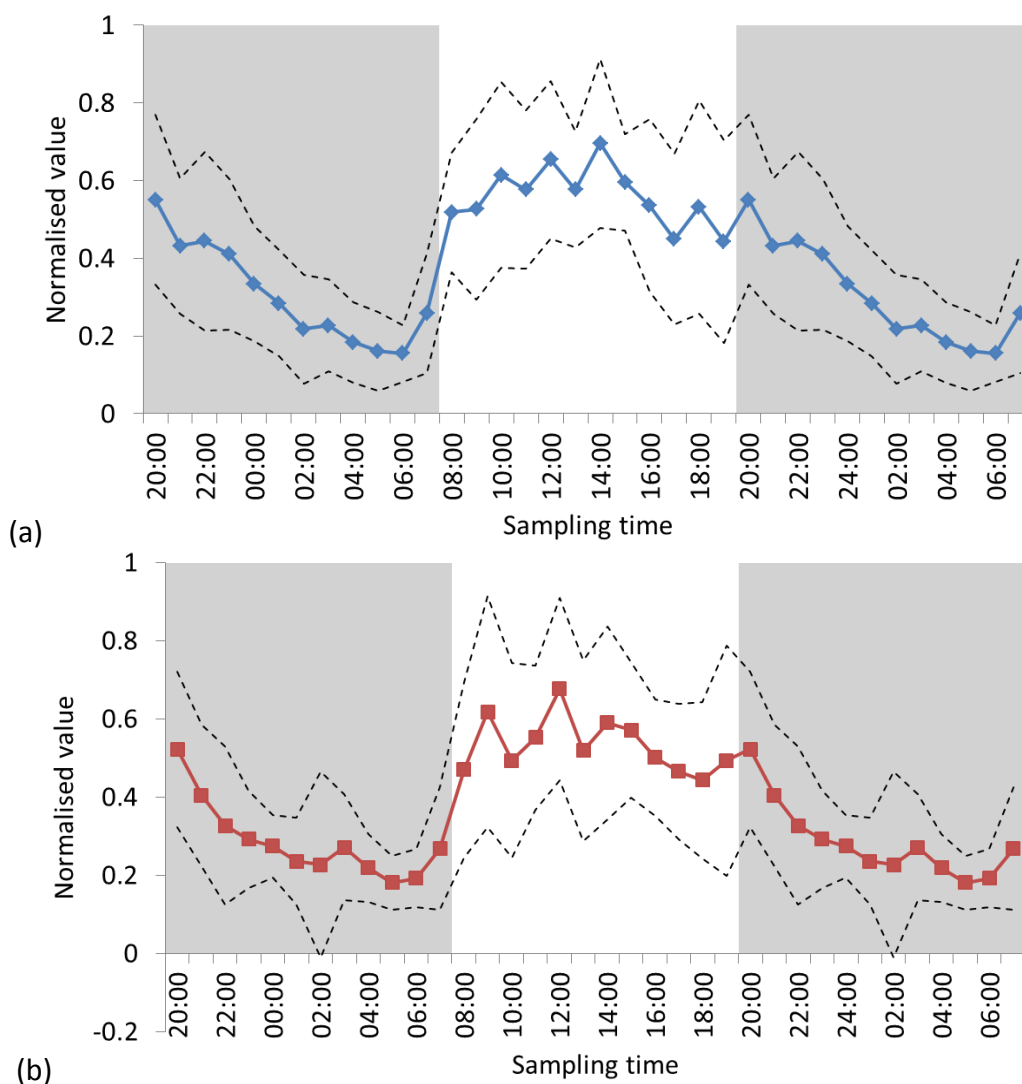
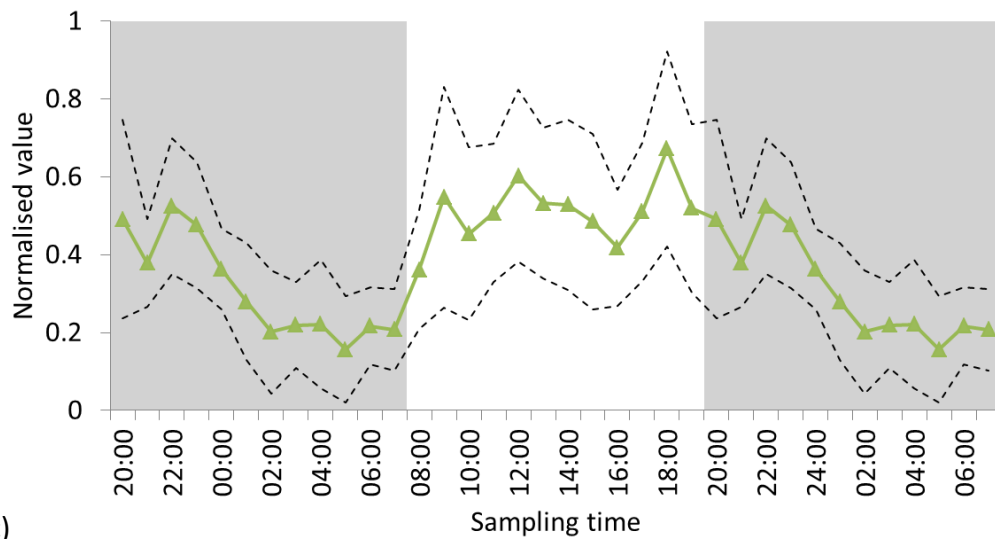


Figure 5.11. Arithmetic means of normalised FIB concentrations of samples obtained from sites exhibiting typical variation in FIB excluding samples from Watten Drive and Verde (Table 5.2) with 95% confidence intervals: (a) TC (b) *E. coli* (c) enterococci



(c) Figure 5.11. (continued) Arithmetic means of normalised FIB concentrations of samples obtained from sites exhibiting typical variation in FIB excluding samples from Watten Drive and Verde (Table 5.2) with 95% confidence intervals: (a) TC (b) *E. coli* (c) enterococci

Figure 5.11 also shows the 95% confidence interval (obtained using Equation 5.1 in Section 5.2.1) represented by dashed lines. The plots in Figure 5.11 have narrower confidence intervals for nighttime samples compared to Figure 5.8. This shows that the FIB concentrations at Punggol Central, Lorong 8 Toa Payoh, Ang Mo Kio Avenue 10, Verde and Watten Drive are high in the nighttime, in addition to being high in the daytime. This has been attributed to activities at construction sites and wet markets during the nighttime at Punggol Central, Lorong 8 Toa Payoh and Ang Mo Kio Avenue 10 (Section 5.2.2). For Verde and Watten Drive, the greater noise in the concentration data masks the weaker FIB signal, if any, from the leaking sewers in these low-density residential areas. Figure 5.11 (b) also shows that the lower bound of the confidence interval of *E. coli* at 02:00 is less than zero. This indicates that it is possible for the 02:00 *E. coli* population mean (true mean) concentration to be lower than the minimum *E. coli* concentrations observed in this study (that is 52 MPN/100mL).

5.2.5. Comparison of FIB concentration between daytime and nighttime

The Wilcoxon-Mann-Whitney test (Section 4.5.11) was conducted at the 5% significance level to test for statistical differences between daytime and nighttime FIB concentrations for the sites that exhibit typical FIB behaviour (Table 5.3). Prior to the analysis, FIB concentrations were first segregated into daytime and nighttime datasets. The results of this analysis are expressed as p values. The p values of all FIB are lower than 0.05 (p values for TC = 2×10^{-4} , *E. coli* = 3×10^{-5} and enterococci = 2×10^{-3}). This indicates that the difference in concentration between daytime and nighttime samples is significant and that daytime geometric mean concentrations are higher than those at nighttime (Figure 5.12). In contrast, Boehm et al. (2002), Traister and Anisfeld (2006), Enns et al. (2012) and Desai and Rifai (2013) observed low FIB concentrations during the daytime due to sunlight-induced die-off. The contrasting observations show that sunlight-induced die-off is not significant in determining variations in FIB concentrations for the sites considered here. Instead, the exfiltration of sewage seems to be the determining factor affecting the observed diurnal patterns of FIB concentrations.

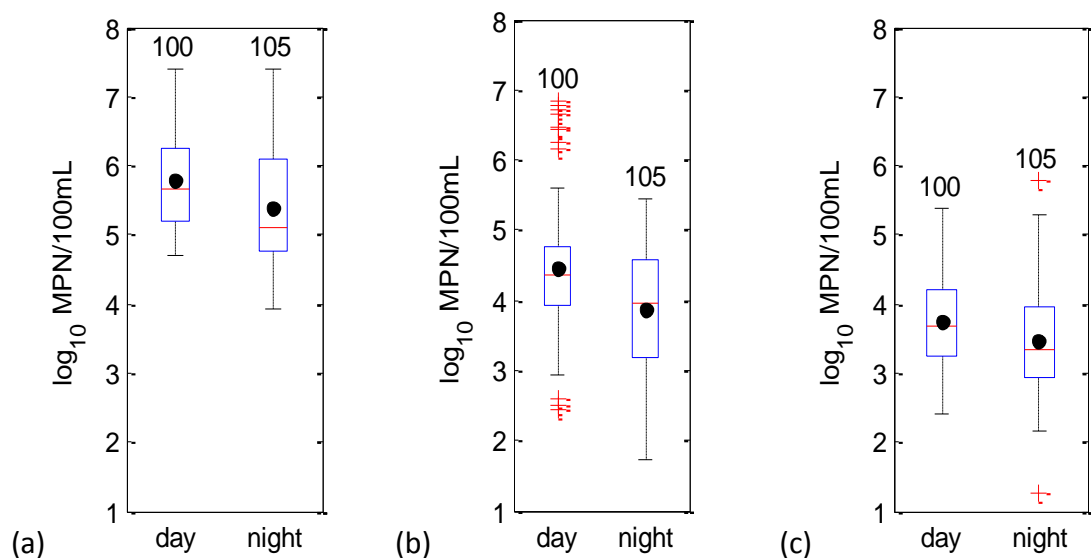


Figure 5.12. Box plots of FIB concentrations grouped based on daytime and nighttime: (a) TC (b) *E. coli* (c) enterococci

Since significant differences exist between daytime and nighttime FIB concentrations and daytime FIB concentrations are higher than nighttime FIB concentrations, it is important to ascertain the time at which samples should be taken when seeking to provide an estimate of the maximum FIB concentration in a given day. Therefore, the daytime and nighttime groups of sampling times were divided into sub-groups with narrower ranges of sampling time. To do this, the 24-hr data were first segregated into four six-hour groups. Next, the Kruskal-Wallis test (Section 4.5.13) was carried out to test if any group of samples would yield results that are significantly (at 5% level) different from the other groups. Different ways of segregating the data into four six-hour groups were tested on the condition that the groups must clearly differentiate between daytime and nighttime. As mentioned in Section 5.1, sampling mostly started at 20:00, therefore 20:00 was taken as the start time of the first of the four six-hour groups (20:00 – 01:00, 02:00 – 07:00, 08:00 – 13:00 and 14:00 – 19:00). The second four six-hour groups consisted of 21:00 – 02:00, 03:00 – 08:00, 09:00 – 14:00 and 15:00 – 20:00 groups of data. Similarly, the last grouping considered was 01:00 – 06:00, 07:00 – 12:00, 13:00 – 18:00 and 19:00 – 24:00. Among the three tested ways of segregating the data into four six-hour groups, the first method (20:00 – 01:00, 02:00 – 07:00, 08:00 – 13:00 and 14:00 – 19:00) resulted in the lowest p values for all three FIB. This implies that this particular choice of four six-hour groups is the most discriminating among the three methods of groupings tested.

When the Kruskal-Wallis tests showed that differences between the groups were significant ($p \leq 0.05$), the Tukey-Kramer test (Section 4.5.14) was conducted to identify the groups of samples that gave significantly (at 5% level) different results. The Kruskal-Wallis test showed significant differences ($p \leq 0.05$) among the four groups for all three FIB, with the significantly different groups for the first four six-hour groups identified by the Tukey-Kramer test as shown in Table 5.4. The results show that TC and *E. coli* concentrations over two six-hour groups (08:00 – 13:00 and 14:00 – 19:00) are significantly different from those at 02:00 – 07:00, as illustrated in Table 5.4 (a). Enterococci concentrations for three six-hour groups (08:00 – 13:00, 14:00 –

19:00 and 20:00 – 01:00) are also significantly different from those collected at 02:00 – 07:00.

A repeat analysis was carried out using six four-hour groups. Similar to the six-hour groupings, three combinations of four-hour groupings were tested and it was found that the grouping 20:00 – 23:00, 00:00 – 03:00, 04:00 – 07:00, 08:00 – 11:00, 12:00 – 15:00 and 16:00 – 19:00 produced the lowest p values for all three FIB and is the most discriminating sampling scheme.

The Kruskal-Wallis test showed significant differences among the six groups ($p \leq 0.05$) for all three FIB and the Tukey-Kramer test results for the first six four-hour groups are shown in Table 5.4 (b). The results indicate that TC and *E. coli* concentrations of samples collected at 08:00 – 11:00 are significantly different from samples collected at 04:00 – 07:00. In addition, the concentrations of TC and *E. coli* in samples collected at 12:00 – 15:00 are significantly different from samples collected at 00:00 – 03:00 and 04:00 – 07:00. Similarly, enterococci concentrations at 12:00 – 15:00 are significantly different from those collected at 04:00 – 07:00.

The analysis in this section showed that daytime (08:00 – 19:00) FIB concentrations are significantly higher than nighttime (20:00 – 07:00) FIB concentrations. Specifically, samples collected at 12:00 – 15:00 are significantly higher than samples collected at 04:00 – 07:00. This implies that sampling times from 12:00 to 15:00 and from 04:00 to 07:00 capture the salient characteristics of daytime and nighttime, respectively.

Table 5.4. Kruskal-Wallis and Tukey-Kramer tests results for (a) six-hr and (b) four-hr groups

(a)

FIB		six-hour groups (20:00 - 01:00, 02:00 - 07:00, 08:00 - 13:00 and 14:00 - 19:00)																						
TC	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00
	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00
<i>E. coli</i>	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00
	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00
Enterococci	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00
	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00

(b)

FIB		four-hour groups (20:00 - 23:00, 00:00 - 03:00, 04:00 - 07:00, 08:00 - 11:00, 12:00 - 15:00 and 16:00 - 19:00)																							
TC	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	
	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	
<i>E. coli</i>	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	
	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	
Enterococci	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	
	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	

Colour represents significantly different results. Blue represents low FIB concentration and red represents high FIB concentration.

5.2.6. Significance of sampling time on FIB concentration

Based on the analysis of the six four-hour groups, sampling should be conducted hourly from 12:00 to 15:00 in order to obtain maximum concentration estimates and from 04:00 to 07:00 to obtain estimates of minimum FIB concentrations. However, it is impractical to sample two ranges of hourly sampling times (04:00 – 07:00 and 12:00 – 15:00) on a routine basis. Therefore, the Wilcoxon-Mann-Whitney test was used to find the minimum number of samples to be collected during the ranges of sampling

times identified so as to obtain FIB concentrations that are significantly different (at 5% level) from the FIB concentrations during the remaining hours of the day and which would be indicative of the maximum and minimum concentrations. The Wilcoxon-Mann-Whitney test was thus conducted for sites showing typical diurnal behaviour (Table 5.3). The FIB concentrations were segregated into two groups based on the sampling time. The first group contained samples from the suggested four-hour sampling time (12:00 – 15:00 or 04:00 – 07:00), while the second group contained samples from the other sampling times throughout the day. An incremental procedure was applied to the first group in order to arrive at the minimum number of sampling times required. Therefore, FIB concentrations for hourly samples collected during the four-hour sampling time were compared, one at a time, to the samples collected at the remaining hours of the day to see if samples collected during a particular hour would result in significantly different FIB concentrations.

The sample collected at 12:00 showed significantly higher *E. coli* ($p = 0.042$) concentrations, but not TC ($p = 0.108$) and enterococci ($p = 0.090$) concentrations. Sampling at 13:00, 14:00 or 15:00 also failed to reveal significantly higher concentrations for all FIB. Therefore, combinations of two sampling times at 12:00 and 13:00, at 12:00 and 14:00 and at 12:00 and 15:00 were tested to see if sampling at two different times would result in significantly higher concentrations for all three FIB. It was found that sampling at 12:00 and 14:00 resulted in significantly higher concentrations for all three FIB (p values for TC = 0.017, *E. coli* = 0.016 and enterococci = 0.043) while sampling at 12:00 and 13:00 or at 12:00 and 15:00, did not. Thus, a sampling program that collects samples at 12:00 and 14:00 is recommended as an efficient sampling strategy to obtain the expected daily maximum FIB concentrations.

In the same way, samples collected at 05:00 showed significantly lower enterococci ($p = 0.048$) concentrations, but not TC ($p = 0.143$) or *E. coli* ($p = 0.103$) concentrations

and similarly sampling at 04:00, 06:00 or 07:00 did not reveal significantly lower concentrations for all FIB. Therefore, combinations of two sampling times at 05:00 and 04:00, at 05:00 and 06:00 and at 05:00 and 07:00 were tested to see if sampling at two different times would result in significantly lower concentrations for all three FIB. It was found that sampling at 05:00 and 04:00 and at 05:00 and 06:00 resulted in significantly lower concentrations for all three FIB. However, sampling at 05:00 and 04:00 resulted in lower p values (p for TC = 0.026, *E. coli* = 0.010 and enterococci = 0.010) than sampling at 05:00 and 06:00 (p for TC = 0.032, *E. coli* = 0.016 and enterococci = 0.014). Samples collected at 05:00 and 07:00 showed significantly lower enterococci (p = 0.015) concentrations, but not for TC (p = 0.110) and *E. coli* (p = 0.077). A sampling program can collect samples at either 05:00 and 04:00 or 05:00 and 06:00. When a sampling is for a routine monitoring program, sampling at 05:00 and 06:00, rather than sampling at 05:00 and 04:00, is recommended as an efficient sampling strategy to obtain the expected daily minimum FIB concentrations.

In general, the suggested sampling times are sufficient to collect samples likely to reveal the presence of bacterial contamination. However, it should be noted that this is applicable to sampling at upstream areas such as those listed in Table 5.1.

5.2.7. Non-conservative FIB behaviour

The diurnal pattern of FIB observed in this study follows a pattern of high daytime FIB and low nighttime FIB concentrations. This is contrary to patterns of FIB observed in studies by Boehm et al. (2002), Whitman and Nevers (2004) and Enns et al. (2012), which are characterised by high nighttime and low daytime FIB concentrations due to the more rapid FIB degradation during the daytime. This in turn has been attributed to sunlight-induced die-off and higher digestion rate of protozoa at higher temperatures (Sherr et al. 1988) causing higher predation rates on the FIB (Walters et al. 2011). The die-off, re-growth, natural occurrence of FIB and the ability for FIB to survive longer in tropical climates have led some authors (Hazen 1988; Rivera et al. 1988; Muñiz et al. 1989) to question the use of FIB in quantifying bacterial

contamination. However, other studies have also observed that daytime and nighttime FIB concentrations were not significantly different in shaded areas (Traister and Anisfeld 2006) and the pattern for samples collected from a site partially covered (underneath a bridge) was not as apparent as the pattern for samples collected from sites exposed to sunlight (Desai and Rifai 2013). As discussed below, the different daytime-and-nighttime pattern observed in the present study arises primarily from the diurnal variation in sewer leakage. However, it may also be influenced by minimal bacterial die-off during the day because the sampling sites are mostly covered or shaded by buildings and trees.

Traister and Anisfeld (2006) and Desai and Rifai (2013) reported a lack of significant daytime-nighttime difference and an absence of diurnal pattern only for samples from shaded or partially covered areas. In contrast, this study reports a distinct daytime-nighttime FIB concentration pattern for samples that are collected from covered or partially covered sites in highly urbanised areas. The diurnal pattern of FIB observed in this study is extracted through a normalisation methodology outline in Section 4.5.7 (Equation 4.13). The repeatability of the observed patterns is confirmed with autocorrelation results that show positive correlation coefficients at 24-hr lag (Section 5.2.3). This diurnal pattern is observed to resemble the daily sewage flow pattern. This was verified in the field and strongly indicates the occurrence of sewer exfiltration and the existence of possible pathways for sewage leaking from sewer lines to storm drains. Therefore, sources such as leaking sewers play a more important role than natural processes (such as bacterial die-off and re-growth) in bacterial contamination of surface drains in highly urban environments and give rise to the distinct diurnal pattern observed in this study.

5.3. Conclusions

Significant background concentrations of FIB were observed at all sampling sites. However, sewage exfiltration from leaking sewers can add to background levels of FIB, giving rise to diurnal patterns in FIB concentrations that are consistent with

sewage flow patterns. Sampling can be designed in an efficient way to find this signal and to determine if there is significant leakage into the drains. The following can be concluded from the work described in this chapter:

(1) Analysis of artificial time series, box plots of normalised FIB concentrations (grouped based on sampling time) and plots of arithmetic means of normalised FIB concentrations (Section 5.2.1) show a diurnal pattern that is characterised by high daytime and low nighttime FIB concentrations. The Wilcoxon-Mann-Whitney test conducted on data obtained from seven sites that exhibited typical diurnal behaviour of FIB shows that daytime FIB concentrations were significantly different from nighttime FIB concentrations and their box plots show that daytime geometric mean concentrations are higher than those at nighttime (Section 5.2.5). At sites where anomalies to the typical variation of FIB concentrations were observed, it was found that activities at construction sites and wet markets (Section 5.2.2) could have been the cause of the high nighttime FIB concentrations observed.

(2) Autocorrelation of chained time series of sets of samples that exhibit typical FIB diurnal pattern confirms the diurnal pattern of FIB concentration, with maximum autocorrelation occurring at intervals of 24 hr (Sections 5.2.3 and 5.2.4). This diurnal variation of FIB closely follows that of sewer flows, as reported by Enfinger and Stevens (2006). Field tracer tests provide qualitative evidence of sewage exfiltration and transport to drains via preferential flow paths. The likely existence of sewage exfiltration and preferential flow paths from sewers to drains indicates the likelihood of surface water contamination due to leaking sewers. Autocorrelation analysis shows that FIB diurnal variation is stationary. This means that the high FIB concentrations occur at about the same time across different sampling sites and different sampling days.

(3) Plots of the arithmetic means of normalised FIB concentrations (Section 5.2.1) and RMSE values of *E. coli* and enterococci (Section 5.2.4) indicate that *E. coli* is a better

indicator than enterococci because *E. coli* is more closely associated with human faecal and sewage contamination as discussed in Section 5.2.1.

(4) The Kruskal-Wallis test followed by the Tukey-Kramer test show that hourly sampling from 04:00-07:00 and from 12:00-15:00 results in significantly different FIB concentration (minimum and maximum, respectively) which can be useful in identifying upstream areas with leaking sewers (Section 5.2.5). The Wilcoxon-Mann-Whitney test shows that sampling at 12:00 and 14:00 results in significantly higher FIB concentrations while sampling at 05:00 and 04:00 or at 05:00 and 06:00 results in significantly lower FIB concentrations than sampling at other hours during the day. This shows the importance of the time of day at which samples are collected at upstream areas.

CHAPTER 6 POTENTIAL CHEMICAL TRACERS

Analysis of Marina catchment data in Section 3.3.1 showed that some chemical compounds together with faecal indicator bacteria (FIB) characteristics could be classified according to their spatial distribution: downstream, midstream and upstream. In addition, analysis of the Toolbox Study data in Section 3.3.2 showed that the chemical compounds seem to be following the same diurnal pattern as FIB. In the field program reported in this chapter, surface water samples were collected from upstream areas with specific land uses in order to further investigate possible associations of the chemical compounds with FIB, land use and age of sewer pipes.

6.1. Results

After a thorough literature review (Section 2.4.2) on different chemical tracers and their use in Singapore, a suite of 18 potential chemical tracers was identified and analysed in this study as described in Chapter 4. A summary of the number of samples analysed for each chemical parameter is presented in Table 6.1. Five chemical parameters were mostly not detected: bromide was not detected in 29 out of 29 samples (100% non-detect), triclosan in 28 out of 29 samples (97% non-detect), DBP in 232 out of 242 samples (96% non-detect), diclofenac in 90 out of 123 samples (73% non-detect) and acesulfame-k in 78 out of 123 samples (63% non-detect). Chemical compounds must be reliably detectable in order to be used as tracers. Hence, bromide, triclosan, DBP, diclofenac and acesulfame-k were excluded from further data analysis.

Table 6.1. Summary of non-detect samples and lab duplicate analysis

Parameter	Number of samples analysed	Number of non-detect samples	% of non-detect samples	Number of duplicates	% of duplicates	Average CV (%)
Chloride	242	0	0	38	16	2
Bromide	29	29	100	1	3	*
Boron	242	1	0	30	12	4
Orthophosphate	242	6	2	30	12	1
Detergent as MBAS	242	18	7	29	12	2
Triclosan	29	28	97	1	3	*
DBP	242	232	96	22	9	3
DEHP	242	51	21	22	9	4
Cholesterol	242	1	0	23	10	2
Cholestanol	242	76	31	23	10	3
Coprostanol	242	75	31	23	10	4
Caffeine	242	2	1	25	10	4
Acetaminophen	242	23	10	25	10	5
Ibuprofen	242	113	47	25	10	11
Diclofenac	123	90	73	6	5	0
Acesulfame-k	123	78	63	6	5	35
Sucralose	242	83	34	25	10	9
Saccharin	242	33	14	25	10	8

* denotes duplicates were non-detect. Average coefficient of variation (CV) in bold letter denotes a possible bias because the number of duplicates is low and some of the duplicates were non-detect.

Each of the thirteen chemical compounds was analysed in 242 samples (Table 6.1). These consist of 233 dry-weather samples, 2 samples compromised by rain, 5 raw sewage samples and 2 non-urban water samples. The dry-weather samples were daytime samples (samples collected between 08:00 and 19:00, both times inclusive) with the exception of 3 samples which were nighttime samples (samples collected between 20:00 and 07:00, both times inclusive). The 2 samples that were compromised by rain are excluded from further data analysis.

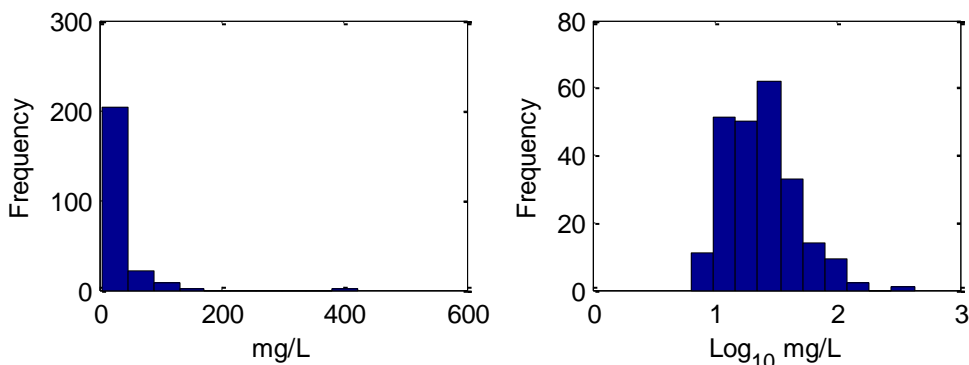
Urban drain water samples were collected from 13 urban catchments: 6 from high-density residential areas, 5 from low-density residential areas, 1 from a commercial area and 1 from an industrial area as described in Section 4.2. Raw sewage samples were collected at manholes in residential areas (Section 4.2). Non-urban water samples were collected from the MacRitchie catchment which is a nature reserve in Singapore (Section 4.2).

Duplicate and blank analyses were done as part of laboratory protocol. For the 13 chemical compounds, duplicates were analysed for 9-16% of the total samples (Table 6.1). The CV of the duplicates averaged from 1-11% (Table 6.1) and the chemical compounds were not detected in the blank samples.

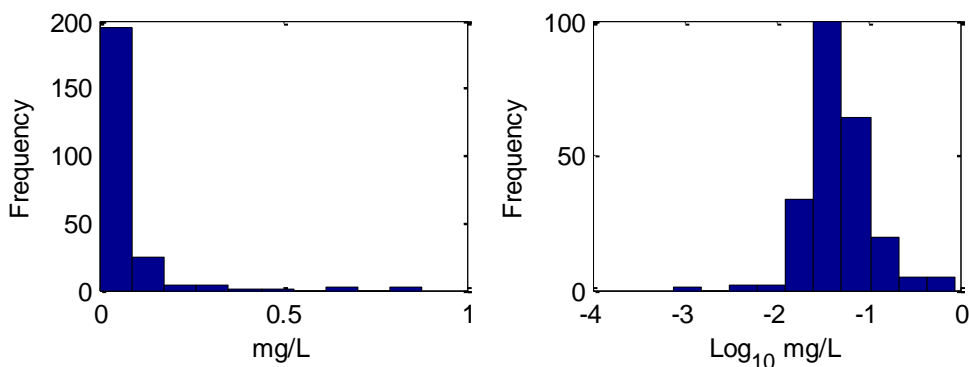
6.1.1. Log-normality of chemical data

Chemical concentrations in the 233 dry-weather samples were plotted as histograms (Figure 6.1) and in probability plots (Figure 6.2). Histograms (Figure 6.1, left) and probability plots (Figure 6.2, left) of raw values generally show that the data are heavily right-skewed. Histograms (Figure 6.1, right) and probability plots (Figure 6.2, right) of log-transformed values mostly show that the data better resemble a normal distribution. However, DEHP, cholestanol and coprostanol still show right-skewed behaviour after log transformation. Because the distributions are not necessarily normal, a non-parametric method will be employed whenever possible, for example by using Spearman correlation instead of Pearson correlation. In addition, histograms

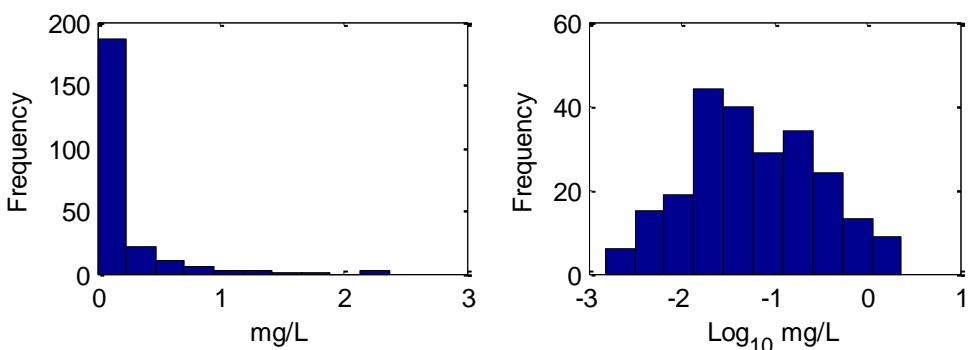
of ibuprofen (Figure 6.1 (k)) and sucralose (Figure 6.1 (l)) show that they have a high frequency of non-detect values. As shown in Table 6.1, ibuprofen and sucralose are the two chemical parameters with the highest frequency of non-detect values after bromide, triclosan, DBP, diclofenac and acesulfame-k which, as discussed in Section 4.1, were dropped from the sampling program early on due to their very infrequent detection.



(a) Chloride

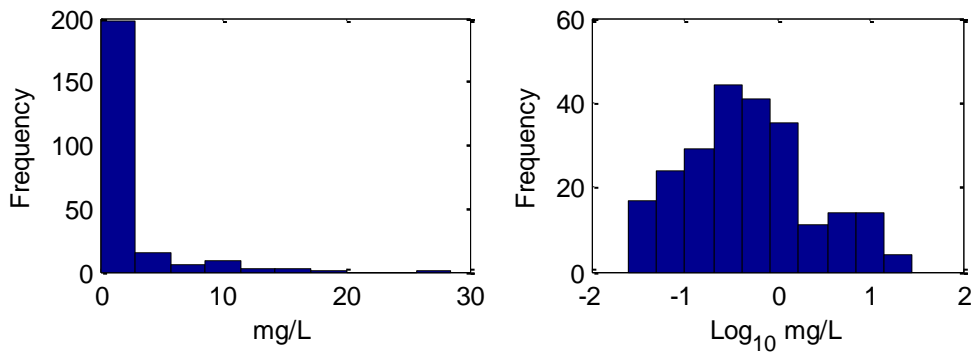


(b) Boron

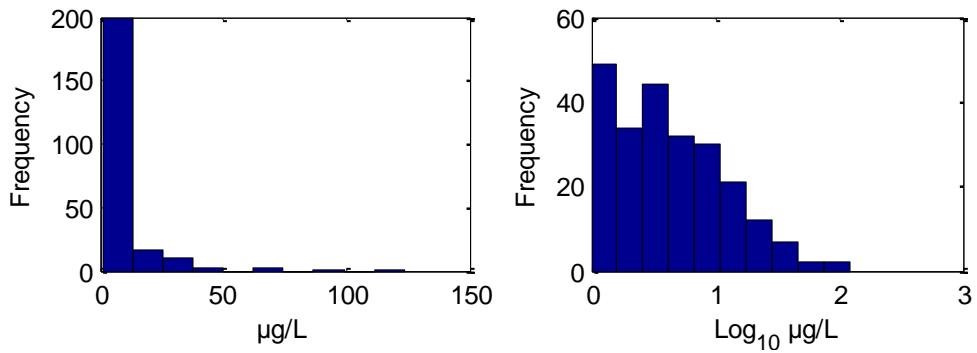


(c) Orthophosphate

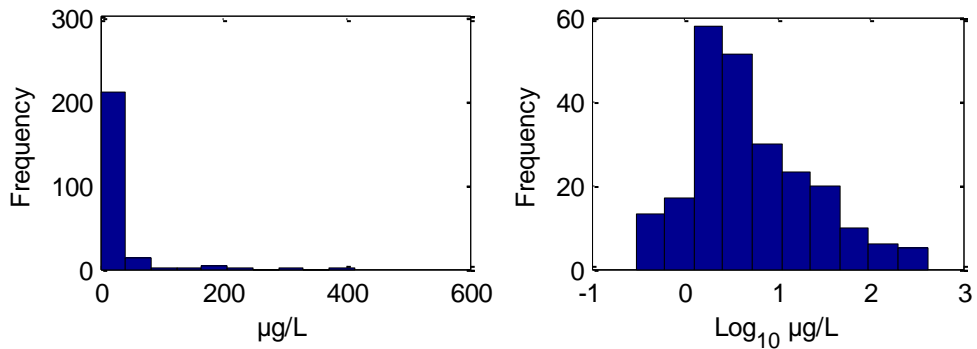
Figure 6.1. Histograms of 13 chemical parameters



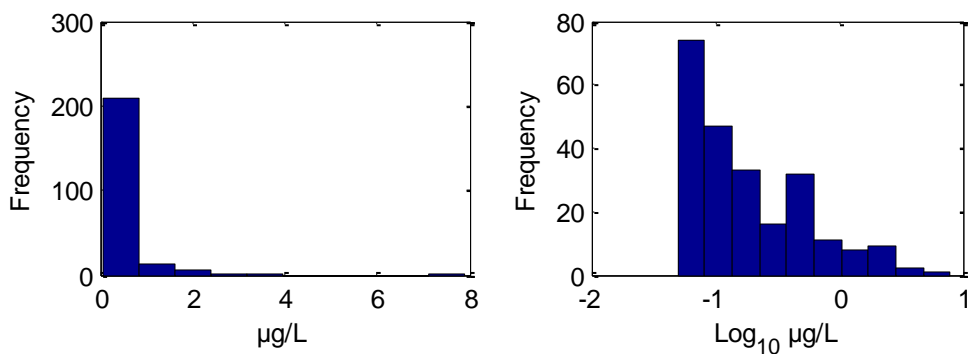
(d) Detergent as MBAS



(e) DEHP

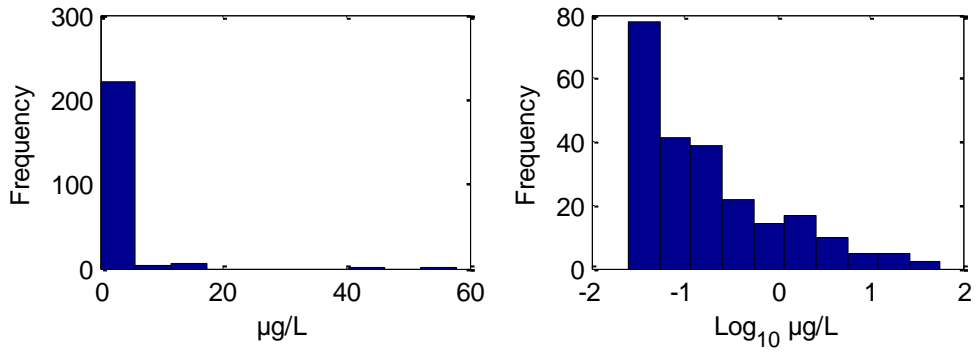


(f) Cholesterol

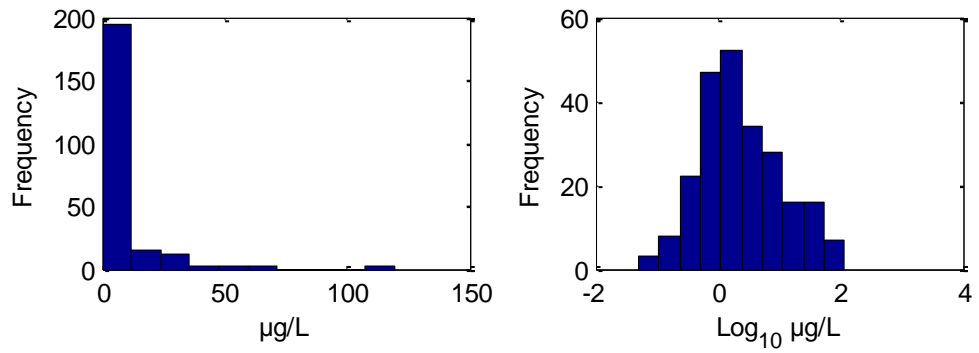


(g) Cholestanol

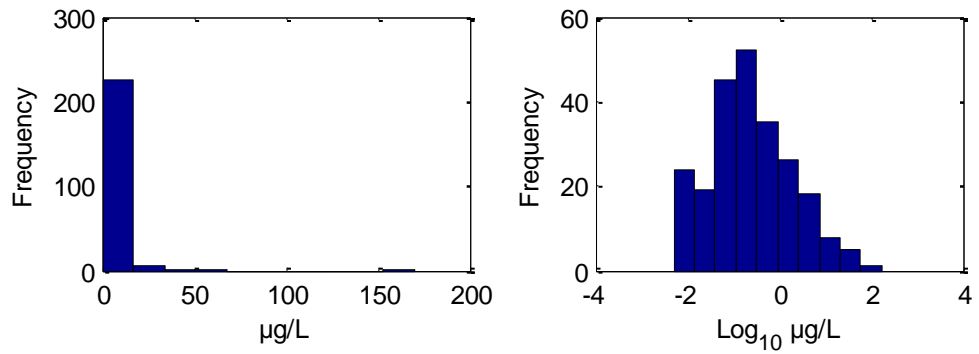
Figure 6.1. (continued) Histograms of 13 chemical parameters



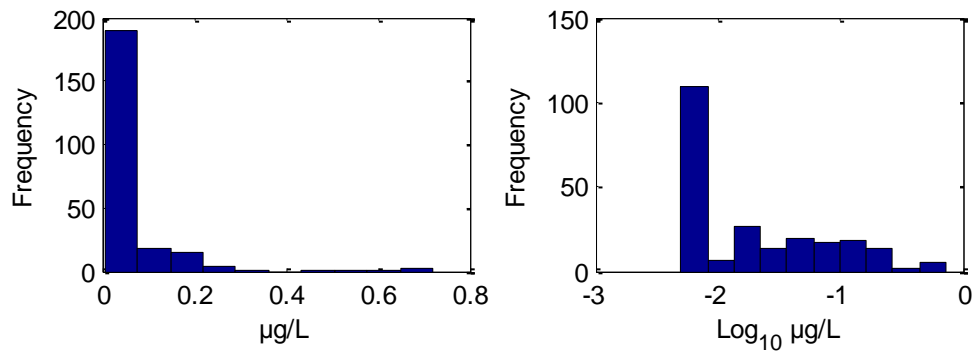
(h) Coprostanol



(i) Caffeine

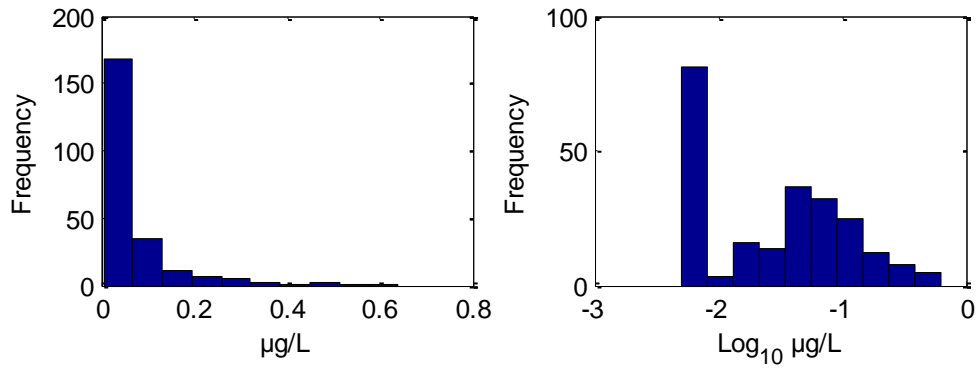


(j) Acetaminophen

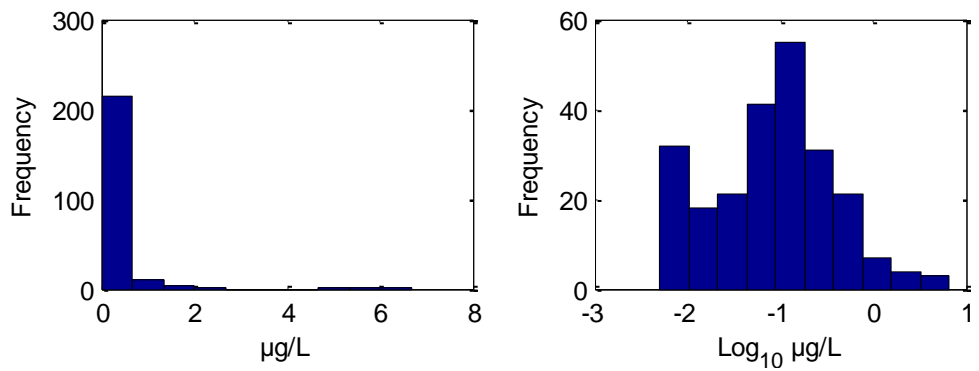


(k) Ibuprofen

Figure 6.1. (continued) Histograms of 13 chemical parameters

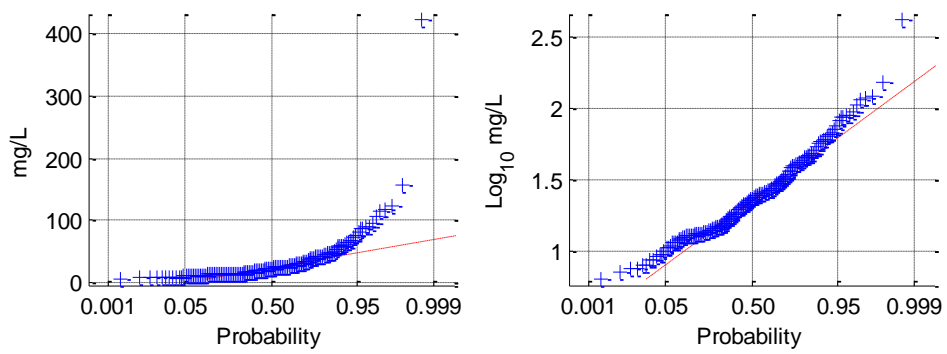


(l) Sucralose



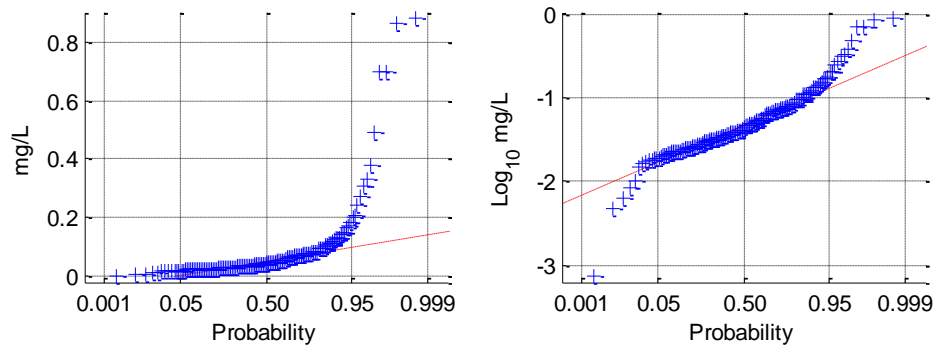
(m) Saccharin

Figure 6.1. (continued) Histograms of 13 chemical parameters

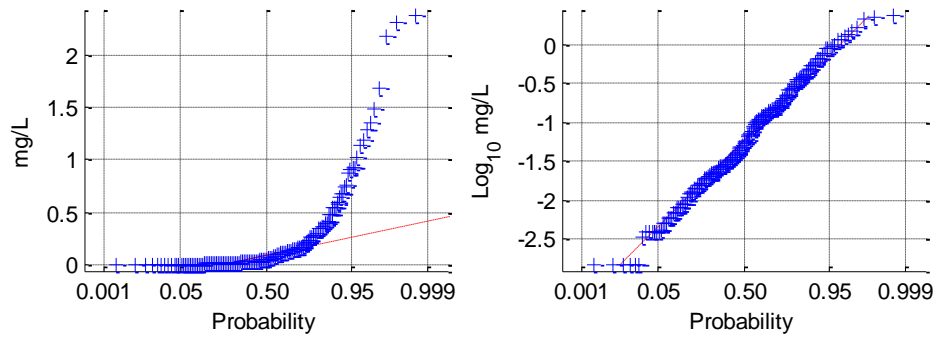


(a) Chloride

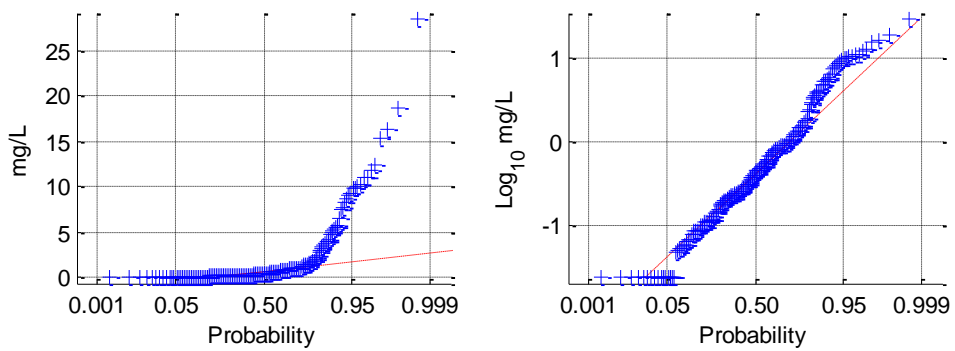
Figure 6.2. Probability plots of 13 chemical parameters



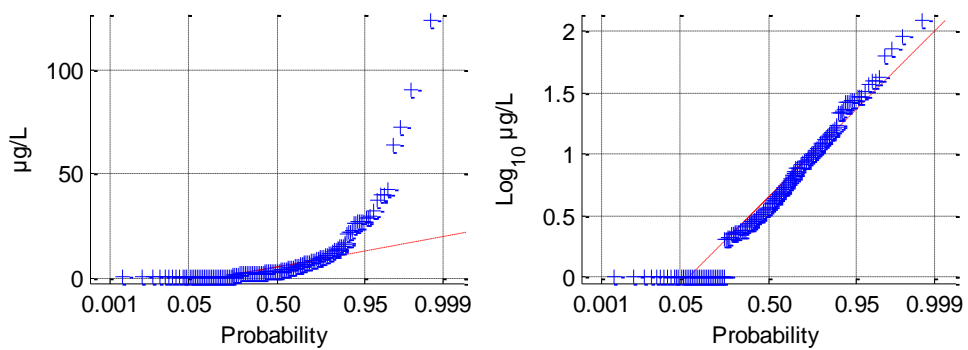
(b) Boron



(c) Orthophosphate

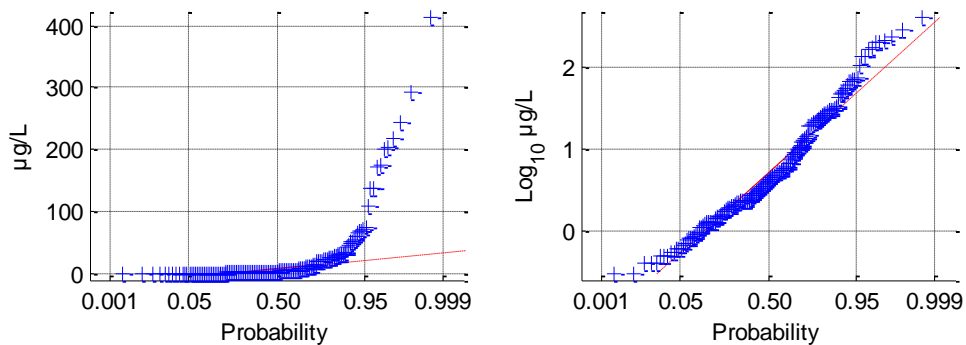


(d) Detergent as MBAS

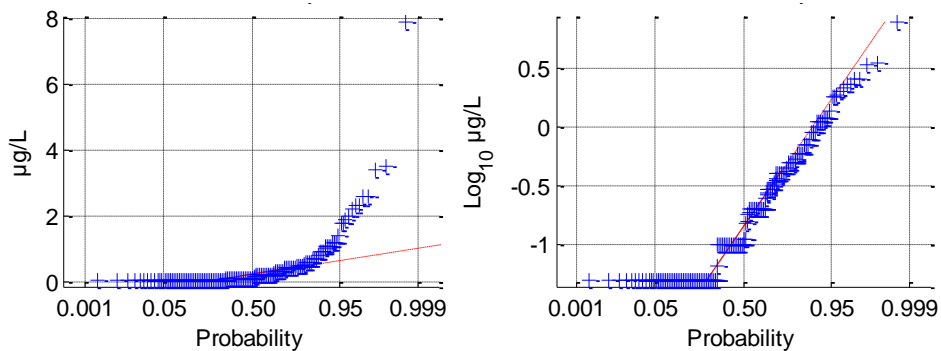


(e) DEHP

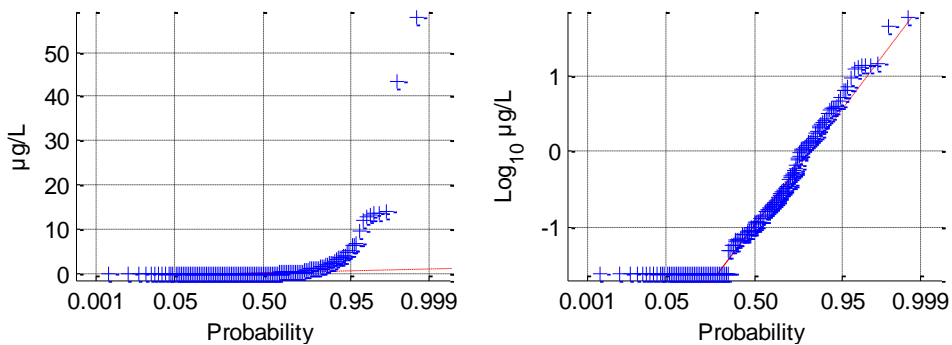
Figure 6.2. (continued) Probability plots of 13 chemical parameters



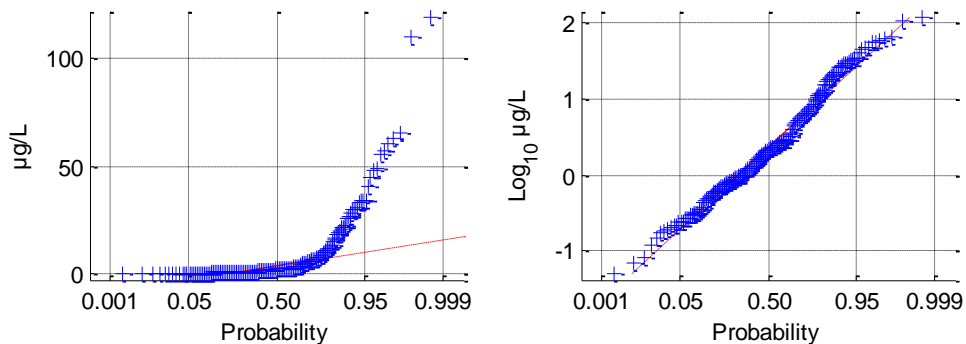
(f) Cholesterol



(g) Cholestanol

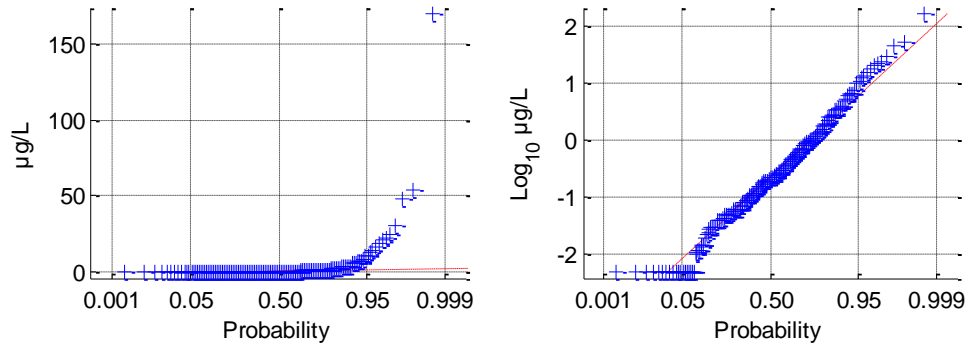


(h) Coprostanol

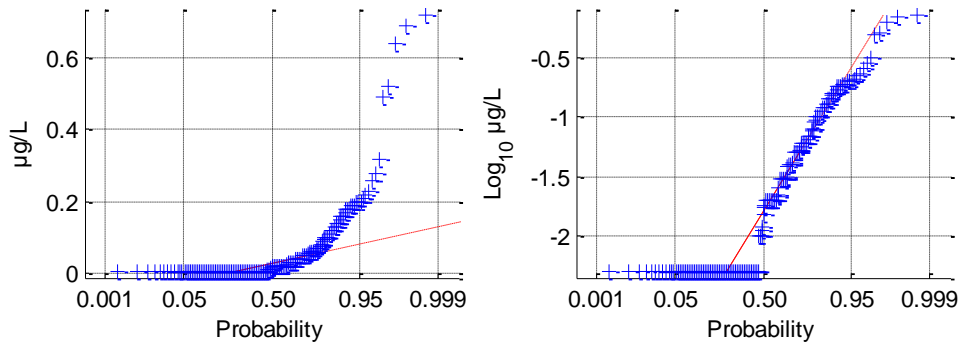


(i) Caffeine

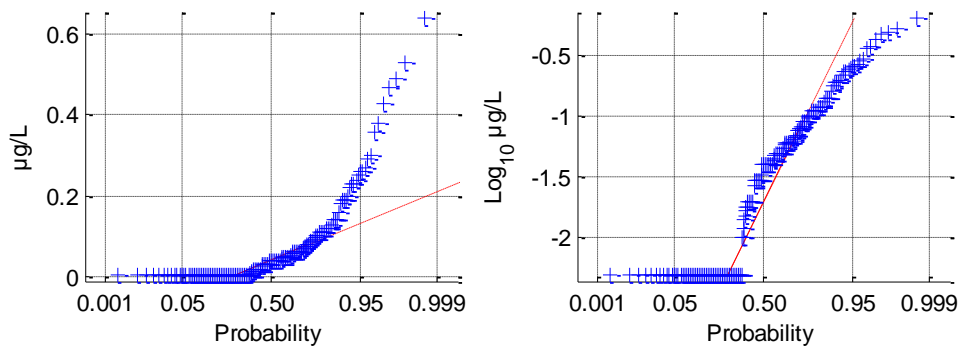
Figure 6.2. (continued) Probability plots of 13 chemical parameters



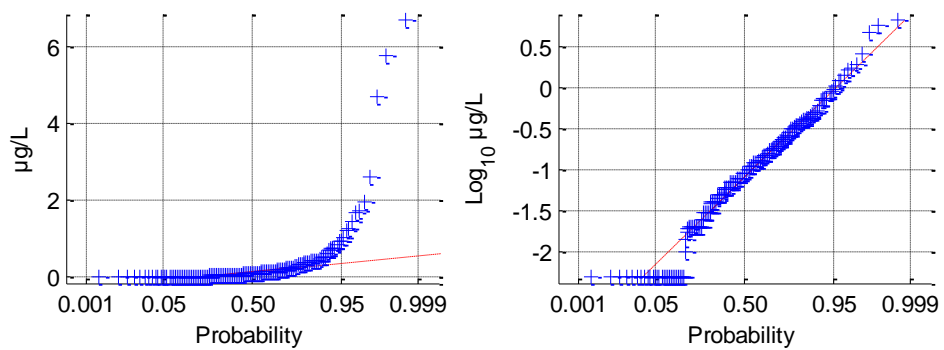
(j) Acetaminophen



(k) Ibuprofen



(l) Sucralose



(m) Saccharin

Figure 6.2. (continued) Probability plots of 13 chemical parameters

6.1.2. Overview of chemical data

The MacRitchie catchment is a nature reserve, not influenced by development. Both non-urban water samples had chemical concentrations below LOD, except for chloride (7.2 and 2.0 mg/L), boron (0.0088 and 0.0028 mg/L), orthophosphate (0.004 and 0.005 mg/L), cholesterol (0.1 µg/L and below LOD) and caffeine (both 0.02 µg/L).

It was hypothesised in Section 3.3.1 that land use may be an important factor for drain water quality. The potential influence of land use was inspected for the 13 chemicals collected in the study reported here as shown from the box plots of chemical concentrations of samples collected during the daytime. As shown in Figure 6.3, the results are classified based on land use (H = high-density residential, L = low-density residential, C = commercial, I = industrial). Concentrations measured in raw sewage samples (rs) are included in the plots for comparison. The geometric mean concentration is superimposed on each box plot with a black circle symbol (●). Detected chemical concentrations of non-urban water samples are illustrated as a dotted line in the box plot.

Some samples from low-density residential areas had chloride and boron concentrations as low as those in the non-urban water samples (Figure 6.3 (a) and (b)) while some samples from both high- and low-density residential areas had orthophosphate concentrations as low as non-urban water samples (Figure 6.3. (c)). The presence of these naturally occurring constituents is not surprising in nominally non-urban samples, for example atmospheric fallout of chloride from rainfall deposition, on average, is 10 kg of chloride per hectare per year on the land surface (Hem 1985). In contrast, cholesterol and caffeine concentrations in non-urban water samples were far below those in urban catchments (Figure 6.3 (f) and (i)). More than half of the number of samples from low-density residential areas had ibuprofen (60 out of 90 samples) and sucralose (48 out of 90 samples) concentrations below their LOD, indicating that ibuprofen and sucralose are not potential tracers for low-density residential areas (Figure 6.3 (k) and (l)). In addition, sucralose was also below the detection limit in 6 out of 10 samples from the industrial area (Figure 6.3 (l)).

It is hypothesised in Chapter 5 that leaking sewers contaminate surface water drains. Therefore, the existence of chemical substances that are used or produced by humans in urban drains, in contrast to non-urban water, is consistent with inflow from leaking sewers into surface water drains. The median and geometric mean concentrations of most of the chemical parameters in high-density residential areas were higher than in low-density residential areas, except for boron, orthophosphate, detergent as MBAS, DEHP and cholesterol. The higher median and geometric mean may indicate that there is a higher degree of contamination in the high-density residential areas than in the low-density residential areas. NTU (2008) collected samples from relatively downstream locations within Kranji catchment, Singapore. Therefore, the samples represented dry-weather flow contributions from a variety of land uses. Although land use impacts on bacterial contamination could not be ascertained with downstream samples, the study reported that the highest bacteria count was obtained from an area of which the land use comprised 70% high-density residential areas. Young and Thackston (1999) also reported that faecal bacteria densities were related to the density of housing.

The higher median and geometric mean of boron, orthophosphate and detergent as MBAS in low-density residential areas (Figure 6.3 (b), (c) and (d)) may indicate additional source of these compounds other than leaking sewers. Since foamy water was observed flowing in drain lines in low-density residential areas, direct discharge of household laundry wastewater into drain lines instead of into sewer lines is the possible additional source of boron, orthophosphate and detergent as MBAS. Section 3.3.2 also discusses this possibility by referring to non-coincident peaks of orthophosphate and detergent as MBAS with peaks of faecal coliform and enterococci. DEHP is also higher in low-density residential areas (Figure 6.3 (e)) and the slightly higher median and geometric mean of cholesterol in low-density residential areas (Figure 6.3 (f)) may be due to plant and animal steroids (Wilkison et al. 2002), however the exact causes of these observations cannot be ascertained.

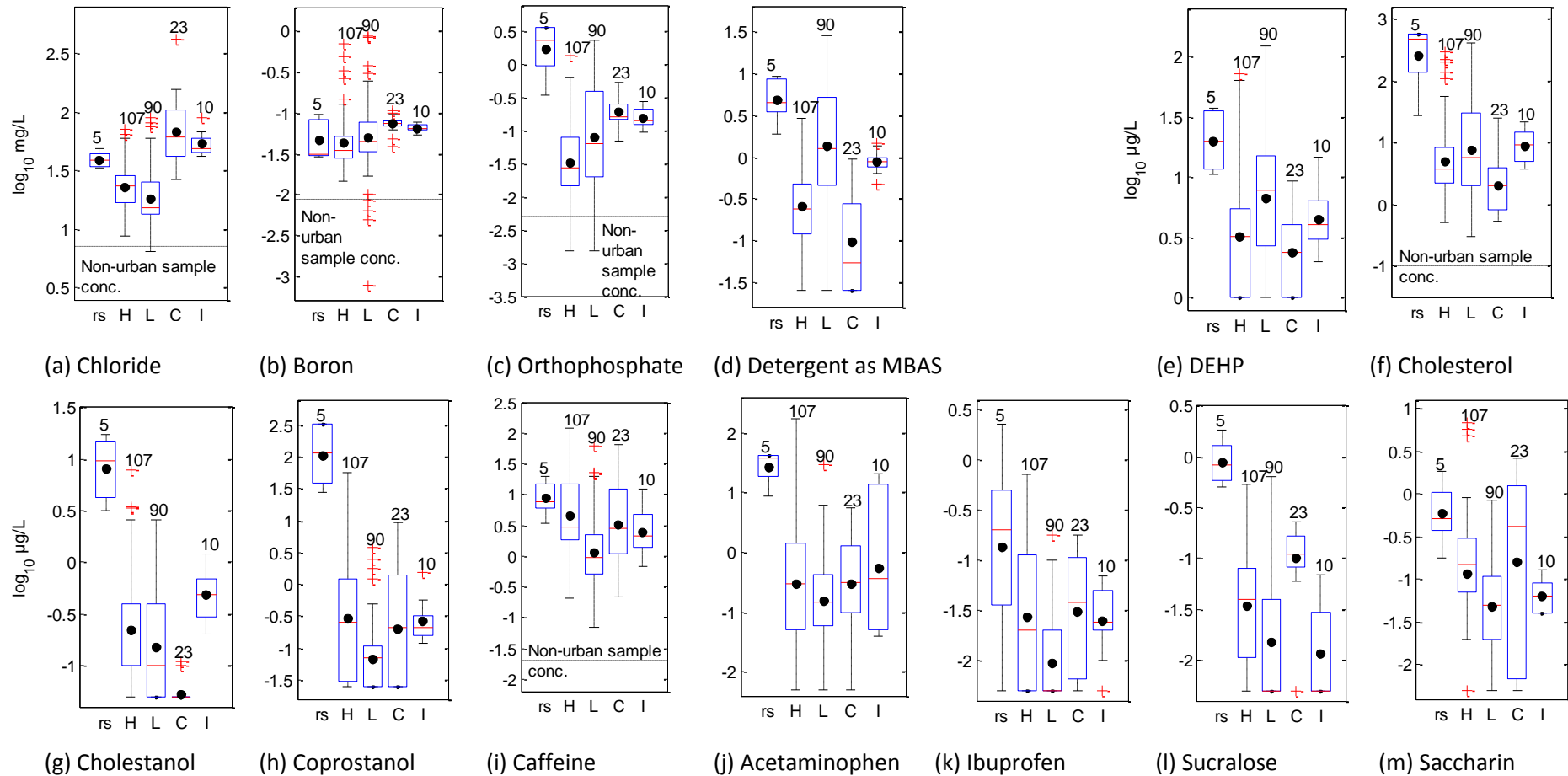


Figure 6.3. Box plots of 13 chemical parameters based on land uses (rs = raw sewage, H = high-density residential, L = low-density residential, C = commercial, I = industrial, • = geometric mean concentration)

Commercial and industrial areas could not be compared objectively with the other land uses because commercial and industrial land use were each represented by only one sampling site: Bras Basah and Sims Place, respectively. However, it is worthwhile to compare the chemical concentrations with raw sewage values. Most of the chemical concentrations in commercial and industrial areas were lower than in raw sewage with the exception of chloride and boron concentrations which were higher than in raw sewage in both commercial and industrial areas (Figure 6.3 (a) and (b)). This may have been caused by seawater intrusion in the subsurface as the Marina Reservoir was formerly an estuary, which has been turned into an impoundment reservoir, and Bras Basah and Sims Place are located in relatively low lying areas close to the Marina Reservoir. In addition, some caffeine and saccharin concentrations in the commercial area were higher than in raw sewage (Figure 6.3 (i) and (m)). This may have been caused by direct discharge from open areas in food courts found in the commercial site (Hans S. Eikaas, PUB, personal communication, 2010).

6.2. Data analysis and discussion

6.2.1. Overall correlation of chemical concentrations

In Section 5.2.5 and elsewhere in Chapter 5, it is shown that daytime samples had higher FIB (total coliforms or TC, *Escherichia coli* or *E. coli* and enterococci) concentrations than nighttime samples. It is therefore likely that this pattern also occurs for chemical parameters. Therefore, daytime chemical data were analysed with Spearman correlation (Section 4.5.5) to find overall associations among the chemical parameters. In addition, FIB and specific conductance measurements are also included in the analysis in order to find their associations with chemical concentrations. Correlations between chemical parameters are reported in Table 6.2. In this table, parameters that are expected to be related are outlined in bold, these include FIB, chloride and specific conductance, the detergent indicators (boron, orthophosphate and MBAS), the faecal sterols, the artificial sweeteners and ibuprofen and acetaminophen. Numerical values shown in bold type indicate correlations significant at $p \leq 0.05$. The magnitude of the correlation coefficients is illustrated with a colour code: red represents high (0.70 – 1.00) positive correlation, yellow represents moderate (0.30 – 0.69) positive correlation and green represents low (0.01 – 0.29) positive correlation or in some cases, negative correlation. N values in Tables 6.2 are the number of pairs of samples used in the correlation analysis and varies for the FIB, specific conductance and the chemical parameters. As mentioned in Section 4.4.2, due to finite resources, only selected samples were analysed for the suite of chemical parameters. Therefore, the number of samples used in the correlation analysis among the FIB is higher than the number of samples used in the correlation analysis between chemical parameters and FIB.

Overall, Table 6.2 shows that *E. coli* has a larger number of significant ($p \leq 0.05$) correlations with other parameters than enterococci. *E. coli* is moderately ($r_s \geq 0.30$) correlated with acetaminophen, coprostanol, cholesterol and orthophosphate. Enterococci is moderately ($r_s \geq 0.31$) correlated with sucralose, caffeine and chloride.

Correlations between parameters that are expected to be related (FIB, chloride and specific conductance, the detergent indicators, the faecal sterols, ibuprofen and acetaminophen and the artificial sweeteners) are all significant ($p \leq 0.05$). Coprostanol, caffeine, acetaminophen, ibuprofen, sucralose and saccharin are weakly to moderately ($r_s \geq 0.13$) correlated with each other. Coprostanol is biodegradable under aerobic conditions and is also hydrophobic and therefore readily incorporated into bottom sediments (Isobe et al. 2002). In this study, drain water samples were sampled during dry-weather with care so as not to disturb the bottom sediments. Hence, the presence of coprostanol in the water samples indicates recent human faecal contamination. Sidhu et al. (2013) reported frequent detection of acetaminophen and caffeine in stormwater runoff in Australia. Sidhu et al. mentioned that caffeine and acetaminophen are biodegradable (and hence labile) indicators of the presence of sewage in environmental water samples. Therefore, the presence of these markers suggests recent contamination by raw sewage which likely originates from leaking sewers as discussed in Chapter 5. However, the use of caffeine and saccharin in Singapore must be interpreted cautiously as they may be discharged directly from open areas in food courts which are commonly found in the high-density residential and commercial areas (Hans S. Eikaas, PUB, personal communication, 2010).

Table 6.2. Correlation matrix for all daytime data

	TC	<i>E. coli</i>	Enterococci	Specific conductance	Chloride	Boron	Orthophosphate	Detergent as MBAS	DEHP	Cholesterol	Cholestanol	Coprostanol	Caffeine	Acetaminophen	Ibuprofen	Sucralose
N = 648																
<i>E. coli</i>	0.62															
Enterococci	0.46	0.35														
N = 629																
Specific conductance	0.24	0.07	0.28													
N = 230																
Chloride	0.22	0.11	0.31	0.76												
Boron	0.15	0.08	0.12	0.67	0.51											
Orthophosphate	0.29	0.30	0.06	0.50	0.42	0.52										
Detergent as MBAS	0.20	0.21	-0.19	0.13	0.00	0.18	0.40									
DEHP	0.22	0.27	0.05	0.19	0.13	0.38	0.36	0.46								
Cholesterol	0.42	0.39	0.08	0.30	0.20	0.34	0.55	0.51	0.52							
Cholestanol	0.21	0.22	0.06	0.10	0.09	0.16	0.28	0.24	0.31	0.65						
Coprostanol	0.28	0.40	0.17	0.27	0.34	0.16	0.29	0.04	0.20	0.52	0.58					
Caffeine	0.17	0.10	0.31	0.12	0.23	0.09	-0.02	-0.21	0.07	0.12	0.32	0.24				
Acetaminophen	0.44	0.43	0.26	-0.02	0.08	0.00	0.16	0.09	0.06	0.16	0.19	0.31	0.27			
Ibuprofen	0.10	0.21	0.22	0.19	0.31	0.15	-0.02	-0.28	0.02	-0.01	0.05	0.39	0.20	0.32		
Sucralose	0.31	0.14	0.40	0.20	0.27	0.18	-0.02	-0.25	-0.05	-0.07	-0.09	0.17	0.29	0.24	0.28	
Saccharin	0.04	0.16	0.02	0.20	0.24	0.16	0.17	-0.06	0.11	0.18	0.03	0.38	0.17	0.13	0.32	0.18

Correlation

High Low

Bold values indicate correlations significant at $p \leq 0.05$.

6.2.2. Influence of land use on chemical concentrations

The daytime chemical data were grouped based on major land use (high-density residential, low-density residential, commercial and industrial) in order to investigate the effect of land use on water quality characteristics. The Kruskal-Wallis test (Section 4.5.13) was applied to the chemical data to test whether there are any significant differences in chemical concentrations associated with different land use. The results are expressed as p values. The p values for all chemical parameters were lower than 0.05, except for acetaminophen ($p = 0.11$). This means that the chemical data show significant differences in at least one land use (compared to other land uses) except for acetaminophen which does not show significant differences across different land uses.

Principal component analysis or PCA (Section 3.1.3) can be applied to the daytime chemical samples to investigate which parameters (out of the 13 chemical parameters) give high variation. The PCA indicates 13 principal components (PCs). The first two PCs explain 47% of the observed variation. Figure 6.4 shows that the first principal component (PC1) is more positively influenced by cholesterol, cholestanol, orthophosphate and coprostanol, while the second principal component (PC2) is most positively influenced by detergent as MBAS.

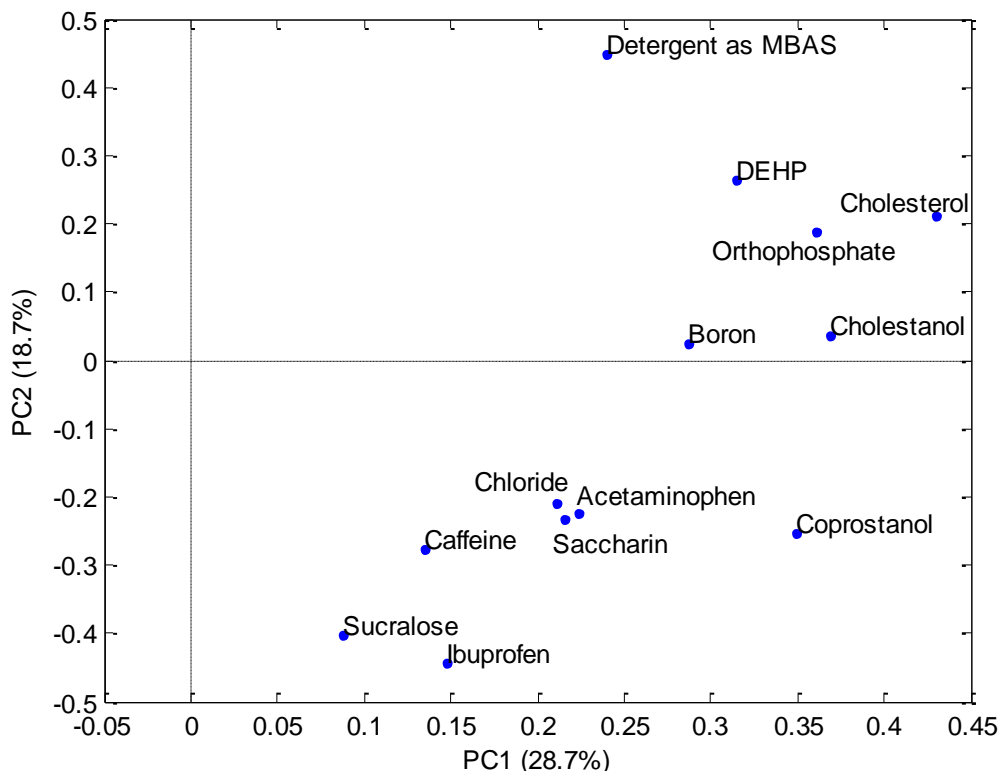


Figure 6.4. Principle component analysis of 13 chemical parameters

The samples are plotted in terms of PC1 and PC2 and symbolised according to the land use to visualise the effect of land use (if any) in Figure 6.5. Samples from high- and low-density residential areas differ in PC2 values. The samples from low-density residential areas are strongly influenced by detergent as MBAS. Samples from the commercial area plot close to samples from high-density residential areas whereas samples from the industrial area plot in between samples from high- and low-density residential areas. However, due to the limited number of commercial and industrial samples, further inferences could not be drawn from these samples.

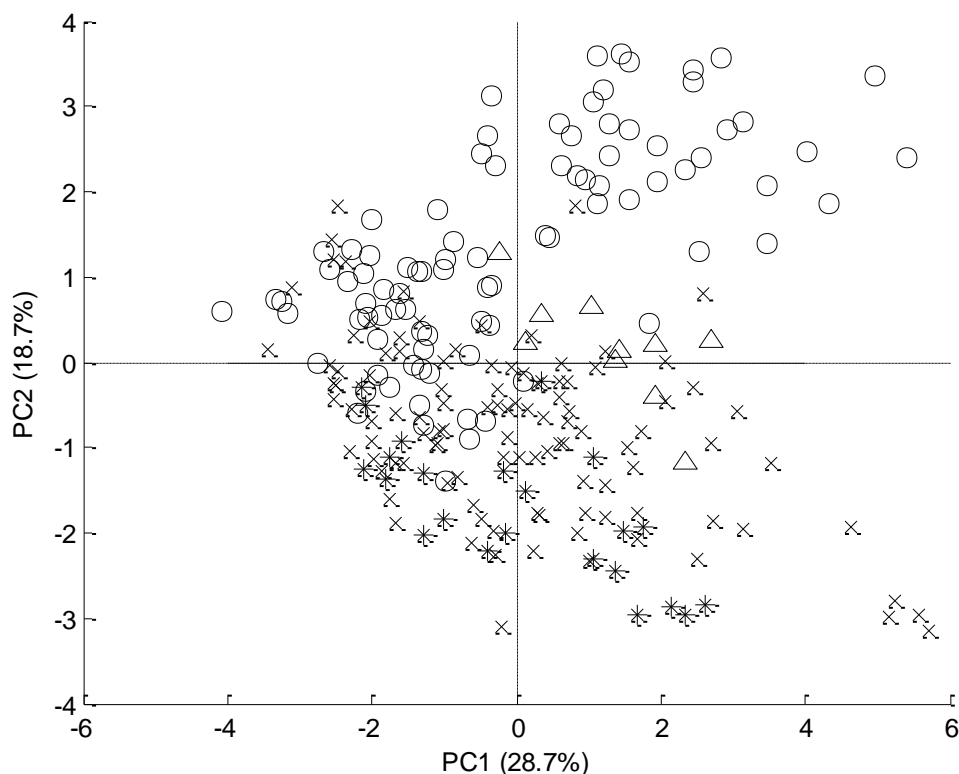


Figure 6.5. PC1 vs. PC2 plot of 230 daytime samples (x = high-density residential, o = low-density residential, * = commercial, Δ = industrial)

The Kruskal-Wallis test showed that most of the chemical parameters are significantly different across the different land uses and PCA shows that samples from different major land uses, especially high- and low-density residential areas, are different. Therefore, the segregated data based on major land use are further analysed with Spearman correlation (Section 4.5.5). The purpose of this analysis is to find associations among the parameters for each land use category. Correlations between parameters are reported in Tables 6.3 through 6.6.

The correlations between cholesterol, cholestanol and coprostanol are the strongest ($r_s \geq 0.79$) in Table 6.3 for samples from high-density residential areas. In addition, cholesterol, cholestanol and coprostanol are weakly to moderately ($r_s \geq 0.19$) correlated with the other parameters except caffeine. The correlations are insignificant between cholestanol/coprostanol and enterococci and also between coprostanol and detergent as MBAS. Correlations of caffeine and detergent as MBAS

with most of the other parameters are also not significant. *E. coli* is better correlated with acetaminophen, coprostanol, cholesterol, orthophosphate, sucralose and ibuprofen and enterococci is also better correlated with sucralose and ibuprofen in the high-density residential dataset (Table 6.3) than in the overall dataset (Table 6.2). *E. coli* is better correlated with most of the parameters than is enterococci; exceptions are specific conductance, chloride, caffeine and sucralose. Unlike low-density residential areas, foamy water was rarely observed flowing in drain lines in high-density residential areas. In addition, orthophosphate is moderately ($r_s \geq 0.54$) correlated with cholesterol, cholestanol and coprostanol. Therefore, this may indicate insufficient soil adsorption such that orthophosphate from leaking sewers is leached out to surface water. Alternatively, orthophosphate from leaking sewers may travel to the surface water with minimal adsorption through macropore pathways such as subsurface voids.

There are higher correlations among TC, specific conductance, chloride, boron, orthophosphate, detergent as MBAS, DEHP and cholesterol in the low-density residential dataset ($r_s \geq 0.46$, Table 6.4) than in the overall dataset ($r_s \geq 0.13$, Table 6.2). However, the correlation of boron, orthophosphate and detergent as MBAS in low-density residential areas must be interpreted cautiously because they may originate from two different sources: leaking sewers and direct discharge of household laundry wastewater into drain lines, as discussed in Section 3.3.2 and 6.1.2. The correlation coefficients among cholesterol, cholestanol and coprostanol are moderate, $r_s \geq 0.47$ (lower than for high-density residential areas). Similarly to the overall and high-density residential datasets, *E. coli* is better correlated with other parameters than is enterococci. However, *E. coli* and coprostanol show lower ($r_s = 0.34$) correlation coefficients with other parameters compared to *E. coli* and coprostanol in the overall ($r_s = 0.40$) and high-density residential datasets ($r_s = 0.48$). Caffeine shows better correlation with other parameters in low-density residential areas than in high-density residential areas. On the other hand, ibuprofen and sucralose show negative correlation with many other parameters. This may occur

because more than half of the samples from low-density residential areas were reported as non-detect for ibuprofen (60 out of 90) and sucralose (48 out of 90). Therefore, the negative correlations do not mean that there is no dependence between ibuprofen or sucralose and the other parameters, but instead that ibuprofen and sucralose are not suitable indicators for low-density residential areas.

For the commercial dataset, the samples were obtained from one commercial area. Therefore, the correlation results may not be a representative of commercial area in general. However, a discussion about the results is still beneficial. There are higher correlations among *E. coli*, specific conductance, chloride, orthophosphate, detergent as MBAS, cholesterol, coprostanol, acetaminophen and ibuprofen in the commercial dataset ($r_s \geq 0.42$, Table 6.5) than overall dataset ($r_s \geq 0.13$, Table 6.2). Similarly to residential areas, *E. coli* is better correlated than enterococci with other parameters except caffeine. Boron, DEHP, cholestanol, caffeine and sucralose do not have any positive correlation (at $p \leq 0.05$) coefficients with other parameters except between caffeine and enterococci.

Table 6.3. Correlation matrix for high-density residential areas

	TC	<i>E. coli</i>	Enterococci	Specific conductance	Chloride	Boron	Orthophosphate	Detergent as MBAS	DEHP	Cholesterol	Cholestanol	Coprostanol	Caffeine	Acetaminophen	Ibuprofen	Sucralose
N = 317																
<i>E. coli</i>	0.67															
Enterococci	0.52	0.40														
N = 298																
Specific conductance	0.15	0.07	0.26													
N = 107																
Chloride	0.07	0.10	0.25	0.63												
Boron	-0.05	0.08	0.07	0.68	0.54											
Orthophosphate	0.15	0.41	-0.04	0.33	0.18	0.45										
Detergent as MBAS	0.28	0.22	0.00	0.17	0.09	-0.03	0.26									
DEHP	0.13	0.30	0.25	0.38	0.34	0.39	0.20	0.08								
Cholesterol	0.43	0.49	0.26	0.29	0.31	0.32	0.57	0.23	0.48							
Cholestanol	0.23	0.35	0.13	0.20	0.28	0.32	0.60	0.20	0.35	0.82						
Coprostanol	0.30	0.48	0.11	0.27	0.27	0.29	0.54	0.14	0.40	0.79	0.80					
Caffeine	0.12	0.04	0.22	0.06	0.32	0.10	-0.07	0.01	0.29	0.14	0.13	-0.04				
Acetaminophen	0.64	0.69	0.33	-0.01	0.11	0.04	0.28	0.32	0.21	0.33	0.19	0.31	0.15			
Ibuprofen	0.18	0.31	0.29	0.51	0.53	0.58	0.21	-0.06	0.40	0.29	0.21	0.32	0.16	0.38		
Sucralose	0.51	0.36	0.47	0.22	0.29	0.19	0.08	0.08	0.32	0.34	0.25	0.33	0.41	0.38	0.28	
Saccharin	-0.17	0.16	-0.06	0.17	0.15	0.25	0.22	0.04	0.44	0.24	0.28	0.30	0.19	0.13	0.29	0.16

Correlation

High Low

Bold values indicate correlations significant at $p \leq 0.05$.

Table 6.4. Correlation matrix for low-density residential areas

	TC	<i>E. coli</i>	Enterococci	Specific conductance	Chloride	Boron	Orthophosphate	Detergent as MBAS	DEHP	Cholesterol	Cholestanol	Coprostanol	Caffeine	Acetaminophen	Ibuprofen	Sucralose
N = 245																
<i>E. coli</i>	0.50															
Enterococci	0.33	0.34														
Specific conductance	0.57	0.34	0.18													
N = 90																
Chloride	0.53	0.29	0.20	0.73												
Boron	0.54	0.18	0.12	0.63	0.49											
Orthophosphate	0.57	0.25	0.09	0.52	0.57	0.55										
Detergent as MBAS	0.66	0.26	0.11	0.75	0.62	0.60	0.72									
DEHP	0.56	0.40	0.20	0.49	0.49	0.55	0.59	0.69								
Cholesterol	0.60	0.31	0.16	0.69	0.46	0.62	0.72	0.74	0.62							
Cholestanol	0.32	0.23	0.13	0.40	0.17	0.44	0.39	0.38	0.36	0.55						
Coprostanol	0.29	0.34	0.23	0.29	0.32	0.34	0.32	0.33	0.41	0.47	0.58					
Caffeine	0.07	0.22	0.18	0.11	-0.01	0.25	0.19	0.13	0.26	0.27	0.61	0.42				
Acetaminophen	0.05	-0.02	0.06	-0.22	-0.12	0.02	0.06	0.03	0.03	-0.09	0.27	0.11	0.37			
Ibuprofen	-0.30	-0.16	-0.07	-0.47	-0.29	-0.32	-0.32	-0.49	-0.18	-0.38	-0.14	0.10	0.03	0.03		
Sucralose	-0.12	-0.11	0.08	-0.14	-0.03	0.03	-0.29	-0.14	-0.12	-0.27	-0.21	-0.14	-0.07	-0.05	0.18	
Saccharin	0.24	-0.06	-0.09	0.15	0.18	0.22	0.20	0.13	0.00	0.23	-0.20	0.01	-0.06	-0.09	-0.07	-0.02
Correlation	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00					
		High									Low					

Bold values indicate correlations significant at $p \leq 0.05$.

Table 6.5. Correlation matrix for commercial area

	TC	<i>E. coli</i>	Enterococci	Specific conductance	Chloride	Boron	Orthophosphate	Detergent as MBAS	DEHP	Cholesterol	Cholestanol	Coprostanol	Caffeine	Acetaminophen	Ibuprofen	Sucralose
N = 48																
<i>E. coli</i>	0.69															
Enterococci	0.52	0.25														
Specific conductance	0.16	0.25	-0.29													
N = 23																
Chloride	0.02	0.10	-0.54	0.90												
Boron	0.22	-0.22	0.30	0.17	0.15											
Orthophosphate	0.04	0.52	-0.29	0.53	0.34	-0.43										
Detergent as MBAS	0.21	0.67	-0.54	0.41	0.39	-0.31	0.58									
DEHP	0.10	0.09	0.11	-0.16	-0.09	0.04	-0.32	0.26								
Cholesterol	0.68	0.87	-0.16	0.47	0.36	-0.01	0.39	0.72	0.25							
Cholestanol	0.12	-0.33	-0.06	0.20	0.22	0.37	-0.33	-0.14	0.35	-0.02						
Coprostanol	0.43	0.79	-0.34	0.53	0.42	-0.18	0.67	0.78	0.11	0.83	-0.16					
Caffeine	0.31	0.20	0.66	-0.50	-0.51	0.29	-0.40	-0.28	0.11	0.04	-0.29	-0.19				
Acetaminophen	0.48	0.82	0.04	0.56	0.40	-0.10	0.61	0.59	0.17	0.83	-0.17	0.81	0.01			
Ibuprofen	0.29	0.67	-0.50	0.63	0.52	-0.33	0.71	0.85	0.05	0.71	-0.17	0.83	-0.33	0.71		
Sucralose	0.07	-0.03	0.29	0.23	0.14	0.15	0.06	-0.30	-0.13	-0.06	0.18	-0.18	-0.01	0.15	-0.03	
Saccharin	0.33	0.78	-0.36	0.58	0.50	-0.32	0.70	0.80	0.10	0.76	-0.14	0.91	-0.27	0.81	0.89	0.01
Correlation	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00					
		High									Low					

Bold values indicate correlations significant at $p \leq 0.05$.

Table 6.6. Correlation matrix for industrial area

	TC	<i>E. coli</i>	Enterococci	Specific conductance	Chloride	Boron	Orthophosphate	Detergent as MBAS	DEHP	Cholesterol	Cholestanol	Coprostanol	Caffeine	Acetaminophen	Ibuprofen	Sucralose
N = 38																
<i>E. coli</i>	0.78															
Enterococci	0.58	0.68														
Specific conductance	-0.31	-0.28	-0.29													
N = 10																
Chloride	-0.14	-0.02	0.27	0.10												
Boron	-0.17	-0.21	-0.43	0.64	-0.11											
Orthophosphate	0.25	-0.29	0.15	-0.02	0.27	-0.19										
Detergent as MBAS	0.20	0.59	0.20	-0.32	-0.26	0.09	-0.08									
DEHP	0.17	-0.30	-0.18	0.01	-0.13	-0.15	0.75	0.29								
Cholesterol	0.55	-0.07	0.18	-0.40	-0.04	-0.35	0.80	-0.07	0.71							
Cholestanol	0.62	0.00	0.37	-0.14	0.24	-0.13	0.78	0.00	0.58	0.67						
Coprostanol	0.72	0.08	0.28	-0.35	-0.01	-0.17	0.74	0.19	0.76	0.85	0.87					
Caffeine	0.75	0.09	0.32	-0.36	0.03	-0.16	0.54	0.05	0.59	0.79	0.71	0.85				
Acetaminophen	0.38	-0.17	0.31	0.06	0.38	0.09	0.47	-0.18	0.30	0.43	0.48	0.45	0.73			
Ibuprofen	0.25	0.80	0.32	-0.65	0.12	-0.17	-0.43	0.50	-0.27	-0.13	-0.24	-0.06	0.01	-0.26		
Sucralose	0.31	-0.06	0.29	-0.25	0.37	-0.18	0.77	0.00	0.55	0.73	0.46	0.61	0.59	0.65	-0.06	
Saccharin	-0.08	-0.28	-0.19	0.16	0.54	-0.08	0.57	-0.12	0.50	0.47	0.52	0.40	0.41	0.32	-0.04	0.38
Correlation	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00					
		High									Low					

Bold values indicate correlations significant at $p \leq 0.05$.

For the single industrial area sampled, 10 samples were analysed for the chemical parameters. Therefore, the correlation results may not be representative of industrial areas in general. Nevertheless, a discussion about the results is still useful. It is observed from Table 6.6 that correlations among orthophosphate, DEHP, cholesterol, cholestanol, coprostanol, caffeine, acetaminophen and sucralose in the industrial dataset are generally higher ($r_s \geq 0.65$) than in the overall dataset ($r_s \geq 0.16$, Table 6.2). Similarly to high-density residential and commercial areas, orthophosphate that is highly ($r_s \geq 0.74$) correlated with DEHP, cholesterol, cholestanol, coprostanol and sucralose may indicate insufficient soil adsorption or minimal adsorption if orthophosphate from leaking sewers travels to the surface water through macropore pathways such as subsurface voids. Therefore, measured orthophosphate in surface water may indicate orthophosphate contribution from leaking sewers. Out of the chemical parameters listed in Table 6.6 (from chloride to saccharin), sucralose has the highest number of non-detect samples (6 out of 10 samples). Unlike other land uses discussed above, there is no significant correlation between *E. coli*/enterococci and other parameters except for *E. coli* and ibuprofen.

In general, the correlations for specific land uses (Tables 6.3 through 6.6) show different characteristics from the overall correlations (Table 6.2). There are better correlations among different parameters in each land-use dataset than in the overall dataset. This shows that land use affects drain water quality and that analysis based on land use enables better data interpretation. Generally there is a greater number of significant ($p \leq 0.05$) correlations between FIB and chemical parameters in the high-density residential areas than in the low-density residential areas. Correlations between *E. coli* and orthophosphate, cholesterol, cholestanol, coprostanol, acetaminophen, ibuprofen and sucralose are higher in the high-density residential areas ($r_s \geq 0.31$) than in the low-density residential areas ($r_s \geq 0.23$). Cholesterol, cholestanol and coprostanol are moderately to highly correlated among each other in each land use except in the commercial area in which only cholesterol is highly correlated with coprostanol. *E. coli* is better correlated than enterococci with other

parameters in each land use. However, this *E. coli* characteristic is not as evident in the industrial area as in other land uses.

6.2.3. Chemical data segregation based on FIB

In Section 6.2.1 and 6.2.2, it is indicated that many of the chemical parameters are correlated with *E. coli* and enterococci, which are widely used surface water quality indicators, including in Singapore. In this section, samples from residential areas are further segregated based on the concentrations of *E. coli* and/or enterococci, but samples from commercial and industrial areas are not because there is insufficient data. *E. coli* and/or enterococci concentrations measured in the residential areas are normalised according to the maximum and minimum concentrations of the 12-hr time series (Section 4.5.7). The samples are then segregated with respect to their normalised *E. coli* and/or enterococci concentrations. Three segregation methods are investigated. Each segregation method classifies data into three groups: (i) “high”, (ii) “moderate” and (iii) “low” groups. Spearman correlation analysis (Section 4.5.5) is then conducted on the segregated datasets. The best segregation method is chosen based on the correlation coefficients between chemical parameters and FIB.

The following segregation schemes were employed:

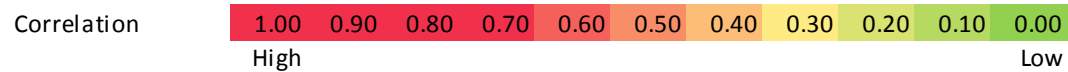
1. The first segregation method grouped data based on the classification of *E. coli* and enterococci concentrations: (i) normalised *E. coli* or enterococci (Ent) concentrations ≥ 0.70 as “high” dataset, (ii) $0.30 < E. coli$ or enterococci < 0.70 as “moderate” dataset and (iii) *E. coli* and enterococci ≤ 0.30 as “low” dataset.
2. The second segregation method grouped data based on the classification of *E. coli* concentrations only: (i) normalised *E. coli* concentrations ≥ 0.70 as “high” dataset, (ii) $0.30 < E. coli < 0.70$ as “moderate” dataset and (iii) *E. coli* ≤ 0.30 as “low” dataset.
3. The third segregation method grouped data based on enterococci concentrations only: (i) normalised enterococci concentrations ≥ 0.70 as “high” dataset, (ii) $0.30 < \text{enterococci} < 0.70$ as “moderate” dataset and (iii) enterococci ≤ 0.30 as “low” dataset.

<enterococci < 0.70 and as “moderate” dataset and (iii) enterococci \leq 0.30 as “low” dataset.

The correlation coefficients between chemical parameters and FIB are presented in Tables 6.7 through 6.9 for the dataset obtained for the high-density residential areas and in Tables 6.10 through 6.12 for the dataset obtained for the low-density residential areas. The magnitude of correlation coefficients is illustrated with the same colour code as used above: red represents high (positive) correlation, yellow represents moderate (positive) correlation, and green represents low (positive) or negative correlation. N is the number of pairs of samples used in the correlation analysis.

Table 6.7. First segregation method for samples from high-density residential areas

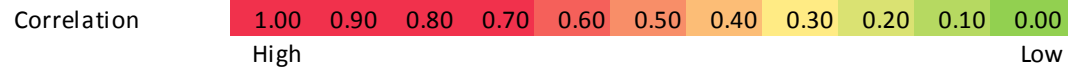
	All N = 107			"High" dataset <i>E. coli</i> or enterococci ≥ 0.70 N = 58			"Moderate" dataset 0.30 < <i>E. coli</i> or enterococci < 0.70 N = 20			"Low" dataset <i>E. coli</i> and enterococci ≤ 0.30 N = 29		
	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent
Chloride	0.07	0.10	0.25	-0.01	0.12	0.26	0.25	0.16	0.36	-0.06	-0.03	0.35
Boron	-0.05	0.08	0.07	-0.05	0.21	0.12	-0.01	0.29	0.17	-0.02	-0.04	0.32
Orthophosphate	0.15	0.41	-0.04	0.15	0.50	-0.05	0.23	0.37	-0.06	0.16	0.21	-0.21
Detergent as MBAS	0.28	0.22	0.00	0.06	-0.01	-0.42	0.14	0.04	0.09	0.18	0.08	-0.20
DEHP	0.13	0.30	0.25	0.25	0.34	0.23	-0.21	0.47	-0.06	-0.03	0.00	0.45
Cholesterol	0.43	0.49	0.26	0.43	0.59	0.18	0.43	0.36	0.10	0.07	0.01	0.02
Cholestanol	0.23	0.35	0.13	0.29	0.50	0.11	0.12	0.10	-0.09	-0.11	-0.09	-0.10
Coprostanol	0.30	0.48	0.11	0.33	0.61	0.03	0.32	0.52	0.05	-0.04	0.15	0.05
Caffeine	0.12	0.04	0.22	0.18	0.16	0.22	-0.07	-0.14	0.29	-0.06	-0.27	0.10
Acetaminophen	0.64	0.69	0.33	0.67	0.76	0.29	0.34	0.67	0.30	0.66	0.71	0.06
Ibuprofen	0.18	0.31	0.29	0.26	0.43	0.29	0.01	0.58	0.17	0.11	0.03	0.48
Sucralose	0.51	0.36	0.47	0.48	0.41	0.37	0.53	0.11	0.67	0.36	0.29	0.34
Saccharin	-0.17	0.16	-0.06	-0.06	0.30	-0.09	-0.58	0.19	-0.14	-0.24	-0.25	-0.01



Bold values indicate correlations significant at p ≤ 0.05.

Table 6.8. Second segregation method for samples from high-density residential areas

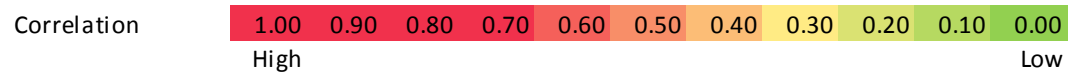
	All N = 107			"High" dataset <i>E. coli</i> ≥ 0.70 N = 41			"Moderate" dataset 0.30 < <i>E. coli</i> < 0.70 N = 30			"Low" dataset <i>E. coli</i> ≤ 0.30 N = 36		
	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent
Chloride	0.07	0.10	0.25	0.02	0.09	0.29	0.18	0.18	0.31	-0.01	-0.03	0.34
Boron	-0.05	0.08	0.07	0.04	0.26	0.30	-0.14	0.07	-0.21	0.00	-0.04	0.19
Orthophosphate	0.15	0.41	-0.04	0.20	0.50	0.17	0.06	0.24	-0.43	0.13	0.19	-0.24
Detergent as MBAS	0.28	0.22	0.00	-0.03	-0.12	-0.46	0.26	0.21	-0.02	0.23	0.13	-0.02
DEHP	0.13	0.30	0.25	0.27	0.41	0.26	0.06	0.21	0.25	-0.11	0.08	0.26
Cholesterol	0.43	0.49	0.26	0.52	0.69	0.32	0.48	0.44	0.06	0.15	0.01	0.13
Cholestanol	0.23	0.35	0.13	0.32	0.55	0.27	0.22	0.11	-0.05	-0.09	-0.11	-0.08
Coprostanol	0.30	0.48	0.11	0.42	0.68	0.22	0.45	0.46	-0.01	-0.11	0.05	0.01
Caffeine	0.12	0.04	0.22	0.11	0.14	0.19	-0.04	0.11	0.17	0.13	-0.21	0.19
Acetaminophen	0.64	0.69	0.33	0.66	0.71	0.45	0.64	0.83	0.25	0.59	0.71	0.09
Ibuprofen	0.18	0.31	0.29	0.18	0.41	0.34	0.34	0.45	0.25	0.02	0.12	0.32
Sucralose	0.51	0.36	0.47	0.48	0.44	0.47	0.52	0.33	0.40	0.43	0.15	0.44
Saccharin	-0.17	0.16	-0.06	0.11	0.33	0.13	-0.50	-0.11	-0.21	-0.38	-0.16	-0.20



Bold values indicate correlations significant at p ≤ 0.05.

Table 6.9. Third segregation method for samples from high-density residential areas

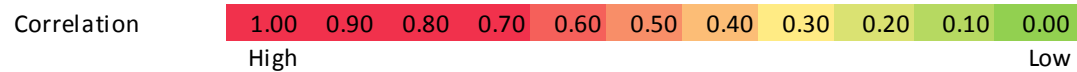
	All N = 107			"High" dataset Enterococci ≥ 0.70 N = 43			"Moderate" dataset 0.30 < Enterococci < 0.70 N = 26			"Low" dataset Enterococci ≤ 0.30 N = 38		
	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent
Chloride	0.07	0.10	0.25	0.01	0.04	0.41	0.20	0.24	0.09	-0.03	0.06	0.28
Boron	-0.05	0.08	0.07	-0.01	0.19	0.13	0.01	0.29	0.22	-0.01	0.02	0.25
Orthophosphate	0.15	0.41	-0.04	0.13	0.50	-0.06	0.22	0.64	0.00	0.04	0.15	-0.30
Detergent as MBAS	0.28	0.22	0.00	0.25	0.18	-0.38	0.01	-0.01	-0.33	0.19	0.19	-0.11
DEHP	0.13	0.30	0.25	0.30	0.41	0.19	0.07	0.44	0.25	-0.01	0.09	0.36
Cholesterol	0.43	0.49	0.26	0.48	0.61	0.12	0.49	0.60	0.20	0.19	0.22	0.01
Cholestanol	0.23	0.35	0.13	0.34	0.55	0.02	0.13	0.29	0.10	-0.01	0.10	-0.16
Coprostanol	0.30	0.48	0.11	0.36	0.63	-0.06	0.24	0.53	0.14	0.10	0.24	-0.02
Caffeine	0.12	0.04	0.22	0.22	0.13	0.30	0.00	0.02	0.16	-0.01	-0.11	0.12
Acetaminophen	0.64	0.69	0.33	0.66	0.74	0.16	0.49	0.68	0.40	0.58	0.62	0.03
Ibuprofen	0.18	0.31	0.29	0.36	0.42	0.35	0.08	0.57	0.34	0.11	0.09	0.34
Sucralose	0.51	0.36	0.47	0.39	0.34	0.35	0.68	0.30	0.69	0.44	0.41	0.37
Saccharin	-0.17	0.16	-0.06	-0.04	0.32	-0.16	-0.38	0.31	-0.04	-0.30	-0.24	-0.07



Bold values indicate correlations significant at $p \leq 0.05$.

Table 6.10. First segregation method for samples from low-density residential areas

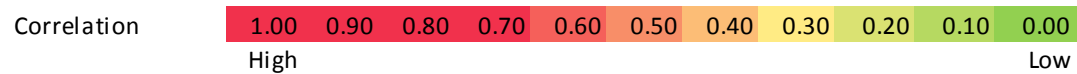
	All N = 90			"High" dataset <i>E. coli</i> or enterococci ≥ 0.70 N = 54			"Moderate" dataset $0.30 < E. coli$ or enterococci < 0.70 N = 23			"Low" dataset <i>E. coli</i> and enterococci ≤ 0.30 N = 13		
	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent
Chloride	0.53	0.29	0.20	0.61	0.38	0.20	0.38	0.22	0.36	0.65	0.49	0.12
Boron	0.54	0.18	0.12	0.56	0.23	0.16	0.67	0.08	-0.09	0.00	-0.08	0.06
Orthophosphate	0.57	0.25	0.09	0.56	0.24	0.19	0.36	0.13	-0.16	0.80	0.37	-0.19
Detergent as MBAS	0.66	0.26	0.11	0.63	0.20	0.19	0.72	0.37	-0.02	0.68	0.29	-0.34
DEHP	0.56	0.40	0.20	0.57	0.44	0.23	0.65	0.50	0.16	0.55	0.62	0.02
Cholesterol	0.60	0.31	0.16	0.64	0.28	0.29	0.33	0.04	-0.36	0.45	0.42	-0.44
Cholestanol	0.32	0.23	0.13	0.40	0.31	0.17	0.00	-0.10	-0.21	-0.03	-0.31	-0.10
Coprostanol	0.29	0.34	0.23	0.36	0.31	0.20	-0.30	-0.01	-0.08	0.30	0.01	-0.30
Caffeine	0.07	0.22	0.18	0.20	0.41	0.15	-0.51	-0.07	0.16	0.02	-0.61	0.34
Acetaminophen	0.05	-0.02	0.06	0.02	-0.08	0.00	0.14	0.04	0.07	-0.24	-0.32	0.43
Ibuprofen	-0.30	-0.16	-0.07	-0.26	-0.21	-0.16	-0.46	-0.08	0.10	-0.31	-0.01	0.11
Sucralose	-0.12	-0.11	0.08	-0.06	-0.18	-0.01	-0.01	-0.17	0.17	-0.47	0.50	0.59
Saccharin	0.24	-0.06	-0.09	0.20	-0.12	-0.01	0.18	-0.14	-0.40	0.33	-0.01	-0.49



Bold values indicate correlations significant at $p \leq 0.05$.

Table 6.11. Second segregation method for samples from low-density residential areas

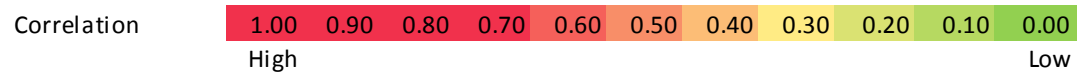
	All N = 90			"High" dataset <i>E. coli</i> ≥ 0.70 N = 36			"Moderate" dataset 0.30 < <i>E. coli</i> < 0.70 N = 24			"Low" dataset <i>E. coli</i> ≤ 0.30 N = 30		
	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent
Chloride	0.53	0.29	0.20	0.70	0.51	0.26	0.25	0.29	0.21	0.55	0.54	0.13
Boron	0.54	0.18	0.12	0.66	0.36	0.03	0.48	0.11	0.17	0.57	0.27	0.20
Orthophosphate	0.57	0.25	0.09	0.59	0.33	0.11	0.43	0.29	0.24	0.72	0.51	-0.02
Detergent as MBAS	0.66	0.26	0.11	0.72	0.34	0.10	0.42	0.40	0.33	0.84	0.50	0.00
DEHP	0.56	0.40	0.20	0.65	0.50	0.23	0.34	0.52	0.30	0.59	0.63	0.11
Cholesterol	0.60	0.31	0.16	0.69	0.48	0.16	0.50	0.16	0.26	0.61	0.45	0.15
Cholestanol	0.32	0.23	0.13	0.59	0.51	0.09	0.10	-0.09	-0.02	0.14	0.15	0.31
Coprostanol	0.29	0.34	0.23	0.51	0.37	0.16	-0.29	0.21	0.03	0.30	0.25	0.37
Caffeine	0.07	0.22	0.18	0.37	0.42	0.15	-0.18	0.19	0.17	-0.19	-0.07	0.32
Acetaminophen	0.05	-0.02	0.06	0.01	-0.19	-0.14	0.23	-0.08	0.14	-0.08	-0.13	0.18
Ibuprofen	-0.30	-0.16	-0.07	-0.31	-0.36	-0.26	-0.39	-0.24	-0.15	-0.38	-0.07	0.07
Sucralose	-0.12	-0.11	0.08	-0.05	-0.30	-0.10	0.01	-0.25	-0.02	-0.37	0.02	0.30
Saccharin	0.24	-0.06	-0.09	0.22	-0.11	0.02	0.14	-0.42	-0.07	0.35	0.07	-0.27



Bold values indicate correlations significant at $p \leq 0.05$.

Table 6.12. Third segregation method for samples from low-density residential areas

	All N = 90			"High" dataset Enterococci ≥ 0.70 N = 35			"Moderate" dataset 0.30 < Enterococci < 0.70 N = 30			"Low" dataset Enterococci ≤ 0.30 N = 25		
	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent	TC	<i>E. coli</i>	Ent
Chloride	0.53	0.29	0.20	0.66	0.36	0.46	0.33	0.12	0.34	0.66	0.50	0.16
Boron	0.54	0.18	0.12	0.50	0.13	0.41	0.67	0.22	-0.03	0.42	0.27	0.06
Orthophosphate	0.57	0.25	0.09	0.54	0.20	0.39	0.52	0.34	0.00	0.60	0.23	0.00
Detergent as MBAS	0.66	0.26	0.11	0.59	0.14	0.51	0.77	0.36	0.02	0.65	0.30	-0.14
DEHP	0.56	0.40	0.20	0.56	0.43	0.60	0.60	0.48	0.03	0.59	0.35	-0.01
Cholesterol	0.60	0.31	0.16	0.65	0.23	0.50	0.56	0.39	0.01	0.52	0.28	-0.19
Cholestanol	0.32	0.23	0.13	0.08	0.01	0.32	0.43	0.48	0.13	0.45	0.07	0.01
Coprostanol	0.29	0.34	0.23	0.23	0.18	0.39	0.23	0.30	0.11	0.43	0.55	0.04
Caffeine	0.07	0.22	0.18	-0.03	0.35	0.21	0.02	0.24	0.25	0.20	-0.09	0.25
Acetaminophen	0.05	-0.02	0.06	-0.08	-0.11	0.06	0.09	0.05	0.23	0.06	-0.07	0.26
Ibuprofen	-0.30	-0.16	-0.07	-0.30	-0.16	-0.18	-0.41	-0.14	-0.20	-0.22	-0.14	-0.15
Sucralose	-0.12	-0.11	0.08	-0.10	-0.17	0.05	-0.15	-0.26	0.18	-0.17	0.13	0.06
Saccharin	0.24	-0.06	-0.09	0.15	-0.10	-0.07	0.05	-0.14	-0.23	0.40	0.02	-0.49



Bold values indicate correlations significant at $p \leq 0.05$.

For high-density residential areas (Tables 6.7 – 6.9), all segregation methods give a greater number of significant ($p \leq 0.05$) correlations for the “high” dataset. A closer look reveals that the second segregation method (Table 6.8) based on only *E. coli* is slightly better than the first segregation method (Table 6.7) in that it results in higher ($r_s \geq 0.44$) correlations between *E. coli* or enterococci and other parameters such as DEHP, cholesterol, cholestanol, coprostanol and sucralose for the “high” dataset. The third segregation method (Table 6.9) based on only enterococci gives slightly weaker ($r_s \geq 0.32$) correlation coefficients between FIB and chemical parameters than the second method, but slightly higher than the first method for the “high” dataset. However, the third segregation method gives a greater number of significant ($p \leq 0.05$) correlations for the “moderate” dataset than the other segregation methods. This indicates enterococci do not differentiate the chemical parameters well enough. In addition, although the third method segregates data based on enterococci, the correlations are better between chemical parameters and *E. coli* than enterococci.

For low-density residential areas (Tables 6.10 – 6.12), chemical parameters have better correlation with *E. coli* than enterococci when using the first and second segregation methods. However, the reverse outcome, in which chemical parameters have better correlation with enterococci than *E. coli*, occurs with the third segregation method. Enterococci has higher ($r_s = 0.39$) correlation with coprostanol for the “high” dataset than the other two datasets with this third segregation method. But, *E. coli* has higher ($r_s = 0.55$) correlation with coprostanol for the “low” dataset with this third segregation method. Similarly to high-density residential, a closer look reveals that the second segregation method ($r_s \geq 0.33$) is better than the first segregation method ($r_s \geq 0.28$) for the “high” dataset. However, the second segregation method gives higher correlation between *E. coli* and some other parameters such as chloride, orthophosphate, detergent as MBAS and DEHP for the “low” dataset ($r_s \geq 0.50$) than the “high” dataset ($r_s \geq 0.33$).

In general, there is a greater number of significant ($p \leq 0.05$) correlations in the “high” dataset than in the “low” dataset. This implies that there is better agreement between FIB and chemical parameters when FIB concentrations are high. The chemical parameters analysed in this study can be associated with wastewater. Therefore, higher FIB concentrations with marked repeatability will most likely have originated from leaking sewers, while lower FIB concentrations are likely background concentrations. The different strength of correlations for different segregated datasets in this study also implies that any sampling program must be designed carefully in order to effectively monitor for human faecal and sewage contamination of surface waters.

Overall, the first and second segregation methods give a roughly similar correlation trend, which is higher correlation between *E. coli* and other parameters for the “high” dataset. For measurements in high-density residential areas, the third segregation method gives higher correlation between *E. coli* and other parameters for the “high” dataset, whereas for low-density residential data, the third segregation method gives higher correlation between enterococci and other parameters for the “high” dataset. Since *E. coli* and enterococci are widely used indicator bacteria, both *E. coli* and enterococci are used to segregate the data for further analysis in this chapter.

6.2.4. Associations among chemical parameters and FIB

Box plots are presented in Figure 6.6 showing the distribution of the segregated chemical data. The data are classified based on land use and normalised *E. coli* or enterococci values as follows: H indicates high-density residential, L low-density residential, 1 indicates the “high” dataset (normalised values of *E. coli* or enterococci ≥ 0.70), 2 the “moderate” dataset ($0.30 < E. coli$ or enterococci < 0.70) and 3 the “low” dataset (*E. coli* and enterococci ≤ 0.30). For example, H1 means samples of high-density residential areas in which normalised *E. coli* or enterococci values are greater or equal to 0.70. The geometric mean concentration is shown on each box plot with a

black circle symbol (●). The number above each box plot indicates the number of samples.

Figure 6.6 shows that generally geometric means of chemical concentrations decrease from the “high” to the “low” dataset. Particularly, the geometric means of cholesterol, cholestanol, coprostanol, acetaminophen and sucralose in high-density residential areas and of coprostanol and saccharin in low-density residential areas decrease from the “high” to the “low” dataset. The “high”, “moderate” and “low” datasets were defined by high, moderate and low normalised FIB concentrations. This observation shows that generally higher normalised FIB concentrations are accompanied by higher chemical concentrations. Since the chemical parameters analysed in this study can be associated with wastewater, high FIB concentrations are likely to result from human faecal and sewage contamination such as from leaking sewers.

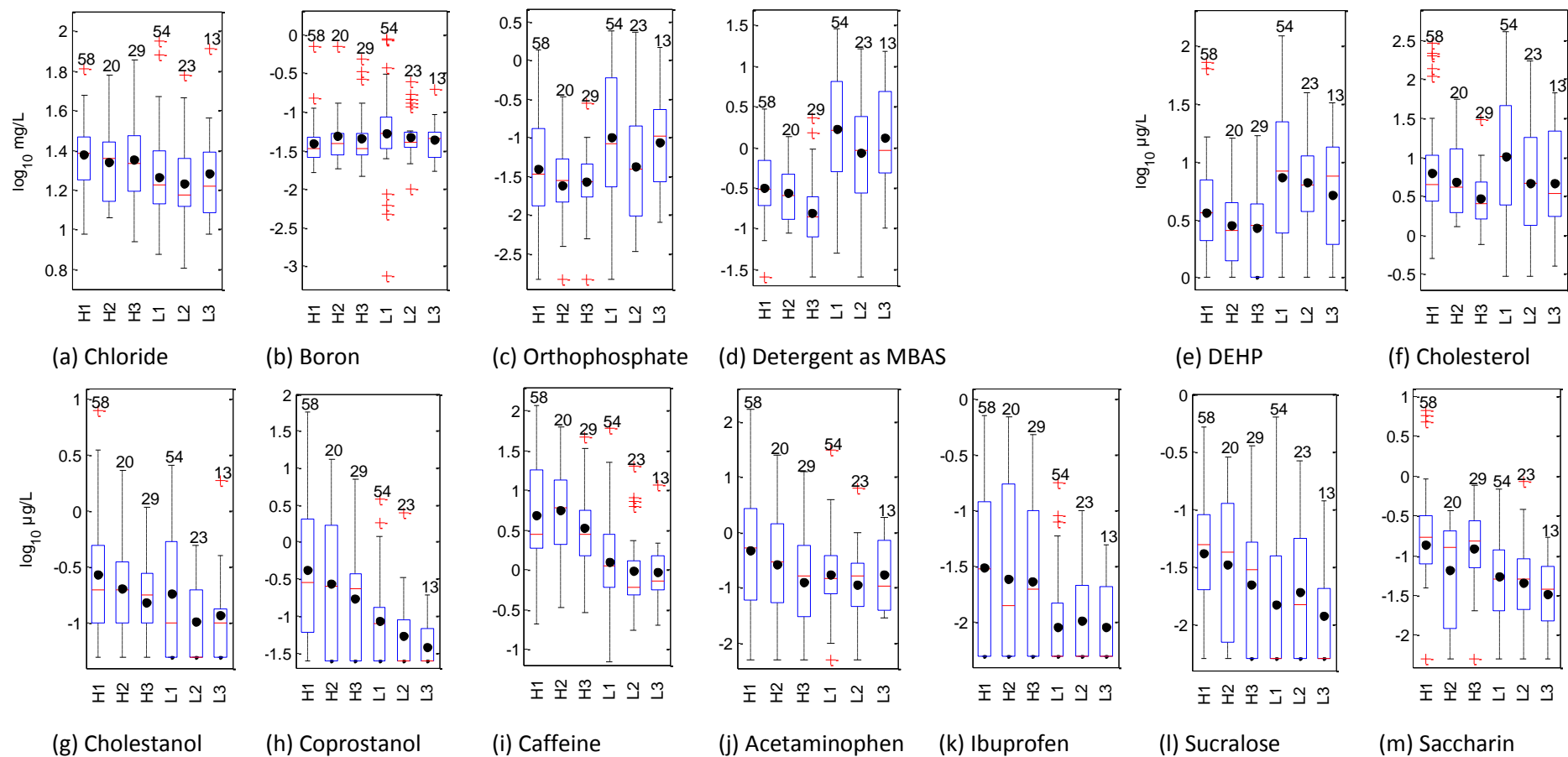


Figure 6.6. Box plots of measured concentrations of 13 chemical parameters based on land use (H or L) and concentrations groups (1, 2 or 3) defined by the first segregation method

In order to investigate this possibility in greater detail, Spearman correlation analysis (Section 4.5.5) was conducted on the segregated datasets. Complete correlation matrices among the chemical parameters, FIB and specific conductance are presented in Tables 6.13 - 6.18. The data are segregated based on the first segregation method: (i) "high" dataset for normalised *E. coli* or enterococci concentrations ≥ 0.70 , (ii) "moderate" dataset for *E. coli* or enterococci < 0.70 and > 0.30 and (iii) "low" dataset for *E. coli* and enterococci ≤ 0.30 . The correlation matrices for high-density residential areas are presented in Tables 6.13 through 6.15 and for low-density residential areas in Tables 6.16 through 6.18.

For high-density residential areas, a greater number of significant ($p \leq 0.05$) correlations among DEHP, cholesterol, cholestanol, coprostanol, acetaminophen, ibuprofen, sucralose and saccharin is found in the "high" dataset (Table 6.13) than in the "moderate" and "low" datasets (Tables 6.14 – 6.15) or in the overall high-density residential dataset (Table 6.3). Orthophosphate is more highly correlated with cholesterol, cholestanol and coprostanol in the "high" dataset ($r_s \geq 0.65$) than the overall high-density residential dataset ($r_s \geq 0.54$). This may further indicate minimal subsurface adsorption. In general, there is a greater number of significant ($p \leq 0.05$) correlations between FIB and chemical parameters in the "high" dataset than in the "moderate" and "low" datasets. *E. coli* is more highly correlated with coprostanol in the "high" dataset ($r_s \geq 0.61$) than in the "moderate" ($r_s \geq 0.52$) and "low" datasets (correlation is not significant). In the "high" dataset, *E. coli* has a greater number of significant ($p \leq 0.05$) correlations with chemical parameters than do enterococci.

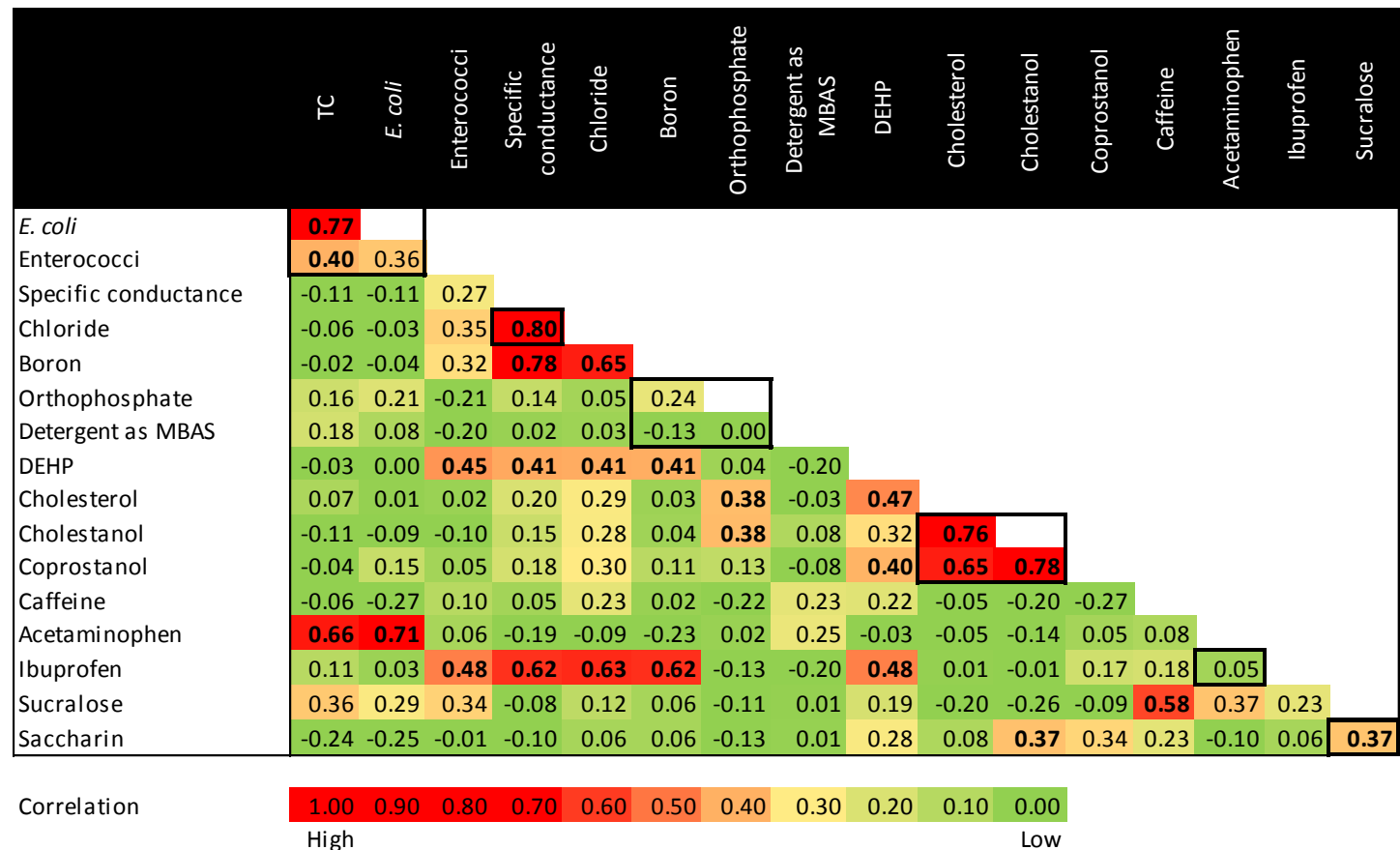
Table 6.14. Correlation matrix of “moderate” dataset (20 samples) from high-density residential areas

	TC	<i>E. coli</i>	Enterococci	Specific conductance	Chloride	Boron	Orthophosphate	Detergent as MBAS	DEHP	Cholesterol	Cholestanol	Coprostanol	Caffeine	Acetaminophen	Ibuprofen	Sucralose
<i>E. coli</i>	0.30															
Enterococci	0.59	-0.01														
Specific conductance	0.20	0.29	0.19													
Chloride	0.25	0.16	0.36	0.57												
Boron	-0.01	0.29	0.17	0.63	0.44											
Orthophosphate	0.23	0.37	-0.06	0.22	0.14	0.23										
Detergent as MBAS	0.14	0.04	0.09	0.14	0.09	-0.11	0.57									
DEHP	-0.21	0.47	-0.06	0.39	0.03	0.38	0.06	0.08								
Cholesterol	0.43	0.36	0.10	0.37	0.34	0.12	0.57	0.61	0.24							
Cholestanol	0.12	0.10	-0.09	0.25	0.24	0.24	0.65	0.45	0.06	0.77						
Coprostanol	0.32	0.52	0.05	0.39	0.29	0.24	0.63	0.45	0.16	0.73	0.67					
Caffeine	-0.07	-0.14	0.29	-0.09	0.31	0.11	-0.07	-0.03	0.19	-0.09	0.01	-0.29				
Acetaminophen	0.34	0.67	0.30	0.05	0.19	0.03	0.13	0.24	0.24	0.24	-0.08	0.28	0.10			
Ibuprofen	0.01	0.58	0.17	0.60	0.61	0.63	0.12	0.02	0.53	0.24	0.03	0.26	0.20	0.55		
Sucralose	0.53	0.11	0.67	0.41	0.29	0.39	0.01	0.03	0.01	0.35	0.27	0.26	0.20	0.10	0.15	
Saccharin	-0.58	0.19	-0.14	0.31	0.21	0.52	0.09	-0.05	0.55	-0.09	-0.05	-0.06	0.22	0.10	0.65	-0.15

Correlation	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00
	High					Low					

Bold values indicate correlations significant at $p \leq 0.05$.

Table 6.15. Correlation matrix of “low” dataset (29 samples) from high-density residential areas



Bold values indicate correlations significant at $p \leq 0.05$.

Table 6.16. Correlation matrix of “high” dataset (54 samples) from low-density residential areas

	TC	<i>E. coli</i>	Enterococci	Specific conductance	Chloride	Boron	Orthophosphate	Detergent as MBAS	DEHP	Cholesterol	Cholestanol	Coprostanol	Caffeine	Acetaminophen	Ibuprofen	Sucralose
<i>E. coli</i>	0.36															
Enterococci	0.19	-0.03														
Specific conductance	0.46	0.28	0.06													
Chloride	0.61	0.38	0.20	0.71												
Boron	0.56	0.23	0.16	0.69	0.59											
Orthophosphate	0.56	0.24	0.19	0.48	0.62	0.70										
Detergent as MBAS	0.63	0.20	0.19	0.74	0.66	0.75	0.73									
DEHP	0.57	0.44	0.23	0.50	0.45	0.58	0.58	0.74								
Cholesterol	0.64	0.28	0.29	0.66	0.59	0.79	0.74	0.81	0.73							
Cholestanol	0.40	0.31	0.17	0.34	0.25	0.48	0.38	0.46	0.46	0.50						
Coprostanol	0.36	0.31	0.20	0.29	0.38	0.44	0.41	0.52	0.54	0.54	0.68					
Caffeine	0.20	0.41	0.15	0.11	0.04	0.31	0.19	0.22	0.43	0.30	0.61	0.44				
Acetaminophen	0.02	-0.08	0.00	-0.34	-0.14	-0.10	0.02	-0.08	-0.04	-0.09	0.30	0.27	0.35			
Ibuprofen	-0.26	-0.21	-0.16	-0.50	-0.45	-0.38	-0.33	-0.40	-0.14	-0.39	-0.18	-0.09	0.02	0.19		
Sucralose	-0.06	-0.18	-0.01	-0.15	-0.07	0.02	-0.21	-0.12	-0.15	-0.26	-0.24	-0.09	-0.11	-0.11	0.29	
Saccharin	0.20	-0.12	-0.01	0.17	0.21	0.25	0.10	0.12	-0.04	0.21	-0.28	-0.10	-0.10	-0.09	-0.13	0.12

Correlation: 1.00 0.90 0.80 0.70 0.60 0.50 0.40 0.30 0.20 0.10 0.00

High Low

Bold values indicate correlations significant at $p \leq 0.05$.

For low-density residential areas, a greater number of significant ($p \leq 0.05$) correlations among specific conductance, chloride, boron, orthophosphate, detergent as MBAS, DEHP, cholesterol, cholestanol and coprostanol is found in the “high” dataset (Table 6.16) than in the “moderate” and “low” datasets (Tables 6.17 – 6.18). Correlations among chloride, boron, orthophosphate, detergent as MBAS, DEHP, cholesterol, cholestanol and coprostanol are higher in the “high” dataset ($r_s \geq 0.38$) than in the overall low-density residential dataset ($r_s \geq 0.32$, Table 6.4). In addition, in the “high” dataset, *E. coli* has significant ($p \leq 0.05$) correlations with specific conductance, chloride, cholesterol, cholestanol, coprostanol and caffeine, while enterococci shows significant ($p \leq 0.05$) correlation with cholesterol. These significant correlations are not found in the “moderate” and “low” datasets. In low-density residential areas, the correlations of boron, orthophosphate and detergent as MBAS must be interpreted cautiously because they may originate from two possible sources: leaking sewers or direct discharge of household washing water into surface drains, as discussed in Sections 3.3.2, 6.1.2 and 6.2.2. Boron, orthophosphate and detergent as MBAS are moderately ($r_s \geq 0.41$) correlated with coprostanol in the “high” dataset. This may indicate that in the “high” dataset, boron, orthophosphate and detergent as MBAS may originate from leaking sewers. Similarly to high-density residential areas, *E. coli* has better correlation with other parameters than enterococci for all segregated datasets, especially the “high” dataset. This further indicates *E. coli* is more closely associated with human faecal and sewage contamination than enterococci. The correlations of ibuprofen or sucralose with other parameters, either positive or negative, in the segregated datasets may be invalid because these compounds were not detected in about half or more of the samples analysed.

6.2.5. Influence of age of sewers on chemical data

Chapter 5 makes the case that leaking sewers contribute to the levels of faecal contamination in storm drains, a conclusion further supported by the findings in Section 6.2.4. Exfiltration from sewer leaks may occur at the most upstream sewers

such as the laterals that connect buildings to the deeper public sewers at the manholes. Leaking sewers are hypothesised to occur due to damage during backfilling, subsidence after backfilling operations or failing old infrastructure. In this section, the influence of the age of sewers in impairing surface water quality is investigated. The “high” datasets (using the first segregation method that is based on the normalised concentration of *E. coli* or enterococci exceeding 0.7), H1 and L1, from the previous section are further segregated based on the age of sewers. Sampling sites were classified into sites with old and new sewers, estimated from the age of the respective area developments. High-density residential areas with old sewers include Lorong 8 Toa Payoh, Lorong 6 Toa Payoh, Ang Mo Kio Avenue 10 and Toa Payoh North whereas the sewers in Punggol Central and Choa Chu Kang (CCK) Crescent are newer. Low-density residential areas with old sewers include Jalan Siantan, MacPherson Lane, Serangoon Garden and Watten Drive whereas the sewers in Verde are relatively newer. Box plots (Figure 6.7) were created in order to visualise the chemical data distribution in each segregated dataset. The geometric mean concentration is superimposed on each box plot with a black circle symbol (●). The number above each box plot indicates the number of samples.

For high-density residential areas, the geometric mean concentration of all chemical parameters, except for caffeine, is higher where the sewers are older. In contrast, for low-density residential areas, most of the geometric mean values are lower where there are old sewers than where there are new sewers, except for acetaminophen, ibuprofen and sucralose. However, this comparison may not be general enough because only one low-density residential sampling site, Verde, is serviced by new sewers. Nonetheless, results for old and new sewers in low-density residential areas are still compared since it might reveal some useful information. Spearman correlation analysis (Section 4.5.5) was conducted on the segregated datasets. The results are presented in correlation matrices (Tables 6.19 – 6.22).

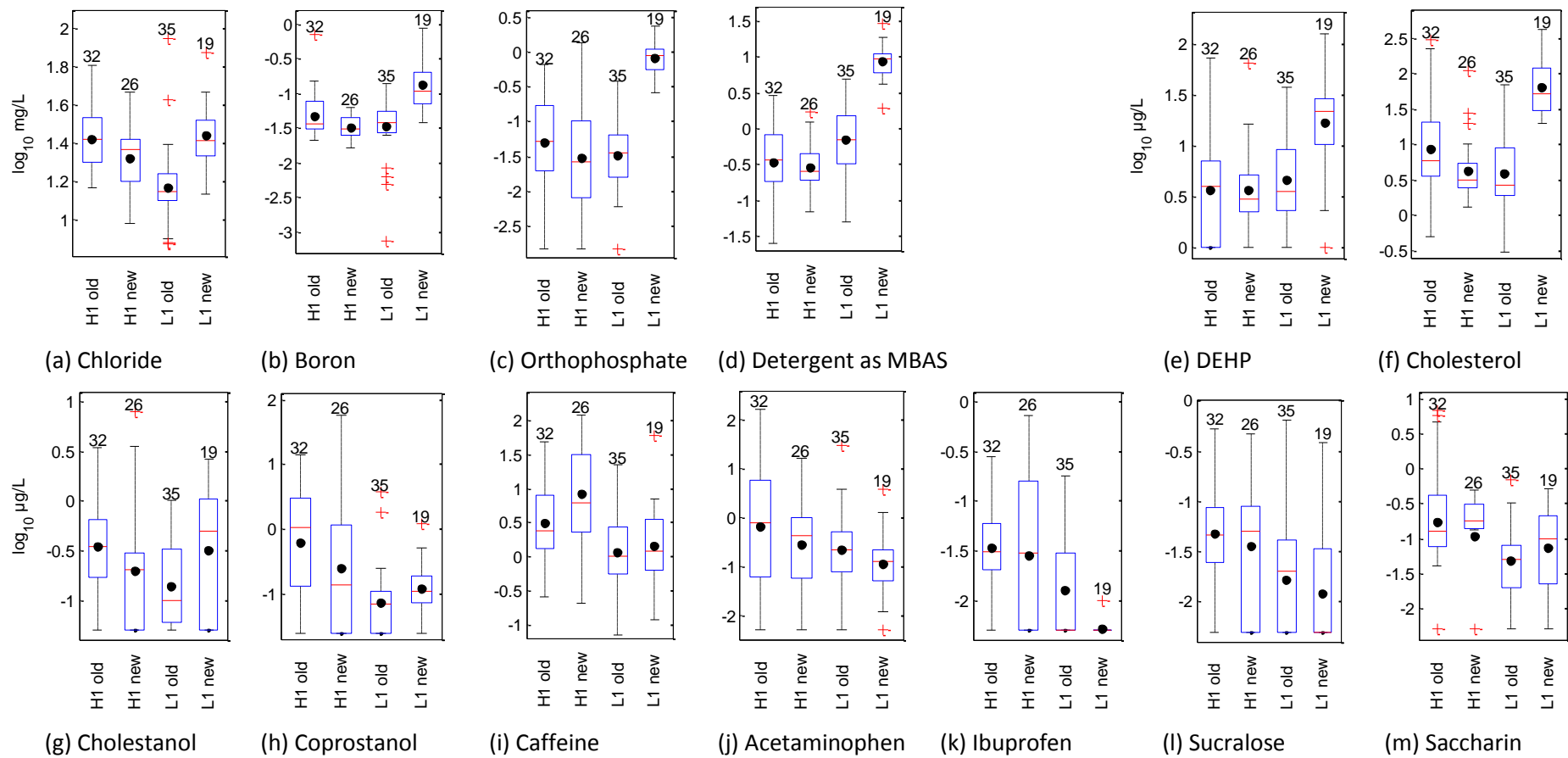


Figure 6.7. Box plots of 13 chemical parameters based on land uses, “high” dataset and age of sewers

Table 6.21. Correlation matrix of 35 samples from low-density residential areas, “high” dataset and old sewers

	TC	<i>E. coli</i>	Enterococci	Specific conductance	Chloride	Boron	Orthophosphate	Detergent as MBAS	DEHP	Cholesterol	Cholestanol	Coprostanol	Caffeine	Acetaminophen	Ibuprofen	Sucralose
<i>E. coli</i>	0.30															
Enterococci	0.27	-0.02														
Specific conductance	0.13	0.20	0.09													
Chloride	0.31	0.30	0.25	0.65												
Boron	0.33	0.09	0.06	0.55	0.35											
Orthophosphate	0.13	0.17	0.14	0.00	0.34	0.33										
Detergent as MBAS	0.28	-0.02	0.18	0.61	0.45	0.34	0.16									
DEHP	0.45	0.41	0.31	0.28	0.40	0.26	0.39	0.54								
Cholesterol	0.45	0.24	0.34	0.45	0.26	0.60	0.34	0.48	0.55							
Cholestanol	0.24	0.25	0.15	0.56	0.27	0.50	0.31	0.50	0.52	0.85						
Coprostanol	0.16	0.18	0.12	0.25	0.31	0.17	0.23	0.44	0.44	0.59	0.63					
Caffeine	0.16	0.32	0.14	0.13	0.00	0.32	0.21	0.20	0.50	0.52	0.49	0.28				
Acetaminophen	0.04	-0.24	-0.14	-0.38	-0.17	-0.15	0.18	-0.03	-0.06	0.00	0.00	0.23	0.20			
Ibuprofen	0.03	-0.16	-0.08	-0.39	-0.24	-0.20	0.01	-0.16	0.15	-0.12	-0.21	0.07	0.04	0.15		
Sucralose	-0.05	-0.41	-0.10	-0.01	0.07	0.12	-0.16	-0.07	-0.30	-0.38	-0.38	-0.17	-0.37	-0.18	0.35	
Saccharin	0.15	-0.30	0.06	0.02	0.00	0.24	-0.01	-0.06	-0.26	0.05	-0.15	-0.18	-0.04	0.18	-0.03	0.09

Correlation

High Low

Bold values indicate correlations significant at $p \leq 0.05$.

For high-density residential areas (Tables 6.19 – 6.20), a greater number of significant ($p \leq 0.05$) correlations among many chemical parameters (DEHP, cholesterol, cholestanol, coprostanol, caffeine, acetaminophen, sucralose and saccharin) are found in areas with old sewers. In contrast, ibuprofen has a greater number of significant ($p \leq 0.05$) correlations with other parameters in areas with new sewers. There are higher correlations between *E. coli* and orthophosphate ($r_s = 0.51$), cholesterol ($r_s = 0.59$), acetaminophen ($r_s = 0.86$) and saccharin ($r_s = 0.37$), and also between enterococci and ibuprofen ($r_s = 0.35$) and between enterococci and sucralose ($r_s = 0.47$) in areas with old sewers. On the other hand, there are higher correlations between *E. coli* and chloride ($r_s = 0.48$), DEHP ($r_s = 0.43$), cholestanol ($r_s = 0.50$), coprostanol ($r_s = 0.70$), ibuprofen ($r_s = 0.47$) and sucralose ($r_s = 0.61$) in areas with new sewers. This pattern of correlation indicates that although leaking sewers appear to be more prevalent in areas of old sewers, which is to be expected, the age of sewers is not the only factor resulting in leaking sewers since some correlations among parameters are still significant and sometimes higher in the new-sewer dataset than in the old-sewer dataset. There are more uncertainties with respect to the age of the sewer as old sewers may have undergone rehabilitation or an area may have a combination of old and new sewers since new sewers have often been added to an area with old sewers to increase the sewer capacity as the area grew in population.

Unlike high-density residential areas, in low-density residential areas (Table 6.21 – 6.22), there is a greater number of significant ($p \leq 0.05$) correlations among parameters in areas with new sewers than in areas with old sewers. There are higher correlations among detergent as MBAS, DEHP, cholestanol, coprostanol, caffeine and acetaminophen in the new-sewer dataset. *E. coli* is also moderately ($r_s \geq 0.47$) correlated with chloride, boron, detergent as MBAS, cholestanol, coprostanol and caffeine in the new-sewer dataset. On the other hand, there are higher correlations between cholesterol and cholestanol ($r_s = 0.85$), cholesterol and coprostanol ($r_s = 0.59$), and cholesterol and caffeine ($r_s = 0.52$) in the old-sewer dataset. It has been

noted earlier that the new-sewer dataset only contains data from one sampling location, Verde. Therefore, further studies are needed to verify this finding.

6.3. Conclusions

Analysis of the Toolbox Study in Section 3.3.2 shows that some chemical parameters have the same diurnal variation as FIB. Therefore, in this study, 13 chemical parameters were investigated in order to find agreement among tracers for human faecal and sewage contamination in urban catchments. It is hypothesised in Section 3.3.1 that land use may be an important factor in surface water characteristics. Hence, environmental surface water samples were collected from sampling locations with distinct major land uses: high-density residential, low-density residential, commercial and industrial. The Kruskal-Wallis test results indicate significant differences in most chemical concentrations between different land uses. In addition, PCA results show that samples from high- and low-density residential areas differ in the second principal component values: low-density residential areas are strongly influenced by detergent as MBAS while high-density residential areas are not. Therefore, in this chapter, the data are analysed to investigate the association among parameters for each land use category. Chapter 5 shows that leaking sewers are a likely source of surface water contamination. This chapter further investigates this possibility. Chapter 5 indicates that there is a greater probability of FIB coming from leaking sewers when normalised FIB concentrations are high. Therefore, the data are segregated based on the normalised FIB concentrations and correlation analysis is conducted on the segregated datasets. Sewers may be damaged during construction or with age. The latter possibility is investigated by further segregating the dataset based on the age of sewers and then correlation analysis is conducted on the segregated datasets.

The following conclusions can be summarised from this study:

(1) There are better correlations among different parameters in each land-use dataset (Tables 6.3 through 6.6) than in the overall dataset (Table 6.2). This shows that land use affects drain water quality. Correlation analysis of chemical data from high-density residential areas (Table 6.3) shows that there are significant correlations between cholesterol, cholestanol or coprostanol and most other chemical parameters whereas detergent as MBAS and caffeine are significantly correlated with few other parameters. On the other hand, correlation analysis of chemical data from low-density residential areas (Table 6.4) shows that there are significant correlations among chloride, boron, orthophosphate, detergent as MBAS, DEHP and cholesterol. In contrast with high-density residential areas, there is a greater number of significant ($p \leq 0.05$) correlations between caffeine and other parameters in low-density residential areas. Generally high-density residential areas have a higher degree of contamination than low-density residential areas since most of the chemical parameters have higher median and geometric mean concentrations in the high-density residential areas than in the low-density residential areas (Figure 6.3).

(2) *E. coli* is better correlated with other parameters than enterococci in each land use (Tables 6.3 through 6.6). However, this *E. coli* characteristic is not as evident in the industrial area as in other land uses. The closer association of *E. coli* with human faecal and sewage contamination than of enterococci is further demonstrated in the segregated datasets from both types of residential areas (Tables 6.7 and 6.10).

(3) Correlation analysis of the "high" datasets reveals that *E. coli* is moderately correlated with orthophosphate ($r_s = 0.50$), DEHP ($r_s = 0.34$), cholesterol ($r_s = 0.59$), cholestanol ($r_s = 0.50$), coprostanol ($r_s = 0.61$), sucralose ($r_s = 0.43$) and saccharin ($r_s = 0.41$) in the high-density residential areas (Table 6.13) while *E. coli* is moderately correlated with chloride ($r_s = 0.38$), DEHP ($r_s = 0.44$), cholestanol ($r_s = 0.31$), coprostanol ($r_s = 0.31$) and caffeine ($r_s = 0.41$) in the low-density residential areas

(Table 6.16). Acetaminophen is highly ($r_s = 0.76$) correlated with *E. coli* in the “high” dataset of high-density residential areas (Table 6.13). Generally, cholesterol, cholestanol and coprostanol can be used as effective indicators of faecal contamination for both types of residential areas since those showed significant ($p \leq 0.05$) correlations with *E. coli* in the “high” dataset for both land uses. Ibuprofen and sucralose are not good tracers for low-density residential areas because they are below the limit of detection for more than half of the samples in the “high” dataset.

(4) Age of sewers seems to influence leaking sewers to a higher degree in high-density residential areas than low-density residential areas since there is a greater number of significant ($p \leq 0.05$) correlations among parameters in the old-sewer dataset from high-density residential areas than in the old-sewer dataset from low-density residential areas. However, it should be borne in mind that the age of sewers is not the sole determinant as other factors, such as the number of building connections to sewers, poor workmanship/poor quality of sewer construction, damage during backfilling, settling immediately after backfilling or subsidence, may also result in leaking sewers.

Overall, Chapter 5 has discussed that FIB concentrations vary diurnally and this diurnal pattern of FIB concentrations closely resembles the daily sewage flow pattern. Chapter 6 discusses that generally there are better correlations among parameters when the data are segregated based on land use and high normalised FIB concentrations. These two lines of evidence established in the two chapters, (i) diurnal pattern of FIB concentrations and (ii) agreement among chemical tracers, support the hypothesis that leaking sewage is contaminating urban storm drains in Singapore.

CHAPTER 7 DATA ANALYSIS BASED ON HF183 GROUPS

Bacteroides-Prevotella is reported as a potential indicator because of its ease of detection, longer survival rate in water and relatively stable abundance in the human faecal matter (Bernhard and Field 2000a). Specifically, HF183 marker which can be found in *Bacteroides-Prevotella* targets a more specific gene that is likely found in humans only (Bernhard and Field 2000b). Moreover, the applicability of HF183 marker has been reported in many areas: North America (the USA, Canada and Puerto Rico), many countries in Europe (Belgium, France, Ireland, Portugal and the United Kingdom), Australia, tropical Africa (Kenya) and Asia (Singapore and Bangladesh) (Gawler et al. 2007; Santoro and Boehm 2007; Jenkins et al. 2009; Nshimyimana 2010; Ahmed et al. 2010b; Santiago-Rodriguez et al. 2012). The HF183 marker can be detected using PCR methods and quantified using QPCR methods (Section 4.4.3). The QPCR technique however has some intrinsic disadvantages such as (i) the use of standards (which requires plasmid DNA as a standard from a single, pure species, accurate pipetting and stability of diluted standards) and (ii) the use of amplification cycles to give quantitative results which is susceptible to inhibition (Baker 2012). Further, the HF183 quantification methodology for application in the tropics is still under development and not yet fully reliable. Although the absolute human factor (HF) values are imprecise, it is reasonable to assume that the HF183 data are sufficiently accurate to carry out a bulk analysis by separating the data into three categories: high, medium and low HF183 concentration groups (Janelle R. Thompson, MIT, personal communication, 2013). Data analysis based on these HF183 groups is useful as it would be able to yield information such as whether FIB and chemical concentrations varied across HF groups, sampling time and land use.

7.1. Results

The HF data (in GE/100mL) were binned into three concentration groups: (1) high, (2) medium and (3) low HF concentration. The high HF group consists of $HF > 3 \times 10^5$ GE/100mL, the medium HF group consists of $2 \times 10^4 \text{ GE/100mL} \leq HF \leq 3 \times$

10^5 GE/100mL and the low HF group consists of $HF < 2 \times 10^4$ GE/100mL. The thresholds were decided based on the distribution of HF concentrations so that all bins include a roughly equal number of data.

HF183 marker was analysed in 563 samples collected from 13 sampling sites (Table 4.1). Out of 563 samples, 183 samples (33%) were categorised as high HF concentration group (group 1), 186 samples (33%) as medium HF concentration group (group 2) and 194 samples (34%) as low HF concentration group (group 3). FIB (total coliform or TC, *Escherichia coli* or *E. coli* and enterococci) were also analysed in all 563 samples, specific conductance was measured in 543 out of 563 samples and 13 chemical parameters (chloride, boron, orthophosphate, detergent as MBAS, DEHP, cholesterol, cholestanol, coprostanol, caffeine, acetaminophen, ibuprofen, sucralose and saccharin) were analysed in 137 out of 563 samples. Since HF183 concentrations were categorised in ranges, the data analysis methodology adopted in this chapter is the analysis of variance and Tukey-Kramer tests.

7.2. Data analysis and discussion

7.2.1. Analysis of all data based on HF group

All FIB and chemical concentrations were log-transformed and then grouped based on their corresponding HF values into group 1 (group with high HF values), 2 (medium HF values) or 3 (low HF values). The distribution of FIB and chemical data in each HF group is illustrated with box plots in Figures 7.1 and 7.2. The geometric mean concentration is superimposed on each box plot with a black circle symbol (●). The number of data in each HF group is indicated above each box plot. Geometric means of TC, *E. coli*, enterococci, cholestanol, coprostanol, acetaminophen, ibuprofen, sucralose and saccharin decrease from the high HF group to the low HF group. This implies that higher HF concentrations are accompanied by higher concentrations of FIB, faecal stanols, pharmaceuticals and artificial sweeteners. Santoro and Boehm (2007) also reported that a higher incidence of HF183 marker was correlated with

higher FIB abundance. In contrast, it can also be observed from the figures that geometric means of specific conductance, chloride, cholesterol and caffeine increase from the high HF group to the low HF group. This implies that higher HF concentrations are accompanied by lower concentrations of specific conductance, chloride, cholesterol and caffeine.

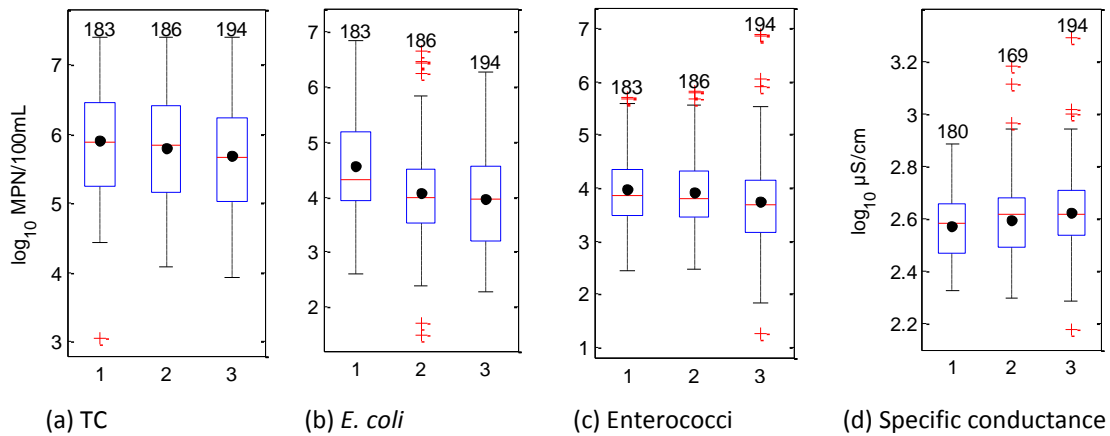


Figure 7.1. Box plots of concentrations of FIB and specific conductance based on HF groups (x axis: 1 = high HF, 2 = medium HF and 3 = low HF concentration group)

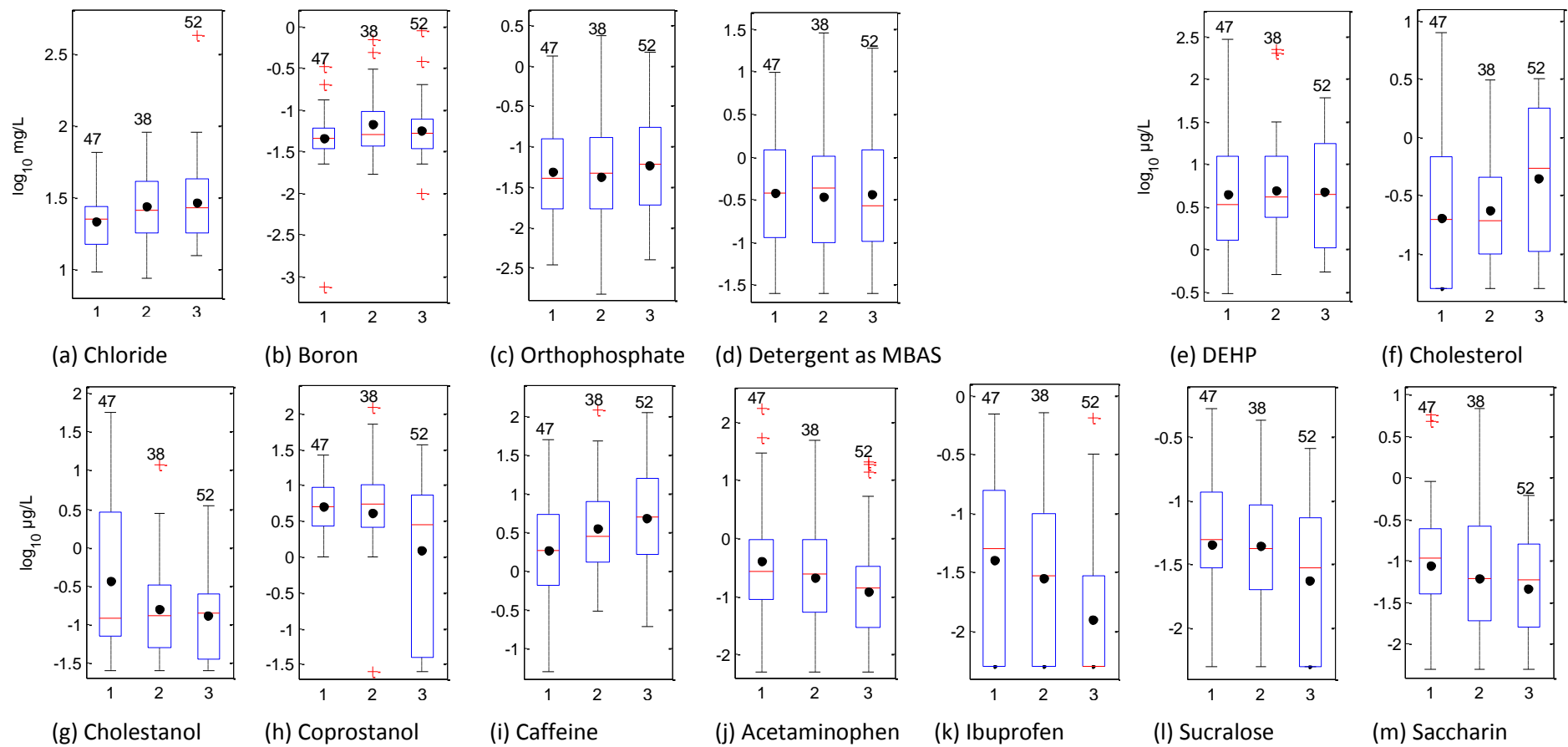


Figure 7.2. Box plots of chemical concentrations based on HF groups (x axis: 1 = high HF, 2 = medium HF and 3 = low HF concentration group)

One-way analysis of variance (ANOVA, Section 4.5.12) followed by Tukey-Kramer (Section 4.5.14) tests were conducted on the data to determine whether FIB and chemical concentrations varied significantly (at 5% level) across the three HF groups. The Tukey-Kramer test results are reported as a connecting letters report (“A”, “B” and “C”). “A” indicates a significantly higher mean than “B” or “C” and “B” indicates a significantly higher mean than “C” (although no C values happen to appear in Table 7.1). “AB” indicates an insignificantly different mean than “A” or “B” where “AB” is a lower mean than “A” but a higher mean than “B”. The means, number of data used in the analysis (N), connecting letters report and p values are summarised in Table 7.1. p values less than 0.05 indicate significant differences and are highlighted in the table with grey shading.

The high HF group (group 1) has significantly ($p \leq 0.05$, Table 7.1, row 8) different TC, *E. coli*, enterococci, specific conductance, chloride, cholesterol, cholestanol, coprostanol, caffeine, acetaminophen, ibuprofen and sucralose means compared to the means in low HF group (group 3). An inspection of the data shows that the means of TC, *E. coli*, enterococci, cholestanol, coprostanol, acetaminophen, ibuprofen and sucralose are higher in the high HF group (“A”, Table 7.1, row 5) than in the low HF group (“B”, Table 7.1, row 7) while the means of specific conductance, chloride, cholesterol and caffeine are lower in the high HF group (“B”, Table 7.1, row 5) than in the low HF group (“A”, Table 7.1, row 7). The mean of *E. coli* is also significantly ($p \leq 0.05$, Table 7.1, row 9) higher in the high HF group than in the medium HF group. Therefore, *E. coli* has a significantly higher mean in the high HF group than in both of the other two HF groups. In addition, the means of enterococci, coprostanol and ibuprofen are significantly ($p \leq 0.05$, Table 7.1, row 10) higher in the medium HF group than in the low HF group.

Table 7.1. Tukey-Kramer test results of all data based on HF groups

Row	HF group	TC	<i>E. coli</i>	Enterococci	Specific conductance	Chloride	Boron	Orthophosphate	Detergent as MBAS	DEHP	Cholesterol	Cholestanol	Coprostanol	Caffeine	Acetaminophen	Ibuprofen	Sucralose	Saccharin		
1																				
2	Mean	1	5.91	4.55	3.98	2.57	1.33	-1.35	-1.31	-0.42	0.65	-0.69	-0.44	0.71	0.25	-0.39	-1.40	-1.34	-1.05	
3		2	5.78	4.06	3.91	2.59	1.44	-1.17	-1.37	-0.47	0.69	-0.63	-0.81	0.61	0.55	-0.67	-1.55	-1.35	-1.22	
4		3	5.68	3.95	3.72	2.62	1.46	-1.24	-1.24	-0.44	0.67	-0.35	-0.89	0.09	0.67	-0.92	-1.90	-1.62	-1.33	
5	Connecting letters report	1	A	A	A	B	B	A	A	A	A	B	A	A	B	A	A	A	A	
6		2	AB	B	A	AB	AB	A	A	A	A	AB	AB	A	AB	AB	A	AB	A	
7		3	B	B	B	A	A	A	A	A	A	A	B	B	A	B	B	B	A	
8	p value	1	3	0.0152	< 0.0001	0.0022	0.0003	0.0259	0.3337	0.8221	0.9862	0.9844	0.0192	0.0135	0.0008	0.0045	0.0188	0.0003	0.0493	0.1234
9		1	2	0.3005	< 0.0001	0.6164	0.2161	0.1412	0.0559	0.9127	0.9507	0.9500	0.8919	0.0856	0.8474	0.0943	0.369	0.5116	0.9985	0.5329
10		2	3	0.3982	0.3697	0.0390	0.0829	0.8568	0.5551	0.5913	0.986	0.9872	0.0907	0.8665	0.0104	0.6428	0.4521	0.0253	0.0771	0.7265
11	N		563	563	563	543	137	137	137	137	137	137	137	137	137	137	137	137	137	

Log-transformed values are used for this analysis. Cells in which p values show a significant difference ($p \leq 0.05$) are shaded.

7.2.2. Analysis of FIB data based on HF group and sampling time (daytime/nighttime)

All (14) sets of 24-hr FIB concentrations and the 9 sets of 24-hr FIB concentrations identified in Table 5.3 as showing typical diurnal FIB variation were log-transformed and grouped based on HF groups and daytime/nighttime samples. The data were then plotted as box plots in Figures 7.3 and 7.4. Generally, FIB shows higher concentrations in the high HF group than in the medium or low HF group for all sets of 24-hr samples and the sets of 24-hr samples that show typical FIB diurnal variation. However, TC and enterococci show higher concentrations in the medium HF group than in the high HF group for sets of 24-hr samples that show typical FIB diurnal variation (Figures 7.4 (a), (c)).

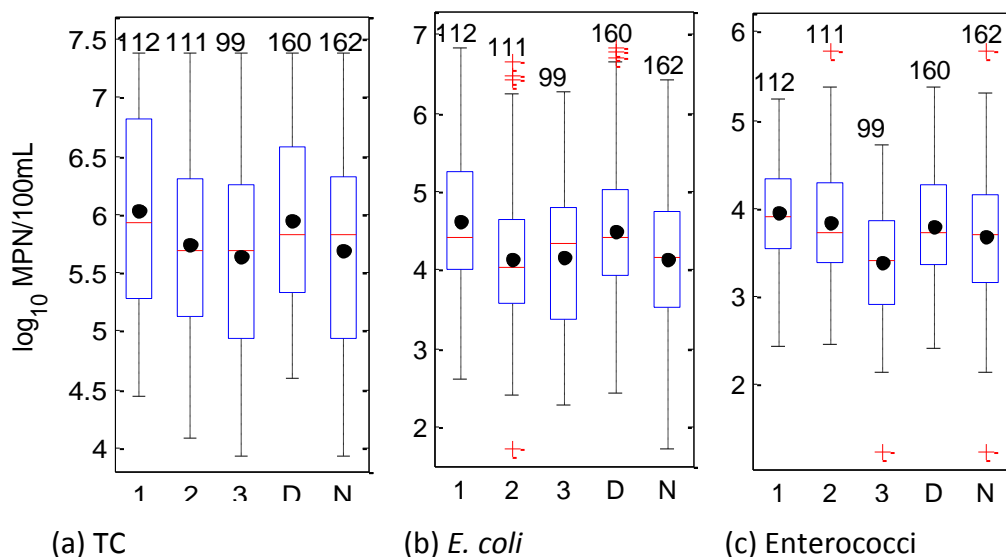


Figure 7.3. Box plots of FIB concentrations of all sets of 24-hr samples based on HF group and daytime/nighttime (x axis: 1 = high HF, 2 = medium HF, 3 = low HF, D = daytime and N = nighttime)

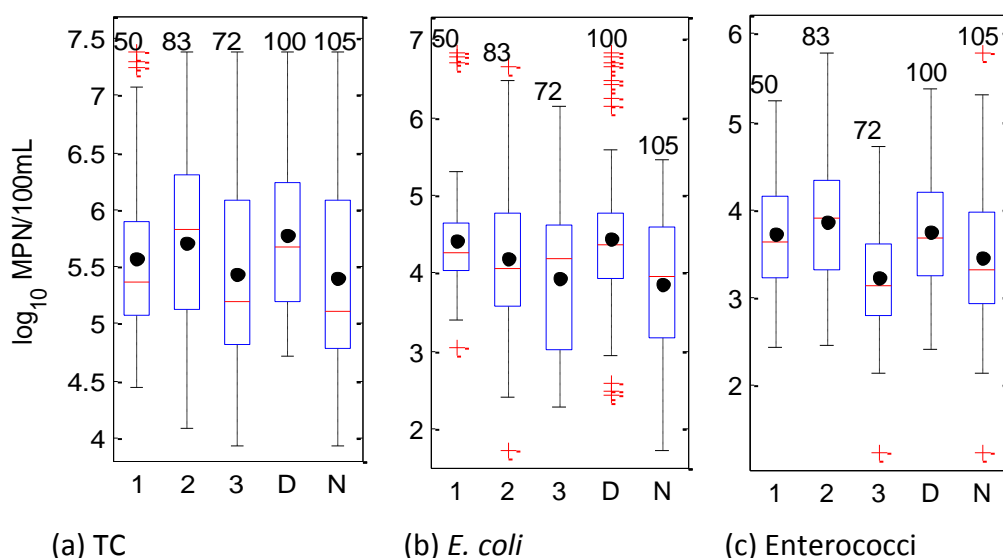


Figure 7.4. Box plots of FIB concentrations of sets of 24-hr samples that show typical FIB diurnal variation (Table 5.3) based on HF group and daytime/nighttime (x axis: 1 = high HF, 2 = medium HF, 3 = low HF, D = daytime and N = nighttime)

Two-way ANOVA (Section 4.5.12) was conducted on the data to determine whether FIB concentrations varied significantly across HF groups and between sampling time (daytime and nighttime). F ratios and p values are summarised in Table 7.2. Two-way ANOVA tests carried out on the mean log-transformed FIB concentrations of all sets of 24-hr samples reveal that TC and *E. coli*, but not enterococci, vary significantly if they are associated with different HF groups and between daytime and nighttime samples ($p \leq 0.05$) but there is not similar variation in the interaction of HF groups and daytime/nighttime samples ($p > 0.05$). In other words, there are significant FIB variations among the HF groups and between daytime and nighttime samples, but the variation within each HF group is not significantly different with respect to the time (daytime/nighttime) at which samples were collected. For example, there is no significant variation between daytime and nighttime TC concentrations in the high HF group. Similarly, two-way ANOVA tests carried out on 24-hr sample sets exhibiting typical diurnal FIB variation reveals differences in the mean log-transformed FIB concentrations for different HF group ($p \leq 0.05$) and between daytime and nighttime ($p \leq 0.05$) but not the interaction of HF groups and daytime/nighttime ($p > 0.05$). The two-way ANOVA results imply that HF183 categorical data do not vary significantly

between daytime and nighttime. In other words, sampling time does not significantly influence high or low HF values detected. Why HF concentrations do not vary significantly between daytime and nighttime cannot be ascertained from this study. However, *Bacteroides-Prevotella* is not expected to behave in the same way as FIB in the environment. Therefore, further studies, such as microcosm studies should be conducted to study the basic processes of HF in the tropics and to assess HF behaviour at different times of the day. The results of such studies can lead to a better understanding of the utility of HF in tropical environments.

Table 7.2. Two-way ANOVA test results of FIB data based on HF group and daytime/nighttime

FIB	Variation with respect to	all (14) sets of 24-hr samples		9 sets of 24-hr samples	
		F	p	F	p
TC	HF group	6.90	0.0012	3.95	0.0207
	Daytime/nighttime	8.69	0.0034	14.80	0.0002
	Interaction of HF group and daytime/nighttime	2.25	0.1075	0.24	0.7871
<i>E. coli</i>	HF group	11.59	< 0.0001	5.94	0.0031
	Daytime/nighttime	14.59	0.0002	24.63	< 0.0001
	Interaction of HF group and daytime/nighttime	1.37	0.2563	0.34	0.7156
Enterococci	HF group	24.32	< 0.0001	24.30	< 0.0001
	Daytime/nighttime	3.59	0.0592	15.28	0.0001
	Interaction of HF group and daytime/nighttime	1.72	0.1799	0.57	0.5655

7.2.3. Analysis of daytime data based on HF groups and land use

Out of the 563 samples mentioned in Section 7.1, 304 samples are from high-density residential areas, 210 from low-density residential areas, 25 from a commercial area and 24 from an industrial area. For the chemical data, out of 137 samples, 119 samples are from residential areas. Out of these 119 residential samples, 116 samples were collected during daytime. Therefore, further analysis in this chapter focuses on

samples collected during the daytime from residential areas only, due to the lack of data from the other land use types.

The daytime data were segregated based on high- and low-density residential areas and plotted as box plots in Figures 7.5 and 7.6. In general, geometric mean concentrations of many parameters are higher in the high-density residential areas than in the low-density residential areas. For high-density residential areas (Figures 7.5 and 7.6, H1 to H3), the geometric means of TC, *E. coli*, specific conductance, orthophosphate, detergent as MBAS, DEHP, cholestanol, coprostanol, acetaminophen, ibuprofen, sucralose and saccharin decrease from the high HF group to the low HF group. This implies that higher HF concentrations are accompanied by higher concentrations of TC, *E. coli*, specific conductance, orthophosphate, detergent as MBAS, DEHP, faecal stanols, pharmaceuticals and artificial sweeteners. On the other hand, the geometric mean of caffeine increases from the high HF group to the low HF group. This implies that higher HF concentrations are accompanied by lower concentrations of caffeine. This indicates that caffeine is not a good indicator of human faecal pollution in high-density residential areas. The co-occurrence of high HF concentrations with high concentrations of orthophosphate and detergent as MBAS and in high-density residential areas indicates that orthophosphate and detergent as MBAS likely originate from leaking sewers. In contrast, in low-density residential areas, orthophosphate and detergent as MBAS do not coincide with high HF. This is consistent with the known practice of direct discharge of household laundry wastewater into drain lines rather than to sewers (Sections 3.3.2 and 6.1.2).

For low-density residential areas (Figures 7.5 and 7.6, L1 to L3), geometric means of enterococci, acetaminophen, ibuprofen and sucralose decrease from the high HF group to the low HF group. This implies that higher HF concentrations are accompanied by higher concentrations of enterococci, pharmaceuticals and sucralose. In contrast, geometric means of TC, specific conductance, chloride, boron, orthophosphate, detergent as MBAS, DEHP, cholesterol, cholestanol, coprostanol and

caffeine increase from high HF group to low HF group. This implies that higher HF concentrations are accompanied by lower concentrations of TC, specific conductance, chloride, boron, orthophosphate, detergent as MBAS, DEHP, faecal sterol and stanols and caffeine.

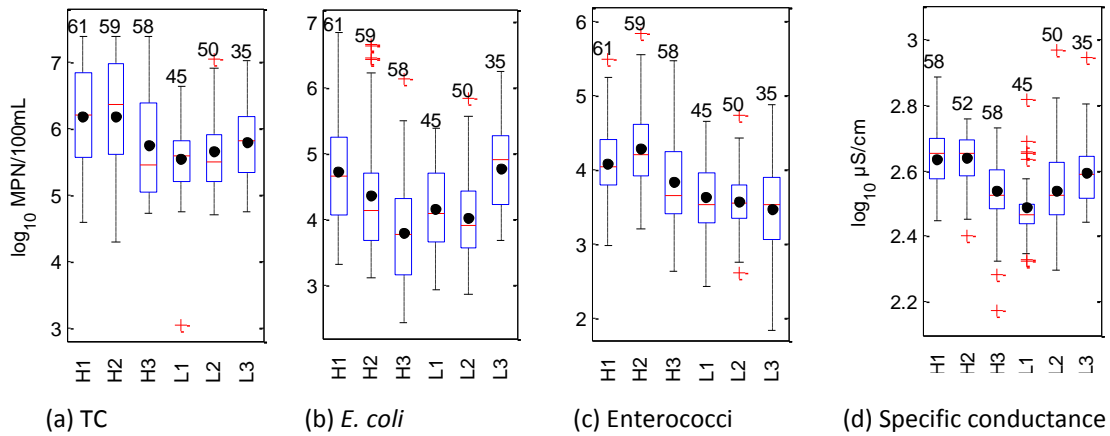


Figure 7.5. Box plots of the concentrations of FIB and specific conductance based on land uses and HF group (x axis: H = high-density residential, L = low-density residential, 1 = high HF, 2 = medium HF and 3 = low HF. For example, H1 means high-density residential data that have high HF concentrations.)

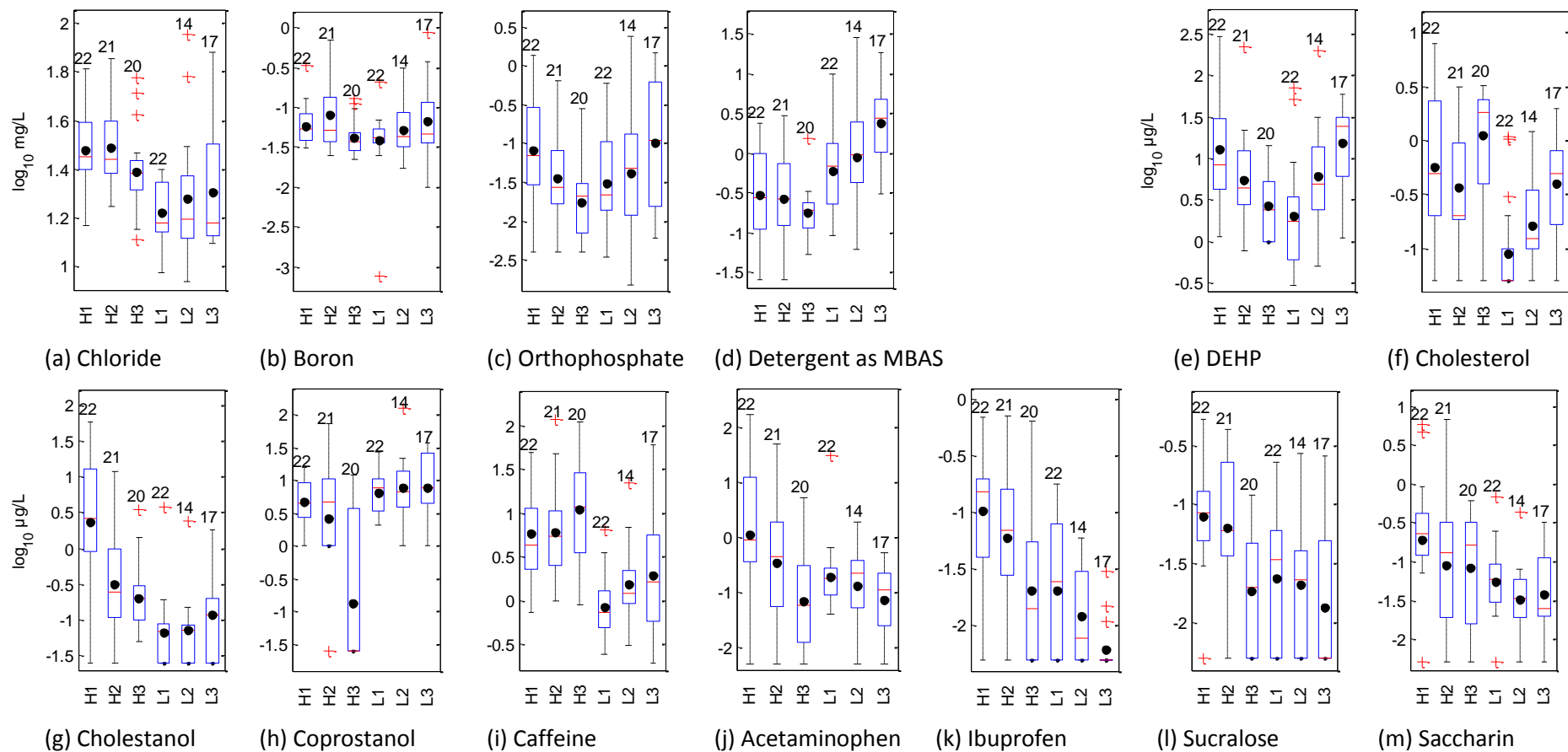


Figure 7.6. Box plots of the chemical concentrations based on land uses and HF group (x axis: H = high-density residential, L = low-density residential, 1 = high HF, 2 = medium HF and 3 = low HF. For example, H1 means high-density residential data that have high HF concentrations.)

One-way ANOVA (Section 4.5.12) followed by Tukey-Kramer (Section 4.5.14) tests were conducted on high-density and low-density residential data to determine whether FIB and chemical concentrations measured within each land use varied significantly (at 5% level) across the three HF groups. Similarly to Section 7.2.1, the Tukey-Kramer test results are reported as a connecting letters report. The connecting letters report and p values are summarised in Tables 7.3 and 7.4 for high-density and low-density residential, respectively.

For high-density residential areas (Table 7.3), the means of TC, *E. coli*, specific conductance, orthophosphate, DEHP, cholestanol, coprostanol, acetaminophen, ibuprofen and sucralose are significantly ($p \leq 0.05$, Table 7.3, row 8) higher in the high HF concentration group (group 1, "A", Table 7.3, row 5) than in the low HF concentration group (group 3, "B or C", Table 7.3, row 7). In addition, the means of *E. coli* and cholestanol are significantly ($p \leq 0.05$, Table 7.3, row 9) higher in the high HF group than in the medium HF group. Therefore, the means of *E. coli* and cholestanol are significantly higher in the high HF group than in the other two HF groups.

The medium HF group has significantly ($p \leq 0.05$, Table 7.3, row 10) different TC, *E. coli*, enterococci, specific conductance, boron, cholesterol, coprostanol, ibuprofen and sucralose mean concentrations compared to the means in the low HF group. The means of TC, *E. coli*, enterococci, specific conductance, boron, coprostanol, ibuprofen and sucralose are higher in the medium HF group ("A or B", Table 7.3, row 6) than in the low HF group ("B or C", Table 7.3, row 7). On the other hand, the mean concentration of cholesterol is lower in the medium HF group ("B", Table 7.3, row 6) than in the low HF group ("A", Table 7.3, row 7).

The means of *E. coli* of the three HF groups are significantly ($p \leq 0.05$) different. On the other hand, enterococci that are associated with the high HF group are not significantly different from enterococci associated with the medium ($p = 0.1844$) and low ($p = 0.0737$) HF groups. This further confirms that *E. coli* is more closely

associated with human faecal contamination as HF183 is known to be specific to humans (Bernhard and Field 2000b; Ahmed et al. 2008).

For low-density residential areas (Table 7.4), the high HF concentration group has significantly ($p \leq 0.05$, Table 7.4, row 8) different *E. coli*, specific conductance, detergent as MBAS, DEHP, cholesterol and ibuprofen means compared to the means in the low HF group. The mean of ibuprofen is higher in the high HF group ("A", Table 7.4, row 5) than in the low HF group ("B", Table 7.4, row 7). In contrast, the mean concentrations of *E. coli*, specific conductance, detergent as MBAS, DEHP and cholesterol are lower in the high HF group ("B", Table 7.4, row 5) than in the low HF group ("A", Table 7.4, row 7). In addition, the mean of *E. coli* is significantly ($p \leq 0.05$, Table 7.4, row 10) lower in the medium HF group than in the low HF group. Therefore, the mean of *E. coli* is significantly higher in the low HF group ("A", Table 7.4, row 7) than in the other two HF groups ("B", Table 7.4, rows 5 and 6). On the other hand, enterococci concentrations are not significantly ($p > 0.05$, Table 7.4, row 8-10) different among the three HF groups.

Table 7.3. Tukey-Kramer test results of high-density residential data based on HF group

Row	HF group	TC	<i>E. coli</i>	Enterococci	Specific conductance	Chloride	Boron	Orthophosphate	Detergent as MBAS	DEHP	Cholesterol	Cholestanol	Coprostanol	Caffeine	Acetaminophen	Ibuprofen	Sucralose	Saccharin		
1																				
2	Mean	1	6.17	4.74	4.09	2.64	1.48	-1.24	-1.10	-0.53	1.11	-0.24	0.37	0.67	0.76	0.05	-0.99	-1.11	-0.72	
3		2	6.17	4.36	4.29	2.64	1.49	-1.09	-1.44	-0.58	0.74	-0.44	-0.51	0.42	0.78	-0.47	-1.22	-1.19	-1.05	
4		3	5.74	3.81	3.85	2.54	1.39	-1.38	-1.76	-0.76	0.42	0.05	-0.70	-0.88	1.03	-1.15	-1.69	-1.73	-1.09	
5	Connecting letters report	1	A	A	AB	A	A	AB	A	A	A	AB	A	A	A	A	A	A	A	
6		2	A	B	A	A	A	A	AB	A	AB	B	B	A	A	AB	A	A	A	
7		3	B	C	B	B	A	B	B	A	B	A	B	B	A	B	B	B	A	
8	p value	1	3	0.0178	< 0.0001	0.0737	< 0.0001	0.1577	0.3419	0.0010	0.3007	0.0005	0.2104	< 0.0001	< 0.0001	0.2224	0.0022	0.0013	0.0008	0.3135
9		1	2	0.9993	0.0373	0.1844	0.9962	0.9798	0.2993	0.1167	0.9326	0.0782	0.4910	0.0003	0.6167	0.9925	0.2734	0.4216	0.8469	0.3814
10		2	3	0.0209	0.0015	0.0004	< 0.0001	0.1133	0.0159	0.1872	0.4994	0.1660	0.0182	0.6536	< 0.0001	0.2778	0.1240	0.0471	0.0048	0.9881
11	N		178	178	178	168	63	63	63	63	63	63	63	63	63	63	63	63		

Log-transformed values are used for this analysis. Cells in which p values show a significant difference ($p \leq 0.05$) are shaded.

Table 7.4. Tukey-Kramer test results of low-density residential data based on HF group

Row	HF group	TC	<i>E. coli</i>	Enterococci	Specific conductance	Chloride	Boron	Orthophosphate	Detergent as MBAS	DEHP	Cholesterol	Cholestanol	Coprostanol	Caffeine	Acetaminophen	Ibuprofen	Sucralose	Saccharin		
1																				
2	Mean	1	5.55	4.16	3.63	2.49	1.22	-1.42	-1.52	-0.22	0.30	-1.05	-1.18	0.81	-0.08	-0.71	-1.69	-1.63	-1.27	
3		2	5.64	4.04	3.58	2.54	1.28	-1.29	-1.38	-0.05	0.78	-0.79	-1.15	0.89	0.18	-0.88	-1.92	-1.68	-1.49	
4		3	5.78	4.79	3.46	2.59	1.31	-1.17	-1.00	0.38	1.18	-0.41	-0.94	0.89	0.28	-1.13	-2.21	-1.87	-1.43	
5	Connecting letters report	1	A	B	A	B	A	A	A	B	B	B	A	A	A	A	A	A	A	
6		2	A	B	A	AB	A	A	A	AB	AB	AB	A	A	A	A	AB	A	A	
7		3	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	A	A	
8	p value	1	3	0.1945	< 0.0001	0.3259	0.0003	0.4277	0.1773	0.0828	0.0070	0.0001	0.0002	0.3365	0.8347	0.0743	0.1300	0.0033	0.3701	0.5717
9		1	2	0.7278	0.6228	0.8731	0.0759	0.7105	0.6360	0.8451	0.6750	0.0616	0.2237	0.9842	0.8641	0.2976	0.7356	0.3365	0.9616	0.3889
10		2	3	0.5344	< 0.0001	0.5750	0.1111	0.9263	0.7320	0.3311	0.1130	0.1582	0.0567	0.5107	0.9997	0.8321	0.5468	0.2026	0.6033	0.9360
11	N		130	130	130	130	53	53	53	53	53	53	53	53	53	53	53	53	53	

Log-transformed values are used for this analysis. Cells in which p values show a significant difference ($p \leq 0.05$) are shaded.

Overall, the means of 10 parameters (TC, *E. coli*, specific conductance, orthophosphate, DEHP, cholestanol, coprostanol, acetaminophen, ibuprofen and sucralose) are significantly higher in the high HF concentration group than in the low HF concentration group in high-density residential areas. On the other hand, only the mean of ibuprofen is higher in the high HF group than in the low HF group in low-density residential areas. The majority of the parameters (11 parameters) in low-density residential areas are indifferent with respect to the HF groups (Table 7.4, row 8-10 without grey shading). This implies that parameters in high-density residential areas show better agreement with the HF183 marker than parameters in low-density residential areas. Section 6.1.2 shows that median and geometric mean concentrations of chloride, cholestanol, coprostanol, caffeine, acetaminophen, ibuprofen, sucralose and saccharin in high-density residential areas were higher than in low-density residential areas. The higher median and geometric mean may indicate that there is a higher degree of contamination in the high-density residential areas than in the low-density residential areas. The finding in this section that parameters in the high-density residential areas show better agreement with HF183 marker than parameters in the low-density residential areas corroborates the finding obtained in Section 6.1.2.

7.3. Conclusions

In this chapter, the variation of FIB and chemical data was analysed as a function of categorical HF concentration and also with respect to time of sampling (daytime vs. nighttime) and land use. The following can be summarised from this chapter:

(1) One-way ANOVA of all data followed by Tukey-Kramer tests reveals that the means of TC, *E. coli*, enterococci, cholestanol, coprostanol, acetaminophen, ibuprofen and sucralose are significantly higher in the high HF group than in the low HF group, while the means of specific conductance, chloride, cholesterol and caffeine are significantly lower in the high HF group than in the low HF group.

(2) Two-way ANOVA of sets of 24-hr samples show that HF183 categorical data do not vary significantly between daytime and nighttime samples. The reason for the lack of distinction in HF data between daytime and nighttime samples cannot be ascertained from this study.

(3) There is a greater number of parameters which show a positive relation with HF group (significantly higher in the high HF group than in the low HF group) in high-density residential areas than in low-density residential areas. This observation confirms that there is a higher degree of contamination in the high-density residential areas than in the low-density residential areas.

CHAPTER 8 CONCLUSIONS AND RECOMMENDATIONS

8.1. Conclusions

There are many sources of surface water contamination of which human faecal sources of contamination pose the most prevalent hazard for human health (Cabelli 1977). Since the cultivation and identification of pathogenic microorganisms are difficult and costly, indicators are used instead. Faecal indicator bacteria (FIB) are widely used as indicators. However, FIB were first developed for temperate climates. The use of FIB in tropical areas is questionable since FIB grow, occur naturally or survive longer in tropical environments (Hazen 1988; Rivera et al. 1988; Muñiz et al. 1989). Therefore, FIB are not specific to the presence of human faecal contamination. These concerns have motivated the present study in Singapore, a tropical and highly urbanised country, in which identification of human-related sources is important. As stated in Section 1.2, this study aimed to achieve the following objectives: (i) to suggest sampling protocols that can be applied for the identification of human faecal and sewage contamination in a highly urbanised catchment, (ii) to identify suitable tracers for human faecal and sewage contamination in a highly urbanised catchment as a function of land use and (iii) to identify evidence of leaking sewers as sources of human faecal and sewage contamination in a highly urbanised catchment.

Multivariate analysis of data collected in the Marina catchment by PUB showed that sampling location positions within a catchment influence the surface water quality since the samples can be grouped into a “drain” cluster for upstream samples, a “river” cluster for mid-stream samples and a “reservoir” cluster for downstream samples (Section 3.3.1). In addition, it is hypothesised that land use affects drain water quality. PUB’s Toolbox Study, which reported hourly dry-weather samples, suggests diurnal variations of surface water quality parameters such as faecal coliform, enterococci, somatic coliphage, pH, temperature, turbidity, total phosphorus, orthophosphate, detergent as MBAS, cholesterol, cholestanol, coprostanol, epicholestanol and epicoprostanol (Section 3.3.2). Therefore, hourly

sampling, instead of the commonly adopted random grab sampling method, is necessary to identify the daily peak concentrations. The findings of these preliminary analyses of PUB's data provided guidance for site selection and sampling procedure of this study: hourly dry-weather sampling at upstream areas with specific land uses (Sections 4.2 and 4.3). Therefore, sampling was conducted hourly over 12- or 24-hr at 13 sites which have the following major land uses: high-density residential (Lorong 8 Toa Payoh, Toa Payoh North, Lorong 6 Toa Payoh, Ang Mo Kio Avenue 10, Choa Chu Kang Crescent and Punggol Central), low-density residential (Serangoon Garden, Jalan Siantan, MacPherson Lane, Watten Drive and Verde), commercial (Bras Basah) and industrial (Sims Place).

Analysis of the findings of the Toolbox Study also shows that different parameters roughly peak at the same time which implies the potential suitability of the parameters as tracers. Therefore, a suite of potential tracers is identified through review of the literature and by considering the practicality of detection. The potential tracers selected include FIB, chemical parameters and HF183 (Section 4.1). Despite the non-conservative behaviour of FIB, they are widely used as indicators in regulatory guidelines for tropical areas (Kenzaka et al. 2001) including Singapore, therefore FIB were also analysed in this study. Possible chemical tracers are selected based on their feasibility which includes the use (consumption) of the chemicals in Singapore and quantification by reasonably facile methods. Selected possible chemical tracers include chloride, bromide, boron, orthophosphate, detergent as MBAS, triclosan, DBP, DEHP, cholesterol, cholestanol, coprostanol, caffeine, acetaminophen, ibuprofen, diclofenac, acesulfame-k, sucralose and saccharin. However, bromide, triclosan, DBP, diclofenac and acesulfame-k were excluded from the suite of potential tracers during the course of this research since they were rarely detected. In addition, water temperature, specific conductance and salinity were also measured. Studies have shown the applicability of the HF183 marker in temperate and tropical areas (Gawler et al. 2007; Santoro and Boehm 2007; Jenkins et al. 2009; Nshimiyimana 2010; Ahmed et al. 2010b; Boehm et al. 2011; Santiago-Rodriguez et al.

2012; Toledo-Hernandez et al. 2013). Nshimiyimana (2010) compared HF183 genes obtained in Singapore with the genes obtained by Santoro and Boehm (2007) in temperate climates and showed that they were clustered within the same evolutionary group. Therefore, Nshimiyimana (2010) confirmed the use of the HF183 assay as a viable technique for human faecal source tracking in Singapore.

The following can be concluded from this study:

(1) The observed diurnal variation of FIB, which is characterised by high daytime and low nighttime concentrations, is an indication of leaking sewers.

In this study, hourly sampling over 24-hr were conducted at eleven residential sites. Significant background concentrations of FIB were observed at all eleven sites. It is hypothesised that sewage exfiltration from leaking sewers can add to background levels of FIB, giving rise to diurnal pattern in FIB concentrations that are consistent with sewage flow patterns. Data analyses including time series and box plots of normalised FIB concentrations which were grouped based on sampling time, plots of arithmetic means of normalised FIB concentrations (Section 5.2.1) and autocorrelation of chained time series (Sections 5.2.3 and 5.2.4) suggest a diurnal pattern of FIB that is characterised by high daytime and low nighttime FIB concentrations. The difference between daytime and nighttime FIB concentrations was significant (determined using the Wilcoxon-Mann-Whitney test, Section 5.2.5). This result is contrary to other studies conducted mainly in temperate, subtropical and tropical areas, which reported that FIB concentrations of samples collected from open drainages (exposed to sunlight) are high at nighttime and low in daytime (Boehm et al. 2002; Whitman and Nevers 2004; Traister and Anisfeld 2006; Enns et al. 2012; Desai and Rifai 2013). In highly urbanised environments where space is at a premium and drainages tend to be covered, flow in surface drains are shielded from sunlight. Therefore bacteria die-off due to sunlight during the daytime may be negligible. Indeed, there is an anecdotal evidence to suggest that environmental samples collected from shaded or partially

covered areas do not exhibit significant differences between daytime and nighttime FIB concentrations (Traister and Anisfeld 2006; Desai and Rifai 2013).

The diurnal variation of FIB in this study closely follows that of sewer flow patterns, as reported by Enfinger and Stevens (2006). Field tracer tests conducted in this study provide qualitative evidence linking sewage exfiltration and transport to surface drains via preferential flow paths and further confirm the likelihood of surface water contamination due to leaking sewers. Therefore, the contrasting diurnal pattern of FIB in this study may be attributed to the existence of leaking sewers as sources of human faecal and sewage contamination.

(2) Land use is an important factor in drain water quality.

Correlations between the various physical, faecal and chemical parameters within land-use specific datasets (Tables 6.3 through 6.6) are stronger than in the combined dataset (Table 6.2), showing the dependence between land use and surface water quality. The median and geometric mean concentrations of most of the chemical parameters were higher in high-density residential areas than in low-density residential areas (Figure 6.3). This shows that high-density residential has a higher degree of contamination than low-density residential. Besides, there is also a greater number of parameters showing positive relations with HF group (significantly higher in the high HF group than in the low HF group) in the high-density residential areas than in the low-density residential areas.

(3) Acetaminophen, orthophosphate, cholesterol, cholestanol, coprostanol, sucralose and saccharin are potential tracers for high-density residential areas. Chloride, cholestanol, coprostanol and caffeine are potential tracers for low-density residential areas.

Acetaminophen is highly ($r_s = 0.76$) correlated with *E. coli*, while orthophosphate, cholesterol, cholestanol, coprostanol, sucralose and saccharin are moderately ($r_s \geq 0.34$) correlated with *E. coli* in the “high” datasets of high-density residential areas (Table 6.13). Chloride, cholestanol, coprostanol and caffeine are moderately ($r_s \geq 0.31$) correlated with *E. coli* in the “high” datasets of low-density residential areas (Table 6.16). Ibuprofen and sucralose are not good tracers for low-density residential areas because they are not detected for more than half of the samples in the “high” dataset. In general, cholesterol, cholestanol and coprostanol can be used as effective indicators of faecal contamination for both types of residential areas since they have significant ($p \leq 0.05$) correlations with *E. coli* in the “high” dataset of both high- and low-density residential areas.

(4) The age of sewers seems to influence leaking sewers to a higher degree in high-density residential areas than in low-density residential areas.

Sewer leakage may occur because of damaged sewer lines or ageing sewer lines. The latter possibility was investigated by further segregating the dataset based on the age of sewers and then conducting correlation analysis on the segregated datasets. There is a greater number of significant ($p \leq 0.05$) correlations among parameters in the old-sewer dataset from high-density residential areas than in the old-sewer dataset from low-density residential areas. Therefore, the age of sewers seems to influence leaking sewers to a higher degree in high-density residential areas than in low-density residential areas. However, it should be borne in mind that the age of sewers is not the sole cause of damaged sewer lines, which may also include backfilling, settling immediately after backfilling or subsidence.

(5) *E. coli* is more closely associated with human faecal and sewage contamination than enterococci.

E. coli was found to be a better indicator than enterococci because *E. coli* is more closely associated with human faecal and sewage contamination. This is shown by (i) a more distinct diurnal pattern of *E. coli* than enterococci (Section 5.2.1), (ii) generally smaller RMSE values of *E. coli* than enterococci (Section 5.2.4), (iii) better correlation between *E. coli* and other parameters than enterococci and other parameters in each land use (Tables 6.3 through 6.6) and (iv) better correlation between *E. coli* and other parameters than enterococci and other parameters in the segregated datasets from both types of residential areas (Tables 6.7 and 6.10).

(6) Higher HF concentrations are accompanied by higher concentrations of FIB, faecal stanols, pharmaceuticals and sucralose.

One-way ANOVA of all data followed by Tukey-Kramer tests reveals that the means of TC, *E. coli*, enterococci, cholestanol, coprostanol, acetaminophen, ibuprofen and sucralose are significantly higher in the high HF group than in the low HF group. However, the means of specific conductance, chloride, cholesterol and caffeine are significantly lower in the high HF group than in the low HF group.

(7) Sampling time is important to reveal the presence of human faecal and sewage contamination.

FIB concentrations obtained of samples collected hourly from 04:00-07:00 were significantly lower from samples that were collected from 12:00-15:00 (Section 5.2.5). Upstream sites with significant sewer leakage can be determined by collecting samples at these times. Samples can be analysed for FIB. Daytime samples can also be analysed for chemical parameters. This is useful to identify upstream areas with significant sewage leaking into the drains. Upstream sewers, such as sewers connecting discharge pipes from buildings or properties to the public sewers, are more likely to be at elevations above the water table and therefore sewage exfiltrate

to surface water drains if sewers are leaking. To resolve this upstream leaking sewer issue, collaboration between the owner of sewer and regulatory body is necessary.

Results obtained from the Wilcoxon-Mann-Whitney test showed that sampling at 12:00 and 14:00 results in significantly higher FIB concentrations while sampling at 05:00 and 04:00 or at 05:00 and 06:00 results in significantly lower FIB concentrations than sampling at other times throughout the day. These results show the importance of sampling time. Therefore, it is suggested that samples for water quality monitoring at upstream areas can be collected at the suggested sampling times to better reveal the presence of human faecal and sewage contamination.

Overall, the diurnal pattern of FIB, which is characterised by high daytime and low nighttime concentrations, closely resembles the daily sewage flow pattern and therefore the diurnal pattern of FIB is an indication of leaking sewers (Sections 5.2.1, 5.2.3 and 5.2.4). Generally there are better correlations among parameters when the data are segregated based on land use and high normalised FIB concentrations (Sections 6.2.2 and 6.2.4). In addition, higher concentrations of FIB, faecal stanols, pharmaceuticals and sucralose are also accompanied by higher HF concentrations (Section 7.2.1). These lines of evidence, (i) diurnal pattern of FIB concentrations and (ii) agreement among tracers, support the hypothesis that leaking sewers are sources of human faecal and sewage contamination in urban storm drains in Singapore.

8.2. Recommendations for future work

Too few samples were collected in commercial and industrial areas or in low-density residential areas with new sewers to characterise areas with those types of land uses and sewers. In addition, further research can be conducted to investigate HF183 in tropical environments in greater detail. The following are suggested as possible future work:

(1) Study of leakage sources can be conducted.

Although this study shows that sewer leakage is likely, the specific locations and the number of leaks are still unknown. These are important information before leakage flow can be quantified. Therefore, more study of leakage sources would be worthwhile. Wolf et al. (2006) tried to assess the problems of sewer-groundwater interaction at a city scale. They used a sewer defect database obtained through CCTV inspection and comprehensive hydrochemical sampling which included a series of marker species including major ions, microbiological parameters, rare earth elements, dissolved organic carbon, boron and pharmaceutical residues. However, investigating specific leakage sources is a challenging task. CCTV inspection cannot directly assess the water tightness of the sewer and can lead to an underestimation of leaks. Manifold parameters affect concentrations of marker species in the groundwater. For example, pharmaceutical residues are subject to microbiological decomposition, adsorption effects during their passage through unsaturated zone and dilution with the unaffected groundwater, all of which lower concentrations—often below detection limits.

When the specific leakage sources have been determined, the mechanism and magnitude of sewer leakage (in terms of either absolute rates or in percentages of total flow) can be further investigated. One way to quantify sewer leakage is through tracer tests to quantify fluxes by means of mass balances.

(2) Sampling campaign can be conducted at commercial and industrial areas.

There are better correlations among different parameters in each land-use dataset (Tables 6.3 through 6.6) than in the overall dataset (Table 6.2). This implies that land use affects water quality. Sufficient numbers of both high- and low-density residential samples, especially those analysed for chemicals, have been investigated in this study. However, this study only investigated one commercial and one industrial area. Further study can be extended to other commercial and industrial areas. Additional commercial and industrial areas can be selected to represent areas with old and new sewers. Possible commercial areas are Jalan Besar (drain was undergoing construction to convert it from an open to a closed drain during this study), Syed Alwi Road near Mustafa Center (flow in the drain was diverted due to nearby construction at the time of this study's sampling campaign), Circular Road (sampling must be conducted when there is no backflow), Lorong 24 Geylang and Temple Street. Many of the drains in commercial areas are covered. Therefore, sampling must be conducted without direct access to the drain, for example, with sampling pole or auto-samplers. The installation of auto-samplers must engage workers qualified for enclosed-space entry. Possible industrial areas are Defu Lane 1, Ubi Avenue 1, Kampong Ampat Road and Playfair Road.

(3) Sampling campaign can be conducted at low-density residential areas with new sewers.

This study only sampled one low-density residential area with new sewers. Therefore, the effect of sewer age in low-density residential areas cannot be ascertained in this study. Further sampling from other low-density residential areas that are served by relatively newer sewers is recommended to better establish the effect of sewer age in low-density residential areas.

(4) Analysis of quantitative HF183 values with other parameters can be conducted.

At the time of this writing, the methodology of HF183 quantification is still under development to account for the day-to-day variability of standard curves and outliers in the standard curves, if any (Section 4.4.3.3). Therefore, the HF183 concentrations were converted into categorical data in this thesis and the most "true" values are assumed to fall within the correct bin. Once the standard curves issue is resolved, quantitative HF183 values can be used and analysis of those quantitative HF183 values with other parameters (FIB and chemicals) can be conducted.

(5) Samples that were collected in this study for HF183 analysis can be further sequenced to determine the community bacteria taxonomy and abundance of pathogens.

Samples collected in this study for HF183 analysis are still kept in extracted form or in filters (non-extracted form). Therefore, further study can be conducted using these samples. They can be sequenced to determine the community bacteria structural phyla/taxonomy and abundance of pathogens in the samples. Analysis then can be conducted to find the association of FIB, chemicals and HF183 analysed with the sequencing results: whether FIB, chemicals and HF183 point out the hazard in surface water quality.

(6) A microcosm study of *Bacteroidales* can be conducted in tropical environments.

Bacteroidales is the order of *Bacteroides-Prevotella*. Human-specific *Bacteroides-Prevotella* marker includes HF183. *Bacteroidales* persists longer at higher salinities (Green et al. 2011; Schulz and Childers 2011), probably because higher salinity delays predator growth. On the other hand, *Bacteroidales* decays more rapidly at higher temperature (Schulz and Childers 2011), probably due to higher predation rate at higher temperature. However, those findings were obtained through microcosm

studies in a temperate area (Green et al. 2011) and a subtropical area (Schulz and Childers 2011). Other microcosm studies of *Bacteroidales* have also been conducted in temperate areas (Walters and Field 2009; Dick et al. 2010). These studies leave open the question of whether *Bacteroidales* in tropical areas persists/decays in a similar fashion to *Bacteroidales* in temperate and subtropical areas. To answer this question, microcosm studies of *Bacteroidales* can be conducted in tropical environments. The microcosms can be differentiated in terms of sunlight, temperature and other environmental factors that may affect the persistence/decay in tropical areas.

(7) Digital PCR may be explored as an alternative to QPCR.

This study uses QPCR to obtain HF183 concentrations because it was the state-of-the-art methodology available when this study was conducted. However, QPCR has two intrinsic disadvantages: (i) the use of standards and (ii) the use of thermal (amplification) cycles to give quantitative results. Inhibitors increase the number of amplification cycles required to reach a given signal (Baker 2012). The use of standards requires (i) plasmid DNA (that is used as positive controls/standards) from a single, pure species (with no contamination); (ii) accurate pipetting because the standards must be diluted over several orders of magnitude and (iii) stability of the diluted standards. To preserve the stability of the standards, diluted standards can be divided into small aliquots, stored at -80 °C and thawed only once before use.

In order to prevent inhibition and to increase efficiency in PCR analysis, an emerging technology, digital PCR, may be explored as an alternative to QPCR. Digital PCR uses the same primers and probes as QPCR but digital PCR is capable of (i) higher sensitivity, (ii) higher specificity (less ambiguity), (iii) higher precision (more accuracy) than QPCR and (iv) quantitative results obtained without referencing to any other material (standards) and without relying on the number of amplification cycles (Baker 2012). Digital PCR works by partitioning a sample into many individual real-time PCR

reactions; some portion of these reactions contain the target molecule (positive) while others do not (negative)(Life 2013).By counting the number of “positive” partitions (in which the sequence is detected) versus “negative” partitions (in which it is not), the number of copies of a DNA molecule in the original sample can be determined (Baker 2012). Digital PCR works like the MPN method for bacteria enumeration. It relies on Poisson statistical analysis to determine the absolute template quantity(Bio-Rad 2013). When the partitions are 80% positive, the precision is good. But, when the partitions are 90% positive, the precision drops(Baker 2012).

However, there are potential problems in using this emerging technology. Although digital PCR does not require calibration and controls like QPCR, there is still plenty of scope for artefacts because digital PCR involves working with tiny volumes and with single-molecule concentrations(Baker 2012). In addition, even though digital PCR does not count cycles, inhibitors could still be a problem if they cause false negatives by preventing reactions from occurring at all. In general, digital PCR is still at an early stage, therefore if it is going to be used, it must be used with caution. On the other hand, QPCR is less expensive and works over a much broader dynamic range than digital PCR(Baker 2012). Unlike QPCR, digital PCR is not yet developed to measure a large number of samples efficiently. In addition, digital PCR is not widely available in laboratories. Therefore, the use of digital PCR may require greater funding or capital resources.

Ideally, both methods (QPCR and digital PCR) can be studied and compared. However, when resources are limited, the pros and cons of using QPCR and digital PCR can be weighed and the method that will better serve the objective of the study can be employed and investigated in detail.

References

- Ahmed, W., Goonetilleke, A. and Gardner, T. (2010a), "Human and bovine adenoviruses for the detection of source-specific fecal pollution in coastal waters in Australia", Water Research, Vol. 44, No. 16, pp. 4662-4673.
- Ahmed, W., Stewart, J., Powell, D. and Gardner, T. (2008), "Evaluation of *Bacteroides* markers for the detection of human faecal pollution", Letters in Applied Microbiology, Vol. 46, No. 2, pp. 237-242.
- Ahmed, W., Yusuf, R., Hasan, I., Goonetilleke, A. and Gardner, T. (2010b), "Quantitative PCR assay of sewage-associated *Bacteroides* markers to assess sewage pollution in an urban lake in Dhaka, Bangladesh", Canadian Journal of Microbiology, Vol. 56, pp. 838-845.
- Al-Odaini, N. A., Zakaria, M. P., Yaziz, M. I. and Surif, S. (2010), "Multi-residue analytical method for human pharmaceuticals and synthetic hormones in river water and sewage effluents by solid-phase extraction and liquid chromatography–tandem mass spectrometry", Journal of Chromatography A, Vol. 1217, No. 44, pp. 6791-6806.
- Amick, R. S. and Burgess, E. H. (2000), "Exfiltration in Sewer Systems", U.S. Environmental Protection Agency, Cincinnati, OH, USA.
- Arnone, R. D. and Walling, J. P. (2007), "Waterborne pathogens in urban watersheds", Journal of Water and Health, Vol. 5, No. 1, pp. 149-162.
- Ashbolt, N. J., Schoen, M. E., Soller, J. A. and Roser, D. J. (2010), "Predicting pathogen risks to aid beach management: The real value of quantitative microbial risk assessment (QMRA)", Water Research, Vol. 44, No. 16, pp. 4692-4703.
- Astel, A., Tsakovski, S., Barbieri, P. and Simeonov, V. (2007), "Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets", Water Research, Vol. 41, No. 19, pp. 4566-4578.
- AVA (2006), "Sale of Food Act (Chapter 283, Section 56 (1)) Food Regulations", Agri-Food and Veterinary Authority of Singapore, Singapore.
- Aydin, A., Kahraman, T. and Aydin, H. (2010), "The microbiological and physico-chemical parameters with trace metal pollution of coastal bathing water in Dardanelles and Thracian Sea", Journal of Animal and Veterinary Advances, Vol. 9, No. 6, pp. 1077-1082.
- Baker, M. (2012), "Digital PCR hits its stride", Nature Methods, Vol. 9, No. 6, pp. 541-544.
- Barone, J. J. and Roberts, H. R. (1996), "Caffeine consumption", Food and Chemical Toxicology, Vol. 34, No. 1, pp. 119-129.
- Benotti, M. J. and Brownawell, B. J. (2007), "Distributions of pharmaceuticals in an urban estuary during both dry- and wet-weather conditions", Environmental Science & Technology, Vol. 41, No. 16, pp. 5795-5802.
- Berkheiser, V. E., Street, J. J., Rao, P. S. C., Yuan, T. L. and Ellis, B. G. (1980), "Partitioning of inorganic orthophosphate in soil-water systems", C R C Critical Reviews in Environmental Control, Vol. 10, No. 3, pp. 179-224.

- Bernhard, A. E. and Field, K. G. (2000a), "Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes", Applied and Environmental Microbiology, Vol. 66, No. 4, pp. 1587-1594.
- Bernhard, A. E. and Field, K. G. (2000b), "A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA", Applied and Environmental Microbiology, Vol. 66, No. 10, pp. 4571-4574.
- Berthouex, P. M. and Brown, L. C. (2002), Statistics for Environmental Engineers, 2nd Edition, Lewis Publishers, Boca Raton, FL, USA.
- Bio-Rad (2013), "Digital PCR", retrieved on 21 August 2013, from <http://www.bio-rad.com/en-sg/applications-technologies/digital-pcr>, Bio-Rad Laboratories, Inc., Hercules, CA, USA.
- Boehm, A., Yamahara, K., Walters, S., Layton, B., Keymer, D., Thompson, R., Knee, K. and Rosener, M. (2011), "Dissolved inorganic nitrogen, soluble reactive phosphorous, and microbial pollutant loading from tropical rural watersheds in Hawaii to the coastal ocean during non-storm conditions", Estuaries and Coasts, Vol. 34, No. 5, pp. 925-936.
- Boehm, A. B., Fuhrman, J. A., Mrše, R. D. and Grant, S. B. (2003), "Tiered approach for identification of a human fecal pollution source at a recreational beach: case study at Avalon Bay, Catalina Island, California", Environmental Science & Technology, Vol. 37, No. 4, pp. 673-680.
- Boehm, A. B., Grant, S. B., Kim, J. H., Mowbray, S. L., McGee, C. D., Clark, C. D., Foley, D. M. and Wellman, D. E. (2002), "Decadal and shorter period variability of surf zone water quality at Huntington Beach, California", Environmental Science & Technology, Vol. 36, No. 18, pp. 3885-3892.
- Boehm, A. B., Shellenbarger, G. G. and Paytan, A. (2004), "Groundwater discharge: potential association with fecal indicator bacteria in the surf zone", Environmental Science & Technology, Vol. 38, No. 13, pp. 3558-3566.
- Bolton, F. J., Hinchliffe, P. M., Coates, D. and Robertson, L. (1982), "A most probable number method for estimating small numbers of *Campylobacters* in water", The Journal of Hygiene Vol. 89, pp. 185-190.
- Bonilla, N., Santiago, T., Marcos, P., Urdaneta, M., Domingo, J. S. and Toranzos, G. A. (2010), "Enterophages, a group of phages infecting *Enterococcus faecalis*, and their potential as alternate indicators of human faecal contamination", Water Science & Technology, Vol. 61, No. 2, pp. 293-300.
- Borrego, J. J., Moriñigo, M. A., de Vicente, A., Córnox, R. and Romero, P. (1987), "Coliphages as an indicator of faecal pollution in water. Its relationship with indicator and pathogenic microorganisms", Water Research, Vol. 21, No. 12, pp. 1473-1480.
- Buerge, I. J., Buser, H. R., Kahle, M., Muller, M. D. and Poiger, T. (2009), "Ubiquitous occurrence of the artificial sweetener acesulfame in the aquatic environment: an ideal chemical marker of domestic wastewater in groundwater", Environmental Science & Technology, Vol. 43, No. 12, pp. 4381-4385.

- Buerge, I. J., Poiger, T., Müller, M. D. and Buser, H. R. (2003), "Caffeine, an anthropogenic marker for wastewater contamination of surface waters", Environmental Science & Technology, Vol. 37, No. 4, pp. 691-700.
- Burnes, B. S. (2003), "Antibiotic resistance analysis of fecal coliforms to determine fecal pollution sources in a mixed-use watershed", Environmental Monitoring and Assessment, Vol. 85, No. 1, pp. 87-98.
- Buser, H. R., Poiger, T. and Müller, M. D. (1998), "Occurrence and fate of the pharmaceutical drug diclofenac in surface waters: rapid photodegradation in a lake", Environmental Science & Technology, Vol. 32, No. 22, pp. 3449-3456.
- Buser, H. R., Poiger, T. and Müller, M. D. (1999), "Occurrence and environmental behavior of the chiral pharmaceutical drug ibuprofen in surface waters and in wastewater", Environmental Science & Technology, Vol. 33, No. 15, pp. 2529-2535.
- Busman, L., Lamb, J., Randall, G., Rehm, G. and Schmitt, M. (1997), "The Nature of Phosphorus in Soils", retrieved on 19 March 2013, from <http://www1.extension.umn.edu/agriculture/nutrient-management/phosphorus/docs/FO-6795-B-1.pdf>, University of Minnesota, Minneapolis, MN, USA.
- Butler, D. and Graham, N. J. D. (1995), "Modeling dry weather wastewater flow in sewer networks", Journal of Environmental Engineering, Vol. 121, No. 2, pp. 161-173.
- Cabelli, V. J. (1977), "Indicators of Recreational Water Quality", Bacterial Indicators/Health Hazards Associated With Water, ASTM STP635, A. W. Hoadley and B. J. Dutka, eds., American Society for Testing and Materials, West Conshohocken, PA, USA
- Cabelli, V. J. (1983), "Health effects criteria for marine recreational waters, EPA-600/1-80-031", U.S. Environmental Protection Agency, Research Triangle Park, NC, USA.
- Cabral, J. and Marques, C. (2006), "Faecal coliform bacteria in Febros River (Northwest Portugal): temporal variation, correlation with water parameters, and species identification", Environmental Monitoring and Assessment, Vol. 118, No. 1, pp. 21-36.
- Carroll, S. P., Dawes, L., Hargreaves, M. and Goonetilleke, A. (2009), "Faecal pollution source identification in an urbanising catchment using antibiotic resistance profiling, discriminant analysis and partial least squares regression", Water Research, Vol. 43, No. 5, pp. 1237-1246.
- Carson, C. A., Shear, B. L., Ellersieck, M. R. and Asfaw, A. (2001), "Identification of fecal *Escherichia coli* from humans and animals by ribotyping", Applied and Environmental Microbiology, Vol. 67, No. 4, pp. 1503-1507.
- Carson, C. A., Shear, B. L., Ellersieck, M. R. and Schnell, J. D. (2003), "Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal *Escherichia coli* from humans and animals", Applied and Environmental Microbiology, Vol. 69, No. 3, pp. 1836-1839.

- Champ, D. R. and Schroeter, J. (1988), "Bacterial transport in fractured rock - a field-scale tracer test at the Chalk River nuclear laboratories", Water Science & Technology, Vol. 20, No. 11-12, pp. 81-87.
- Chesner, W. H. and Pai, M. (1981), "Hourly Diurnal Flow Variations in Publicly-Owned Wastewater Treatment Facilities", U.S. Environmental Protection Agency, Cincinnati, OH, USA.
- Cheung, W. H. S., Chang, K. C. K. and Hung, R. P. S. (1991), "Variations in Microbial Indicator Densities in Beach Waters and Health-Related Assessment of Bathing Water Quality", Epidemiology and Infection, Vol. 106, No. 2, pp. 329-344.
- Chua, L. H. C., Shanahan, P., Lo, E. Y. M., Shuy, E. B., Thompson, J. R., Dixon, C. C., Kerigan, K. B., Nshimiyimana, J. P., Yeager, J. M., Lee, L. J. and Por, Y. L. (2010), "Dry Weather Bacteria Monitoring and Variation with Land Use for Kranji Reservoir Catchment, Singapore", 17th Congress of the Asia and Pacific Division of the International Association of Hydraulic Engineering and Research, Auckland, New Zealand, February 21-24, 2010.
- Cochran, W. G. (1950), "Estimation of bacterial densities by means of the 'most probable number'", Biometrics, Vol. 6, No. 2, pp. 105-116.
- Cohen, J. and Shuval, H. I. (1973), "Coliforms, fecal coliforms, and fecal streptococci as indicators of water pollution", Water, Air, & Soil Pollution, Vol. 2, No. 1, pp. 85-95.
- Cole, D., Long, S. C. and Sobsey, M. D. (2003), "Evaluation of F+ RNA and DNA coliphages as source-specific indicators of fecal contamination in surface waters", Applied and Environmental Microbiology, Vol. 69, No. 11, pp. 6507-6514.
- Critchley, J. A. J. H., Critchley, L. A. H., Anderson, P. J. and Tomlinson, B. (2005), "Differences in the single-oral-dose pharmacokinetics and urinary excretion of paracetamol and its conjugates between Hong Kong Chinese and Caucasian subjects", Journal of Clinical Pharmacy and Therapeutics, Vol. 30, No. 2, pp. 179-184.
- Cumming, G., Fidler, F. and Vaux, D. L. (2007), "Error bars in experimental biology", The Journal of Cell Biology, Vol. 177, No. 1, pp. 7-11.
- Davies-Colley, R. J., Nagels, J. W., Smith, R. A., Young, R. G. and Phillips, C. J. (2004), "Water quality impact of a dairy cow herd crossing a stream", New Zealand Journal of Marine and Freshwater Research, Vol. 38, No. 4, pp. 569-576.
- Decon (undated), "Decon[®]90", retrieved on 20 February 2013, from <http://www.decon.co.uk/english/decon90.asp>, Decon Laboratories Ltd, East Sussex, UK.
- Desai, A. M. and Rifai, H. S. (2010), "Variability of *Escherichia coli* concentrations in an urban watershed in Texas", Journal of Environmental Engineering, Vol. 136, No. 12, pp. 1347-1359.
- Desai, A. M. and Rifai, H. S. (2013), "*Escherichia coli* concentrations in urban watersheds exhibit diurnal sag: implications for water-quality monitoring and assessment", Journal of the American Water Resources Association, Vol. 49, No. 4, pp. 766-779.

- Diagne, N. A. (2013), "Evaluation of Sewer Leakage into the Stormwater Drainage System in Singapore", Master of Engineering Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
- Dick, L. K., Stelzer, E. A., Bertke, E. E., Fong, D. L. and Stoeckel, D. M. (2010), "Relative decay of *Bacteroidales* microbial source tracking markers and cultivated *Escherichia coli* in freshwater microcosms", Applied and Environmental Microbiology, Vol. 76, No. 10, pp. 3255-3262.
- Dickenson, E. R. V., Snyder, S. A., Sedlak, D. L. and Drewes, J. E. (2011), "Indicator compounds for assessment of wastewater effluent contributions to flow and water quality", Water Research, Vol. 45, No. 3, pp. 1199-1212.
- Dixon, C., Kerigan, K., Nshimyimana, J. P. and Yeager, J. (2009), "Water Quality Monitoring, Modelling and Management for the Kranji Catchment/Reservoir System", Massachusetts Institute of Technology, Cambridge, MA, USA.
- Dombek, P. E., Johnson, L. K., Zimmerley, S. T. and Sadowsky, M. J. (2000), "Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources", Applied and Environmental Microbiology, Vol. 66, No. 6, pp. 2572-2577.
- Doshi, J. (2012), "An Investigation of Leaky Sewers as a Source of Fecal Contamination in the Stormwater Drainage System in Singapore", Master of Engineering Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
- Dreyer, A., Matthias, V., Weinberg, I. and Ebinghaus, R. (2010), "Wet deposition of poly- and perfluorinated compounds in Northern Germany", Environmental Pollution, Vol. 158, No. 5, pp. 1221-1227.
- Dufour, A. P. (1977), "Escherichia coli: The Fecal Coliform", Bacterial Indicators/Health Hazards Associated With Water, ASTM STP635, A. W. Hoadley and B. J. Dutka, eds., American Society for Testing and Materials, West Conshohocken, PA, USA.
- Dufour, A. P. (1984), "Health effects criteria for fresh recreational waters, EPA-600/1-84-004", Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH, USA.
- Duris, J. W., Reif, A. G., Krouse, D. A. and Isaacs, N. M. (2013), "Factors related to occurrence and distribution of selected bacterial and protozoan pathogens in Pennsylvania streams", Water Research, Vol. 47, No. 1, pp. 300-314.
- Dutka, B. J. and El-Shaarawi, A. (1975), "Relationship between various bacterial populations and coprostanol and cholesterol", Canadian Journal of Microbiology, Vol. 21, pp. 1386-1398.
- Eaton, A. D., Clesceri, L. S., Rice, E. W. and Greenberg, A. E. (2005), "Standard Methods for the Examination of Water and Wastewater, 21st Edition", American Public Health Association, Washington, D.C., USA.
- Eiswirth, M., Hötzl, H., Kramp, J., Lazar, C. and Merkler, G. P. (1995), "Qualifying Methods for Leakage Detection", Proceedings of the 12. Int. No-Dig conference, Dresden, Germany, September 19-22, 1995, pp. 195-208.
- Ellis, J. B., Revitt, D. M., Vollertsen, J. and Blackwood, D. J. (2009), "Sewer exfiltration and the colmation layer", Water Science & Technology, Vol. 59, No. 11, pp. 2273-2280.

- Enfinger, K. L. and Stevens, P. L. (2006), "Sewer Sociology - The Days of Our (Sewer) Lives", Proceedings of the Water Environment Federation, WEFTEC 2006, Dallas, TX, USA, October 21-25, 2006, pp. 6962-6974.
- Enns, A. A., Vogel, L. J., Abdelzaher, A. M., Solo-Gabriele, H. M., Plano, L. R. W., Gidley, M. L., Phillips, M. C., Klaus, J. S., Piggot, A. M., Feng, Z., Reniers, A. J. H. M., Haus, B. K., Elmir, S. M., Zhang, Y., Jimenez, N. H., Abdel-Mottaleb, N., Schoor, M. E., Brown, A., Khan, S. Q., Dameron, A. S., Salazar, N. C. and Fleming, L. E. (2012), "Spatial and temporal variation in indicator microbe sampling is influential in beach management decisions", Water Research, Vol. 46, No. 7, pp. 2237-2246.
- Evison, L. M. and James, A. (1973), "A comparison of the distribution of intestinal bacteria in British and East African water sources", Journal of Applied Microbiology, Vol. 36, No. 1, pp. 109-118.
- Farmer, J. J. (2006), "Bacterial Source Tracking in the Karstic Duck River Watershed of Middle Tennessee: a Comparison of Methods", Tennessee State University, Nashville, TN, USA.
- Feachem, R. (1975), "An improved role for faecal coliform to faecal streptococci ratios in the differentiation between human and non-human pollution sources", Water Research, Vol. 9, No. 7, pp. 689-690.
- Feng, P., Weagant, S. D. and Grant, M. A. (2002), "BAM: Enumeration of Escherichia coli and the Coliform Bacteria", retrieved on 19 July 2010, from <http://www.fda.gov/food/scienceresearch/LaboratoryMethods/BacteriologicaAnalyticalManualBAM/ucm064948.htm>, U.S. Food and Drug Administration, Silver Spring, MD, USA.
- Férézou, J., Gouffier, E., Coste, T. and Chevallier, F. (1978), "Daily elimination of fecal neutral sterols by humans", Digestion, Vol. 18, pp. 201-212.
- Field, K. G., Bernhard, A. E. and Brodeur, T. J. (2003), "Molecular approaches to microbiological monitoring: fecal source detection", Environmental Monitoring and Assessment, Vol. 81, No. 1, pp. 313-326.
- Field, R., Pitt, R., Lalor, M., Brown, M., Vilkelis, W. and Phackston, E. (1994), "Investigation of dry weather pollutant entries into storm drainage systems", Journal of Environmental Engineering, Vol. 120, No. 5, pp. 1044-1066.
- Fiss, E. M., Rule, K. L. and Vikesland, P. J. (2007), "Formation of chloroform and other chlorinated byproducts by chlorination of triclosan-containing antibacterial products", Environmental Science & Technology, Vol. 41, No. 7, pp. 2387-2394.
- Foley, K., Granger, E. C., Sagel, A. M., Nshimiyimana, J. P. and Sudhalkar, A. (2010), "Characterization of Water Quality in Kranji Reservoir Catchment", Massachusetts Institute of Technology, Cambridge, MA, USA.
- Fong, T. T., Griffin, D. W. and Lipp, E. K. (2005), "Molecular assays for targeting human and bovine enteric viruses in coastal waters and their application for library-independent source tracking", Applied and Environmental Microbiology, Vol. 71, No. 4, pp. 2070-2078.

- Fricker, E. J., Illingworth, K. S. and Fricker, C. R. (1997), "Use of two formulations of Colilert and QuantiTray™ for assessment of the bacteriological quality of water", Water Research, Vol. 31, No. 10, pp. 2495-2499.
- Gannon, J. J. and Busse, M. K. (1989), "E. coli and enterococci levels in urban stormwater, river water and chlorinated treatment plant effluent", Water Research, Vol. 23, No. 9, pp. 1167-1176.
- Gawler, A. H., Beecher, J. E., Brandão, J., Carroll, N. M., Falcão, L., Gourmelon, M., Masterson, B., Nunes, B., Porter, J., Rincé, A., Rodrigues, R., Thorp, M., Martin Walters, J. and Meijer, W. G. (2007), "Validation of host-specific Bacteroidales 16S rRNA genes as markers to determine the origin of faecal pollution in Atlantic Rim countries of the European Union", Water Research, Vol. 41, No. 16, pp. 3780-3784.
- Geldreich, E. E. and Kenner, B. A. (1969), "Concepts of fecal streptococci in stream pollution", Journal (Water Pollution Control Federation), Vol. 41, No. 8, pp. R336-R352.
- Grant, S. B., Sanders, B. F., Boehm, A. B., Redman, J. A., Kim, J. H., Mrše, R. D., Chu, A. K., Gouldin, M., McGee, C. D., Gardiner, N. A., Jones, B. H., Svejkský, J., Leipzig, G. V. and Brown, A. (2001), "Generation of Enterococci Bacteria in a Coastal Saltwater Marsh and Its Impact on Surf Zone Water Quality", Environmental Science & Technology, Vol. 35, No. 12, pp. 2407-2416.
- Green, H. C., Shanks, O. C., Sivaganesan, M., Haugland, R. A. and Field, K. G. (2011), "Differential decay of human faecal *Bacteroides* in marine and freshwater", Environmental Microbiology, Vol. 13, No. 12, pp. 3235-3249.
- Griffin, D. W., Stokes, R., Rose, J. B. and III, J. H. P. (2000), "Bacterial indicator occurrence and the use of an F⁺ specific RNA coliphage assay to identify fecal sources in Homosassa Springs, Florida", Microbial Ecology, Vol. 39, pp. 56-64.
- Gronewold, A. D. and Wolpert, R. L. (2008), "Modeling the relationship between most probable number (MPN) and colony-forming unit (CFU) estimates of fecal coliform concentration", Water Research, Vol. 42, No. 13, pp. 3327-3334.
- Hach (1999), "Coliforms: Membrane Filtration (simultaneous detection)", retrieved on 16 November 2010, from <http://www.water-research.net/Waterlibrary/watermanual/MCOLI.PDF>, Hach Company, Loveland, CO, USA.
- Hach (2002), "Expected Reactions of Various Microorganisms with m-ColiBlue24[®]", retrieved on 16 November 2010, from <http://www.hach.com/fmmimghach?/CODE%3AMCOLIBLUE24%20TS1573%7C1>, Hach Company, Loveland, CO, USA.
- Hach (2006), "EPA Approves New Test Procedures for the Analysis of Microbiological Pollutants in Wastewater and Sludge", retrieved on 16 November 2010, from <http://www.hach.com/fmmimghach?/CODE%3AMCB24CRWAPUBLICATION12892%7C1>, Hach Company, Loveland, CO, USA.
- Hach (2008), "m-ColiBlue24[®] Broth: Data Sheet", retrieved on 16 November 2010, from <http://www.hach.com/fmmimghach?/CODE%3AL2628-COLOR-WEB-200816820%7C1>, Hach Company, Loveland, CO, USA.

- Hagedorn, C., Crozier, J., Mentz, K., Booth, A., Graves, A., Nelson, N. and Reneau, R. (2003), "Carbon source utilization profiles as a method to identify sources of faecal pollution in water", Journal of Applied Microbiology, Vol. 94, No. 5, pp. 792-799.
- Hagedorn, C., Robinson, S. L., Filtz, J. R., Grubbs, S. M., Angier, T. A. and Reneau, R. B. (1999), "Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci", Applied and Environmental Microbiology, Vol. 65, No. 12, pp. 5522-5531.
- Hagedorn, C. and Weisberg, S. (2009), "Chemical-based fecal source tracking methods: current status and guidelines for evaluation", Reviews in Environmental Science and Biotechnology, Vol. 8, No. 3, pp. 275-287.
- Harris, D. C. (2007), Quantitative Chemical Analysis, 7th Edition, W.H. Freeman, New York, NY, USA.
- Hartel, P. G., Frick, E. A., Funk, A. L., Hill, J. L., Summer, J. D. and Gregory, M. B. (2004), "Sharing of ribotype patterns of *Escherichia coli* isolates during baseflow and stormflow conditions", U.S. Department of the Interior, U.S. Geological Survey, Reston, VA, USA.
- Harwood, V. J., Whitlock, J. and Withington, V. (2000), "Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of fecal contamination in subtropical waters", Applied and Environmental Microbiology, Vol. 66, No. 9, pp. 3698-3704.
- Harwood, V. J., Wiggins, B., Hagedorn, C., Ellender, R. D., Gooch, J., Kern, J., Samadpour, M., Chapman, A. C. H., Robinson, B. J. and C. Thompson, B. (2003), "Phenotypic library-based microbial source tracking methods: efficacy in the California collaborative study", Journal of Water and Health, Vol. 01, pp. 153-166.
- Hazen, T. C. (1988), "Fecal coliforms as indicators in tropical waters: a review", Toxicity Assessment: An International Journal, Vol. 3, No. 5, pp. 461-477.
- Held, I., Wolf, L., Eiswirth, M. and Hötzl, H. (2006), "Impacts of Sewer Leakage on Urban Groundwater", Urban Groundwater Management and Sustainability, J. H. Tellam, M. O. Rivett, R. G. Israfilov, and L. G. Herringshaw, eds., Springer Netherlands, pp. 189-204.
- Helsel, D. R. and Hirsch, R. M. (2002), Statistical Methods in Water Resources, U.S. Geological Survey, USA.
- Hem, J. D. (1985), "Study and Interpretation of the Chemical Characteristics of Natural Water, Third Edition", U.S. Geological Survey, Washington, D.C., USA.
- Henry, P. D. (1980), "Comparative pharmacology of calcium antagonists: nifedipine, verapamil and diltiazem", The American Journal of Cardiology, Vol. 46, No. 6, pp. 1047-1058.
- Hutson, S. (2010), "Painkiller concerns grow ahead of new guidelines", Nature Medicine, Vol. 16, No. 1, pp. 10-10.
- Hyer, K. E. (2007), "A Multiple-tracer Approach for Identifying Sewage Sources to an Urban Stream System", U.S. Department of the Interior, U.S. Geological Survey, Reston, VA, USA.

- Hyer, K. E. and Moyer, D. L. (2003), "Patterns and Sources of Fecal Coliform Bacteria in Three Streams in Virginia, 1999-2000", U.S. Department of the Interior, U.S. Geological Survey, Richmond, VA, USA.
- IDEXX (2007), "Colilert[®] Test Kit Instructions", IDEXX Laboratories, Westbrook, ME, USA.
- IDEXX (2008), "Enterolert[™] Test Kit Instructions", IDEXX Laboratories, Westbrook, ME, USA.
- IDEXX (2010a), "Water Microbiology Testing", retrieved on 9 November 2010, from http://www.idexx.com/view/xhtml/en_us/water-microbiology.isf, IDEXX Laboratories, Westbrook, ME, USA.
- IDEXX (2010b), "Quanti-Tray[®]/2000 with 95% Confidence Limits", retrieved on 16 November 2010, from http://www.idexx.com/pubwebresources/pdf/en_us/water/qt2k95.pdf, IDEXX Laboratories, Westbrook, ME, USA.
- IDEXX (undated), "IDEXX Quanti-Tray/2000: Insert and Most Probable Number (MPN) Table", retrieved on 18 February 2013, from http://www.idexx.com/pubwebresources/pdf/en_us/water/0602320.pdf, IDEXX Laboratories, Westbrook, ME, USA.
- Isobe, K. O., Tarao, M., Chiem, N. H., Minh, L. Y. and Takada, H. (2004), "Effect of environmental factors on the relationship between concentrations of coprostanol and fecal indicator bacteria in tropical (Mekong Delta) and temperate (Tokyo) freshwaters", Applied and Environmental Microbiology, Vol. 70, No. 2, pp. 814-821.
- Isobe, K. O., Tarao, M., Zakaria, M. P., Chiem, N. H., Minh, L. Y. and Takada, H. (2002), "Quantitative application of fecal sterols using gas chromatography–mass spectrometry to investigate fecal pollution in tropical waters: Western Malaysia and Mekong Delta, Vietnam", Environmental Science & Technology, Vol. 36, No. 21, pp. 4497-4507.
- Jenkins, M. W., Tiwari, S., Lorente, M., Gichaba, C. M. and Wuertz, S. (2009), "Identifying human and livestock sources of fecal contamination in Kenya with host-specific *Bacteroidales* assays", Water Research, Vol. 43, pp. 4956-4966.
- Kacar, A. (2011), "Analysis of spatial and temporal variation in the levels of microbial fecal indicators in the major rivers flowing into the Aegean Sea, Turkey", Ecological Indicators, Vol. 11, No. 5, pp. 1360-1365.
- Kannel, P. R., Lee, S., Kanel, S. R. and Khan, S. P. (2007), "Chemometric application in classification and assessment of monitoring locations of an urban river system", Analytica Chimica Acta, Vol. 582, No. 2, pp. 390-399.
- Karpf, C., Traenckner, J. and Krebs, P. (2009), "Hydraulic modelling of sewage exfiltration", Water Science & Technology, Vol. 59, No. 8, pp. 1559-1565.
- Kaushik, R., Balasubramanian, R. and de la Cruz, A. A. (2012), "Influence of air quality on the composition of microbial pathogens in fresh rainwater", Applied and Environmental Microbiology, Vol. 78, No. 8, pp. 2813-2818.
- Kenzaka, T., Yamaguchi, N., Prapagdee, B., Mikami, E. and Nasu, M. (2001), "Bacterial community composition and activity in urban rivers in Thailand and Malaysia", Journal of Health Science, Vol. 47, No. 4, pp. 353-361.

- Khaiat, A. (2005), "Triclosan in Cosmetic Products", retrieved on 23 July 2013, from http://www.ctfas.org.sg/Files/docs/Triclosan_CTFAS.pdf, The Cosmetic, Toiletries and Fragrance Association of Singapore (CTFAS), Singapore.
- Kinzelman, J., Ng, C., Jackson, E., Gradus, S. and Bagley, R. (2003), "Enterococci as indicators of Lake Michigan recreational water quality: comparison of two methodologies and their impacts on public health regulatory events", Applied and Environmental Microbiology, Vol. 69, No. 1, pp. 92-96.
- Klinger, J., Turkovic, R., Wolf, L. and Hoetzel, H. (2007), "Long term investigations of the colmation processes at a real sewer defect", Geophysical Research Abstracts Vol. 9.
- Koirala, S. R., Gentry, R. W., Perfect, E., Schwartz, J. S. and Sayler, G. S. (2008), "Temporal variation and persistence of bacteria in streams", Journal of Environmental Quality, Vol. 37, No. 4, pp. 1559-1566.
- Kolpin, D. W., Furlong, E. T., Meyer, M. T., Thurman, E. M., Zaugg, S. D., Barber, L. B. and Buxton, H. T. (2002), "Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. streams, 1999–2000: a national reconnaissance", Environmental Science & Technology, Vol. 36, No. 6, pp. 1202-1211.
- Kreder, C. A. (1998), "Persistence of PCR-detectable *Bacteroides distasonis* from human feces in river water", Applied and Environmental Microbiology, Vol. 64, No. 10, pp. 4103-4105.
- Kwasi, O.-D., Keith, J. and Nigel, P. (1999), "The effect of the time of sampling on the compliance of bathing water in NW England with the EU Directive on bathing water quality", Journal of Coastal Conservation, Vol. 5, No. 1, pp. 51-58.
- Leecaster, M. K. and Weisberg, S. B. (2001), "Effect of sampling frequency on shoreline microbiology assessments", Marine Pollution Bulletin, Vol. 42, No. 11, pp. 1150-1154.
- Leeming, R., Ball, A., Ashbolt, N. and Nichols, P. (1996), "Using faecal sterols from humans and animals to distinguish faecal pollution in receiving waters", Water Research, Vol. 30, No. 12, pp. 2893-2900.
- Leeming, R. and Nichols, P. D. (1996), "Concentrations of coprostanol that correspond to existing bacterial indicator guideline limits", Water Research, Vol. 30, No. 12, pp. 2997-3006.
- Life (2013), "Digital PCR: a New Approach to Nucleic Acid Detection and Quantification", retrieved on 21 August 2013, from <http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/PCR/digital-pcr.html>, Life Technologies Corporation, Carlsbad, CA, USA.
- Lim, K. H. J. and Yap, K. B. (1999), "The prescribing pattern of outpatient polyclinic doctors", Singapore Medical Journal, Vol. 40, No. 6.
- Lindström, A., Buerge, I. J., Poiger, T., Bergqvist, P. A., Müller, M. D. and Buser, H. R. (2002), "Occurrence and environmental behavior of the bactericide triclosan and its methyl derivative in surface waters and in wastewater", Environmental Science & Technology, Vol. 36, No. 11, pp. 2322-2329.

- Lundy, D. and Gogel, T. (1988), "Capabilities and Limitations of Wells for Detecting and Monitoring Liquid Phase Hydrocarbons", Proceedings of the Second National Outdoor Action Conference on Aquifer Restoration, Ground Water Monitoring and Geophysical Methods, Las Vegas, Nevada, U.S. May 23-26, 1988, pp. 343-362.
- Madigan, M. T., Martinko, J. M., Stahl, D. A. and Clark, D. P. (2011), Brock Biology of Microorganisms, 13th Edition, Pearson Education, Inc., San Francisco, CA, USA.
- Maier, R. M., Pepper, I. L. and Gerba, C. P. (2000), Environmental Microbiology, Elsevier (Singapore) Pte Ltd, Singapore.
- Manly, B. F. J. (2005), Multivariate Statistical Methods : a Primer, Chapman & Hall/CRC, Boca Raton, FL, USA.
- MathWorks (2013), "Cophenet", retrieved on 23 March 2013, from <http://www.mathworks.com/help/stats/cophenet.html>, The MathWorks, Inc., Natick, MA, USA.
- McCarthy, D. T., Deletic, A., Mitchell, V. G., Fletcher, T. D. and Diaper, C. (2008), "Uncertainties in stormwater E. coli levels", Water Research, Vol. 42, No. 6-7, pp. 1812-1824.
- McDaniel, R. L., Soupir, M. L., Tuttle, R. B. and Cervantes, A. E. (2013), "Release, dispersion, and resuspension of *Escherichia coli* from direct fecal deposits under controlled flows", Journal of the American Water Resources Association, Vol. 49, No. 2, pp. 319-327.
- McGrath, D., Zhang, C. and Carton, O. T. (2004), "Geostatistical analyses and hazard assessment on soil lead in Silvermines area, Ireland", Environmental Pollution, Vol. 127, No. 2, pp. 239-248.
- McKillup, S. and Dyar, M. D. (2010), Geostatistics Explained: An Introductory Guide for Earth Scientists, Cambridge University Press, UK.
- McQuaig, S., Griffith, J. and Harwood, V. J. (2012), "The association of fecal indicator bacteria with human viruses and microbial source tracking markers at coastal beaches impacted by nonpoint source pollution", Applied and Environmental Microbiology, Vol. 78, No. 18, pp. 6423-6432.
- Meays, C. L. (2005), "Bacterial Source Tracking and Survival of *Escherichia coli*", Ph.D. dissertation, University of Victoria, British Columbia, Canada.
- Mendez-Sagel, A. (2010), "Water Quality Studies in Kranji Catchment, Singapore: Use of Organic Tracer and PEDs for Identifying Potential Sewage Sources", Master of Engineering Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
- Miller, J. M. and Rhoden, D. L. (1991), "Preliminary evaluation of Biolog, a carbon source utilization method for bacterial identification", Journal of Clinical Microbiology, Vol. 29, No. 6, pp. 1143-1147.
- Muñiz, I., Jiménez, L., Toranzos, G. A. and Hazen, T. C. (1989), "Survival and activity of *Streptococcus faecalis* and *Escherichia coli* in tropical freshwater", Microbial Ecology, Vol. 18, No. 2, pp. 125-134.
- Nasco (2010), "Instruction for Nasco Sampling Pole B01367WA", Nasco, Fort Atkinson, WI, USA.

- NEA (2010), "Recreational Water Quality", retrieved on 27 January 2011, from http://app2.nea.gov.sg/topics_rwq.aspx, National Environment Agency, Singapore.
- Nguyen, V. T., Reinhard, M. and Gin, K. Y. H. (2011), "Occurrence and source characterization of perfluorochemicals in an urban watershed", Chemosphere, Vol. 82, No. 9, pp. 1277-1285.
- Noble, R. T., Lee, I. M. and Schiff, K. C. (2004), "Inactivation of indicator micro-organisms from various sources of faecal contamination in seawater and freshwater", Journal of Applied Microbiology, Vol. 96, No. 3, pp. 464-472.
- Novotny, V. (2003), Water Quality: Diffuse Pollution and Watershed Management 2nd ed., J. Wiley, Hoboken, NJ, USA.
- Nshimiyimana, J. P. (2010), "Evaluating Human Fecal Contamination Sources in Kranji Reservoir Catchment, Singapore", Master of Science Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
- Nshimiyimana, J. P., Ekklesia, E., Shanahan, P., Chua, L. H. C. and Thompson, J. R. (2014), "Distribution and abundance of human specific *Bacteroides* and relation to traditional indicators in an urban tropical catchment", Journal of Applied Microbiology, Vol. 116, No. 5, pp. 1369-1383.
- NTU (2008), "Water Quality Monitoring, Modelling and Management for Kranji Catchment/Reservoir System - Phases 1 and 2, May 2004 to December 2007", Nanyang Technological University, Singapore.
- Odegaard, A. O., Pereira, M. A., Koh, W. P., Arakawa, K., Lee, H. P. and Yu, M. C. (2008), "Coffee, tea, and incident type 2 diabetes: the Singapore Chinese health study", The American Journal of Clinical Nutrition, Vol. 88, No. 4, pp. 979-985.
- Panda, U. C., Sundaray, S. K., Rath, P., Nayak, B. B. and Bhatta, D. (2006), "Application of factor and cluster analysis for characterization of river and estuarine water systems - A case study: Mahanadi River (India)", Journal of Hydrology, Vol. 331, No. 3-4, pp. 434-445.
- Pang, L., Close, M. and Noonan, M. (1998), "Rhodamine WT and *Bacillus subtilis* transport through an alluvial gravel aquifer", Ground Water, Vol. 36, No. 1, pp. 112-122.
- Parkhurst, D. F. (1998), "Arithmetic Versus Geometric Means for Environmental Concentration Data", Environmental Science & Technology, Vol. 32, No. 3, pp. 92A-98A.
- Parveen, S., Hodge, N. C., Stall, R. E., Farrah, S. R. and Tamplin, M. L. (2001), "Phenotypic and genotypic characterization of human and nonhuman *Escherichia coli*", Water Research, Vol. 35, No. 2, pp. 379-386.
- Parveen, S., Murphree, R. L., Edmiston, L., Kaspar, C. W., Portier, K. M. and Tamplin, M. L. (1997), "Association of multiple-antibiotic-resistance profiles with point and nonpoint sources of *Escherichia coli* in Apalachicola Bay", Applied and Environmental Microbiology, Vol. 63, No. 7, pp. 2607-2612.
- Parveen, S., Portier, K. M., Robinson, K., Edmiston, L. and Tamplin, M. L. (1999), "Discriminant analysis of ribotype profiles of *Escherichia coli* for differentiating

- human and nonhuman sources of fecal pollution", Applied and Environmental Microbiology, Vol. 65, No. 7, pp. 3142-3147.
- Paul, M. J. and Meyer, J. L. (2008), "Streams in the Urban Landscape", Urban Ecology, J. M. Marzluff, E. Shulenberger, W. Endlicher, M. Alberti, G. Bradley, C. Ryan, U. Simon, and C. ZumBrunnen, eds., Springer US, pp. 207-231.
- Peng, X., Yu, Y., Tang, C., Tan, J., Huang, Q. and Wang, Z. (2008), "Occurrence of steroid estrogens, endocrine-disrupting phenols, and acid pharmaceutical residues in urban riverine water of the Pearl River Delta, South China", Science of The Total Environment, Vol. 397, No. 1-3, pp. 158-166.
- Pereira, W. E. and Hostettler, F. D. (1993), "Nonpoint source contamination of the Mississippi River and its tributaries by herbicides", Environmental Science & Technology, Vol. 27, No. 8, pp. 1542-1552.
- Peters, D. G., Hayes, J. M. and Hieftje, G. M. (1974), Chemical Separations and Measurements: Theory and Practice of Analytical Chemistry, Saunders Golden Sunburst Series, Bloomington, IN, USA.
- Pipes, W. O., Ward, P. and Ahn, S. H. (1977), "Frequency distributions for coliform bacteria in water", Journal of American Water Works Association, Vol. 69, No. 12, pp. 664-668.
- Poiger, T., Field, J. A., Field, T. M. and Giger, W. (1996), "Occurrence of fluorescent whitening agents in sewage and river water determined by solid-phase extraction and high-performance liquid chromatography", Environmental Science & Technology, Vol. 30, No. 7, pp. 2220-2226.
- Pope, M. L., Bussen, M., Feige, M. A., Shadix, L., Gonder, S., Rodgers, C., Chambers, Y., Pulz, J., Miller, K., Connell, K. and Standridge, J. (2003), "Assessment of the effects of holding time and temperature on *Escherichia coli* densities in surface water samples", Applied and Environmental Microbiology, Vol. 69, No. 10, pp. 6201-6207.
- Prüss, A. (1998), "Review of epidemiological studies on health effects from exposure to recreational water", International Journal of Epidemiology, Vol. 27, No. 1, pp. 1-9.
- PUB (2004), "Code of Practice on Sewerage and Sanitary Works", retrieved on 12 March 2013, from <http://www.pub.gov.sg/general/Documents/COPSSW%20nov2004.pdf>, Public Utilities Board, Singapore.
- PUB (2009), "Standard Details of Minor Sewers and Manholes", retrieved on 12 March 2013, from http://www.pub.gov.sg/general/Documents/WRN_STD_002A.pdf.
- PUB (2010a), "The Singapore Water History", retrieved on 27 April 2011, from <http://www.pub.gov.sg/water/Pages/singaporewaterstory.aspx>, Public Utilities Board, Singapore.
- PUB (2010b), "Marina Barrage", retrieved on 21 July 2013, from <http://www.pub.gov.sg/Marina/Pages/default.aspx>, Public Utilities Board, Singapore.

- PUB (2010c), "ABC Waters Programme", retrieved on 16 August 2013, from <http://www.pub.gov.sg/abcwaters/Pages/default.aspx>, Public Utilities Board, Singapore.
- PUB (2010d), "Used Water", retrieved on 16 February 2014, from <http://www.pub.gov.sg/about/historyfuture/Pages/UsedWater.aspx>, Public Utilities Board, Singapore.
- PUB (2010e), "Sewer Rehabilitation: Outreach Programme", retrieved on 22 February 2014, from http://www.pub.gov.sg/sewerrehab/private_sewerrehab_prog/Pages/outreach_%20prog.aspx, Public Utilities Board, Singapore.
- Rapaport, R. A. and Eckhoff, W. S. (1990), "Monitoring linear alkyl benzene sulfonate in the environment: 1973–1986", Environmental Toxicology and Chemistry, Vol. 9, No. 10, pp. 1245-1257.
- Reynolds, J. H. and Barrett, M. H. (2003), "A review of the effects of sewer leakage on groundwater quality", Water and Environment Journal, Vol. 17, No. 1, pp. 34-39.
- Rivera, S. C., Hazen, T. C. and Toranzos, G. A. (1988), "Isolation of fecal coliforms from pristine sites in a tropical rain forest", Applied and Environmental Microbiology, Vol. 54, No. 2, pp. 513-517.
- Rodgers, P., Soulsby, C., Hunter, C. and Petry, J. (2003), "Spatial and temporal bacterial quality of a lowland agricultural stream in northeast Scotland", The Science of The Total Environment, Vol. 314-316, pp. 289-302.
- Rompré, A., Servais, P., Baudart, J., de-Roubin, M.-R. and Laurent, P. (2002), "Detection and enumeration of coliforms in drinking water: current methods and emerging approaches", Journal of Microbiological Methods, Vol. 49, No. 1, pp. 31-54.
- Rutkowski, A. A. and Sjogren, R. E. (1987), "Streptococcal population profiles as indicators of water quality", Water, Air, & Soil Pollution, Vol. 34, No. 3, pp. 273-284.
- Rutsch, M., Rieckermann, J. and Krebs, P. (2006), "Quantification of sewer leakage: a review", Water Science & Technology, Vol. 54, No. 6-7, pp. 135-144.
- Santiago-Rodríguez, T. M., Dávila, C., González, J., Bonilla, N., Marcos, P., Urdaneta, M., Cadete, M., Monteiro, S., Santos, R., Domingo, J. S. and Toranzos, G. A. (2010), "Characterization of *Enterococcus faecalis*-infecting phages (enterophages) as markers of human fecal pollution in recreational waters", Water Research, Vol. 44, No. 16, pp. 4716-4725.
- Santiago-Rodriguez, T. M., Tremblay, R. L., Toledo-Hernandez, C., Gonzalez-Nieves, J. E., Ryu, H., Domingo, J. W. S. and Toranzos, G. A. (2012), "Microbial quality of tropical inland waters and effects of rainfall events", Applied and Environmental Microbiology, Vol. 78, No. 15, pp. 5160-5169.
- Santoro, A. E. and Boehm, A. B. (2007), "Frequent occurrence of the human-specific *Bacteroides* fecal marker at an open coast marine beach: relationship to waves, tides and traditional indicators", Environmental Microbiology, Vol. 9, No. 8, pp. 2038-2049.

- Sauer, E. P., VandeWalle, J. L., Bootsma, M. J. and McLellan, S. L. (2011), "Detection of the human specific *Bacteroides* genetic marker provides evidence of widespread sewage contamination of stormwater in the urban environment", Water Research, Vol. 45, No. 14, pp. 4081-4091.
- Savichtcheva, O. and Okabe, S. (2006), "Alternative indicators of fecal pollution: relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives", Water Research, Vol. 40, pp. 2463-2476.
- Scheurer, M., Brauch, H. J. and Lange, F. (2009), "Analysis and occurrence of seven artificial sweeteners in German waste water and surface water and in soil aquifer treatment (SAT)", Analytical and Bioanalytical Chemistry, Vol. 394, No. 6, pp. 1585-1594.
- Schilling, K. E., Zhang, Y.-K., Hill, D. R., Jones, C. S. and Wolter, C. F. (2009), "Temporal variations of *Escherichia coli* concentrations in a large Midwestern river", Journal of Hydrology, Vol. 365, No. 1-2, pp. 79-85.
- Schulz, C. J. and Childers, G. W. (2011), "Fecal *Bacteroidales* diversity and decay in response to variations in temperature and salinity", Applied and Environmental Microbiology, Vol. 77, No. 8, pp. 2563-2572.
- Scott, T. M., Rose, J. B., Jenkins, T. M., Farrah, S. R. and Lukasik, J. (2002), "Microbial source tracking: current methodology and future directions", Applied and Environmental Microbiology, Vol. 68, No. 12, pp. 5796-5803.
- Sercu, B., Van De Werfhorst, L. C., Murray, J. and Holden, P. A. (2009), "Storm drains are sources of human fecal pollution during dry weather in three urban Southern California watersheds", Environmental Science & Technology, Vol. 43, No. 2, pp. 293-298.
- Sercu, B., Van De Werfhorst, L. C., Murray, J. L. S. and Holden, P. A. (2011a), "Cultivation-independent analysis of bacteria in IDEXX Quanti-Tray/2000 fecal indicator assays", Applied and Environmental Microbiology, Vol. 77, No. 2, pp. 627-633.
- Sercu, B., Van De Werfhorst, L. C., Murray, J. L. S. and Holden, P. A. (2011b), "Sewage exfiltration as a source of storm drain contamination during dry weather in urban watersheds", Environmental Science & Technology, Vol. 45, No. 17, pp. 7151-7157.
- Seurinck, S., Defoirdt, T., Verstraete, W. and Siciliano, S. D. (2005a), "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, Vol. 7, No. 2, pp. 249-259.
- Seurinck, S., Verstraete, W. and Siciliano, S. (2005), "Microbial source tracking for identification of fecal pollution", Reviews in Environmental Science and Biotechnology, Vol. 4, No. 1, pp. 19-37.
- Severino, P. and Brisse, S. (2005), "Ribotyping in Clinical Microbiology", Medical Biomethods Handbook, J. M. Walker and R. Rapley, eds., Humana Press Inc.
- Shah, V. G., Hugh Dunstan, R., Geary, P. M., Coombes, P., Roberts, T. K. and Rothkirch, T. (2007), "Bacterial source tracking from diverse land use catchments by sterol ratios", Water Research, Vol. 41, No. 16, pp. 3667-3674.

- Shanahan, P. (2009), "Groundwater in the Urban Environment", The Water Environment of Cities, L. A. Baker, ed., Springer Science+Business Media, LLC, New York, NY, pp. 29-48.
- Shanks, O. C., Nietch, C., Simonich, M., Younger, M., Reynolds, D. and Field, K. G. (2006), "Basin-wide analysis of the dynamics of fecal contamination and fecal source identification in Tillamook Bay, Oregon", Applied and Environmental Microbiology, Vol. 72, No. 8, pp. 5537-5546.
- Sherr, B. F., Sherr, E. B. and Rassoulzadegan, F. (1988), "Rates of digestion of bacteria by marine phagotrophic protozoa: temperature dependence", Applied and Environmental Microbiology, Vol. 54, No. 5, pp. 1091-1095.
- Shin, S. (2012), "An Analysis of Spatial and Temporal Variation in Fecal Indicator Concentrations in Singapore", Master of Engineering Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
- Shumway, R. H. and Stoffer, D. S. (2011), Time Series Analysis and Its Applications, Springer Science+Business Media, LLC, New York, NY, USA.
- Sidhu, J. P. S., Ahmed, W., Gernjak, W., Aryal, R., McCarthy, D., Palmer, A., Kolotelo, P. and Toze, S. (2013), "Sewage pollution in urban stormwater runoff as evident from the widespread presence of multiple microbial and chemical source tracking markers", Science of The Total Environment, Vol. 463–464, No. 0, pp. 488-496.
- Simmons, G. M., Jr. , Herbein, S. A. and James, C. M. (1995), "Managing nonpoint fecal coliform sources to tidal inlets", Universities Council on Water Resources, Water Resources Update 100, pp. 64-74.
- Simmons, G. M., Jr. , Waye, D. F., Herbein, S., Myers, S. and Walker, E. (2000), "Estimating Nonpoint Fecal Coliform Sources in Northern Virginia's Four Mile Run Watershed", Virginia Water Research Symposium 2000.
- Singer, H., Müller, S., Tixier, C. and Pillonel, L. (2002), "Triclosan: occurrence and fate of a widely used biocide in the aquatic environment: field measurements in wastewater treatment plants, surface waters, and lake sediments", Environmental Science & Technology, Vol. 36, No. 23, pp. 4998-5004.
- Sinton, L. W., Donnison, A. M. and Hastie, C. M. (1993), "Faecal streptococci as faecal pollution indicators: a review. Part I: taxonomy and enumeration", New Zealand Journal of Marine and Freshwater Research, Vol. 27, pp. 101-115.
- Sinton, L. W., Finlay, R. K., Pang, L. and Scott, D. M. (1997), "Transport of bacteria and bacteriophages in irrigated effluent into and through an alluvial gravel aquifer", Water, Air, & Soil Pollution, Vol. 98, pp. 17-42.
- Solo-Gabriele, H. M., Wolfert, M. A., Desmarais, T. R. and Palmer, C. J. (2000), "Sources of Escherichia coli in a Coastal Subtropical Environment", Applied and Environmental Microbiology, Vol. 66, No. 1, pp. 230-237.
- Spongberg, A. L., Witter, J. D., Acuña, J., Vargas, J., Murillo, M., Umaña, G., Gómez, E. and Perez, G. (2011), "Reconnaissance of selected PPCP compounds in Costa Rican surface waters", Water Research, Vol. 45, No. 20, pp. 6709-6717.
- Sutherland, B., Woodhouse, L. R., Strong, P. and King, J. C. (1999), "Boron balance in humans", The Journal of Trace Elements in Experimental Medicine, Vol. 12, No. 3, pp. 271-284.

- Tamhane, A. C. and Dunlop, D. D. (2000), Statistics and Data Analysis: from Elementary to Intermediate, Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Tan, Y. S., Lee, T. J. and Tan, K. (2009), Clean, Green and Blue: Singapore's Journey Towards Environmental and Water Sustainability, Institute of Southeast Asian Studies (ISEAS) Publishing, Pasir Panjang, Singapore.
- Taylor, J. A., Jakeman, A. J. and Simpson, R. W. (1986), "Modeling distributions of air pollutant concentrations--I. Identification of statistical models", Atmospheric Environment (1967), Vol. 20, No. 9, pp. 1781-1789.
- Templ, M., Filzmoser, P. and Reimann, C. (2008), "Cluster analysis applied to regional geochemical data: Problems and possibilities", Applied Geochemistry, Vol. 23, No. 8, pp. 2198-2213.
- Tixier, C., Singer, H. P., Oellers, S. and Müller, S. R. (2003), "Occurrence and fate of carbamazepine, clofibric acid, diclofenac, ibuprofen, ketoprofen, and naproxen in surface waters", Environmental Science & Technology, Vol. 37, No. 6, pp. 1061-1068.
- Toledo-Hernandez, C., Ryu, H., Gonzalez-Nieves, J., Huertas, E., Toranzos, G. A. and Santo Domingo, J. W. (2013), "Tracking the primary sources of fecal pollution in a tropical watershed in a one-year study", Applied and Environmental Microbiology, Vol. 79, No. 5, pp. 1689-1696.
- Tortajada, C., Joshi, Y. K. and Biswas, A. K. (2013), The Singapore Water Story: Sustainable Development in an Urban City-State, Routledge, New York, NY, USA.
- Traister, E. and Anisfeld, S. C. (2006), "Variability of indicator bacteria at different time scales in the Upper Hoosic River watershed", Environmental Science & Technology, Vol. 40, No. 16, pp. 4990-4995.
- Trakhtenbrot, A. and Kadmon, R. (2006), "Effectiveness of Environmental Cluster Analysis in Representing Regional Species Diversity", Conservation Biology, Vol. 20, No. 4, pp. 1087-1098.
- USEPA (1983), "Results of the Nationwide Urban Runoff Program: Volume I - Final Report", EPA-832-R-83-112, U.S. Environmental Protection Agency, Washington, D.C., USA.
- USEPA (1986), "Ambient Water Quality Criteria for Bacteria", EPA440/5-84-002, U.S. Environmental Protection Agency, Washington, D.C., USA.
- USEPA (1999), "Collection Systems, O&M Fact Sheet: Trenchless Sewer Rehabilitation", retrieved on 22 February 2014, from http://water.epa.gov/scitech/wastetech/upload/2002_06_28_mtb_rehabl.pdf, U.S. Environmental Protection Agency, Washington, D. C., USA.
- USEPA (2001), "Method 1602: Male-specific (F+) and Somatic Coliphage in Water by Single Agar Layer (SAL) Procedure", EPA 821-R-01-029, U.S. Environmental Protection Agency, Washington, D. C., USA.
- USEPA (2005), "Guidelines Establishing Test Procedures for the Analysis of Pollutants; Analytical Methods for Biological Pollutants in Ambient Water; Proposed Rule", Federal Register, U.S. Environmental Protection Agency, Washington, D. C., USA, Vol. 70, No. 157, pp. 48256-48268.

- USEPA (2006), "Data Quality Assessment: Statistical Methods for Practitioners", U.S. Environmental Protection Agency, Washington, D. C., USA.
- USEPA (2007), "Method 1694: Pharmaceuticals and Personal Care Products in Water, Soil, Sediment, and Biosolids by HPLC/MS/MS", retrieved on 3 March 2014, from http://water.epa.gov/scitech/methods/cwa/bioindicators/upload/2008_01_03_methods_method_1694.pdf, U.S. Environmental Protection Agency, Washington, D. C., USA.
- USEPA (2011), "Biomonitoring: Perfluorochemicals (PFCs)", retrieved on 24 July 2013, from http://www.epa.gov/ace/ace3draft/draft_pdfs/ACE3PFCsReviewPackage3-02-11.pdf, U.S. Environmental Protection Agency, USA.
- Vandenberg, J. A., Ryan, M. C., Nuell, D. D. and Chu, A. (2005), "Field attenuation of mixing length and attenuation of nutrients and fecal coliform in a wastewater effluent plume", Environmental Monitoring and Assessment, Vol. 107, pp. 45-57.
- Vollertsen, J. and Hvitved-Jacobsen, T. (2003), "Exfiltration from gravity sewers: a pilot scale study", Water Science & Technology, Vol. 47, No. 4, pp. 69-76.
- Wade, T. J., Calderon, R. L., Sams, E., Beach, M., Brenner, K. P., Williams, A. H. and Dufour, A. P. (2006), "Rapidly measured indicators of recreational water quality are predictive of swimming-associated gastrointestinal illness", Environmental Health Perspectives, Vol. 114, No. 1, pp. 24.
- Wallis, J. L. and Taylor, H. D. (2003), "Phenotypic population characteristics of the enterococci in wastewater and animal faeces: implications for the New European Directive on the quality of bathing waters", Water Science & Technology, Vol. 47, No. 3, pp. 27-32.
- Walter, D. A. R., Brigid A.; Stollenwerk, Kenneth G.; Savoie, Jennifer (1996), "Geochemical and Hydrologic Controls on Phosphorus Transport in a Sewage-Contaminated Sand and Gravel Aquifer near Ashumet Pond, Cape Cod, Massachusetts", U.S. Department of the Interior, U.S. Geological Survey, Denver, CO, USA.
- Walters, S. P. and Field, K. G. (2009), "Survival and persistence of human and ruminant-specific faecal *Bacteroidales* in freshwater microcosms", Environmental Microbiology, Vol. 11, No. 6, pp. 1410-1421.
- Walters, S. P., Thebo, A. L. and Boehm, A. B. (2011), "Impact of urbanization and agriculture on the occurrence of bacterial pathogens and stx genes in coastal waterbodies of central California", Water Research, Vol. 45, No. 4, pp. 1752-1762.
- Wams, T. J. (1987), "Diethylhexylphthalate as an environmental contaminant -- A review", The Science of The Total Environment, Vol. 66, pp. 1-16.
- Whitlock, J. E., Jones, D. T. and Harwood, V. J. (2002), "Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis", Water Research, Vol. 36, No. 17, pp. 4273-4282.

- Whitman, R. L. and Nevers, M. B. (2004), "*Escherichia coli* sampling reliability at a frequently closed Chicago Beach: monitoring and management implications", Environmental Science & Technology, Vol. 38, No. 16, pp. 4241-4246.
- WHO (2003), Guidelines for safe recreational waters, World Health Organization, Geneva, Switzerland.
- Wiggins, B. A. (1996), "Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters", Applied and Environmental Microbiology, Vol. 62, No. 11, pp. 3997-4002.
- Wiggins, B. A., Andrews, R. W., Conway, R. A., Corr, C. L., Dobratz, E. J., Dougherty, D. P., Eppard, J. R., Knupp, S. R., Limjoco, M. C., Mettenburg, J. M., Rinehardt, J. M., Sonsino, J., Torrijos, R. L. and Zimmerman, M. E. (1999), "Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution", Applied and Environmental Microbiology, Vol. 65, No. 8, pp. 3483-3486.
- Wiggins, B. A., Cash, P. W., Creamer, W. S., Dart, S. E., Garcia, P. P., Gerecke, T. M., Han, J., Henry, B. L., Hoover, K. B., Johnson, E. L., Jones, K. C., McCarthy, J. G., McDonough, J. A., Mercer, S. A., Noto, M. J., Park, H., Phillips, M. S., Purner, S. M., Smith, B. M., Stevens, E. N. and Varner, A. K. (2003), "Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries", Applied and Environmental Microbiology, Vol. 69, No. 6, pp. 3399-3405.
- Wilkison, D. H., Armstrong, D. J. and Blevins, D. W. (2002), "Effects of Wastewater and Combined Sewer Overflows on Water Quality in the Blue River Basin, Kansas City, Missouri and Kansas, July 1998-October 2000", U.S. Department of the Interior, U.S. Geological Survey, Rolla, MO, USA.
- Wolf, L., Eiswirth, M. and Hötzl, H. (2006), "Assessing sewer-groundwater interaction at the city scale based on individual sewer defects and marker species distributions", Environmental Geology, Vol. 49, No. 6, pp. 849-857.
- Wong, T. Y., Foster, P. J., Johnson, G. J., Klein, B. E. K. and Seah, S. K. L. (2001), "The relationship between ocular dimensions and refraction with adult stature: the Tanjong Pagar survey", Investigative Ophthalmology & Visual Science, Vol. 42, No. 6, pp. 1237-1242.
- Wu, J., Yue, J., Hu, R., Yang, Z. and Zhang, L. (2008), "Use of caffeine and human pharmaceutical compounds to identify sewage contamination", World Academy of Science, Engineering and Technology Vol. 44, pp. 438-442.
- Wymer, L. J. and Wade, T. J. (2007), "The lognormal distribution and use of the geometric mean and the arithmetic mean in recreational water quality measurement", Statistical Framework for Recreational Water Quality Criteria and Monitoring, John Wiley & Sons Ltd, West Sussex, UK, pp. 91-112.
- Xu, Y., Luo, F., Pal, A., Gin, K. Y. H. and Reinhard, M. (2011), "Occurrence of emerging organic contaminants in a tropical urban catchment in Singapore", Chemosphere, Vol. 83, No. 7, pp. 963-969.
- Yakub, G. P., Castric, D. A., Stadterman-Knauer, K. L., Tobin, M. J., Blazina, M., Heineman, T. N., Yee, G. Y. and Frazier, L. (2002), "Evaluation of colilert and enterolert defined substrate methodology for wastewater applications", Water Environment Research, Vol. 74, No. 2, pp. 131-5.

- Yap, K. B. and Chan, K. M. (1998), "The prescribing pattern of hospital doctors", Singapore Medical Journal, Vol. 39, No. 11, pp. 496-500.
- Ying, G. G. (2004), "Distribution, Behavior, Fate, and Effects of Surfactants and Their Degradation Products in the Environment", Handbook of Detergents, Part B, U. Zoller, ed., CRC Press, pp. 77-109.
- Young, K. D. and Thackston, E. L. (1999), "Housing density and bacterial loading in urban streams", Journal of Environmental Engineering, Vol. 125, No. 12, pp. 1177-1180.
- YSI (2007), "YSI Model 30 and 30M Handheld Salinity, Conductivity and Temperature System: Operations Manual", retrieved on 15 December 2010, from <http://www.yei.com/media/pdfs/030136-YSI-Model-30-Operations-Manual-RevE.pdf>, YSI, Inc., Yellow Springs, OH, USA.
- Zheng, Y., Kelly, W. R., Panno, S. V. and Liu, W. T. (2013), "Identification of Sources of Fecal Pollution of Karst Waters", Illinois State Water Survey, Prairie Research Institute, University of Illinois at Urbana-Champaign. Champaign, IL, USA.

Appendix A. FIB and HF183 concentrations of raw sewage and non-urban water samples

Sample name	Sampling date and time	TC (MPN/100mL)	<i>E. coli</i> (MPN/100mL)	Enterococci (MPN/100mL)	HF183 (GE/100mL)
RS1	2 Feb 2012 13:30	8,664,000	3,873,000	414,000	31,401,150
RS2	7 Jun 2012 02:31	4,884,000	1,246,000	728,000	422,561,739
RS3	11 Jun 2012 23:55	10,462,000	3,255,000	1,198,000	27,483,941
RS4	12 Jun 2012 00:35	2,098,000	<u>241,960</u>	546,000	32,154,378
RS5	12 Jun 2012 02:10	24,900,000	1,211,000	393,000	3,888,545
PS1	28 Feb 2012 15:00	9,590	225	1,733	611
PS2	26 Mar 2012 15:50	4,960	126	10	1,245
PS3	3 Dec 2012 10:40	8,800	727	162	NA
PS4	3 Dec 2012 11:10	2,010	19	3	NA

RS denotes raw sewage samples and PS denotes non-urban water samples. Sampling locations of each sample in table above can be found in Table 4.2. PS3 and PS4 were also analysed for chemical parameters. The results are described in Section 6.1.2. Underlined concentration indicates an assigned value because the result is >241,960 MPN/100mL. NA denotes the samples were not collected for HF183 analysis because there was insufficient time for HF183 analysis.

Appendix B. Field blank samples

Field blank name	Collection date and time	TC (MPN/100mL)	<i>E. coli</i> (MPN/100mL)	Enterococci (MPN/100mL)
KC1B1	4 Jan 2011 08:20	<1	<1	<1
KC1B2	4 Jan 2011 12:06	2,420	<1	<1
KC1B3	4 Jan 2011 17:12	11	<1	<1
KV1B1	6 Jan 2011 08:15	2	<1	<1
KV1B2	6 Jan 2011 13:07	<1	<1	<1
MB1B1	10 Jan 2011 07:50	1	<1	<1
MB1B2	10 Jan 2011 13:10	<1	<1	<1
MB1B3	10 Jan 2011 16:30	3	3	<1
KV2B1	12 Jan 2011 07:53	1	<1	<1
KV2B2	12 Jan 2011 12:20	<1	<1	<1
KV2B3	12 Jan 2011 18:15	<1	<1	<1
MB2B1	18 Jan 2011 08:34	<1	<1	<1
MB2B2	18 Jan 2011 12:17	<1	<1	<1
KC2B1	19 Jan 2011 12:00	<1	<1	<1
MT1B1	7 Jun 2011 9:30	<1	<1	<1
MT1B2	7 Jun 2011 12:30	26	3	2
MT1B3	7 Jun 2011 19:15	44	<1	2
MS1B1	8 Jun 2011 07:50	<1	<1	<1
MS1B2	8 Jun 2011 12:30	<1	<1	<1
MS1B3	8 Jun 2011 16:30	<1	<1	<1
MB3B1	27 Jun 2011 07:50	<1	<1	<1
MB4B1	28 Jun 2011 07:50	2	<1	1
MB4B2	28 Jun 2011 14:30	42	<1	3
MB5B1	29 Jun 2011 07:50	<1	<1	<1
MB5B2	29 Jun 2011 13:30	<1	<1	<1
MT2B1	4 Jul 2011 08:30	<1	<1	<1
MT2B2	4 Jul 2011 12:30	<1	<1	<1
MS2B1	5 Jul 2011 07:50	1	<1	<1
MS2B2	5 Jul 2011 12:30	<1	<1	<1
MT3B1	6 Jul 2011 08:30	<1	<1	<1
MT3B2	6 Jul 2011 15:30	2	<1	<1
MS3B1	7 Jul 2011 08:30	2	<1	<1
MS3B2	7 Jul 2011 13:30	<1	1	<1
KC3B1	11 Jan 2012 17:30	<1	<1	<1
KC4B1	16 Jan 2012 17:30	3	1	<1
KV4B1	17 Jan 2012 15:30	2	<1	<1
KV5B1	8 May 2012 19:10	44	1	<1
TA1B1	10 May 2012 08:40	16	<1	<1
TA2B1	14 May 2012 08:45	1	<1	<1
KC5B1	21 May 2012 07:50	<1	<1	<1
KC5B2	21 May 2012 14:27	20	3	4
KC6B1	22 May 2012 08:31	<1	<1	<1
KC6B2	22 May 2012 09:30	96	2	<1
MP1B1	24 May 2012 08:40	<1	<1	<1

Field blank name	Collection date and time	TC (MPN/100mL)	<i>E. coli</i> (MPN/100mL)	Enterococci (MPN/100mL)
MP2B1	28 May 2012 08:30	<1	<1	<1
TA3B1	30 May 2012 08:40	<1	<1	<1
MP3B1	5 Jun 2012 08:40	3	<1	<1
WD1B1	6 Jun 2012 09:40	<1	<1	<1
WD2B1	11 Jun 2012 08:45	<1	<1	<1
WD3B1	13 Jun 2012 09:30	<1	<1	<1
MK1B1	20 Jun 2012 08:33	<1	<1	<1
KV6B1	20 Jun 2012 08:35	<1	<1	<1
MK2B1	27 Jun 2012 14:30	1	<1	<1
MK3B1	2 Jul 2012 08:30	<1	<1	<1
MK4B1	4 Jul 2012 14:20	1	<1	<1
SP2B1	9 Jul 2012 17:00	<1	<1	<1
SP3B1	11 Jul 2012 13:00	<1	<1	<1
SP4B1	16 Jul 2012 15:00	1	<1	<1
MK5B1	23 Jul 2012 15:20	<1	<1	<1
TL5B1	25 Jul 2012 14:15	<1	<1	<1
KC7B1	30 Jul 2012 15:30	1	<1	<1
KV7B1	31 Jul 2012 17:30	<1	<1	<1

Appendix C. List of publications or communications

- Ekklesia, E., Shanahan, P., Chua, L. H. C. and Eikaas, H. S. (submitted), "Associations of potential chemical tracers with faecal indicator bacteria in tropical urban catchments".
- Ekklesia, E., Shanahan, P., Chua, L. H. C. and Eikaas, H. S. (submitted), "Temporal variation of faecal indicator bacteria in tropical urban storm drains".
- Nshimiyimana, J. P., Ekklesia, E., Shanahan, P., Chua, L. H. C. and Thompson, J. R. (2014), "Distribution and abundance of human specific Bacteroides and relation to traditional indicators in an urban tropical catchment", Journal of Applied Microbiology, Vol. 116, No. 5, pp. 1369-1383.
- Ekklesia, E., Shanahan, P., Chua, L. H. C. and Eikaas, H. S. (2013), "Associations of Potential Alternative Tracers with Fecal Indicator Bacteria in Tropical Urban Catchments", 8th IWA Specialized Conference on Assessment and Control of Micropollutants and Hazardous Substances in Water: Micropol & Ecohazard, Zurich, Switzerland, June 16-20, 2013.
- Ekklesia, E., Shanahan, P. and Chua, L. H. C. (2012), "Fecal Indicator Bacteria Variability in Urban Drainages", 2012 Asia-Oceania Top University League on Engineering Student Conference, Kuala Lumpur, Malaysia, November 24-25, 2012.
- Ekklesia, E. (2013), "Associations of Potential Alternative Tracers with Fecal Indicator Bacteria in Tropical Urban Catchments", Center for Environmental Sensing and Modeling (CENSAM) Seminar Series, Singapore, August 7, 2013.
- Ekklesia, E. (2012), "Identification of Suitable Indicators for Tracing Human Fecal Contamination and Sewage in Urban Catchments", CENSAM Seminar Series, Singapore, February 22, 2012.
- Ekklesia, E. (2011), "Identification of Alternative Indicators for Human Faecal Contamination and Sewage in Singapore's Catchments", CENSAM Fourth Annual Workshop, Singapore, January 17-18, 2011.