Trend Analysis of Large Physical and Biological Data Sets, Related to Water Bodies Using R Statistical Programming.

Maria O'Neill B.Sc. (Hons)

Thesis submitted in fulfilment of the requirement for the

degree of

Master of Science (M.Sc.)

to

Dublin City University

July 2019

Faculty of Science and Health

School of Chemical Sciences

Supervisor:

Prof. Fiona Regan

School of Chemical Sciences

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Master of Science is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ (Candidate) ID No.: _____ Date: _____

Acknowledgements

A big thanks to everyone who worked with me during my masters. Especially to my supervisor Prof. Fiona Regan who was a great source of knowledge and guidance. Thanks to Dr. Eleanor Jennings, and Patricia Antunes for all their great work in our joint project. Thanks to Dr. Ciprian Briciu-Burghina, Dr. Gill Duffy, Dr. Lisa Jones, and Alan Barrett for all their help and support.

Thanks to Louise Spellacy who was a great housemate during my student life and always made sure I had a hot meal. Thanks especially to my parents who stood by me throughout.

Contents

List of Figures

List of Tables

List of Abbreviations

IoT	Internet of Things
GPRS	General Pocket Radio Service
MxS	Maximum Spearman
CCF	Cross-Correlation Function
MPN	Most Probable Number
KS-test	Kolmogorov-Smirnov test
DCC	Dublin City Council
CSV	Comma Separated Values
BN	Bellewstown
DA	Dublin Airport
GV	Glasnevin
MS	Merrion Square
CT	Casement
BF	Ballyedmonduff House
DL	Dun Laoghaire
NTU	Nephelometric Turbidity Units
m^3/s	Cubic Meter per Second

Abstract

Trend Analysis of Large Physical and Biological Data Sets, Related to Water Bodies Using R Statistical Programming.

Maria O'Neill

Due to the amount of monitoring undertaken in the last 10 years, as well as the increased use of real time data acquisition, many large data stets now exist. Trend analysis is used to collect information from large data sets to understand how the process works and enables us to make informed decisions about what to do when faced with similar situations in the future. There is a growing need for programming skills amongst scientist so that they can affectively process large data sets from sensors and historical monitoring.

This thesis deals with how sensor technology coupled with statistical analysis can provide an invaluable tool for monitoring water quality and managing this vital natural resource. Statistical analysis of the sensor data was performed to determine the optimum conditions for deployment, future monitoring, and determine what precautions, if any, are to be taken. R statistical programming was chosen as a method of analysis as it is capable of processing large complex time series data sets whilst making analysis easily reproducible by other scientists. This thesis examined trends found in historical data and real time data sets of physical and biological properties related to water quality.

The physical properties of water were examined in the Dodder catchment. A network of affordable high frequency ultrasonic water level gauges were deployed throughout the catchment.

The historical biological properties of the Grand Canal Basin was also examined. A bacterial level sensor capable of rapid detection was tested in the basin to determine its use for future monitoring.

Lastly the real time impact of cattle entering a stream was examined against real time turbidity data to examine trends due to monitored external forces. The impact of the cattle on the stream was examined by looking at the difference between the upstream and downstream sensor data.

Chapter 1: Introduction

1.1 Examining the importance of water monitoring.

1.1.1 Problems with our water

Water is a vital resource for humankind, it is used for drinking, bathing, recreation, agriculture, pharmaceuticals, and the food industry (Ridoutt, B., Sanguansri, P., Bonney, L., Crimp, S., Lewis, G., Lim-Camacho, 2013; Malcangio, 2018; Ncibi and Matilainen, 2018; Norlaila *et al.,* 2018; Takatsuka *et al.,* 2018). There has been a large increase in the number of people living in urban areas in recent times. This results in an increased strain on water resources in the city such as drinking water, sewage, and water used recreationally (Flörke, Schneider and Mcdonald, 2018). Urban storm water runoff is a huge contributor to flooding in urban areas which can cause expensive to damage to infrastructure and private homes and businesses (Hellman *et al.,* 2018). In the countryside run off and contamination from livestock is a huge contributing factor to water pollution (Givens *et al.,* 2016; Lefrancq *et al.,* 2017). Although changes made to water, infrastructure and treatment, can be costly to government organises and the public; it is hugely important that safe water supplies are provided for people to protect them from disease as well as the destruction of their property (Prüss-Üstün *et al.,* 2008; Aerts *et al.,* 2013).

1.1.2 Sources of pollution

Humans account for a large proportion of water contamination. Untreated discharge from residential and commercial properties, accidental contamination from spills or leakages, poor infrastructure and treatment practices (Chapman *et al.,* 1996; Singh *et al*., 2004; Messinger and Silman, 2016; Goovaerts, 2019). Natural sources of water contamination can be due to runoff during periods of heavy rainfall or from animals defecating and urinating directly into water bodies (Ribarova, Ninov and Cooper, 2008; Conroy *et al.,* 2016).

Any untreated waste water that enters a waterbody increases the risk to the health of those who bath in these waters (Ahmed *et al*., 2018; Krogh *et al.,* 2018). It also increases the chances of eutrophication which in turn decreases the amount of available oxygen in aquatic environments which can kill fish and other organisms in the water (Naumann *et al.,* 2015). It puts increased pressure on organisations to effectively treat water before it is used in industry and for human consumption (Teodosiu *et al.,* 2018).

There are many ways to approach harmful pollution events. This can be done by taking preventative measures, by monitoring for pollution events, and by determining the probability of future pollution occurrences (Cheng, 2018). These approaches can only be reliable if a sufficient data is obtained to determine the cause and effects of pollution events. Real-time event analysis is vital to determining the best action to take and is the most effective way to protect water bodies and the people who rely on them. Data acquisition through sensors is becoming greater and greater a part of environmental monitoring due to global movements such as the Internet of Things (IoT) (Atzori, Iera and Morabito, 2010; Wong and Kerkez, 2016).

1.1.3 What needs to be monitored in water?

It is important to monitor changing properties of water to ensure water quality. Physical properties such as river height and turbidity, and biological properties such as coliform and E. coli levels can all have an effect on the ecosystem of the water body and the safety of the people who use them (Lymer, Weinberg and Clausen, 2018). The EPA have created strict monitoring levels to indicate at what point the physical, chemical, and biological properties of water become a danger to human health (The Environmental and Protection Agency, 2001). The ideal monitoring system would have sensors deployed in a network at key locations, capable of long term deployment, and notify of breaches of water quality levels immediately (O'Flynn *et al.,* 2010).

1.2 Sensors as a monitoring method for water

1.2.1 Why sensors as a method?

Water quality measurements are usually taken by grab sampling and field measurements. This process is usually labour and time intensive, costly, and thus is usually performed infrequently. Samples are often required to be transported from site to a laboratory with results sometimes taking a few days to be determined especially for the determination of bacterial levels. Sensors provide a way of obtaining water quality results rapidly, remotely, and in real time. This type of monitoring can be useful to providing early warnings for contamination and flooding events, as well as providing an abundance of data points for historical monitoring (O'Flynn *et al.,* 2010). The advantages and disadvantages of spot sampling verses sensors is examined in Table 1.1.

1.2.2 Types of water quality sensors on the market today.

A wide range of sensors are available on the market today. Sensors can be purchased with single or multi parameters. They can be capable of long term deployment or single use. Parameters measured include water level, flow rate, pH, dissolved oxygen, oxidation-reduction potential, oil in water, conductivity, turbidity, temperature, chlorophyll-A, blue-green algae, and dissolved ions such as fluoride, calcium, nitrate and chloride (Raich, 2013).

Table 1.1: Comparison of grab and sensing approaches to sampling.

Parameter	Advantages	Disadvantages
Grab Sampling	• Analysis is carried out in a sterilized laboratory environment. • Laboratories can test for a wide array of parameters. • Instrumentation used is usually has a good sensitivity.	• A large delay between sampling and analysis can mean results are not true to the water bodies' current condition. • It is labor intensive and therefore is usually performed sparsely.
Sensor Sampling	• Provides real time results. • Can use GPS signaling to provide information of immediate breaches.	• On site analysis in water bodies can lead to biofouling and decreased sensitivity of the sensor. • Increasing the frequency of measurements increases the demand on the available power supply.

Sensors can be used for once off monitoring, long-term monitoring, and in a network. The greatest advantage of sensors is the speed at which they obtain results. The type and placement of sensor is dependent on what is being monitored and where. Single use sensors cannot be deployed for long periods of time, but they greatly reduce the labour involved from grab samples meaning it can be performed a lot more frequently. As well as this, they allow for rapid results where as grab samples may take days for results to be obtained when hours for transportation and analysis is accounted for. Obtaining timely results is a highly valued property for water monitoring as it can greatly reduce mitigation time (Gonzalez, Greenwood and Quevauviller, 2009). Sensors intended for long term deployment measure physical and optical properties of water such as pH, temperature, water level, dissolved oxygen, and turbidity. These sensors can be deployed by themselves or as a network across a catchment. They provide results in real time meaning breaches of threshold values are recorded immediately. Real time results also allow for data to be examined with relation to seasonal and yearly trends. They also allow for correlation with other environmental data sources measured in real time such as rainfall. Long term deployed sensors can also safely record data in adverse weather conditions such as storms which may be too dangerous for manual grab samples to be taken. A sensor network provides valuable information on how different parts of a catchment reacts to the same conditions at the same time. Sensors connected to the internet can immediately inform the user when and where a trigger level has been exceeded allowing for the most amount of time possible for the user to examine and determine the best course of action to take (Hart and Martinez, 2006).

There is much research into improving sensor technology and creating novel sensors in recent years, such is the demand (Han *et al.,* 2017; Promchat, Rashatasakhon and Sukwattanasinitt, 2017; Gayathri, Selvan and Sangilimuthu, 2018; Quevedo Casín *et al.,* 2018). Sensors are being improved to increase the battery life and some are being developed to reduce the need for cleaning by investigating ways to reduce biofouling (Derek, Chang and Dickey, 2004). Sensors are also being constantly redeveloped to reduce their cost which in turn will make them more affordable for network deployment (Cheng *et al.,* 2018).

1.2.3 Cleaning and maintenance

The sensitivity of sensors depletes when they are left in the water for long periods of time due to biofouling, an example of this can be seen in Figure 1.1, (O'Flynn *et al.,* 2010). The extent of loss of sensitivity can be determined by noting the measured level of all the sensor parameters, under the same water conditions, immediately before and then immediately after cleaning. Then the following algorithm can be applied to correct for any loss of sensitivity:

$$V_C = V + (V_f - V_f)\left(\frac{T_t - T}{T_t}\right)$$

V_c is the drift corrected value, V is the original measured value, V_f is the response of the sensor immediately before cleaning and validation at the end of the correction interval; V_s is the response of the sensor after cleaning and calibration; T_s is the total time interval for which the correction is applied and T is the time between the end of deployment and the time when the value is measured (Wagner *et al.*, 2006).

Figure 1.1: A sensor left out in salt water which has examples of biofouling around the body and sensor nodes.

The frequency of cleaning to be performed varies with sensor type and with water body type. This is because biofouling affects different materials at different rates. The pH value, temperature, metals present, nutrients, oxygen, organics, velocity and turbulence of the water can also have an effect. Biofouling has also been seen to occur at different rates during different seasons (Kruse, 2018).

1.3 How scientists examine large data sets

1.3.1 How do we get a large data set?

Data sets in scientific studies are becoming larger (Chen and Zhang, 2014). Sensors are an invaluable source for data acquisition. With sensors data can be obtained with more ease, with greater speed, and more frequently than traditional grab sampling techniques. The increase of efficiency of data collection has resulted in a great deal more data being collected. A sensor set to take a reading every 15 min will create, 4 data points in an hour, 96 data points in a day, 672 data points in a week, 2,920 data points in a month, and 35,040 data points in a year.

1.3.2 The need for programming skills amongst scientist

Environmental scientists are frequently required to use data sets from many different sources such as biological, physical, geological, climatic, and chemical. However, there is rarely formal computer programming training provided for environmental scientists (Lowndes *et al.,* 2017).

Programming skills benefit scientists as coding languages can effectively deal with larger, more complex data sets. They also benefit the scientific community as programming scripts are much more easily reproduced. Following a clear system of cleaning and coding data and providing programming scripts can allow other scientists to reproduce the same results with ease. It also allows the scientific community to see the entire method of analysis.

1.3.4 R statistical programming as a tool for analysis of sensor data

R is a programming language that is free to source and is used primarily for statistical analysis and graphical representation of time series data (R Core Team, 2016). It has a large amount of free software packages that can be downloaded and run on the interface to deal with the individual needs for the programmer. Free software that is

open and available to source by all scientists makes it much easier for research to be reproduced and improves conditions for collaboration. R comes with a user-friendly interface called R Studio (Figure 1.2) which allows users to see their code, packages, raw data, and results all in the same space (RStudio Team, 2016).

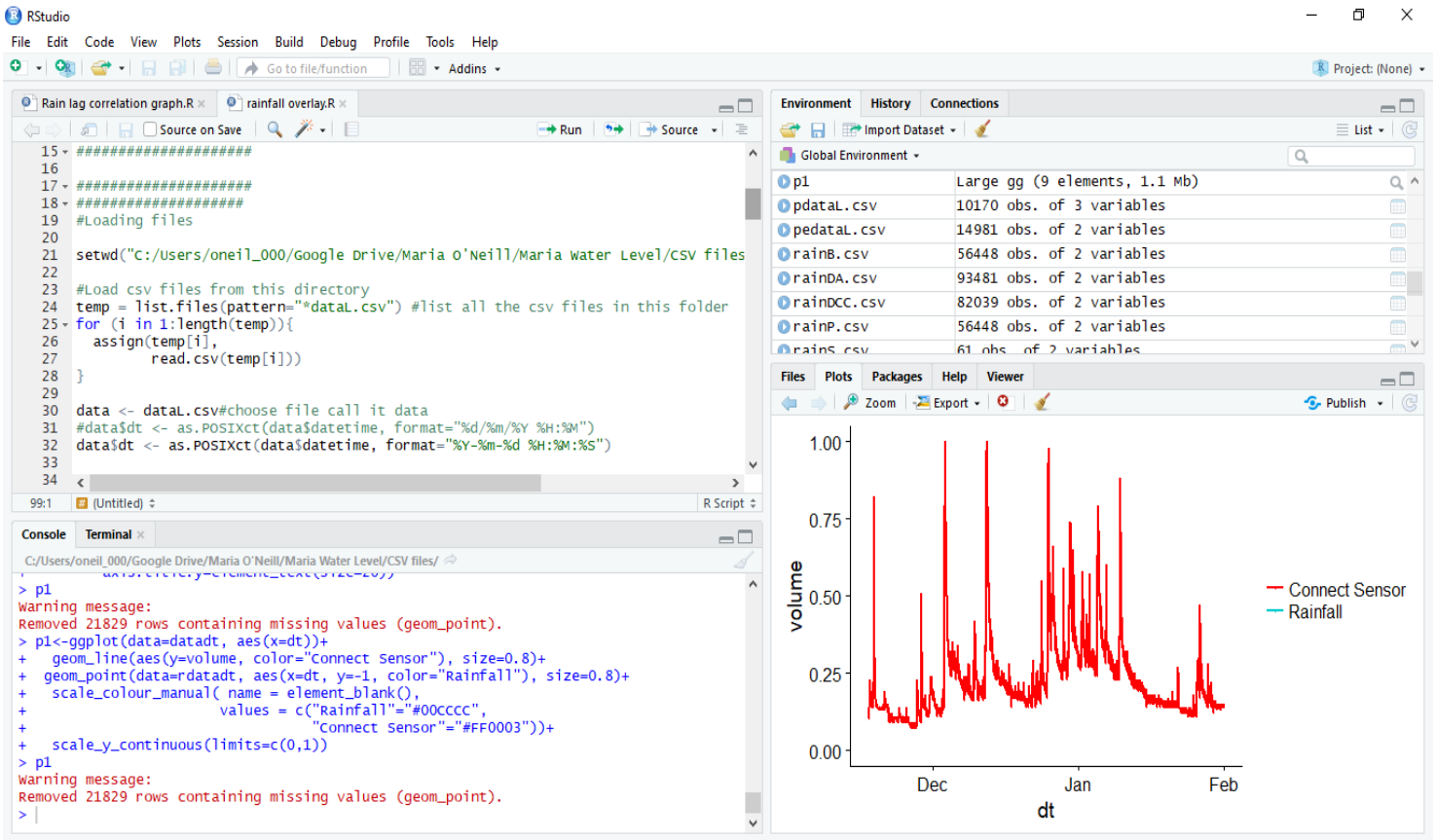

Figure 1.2: The R Studio interface which allows users to see their code, packages, raw data, and results all in the same space (RStudio Team, 2016).

As R can deal with large amounts of homogenous data it is ideal for sensor data which may have multiple parameters recorded for each reading. It is also capable for merging and combining different time series data from different sources which is vital for trend and correlation analysis.

1.4 Summary of the need for water monitoring and analysis by means of R statistical programming

Water is an invaluable resource for humans and it is therefore vital that it is monitored continually so it can be understood better and so that any harmful pollution or flooding events can be rectified in the quickest time possible. Sensors provide a valuable source for data acquisition of environmental properties of water. Their ability to provide rapid real time results allows for a large amount of data to be collected with relative ease. As data sets are increasing in size there is a greater need for the use of programming software and a greater need for programming to be taught to environmental scientists. R statistical programming is a free to source programming language which deals very well with large time series data sets. Environmental scientist can use R programming as an effective tool for analysing complex homogeneous sensor data which is also easily reproducible.

1.5 Outline of the areas discussed in this thesis

In this thesis 3 different areas and approaches to water analysis were examined. This was done to obtain a wide scope for the potential of trend analysis of large data sets related to water bodies. Realtime physical data analysis was examined in the Dodder catchment using the Sonic Signalman. The benefits of network deployment of sensors were also examined in this study. The benefits of historical monitoring of biological data was examined in the Grand Canal Basin study. Use of a rapid single use detection sensor, the Colisense, was also examined in this study. By studying cattle access, in the Dunleer project, the direct impact of an external body entering a water body was examined. The physical change of water for specific events in a specific location captured by real time sensors, the YSI Sonde, was studied here. R programming was used for the statistical analysis under various methods such as Pearson, Spearman correlation, and Kolmogorov-Smirnov test.

References

Aerts, J. C. J. H. *et al.* (2013) 'Cost estimates for flood resilience and protection strategies in New York City', *Annals of the New York Academy of Sciences*, 1294(1), pp. 1–104. doi: 10.1111/nyas.12200.

Ahmed, W. *et al.* (2018) 'Quantitative microbial risk assessment of microbial source tracking markers in recreational water contaminated with fresh untreated and secondary treated sewage', *Environment International*. Elsevier, 117(April), pp. 243–249. doi: 10.1016/j.envint.2018.05.012.

Atzori, L., Iera, A. and Morabito, G. (2010) 'The Internet of Things : A survey', *Computer Networks*. Elsevier B.V., 54(15), pp. 2787–2805. doi: 10.1016/j.comnet.2010.05.010.

Chapman, P. M. *et al.* (1996) 'A Triad Study of Sediment Quality Associated with a Major , Relatively Untreated Marine Sewage Discharge', *Marine Pollution Bulletin*, 32(1), pp. 47–64.

Chen, C. L. P. and Zhang, C.-Y. (2014) 'Data-intensive applications, challenges, techniques and technologies: A survey on Big Data', *Information Sciences*, 275, pp. 314–347. doi: https://doi.org/10.1016/j.ins.2014.01.015.

Cheng, L. *et al.* (2018) 'Towards minimum-delay and energy-efficient flooding in low-duty-cycle wireless sensor networks', *Computer Networks*, 134, pp. 66–77. doi: 10.1016/j.comnet.2018.01.012.

Cheng, T. (2018) 'Risk scenario prediction for sudden water pollution accidents based on Bayesian networks', *International Journal of System Assurance Engineering and Management*. Springer India, 9(5), pp. 1165–1177. doi: 10.1007/s13198-018-0724-y.

Conroy, E. *et al.* (2016) 'The impact of cattle access on ecological water quality in streams: Examples from agricultural catchments within Ireland', *Science of the Total Environment*. Elsevier B.V., 547, pp. 17–29. doi: 10.1016/j.scitotenv.2015.12.120.

Derek, V. M., Chang, G. C. and Dickey, T. D. (2004) 'Methods for Reducing Biofouling of Moored Optical Sensors', *Journal of Atmospheric and Oceanic Technology*, 21, pp. 958–968.

Flörke, M., Schneider, C. and Mcdonald, R. I. (2018) ‘Water competition between cities and agriculture driven by climate change and urban growth’, *Nature Sustainability*. Springer US, 1(1), p. 51. doi: 10.1038/s41893-017-0006-8.

Gayathri, J., Selvan, K. S. and Sangilimuthu, S. N. (2018) ‘A novel sensor for the determination of Hg 2 + in waters based on octadentate ligand immobilized multi-walled carbon nanotube attached to paraffin wax impregnated graphite electrodes (PIGE)’, *Journal of Solid State Electrochemistry*. Journal of Solid State Electrochemistry, 22, pp. 2879–2888.

Givens, C. E. *et al.* (2016) ‘Detection of hepatitis E virus and other livestock-related pathogens in Iowa streams’, *Science of the Total Environment, The*. Elsevier B.V., 566–567, pp. 1042–1051. doi: 10.1016/j.scitotenv.2016.05.123.

Gonzalez, C., Greenwood, R. and Quevauviller, P. (2009) *Rapid chemical and biological techniques for water monitoring*. John Wiley & Sons.

Goovaerts, P. (2019) ‘Geostatistical prediction of water lead levels in Flint , Michigan : A multivariate approach’, *Science of the Total Environment*. Elsevier B.V., 647, pp. 1294–1304. doi: 10.1016/j.scitotenv.2018.07.459.

Han, C. *et al.* (2017) ‘A novel hybrid sensor for combined imaging of dissolved oxygen and labile phosphorus flux in sediment and water’, *Water Research*, 108, pp. 179–188. doi: 10.1016/j.watres.2016.10.075.

Hart, J. K. and Martinez, K. (2006) ‘Environmental Sensor Networks : A revolution in the earth system science ?’, *Earth-Science Reviews*, 78, pp. 177–191. doi: 10.1016/j.earscirev.2006.05.001.

Hellman, K. *et al.* (2018) ‘Estimating the Economic Impact of Stormwater Runo ff in the Allen Creek Watershed’, *Ecological Economics*, 145, pp. 420–429. doi: 10.1016/j.ecolecon.2017.11.022.

Krogh, J. *et al.* (2018) ‘Risks of hypoxia and acidification in the high energy coastal environment near Victoria, Canada’s untreated municipal sewage outfalls’, *Marine Pollution Bulletin*. Elsevier, 133(August 2017), pp. 517–531. doi: 10.1016/j.marpolbul.2018.05.018.

Kruse, P. (2018) ‘Review on water quality sensors’, *Journal of Physics D: Applied Physics*. IOP Publishing, 51(20), p. 203002

Lefrancq, M. *et al.* (2017) ‘High frequency monitoring of pesticides in runoff water to improve understanding of their transport and environmental impacts’, *Science of the Total Environment*. Elsevier B.V., 587–588, pp. 75–86. doi: 10.1016/j.scitotenv.2017.02.022.

Lowndes, J. S. S. *et al.* (2017) ‘Our path to better science in less time using open data science tools’, *Nature Ecology & Evolution*, 1(6), p. 0160. doi: 10.1038/s41559-017-0160.

Lymer, B. L., Weinberg, J. and Clausen, T. J. (2018) ‘Water quality management from source to sea : from global commitments to coordinated implementation’, *Water International*. Routledge, 43(3), pp. 349–360. doi: 10.1080/02508060.2018.1433782.

Malcangio, D. (2018) ‘Statistical Analysis of Bathing Water Quality in Puglia Region (Italy)’, *International journal of environmental research and public health*, 15(5), p. 1010. doi: 10.3390/ijerph15051010.

Messinger, M. and Silman, M. (2016) ‘Unmanned aerial vehicles for the assessment and monitoring of environmental contamination : An example from coal ash spills *’, *Environmental Pollution*. Elsevier Ltd, 218, pp. 889–894. doi: 10.1016/j.envpol.2016.08.019.

Naumann, M. S. *et al.* (2015) ‘Monitoring of coastal coral reefs near Dahab (Gulf of Aqaba, Red Sea) indicates local eutrophication as potential cause for change in benthic communities’, *Environmental monitoring and assessment*, 2(187), p. 44. doi: 10.1007/s10661-014-4257-9.

Ncibi, M. C. and Matilainen, A. (2018) ‘Removal of natural organic matter in drinking water treatment by coagulation : A comprehensive review’, *Chemosphere*, 190, pp. 54–71. doi: 10.1016/j.chemosphere.2017.09.113.

Norlaila, S. *et al.* (2018) ‘Value-added Waste Potential of Wastewater Sludge from Pharmaceutical Industry : A Review’, *CHEMICAL ENGINEERING TRANSACTIONS*, 63, pp. 493–498. doi: 10.3303/CET1863083.

O'Flynn, B. *et al.* (2010) 'Experiences and recommendations in deploying a real-time, water quality monitoring system', *Measurement Science and Technology*, 21(12). doi: 10.1088/0957-0233/21/12/124004.

Promchat, A., Rashatasakhon, P. and Sukwattanasinitt, M. (2017) 'A novel indolium salt as a highly sensitive and selective fluorescent sensor for cyanide detection in water', *Journal of Hazardous Materials*. Elsevier B.V., 329, pp. 255–261. doi: 10.1016/j.jhazmat.2017.01.024.

Prüss-Üstün, A. *et al.* (2008) *Safer water, better health: costs, benefits and sustainability of interventions to protect and promote health*. World Health Organization.

Quevedo Casín, J. J. *et al.* (2018) *An improved tool of water data analytics for flowmeters data*, *CySWter 2018: 4th International Workshop on Cyber-Physical Systems for Smart Water Networks: Porto, Portugal, April 10-13: proceedings book50-253 Porto, Portugal*.

R Core Team (2016) 'R: A Language and Environment for Statistical Computing'. Vienna, Austria. Available at: https://www.r-project.org/.

Raich, J. (2013) 'Review of sensors to monitor water quality', *European reference network for critical infrastructure protection (ERNCIP) project*. Citeseer.

Ribarova, I., Ninov, P. and Cooper, D. (2008) 'Modeling nutrient pollution during a first flood event using HSPF software: Iskar River case study, Bulgaria', *Ecological Modelling*, 211(1–2), pp. 241–246. doi: 10.1016/j.ecolmodel.2007.09.022.

Ridoutt, B., Sanguansri, P., Bonney, L., Crimp, S., Lewis, G., Lim-Camacho, L. (2013) 'Climate Change Adaptation Strategy in the Food Industry — Insights from Product Carbon and Water Footprints', *Climate Change*, 4(2), p. 26. doi: 10.3390/cli4020026.

RStudio Team (2016) 'RStudio: Integrated Development Environment for R'. Boston, MA. Available at: http://www.rstudio.com/.

Singh, K. P. *et al.* (2004) 'Impact assessment of treated / untreated wastewater toxicants discharged by sewage treatment plants on health , agricultural , and environmental quality in the wastewater disposal area', *Chemosphere*, 55(2), pp. 227–255. doi: 10.1016/j.chemosphere.2003.10.050.

Takatsuka, Y. *et al.* (2018) ‘Science of the Total Environment Value of irrigation water usage in South Florida agriculture’, *Science of the Total Environment*. Elsevier B.V., 626, pp. 486–496. doi: 10.1016/j.scitotenv.2017.12.240.

Teodosiu, C. *et al.* (2018) ‘Emerging pollutants removal through advanced drinking water treatment : A review on processes and environmental performances assessment’, *Journal of Cleaner Production*. Elsevier Ltd, 197, pp. 1210–1221. doi: 10.1016/j.jclepro.2018.06.247.

The Environmental and Protection Agency (2001) ‘Parameters of water quality’, *Environmental Protection*, p. 133. doi: 10.1017/CBO9781107415324.004.

Wagner, R. J. *et al.* (2006) *Guidelines and standard procedures for continuous water-quality monitors: station operation, record computation, and data reporting*.

Wong, B. P. and Kerkez, B. (2016) ‘Real-time environmental sensor data : An application to water quality using web services’, *Environmental Modelling and Software*. Elsevier Ltd, 84, pp. 505–517. doi: 10.1016/j.envsoft.2016.07.020.

Aims and objectives of this thesis

The aim of this research is to use R statistical programming as an effective tool for trend analysis of sensor and historical data to determine the optimum conditions for deployment, future monitoring, and determine what precautions, if any, are to be taken. R statistical programming was chosen as a method of analysis as it is capable of dealing with large data sets whilst making analysis easily reproducible by other scientists. A large increase in data acquisition due to the implementation of sensors and the availability of historical data means there is a need for programming literacy for scientists so they can examine large, homogenous data effectively. The trend analysis performed in this research will be used to determine the best practice for monitoring and management of water bodies in the future.

The objectives of this research are:

A. Determine the benefit of deployment of a network of affordable river level sensors in a catchment.
B. Determine the benefit of using trend analysis of historical monitoring of bacterial concentration data to decide the best practice for future monitoring and the benefits of using an instrument capable of rapid results for this analysis.
C. Determine the benefit of using sensor turbidity data as a method of gaining real time high frequency results to establish the impact of cattle entering a stream.

Chapter 2: Affordable Water Level Monitoring Sensors for Network Deployment

Abstract

Flooding has and can cause property damage, injury to people, and loss of life. The use of real-time water level monitoring sensors with automated data collection; can reduce mitigation time by allowing for earlier evaluation of data. Most current systems in use are expensive and therefore are usually deployed sparsely throughout a catchment. By using technology that has been currently employed by the Sonic SignalMan, sold by Kingspan, a water level sensor that is both economical and effective was developed and tested. The system has a stable platform capable of supporting a network of sensors for deployment throughout a catchment. It allows for personalized threshold warnings which can be used to aid in the risk assessment of current water levels. For this study four sensors were deployed in areas where existing water level monitoring sensors (provided by Dublin City Council) are deployed. This was done to validate the sensors against sensors that are already in use. To study the behaviour of water in catchment 8 sensors were deployed in the River Dodder catchment. The system can detect anomalies and notify operators. A highly specialised sensor network can provide a more localized water level monitoring system where different sections of the catchment are treated individually and compared.

2.1 Introduction

2.1.1 Flooding challenges

Increased frequency of storms and storm surges in Europe has significantly increased the risk of flooding (Christensen and Christensen, 2003). Flooding is a global problem which causes property damage, injury to people, and loss of life. These events have been documented worldwide. Some examples of which have occurred in North West Europe where the UK was subject to coastal flood events on the 31st of January to the 1st of February 1953 and 5th to the 6th of December 2013 (Wadey *et al.,* 2015), in America where over 1,100 lives were lost during Hurricane Katrina on the 23rd to the 31st of August 2005 (Jonkman *et al.,* 2009), and in Bangladesh in 1988 and 1998 where major structural damage was during their monsoon season (Faisal, Kabir and Nishat, 1999).

While some have advocated for the adaptation and improvement of infrastructure to combat flooding, these methods do not take immediate effect, require years to benefit from the endured cost (Arnbjerg-Nielsen and Fleischer, 2009). This is because flooding is unpredictable, and many urban areas go years without a flood occurring meaning the cost of the adaptions do not immediately outweigh the cost of repairing damage by flooding. Structural flood protection construction is for the most part only implemented when rebuilding after extensive flood damage has occurred (Ridoutt, B., Sanguansri, P., Bonney, L., Crimp, S., Lewis, G., Lim-Camacho, 2013), as this highlights the buildings that are most vulnerable to flooding and therefore would benefit the most from renovation and as in can be incorporated easily into the construction that is used to rebuild.

2.1.2 Real time data collection as a method of flood monitoring

This project studies real-time water level sensors for immediate risk management. When effectively used a real time data sensor with automated data collection reduces mitigation time by allowing for earlier evaluation of data. In addition to this when historical real time data is studied it can be used to identified thresholds of flooding that then can be used to aid in the risk assessment of current water levels (Carsell, Pingel and Ford, 2004). Thresholds are used as an indication of the potential of a flood

occurring. If water level exceeds a threshold an alarm can be sent to the public and emergency services to give them some time to prepare for the possible flood event.

2.1.3 Benefits of thresholds for water level monitoring

Water level monitoring is more efficient when sensors are used for querying events rather than monitoring long term processes. By setting thresholds the data analyser can be notified when the water level is high and the likelihood of an event happening is increased. In threshold systems a river height downstream is chosen as the height at which an event is likely to occur upstream sometime after. This height is chosen by examining historical data and determining the length of time it takes for an event to happen at the site or where a network has been deployed it can determine the time it takes for an event to occur upstream after this height is breached downstream. When the water level is low river water level changes are not considered to be a large threat to a flooding event. The movement of the river is better understood when large events are occurring. The current market requires a smart interface where relevant results can be displayed and understood by all people (He *et al.,* 2014). Allowing the public to have their own access to water level monitoring data means that they can make their own decisions as to when and how to respond to a potential flood.

2.1.4 The Sonic Signalman

This project tests the performance of The Sonic Signalman as a real-time water level monitoring system transmitting data through General Pocket Radio Service (GPRS). GPRS wireless transmission systems have been used to obtain real time data from sources on rivers relatively successfully in other studies (Gervais-Ducouret, 2011; Keoduangsine and Goodwin, 2012). The technology for water level monitoring is based on ultrasonic detection and has previously been used for oil level detection in storage tanks. An ultrasonic signal is sent from the transducer and travels through the pipe where it reaches the river surface. The time for the signal to be reflected back to the transducer is then converted to length by the transducer. This distance is then converted to river height by the sensor by subtracting it from the length of the pipe.

The Sonic Signalman connects to a web interface (www.connectsensor.com) where the data from each node can be displayed on graphs as well as in the raw format for each individual customer.

A network of sensors was deployed throughout The Dodder Catchment in Co. Dublin, Ireland. Deploying many sensors, throughout a catchment, provides ubiquitous systems which are used to obtain localised data for water level in real time. This information is especially valuable for flash flooding, which can appear rapidly after sudden heavy rainfall, in providing immediate information to people, in geographically relevant locations, warnings through the use of SMS/email notification (Bonnet, Gehrke and Seshadri, 2000; Gourbesville *et al.,* 2012).

2.2 Methods

2.2.1 Catchment

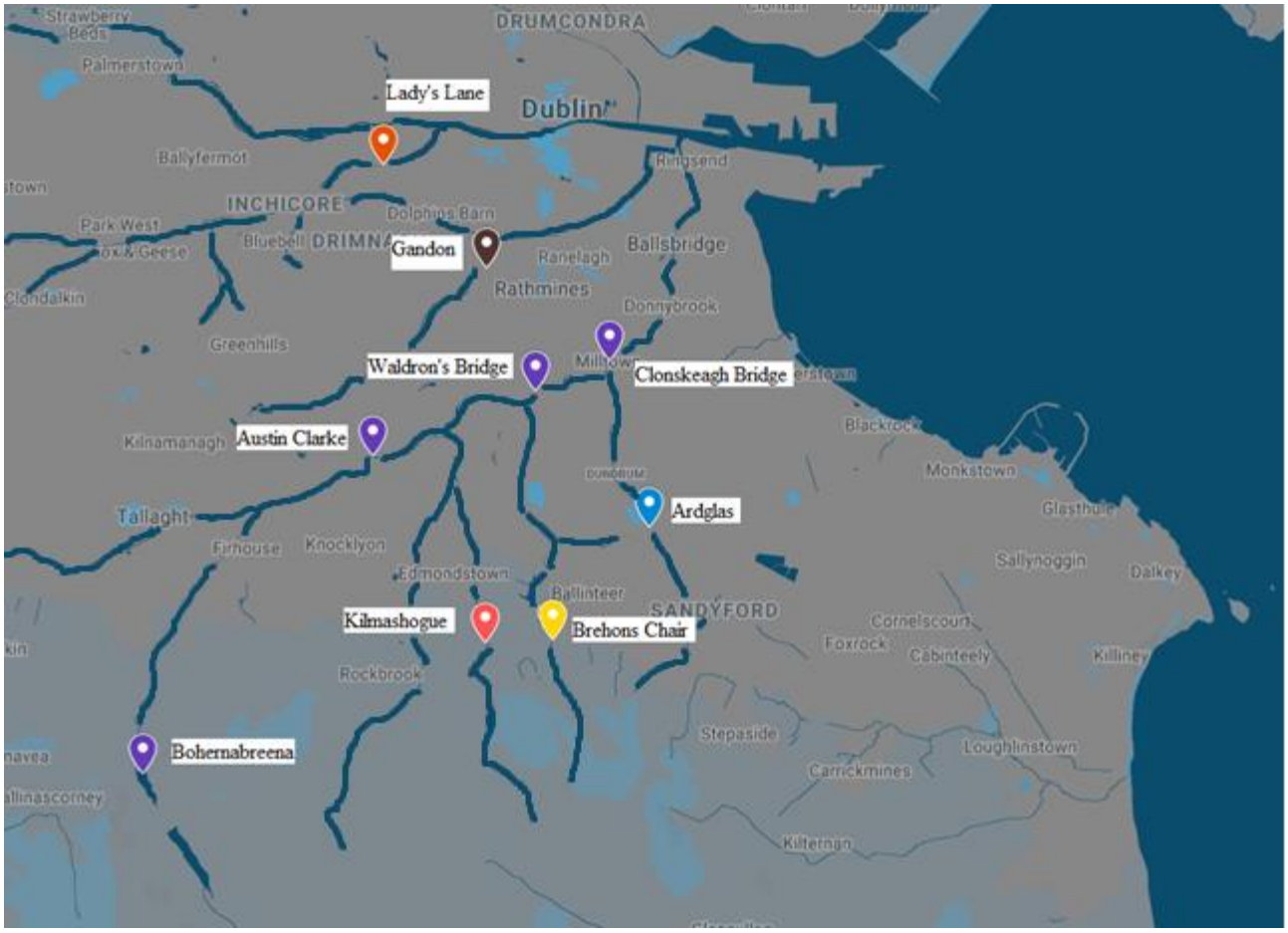

Figure 2.1: The locations of each sensor deployed in the Dodder catchment. A total of 9 sensors were deployed across the catchment successfully. Sites located on the Dodder are represented in purple. Sites on the Camac are in orange. Sites located on the Poddle are in brown. Sites located on the Little Dargle are in yellow. Sites located on the Whitechurch Stream are in pink.

The River Dodder (Figure 2.1) begins its life in the Kippure Mountain. It is the largest tributary of the River Liffey. The entire catchment occupies 12,081 hectares. In the early stages of its life it passes through bog-land, sparse forest, and agricultural land before it reaches urban Dublin. At Bohernabreena, in the upper stage of the river, there are two reservoirs that supply water to the southern part of Dublin city. The river has five main tributaries- Dundrum Slang, Whitechurch, Little Dargle, Owendoher, and Tallaght. All these tributaries flow through prominently urban areas. It has experienced extreme flooding events in the past. Most notably in 1986 when Hurricane Charley brought severe rainfall across Ireland (De Bruijn and Brandsma, 2000). Two locations are not strictly part of the Dodder catchment as they are just being used in locations the Dublin City Council (DCC) has reference sensors. These are the sensors at Lady's Lane and at Gandon.

Ten sensors were deployed in this study (Table 2.1). Four are located along the River Dodder. One was deployed on Whitechurch Stream, one was deployed on the Dundrum Slang, and another on the Little Dargle tributary. Two more were deployed in locations next to existing water level monitors on the Cammac and Poddle rivers. To compare the performance, Pearson's Correlation is used to measure the linear correlation between Kingspan sensor and DCC station readings.

Table 2.1: Sensor location and status

Site	Site Name	Dates Active	Reference Sensor	River
1	LADY'S LANE	24/07/2015 to Present	Yes	Camac
2	BOHERNABREENA	24/07/2015 to Present	Yes	Dodder
3	GANDON	24/07/2015 to Present	Yes	Poddle
4	WALDRON'S BRIDGE	24/07/2015 to Present	Yes	Dodder
5	CLONSKEAGH BRIDGE	24/07/2015 to Present	No	Dodder
6	BREHONS CHAIR	26/08/2015 to Present	No	Little Dargle
7	KILMASHOGUE	26/08/2015 to Present	No	Whitechurch Stream
8	AUSTIN CLARKE	26/08/2015 to Present	No	Dodder
9	ARDGLAS	26/08/2015 to Present	No	Dundrum Slang

2.2.2 Instrumentation

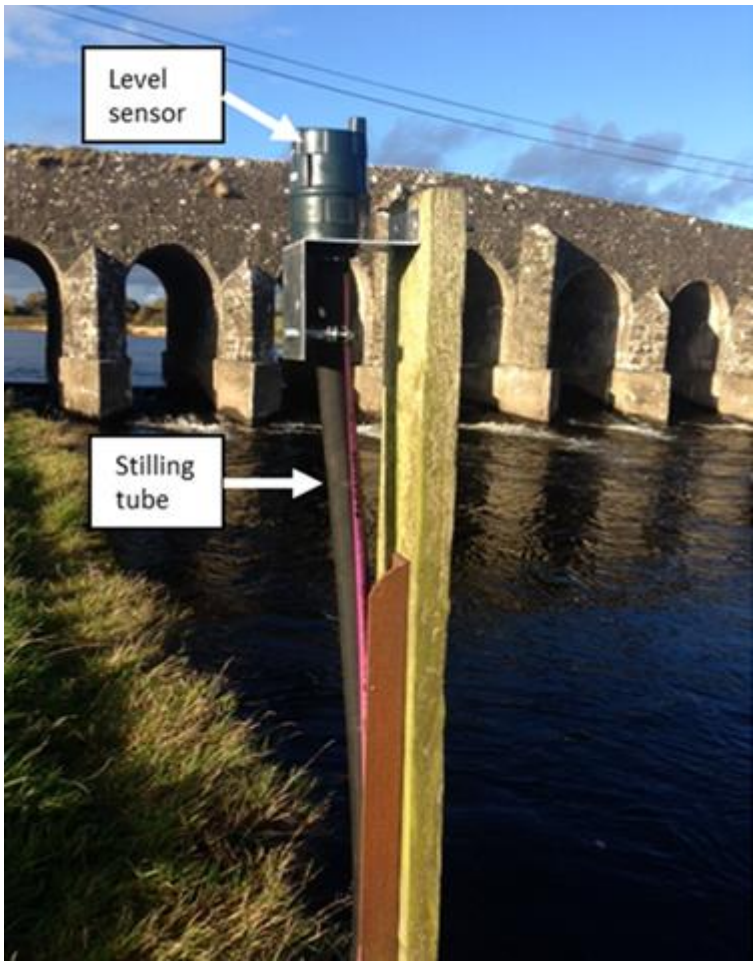

Figure 2.2: A deployed Connect Sensor displaying the ultrasonic transducer which calculates the level at the top of a stilling tube. The sensor is secured to a piece of rigid timber which was lodged in the riverbed.

In Figure 2.2 the sensor that consisted of an ultrasonic transducer, positioned at the top of a stilling tube and was deployed in the river by being secured to a piece of timber that which was lodged in the riverbed. The length of the plastic pipe must be taller than the expected rise of river water. The bottom of the plastic piping was placed in the river 3 centimetres from the river bed enabling entry of the river water. When the river water rose, the water is the pipe rose with it.

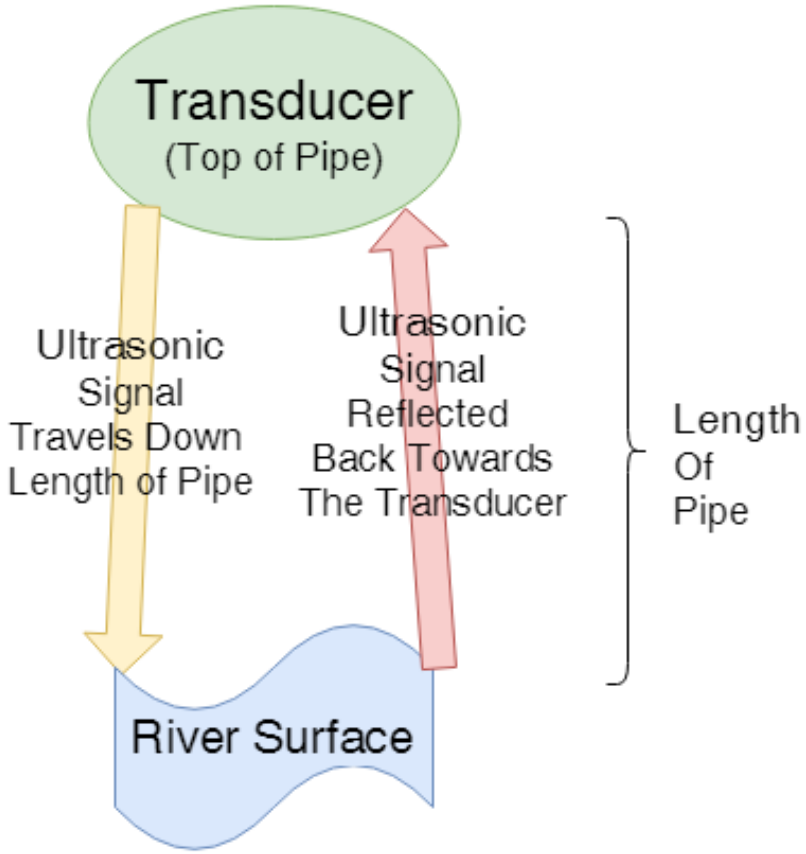

Figure 2.3: Schematic of the connect sensor. An ultrasonic signal is sent from the transducer and travels through the pipe where it reaches the river surface. The time for the signal to be reflected back to the transducer is then converted to length by the transducer. This distance is then converted to river height by the sensor by subtracting it from the length of the pipe.

In Figure 2.3 an ultrasonic signal was sent from the transducer at the top of the pipe, travelled down the pipe, till it reached the water and was refracted as an echo signal. This echo signal was received by the transducer sensor. The time of flight and speed of the transmitted ultrasonic signal was used by the sensor to calculate the distance of the water from the sensor. This distance is then converted to river height by the sensor by subtracting it from the length of the pipe.

The reference sensors provided by DCC for calibration consisted of an Isodaq 'The Frog RX' Telemetry Data Logger with a Loop powered 4-20 mA submersible level sensor, at Gandon Close and Lady's Lane, and a Siemens Logosense at Waldron's Bridge and Bohernabreena. The Frog RX was used with an impress depth level pressure sensor. The Siemens Logosense was also a pressure-based sensor.

2.3.3 Data and statistical analysis

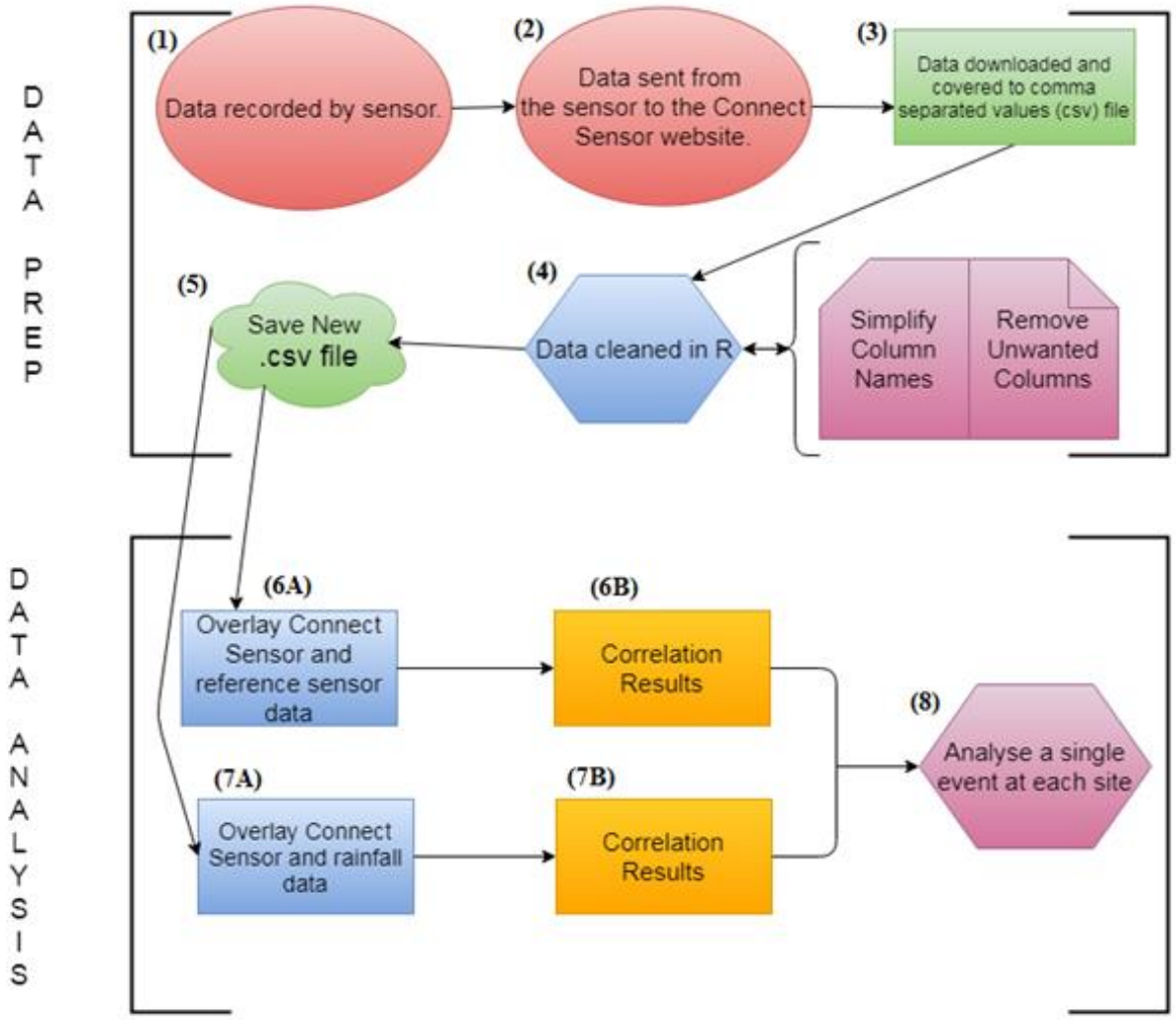

Figure 2.4: Experimental schematic outlining the order analysis was taken from data collection and preparation to data analysis.

Figure 2.4 outlines the general order for all data preparation and analysis that took place in this project. (1) The date is first recorded by the sensor. (2) The data is sent wirelessly by GRPS signal to the Connect Sensor Website. (3) The data is downloaded and converted to comma separated values (CSV) file. (4) the data is cleaned in R, so it can be affectively analysed; column names are simplified for ease of use and any columns not being analysed are removed. (5) The clean ready to use files are saved as csv. (6 A and 6B) The Connect Sensor data is laid over the reference sensor, provided by Dublin City Council, data in order to determine visually if both sensors are recording the same levels. A correlation test is done to determine this numerically. Correlation between each site is then analysed. (7A and 7B) The Connect Sensor data

is laid over rainfall data, taken from Met Eireann, data to determine visually what affect rainfall is having on river height. A correlation test is done to determine this numerically. (8) The effect of a severe rainfall event was analysed using rainfall data from Storm Desmond.

2.3 Results and discussion

2.3.1 Sensor validation

The sensor performance at hourly log was evaluated for the initial 118 days of the study. It was noted, when the data was overlaid, that the water level at the sites was increasing faster than the data logs were recording; resulting in the sensor being unable to record the maximum height of the water level in certain situations (Figure 2.5). Ultrasonic sensors are more accurate when the distance between the sensor and the point of measurement is short i.e. when the water level is high (Tekle, 2014). When the distance is lowered the sensor is less affected by changes in air temperature and therefore takes a more accurate reading. Higher river levels are associated with flooding; therefore, a water level monitoring sensor needs to be accurate when the water level is high. These tests were performed using R programming (R Core Team, 2016) and the ‘corrplot’ package (Wei *et al.,* 2017).

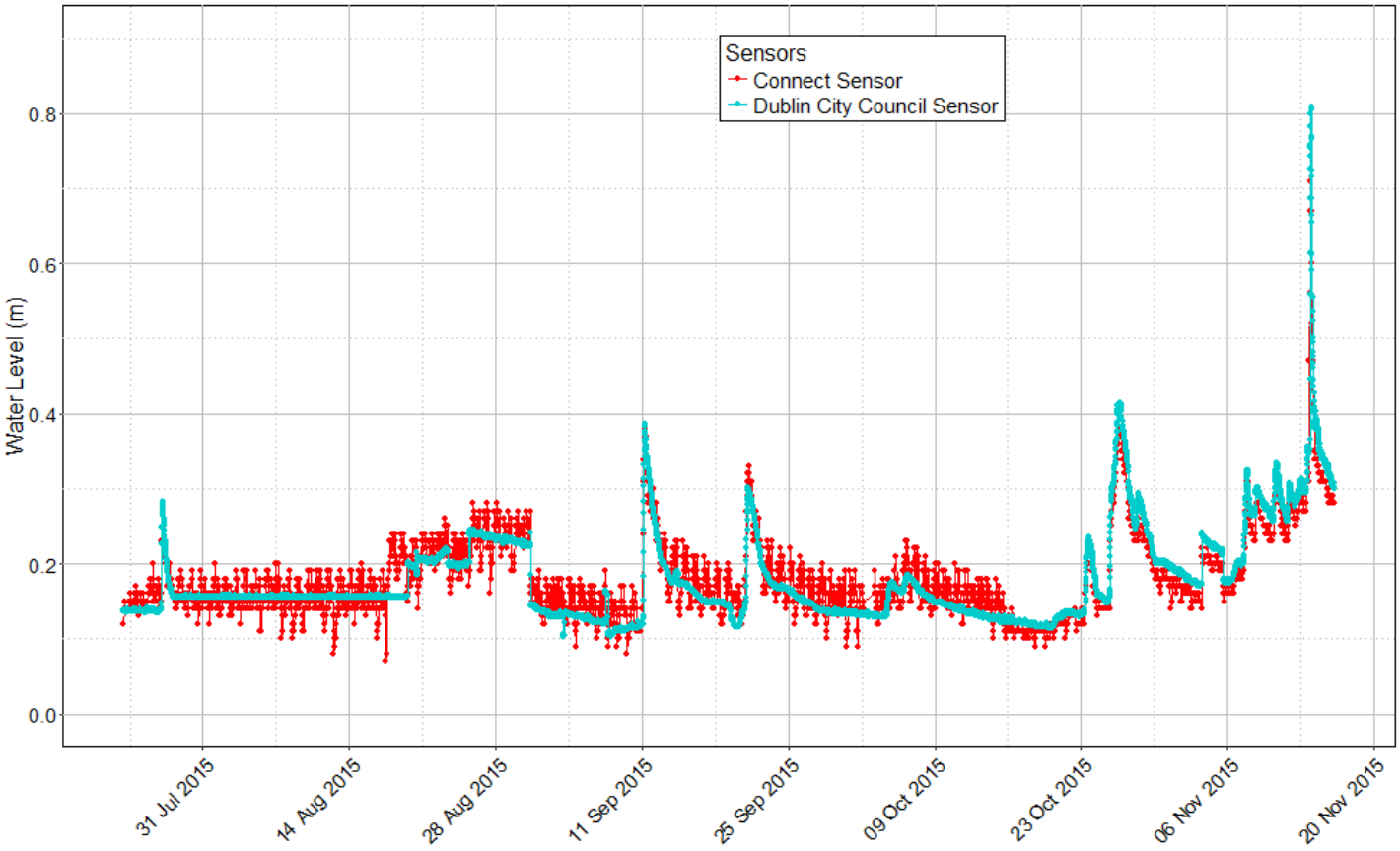

Figure 2.5: Overlay of the Connect Sensor and the reference sensor, provided by DCC, at Bohernabreena when the data collection logs were set to record once every hour.

In Figure 2.5 the data follows the same trend for both sensors, however the connect sensor data does not reach the maximum value of the DCC Sensor readings. It was hypothesized that due to the hourly reading logs were missing the upper levels of river level due to the infrequency of measurements taken.

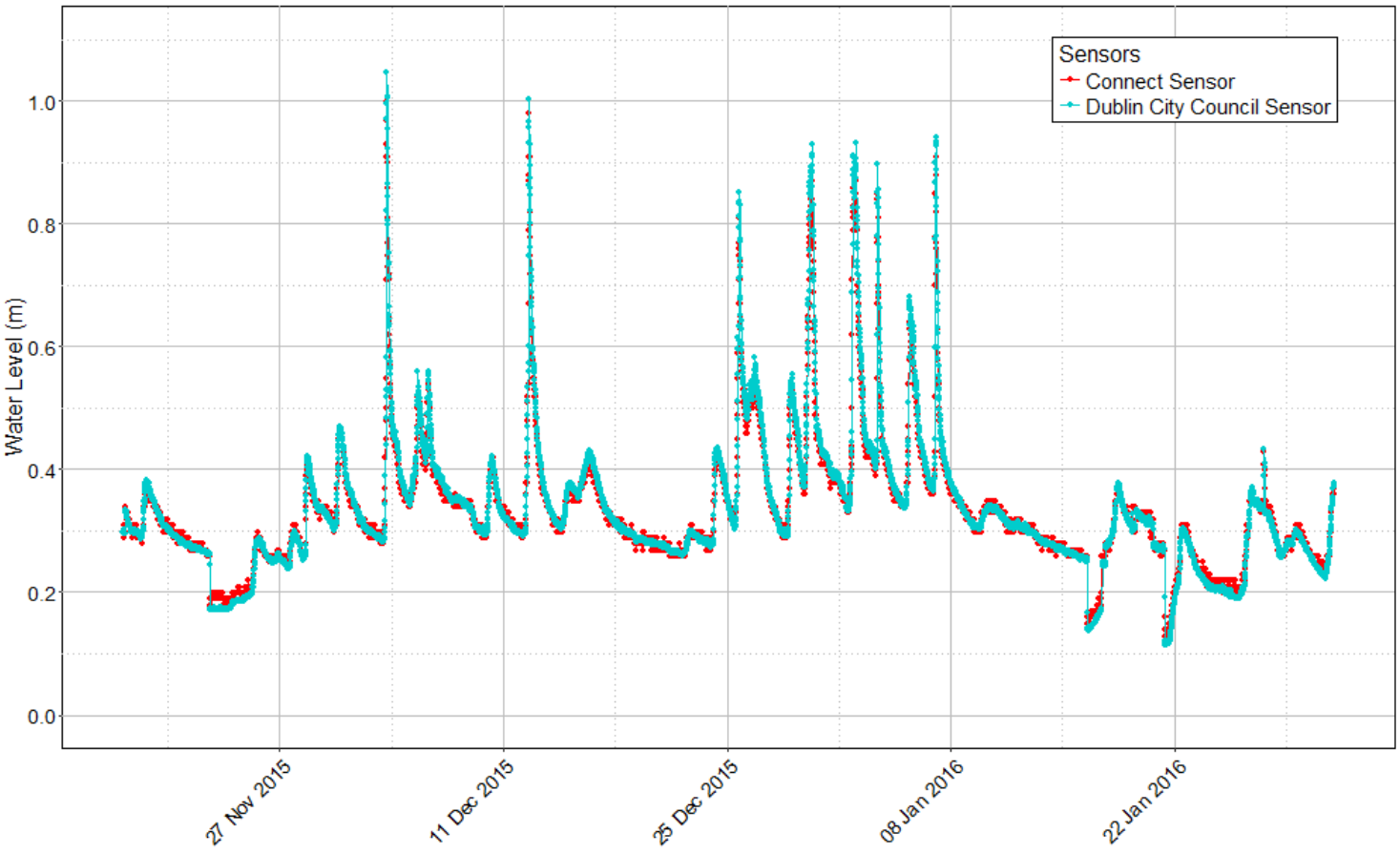

Figure 2.6: Overlay of the Connect Sensor and the reference sensor, provided by DCC, at Bohernabreena when the data collection logs were set to record every 15 min.

In Figure 2.6 the data now reached the same max and min as the DCC sensor. Person and Spearman's correlation coefficient were performed before and after a switch to 15 min data logs to determine if this improved correlation between the sensors.

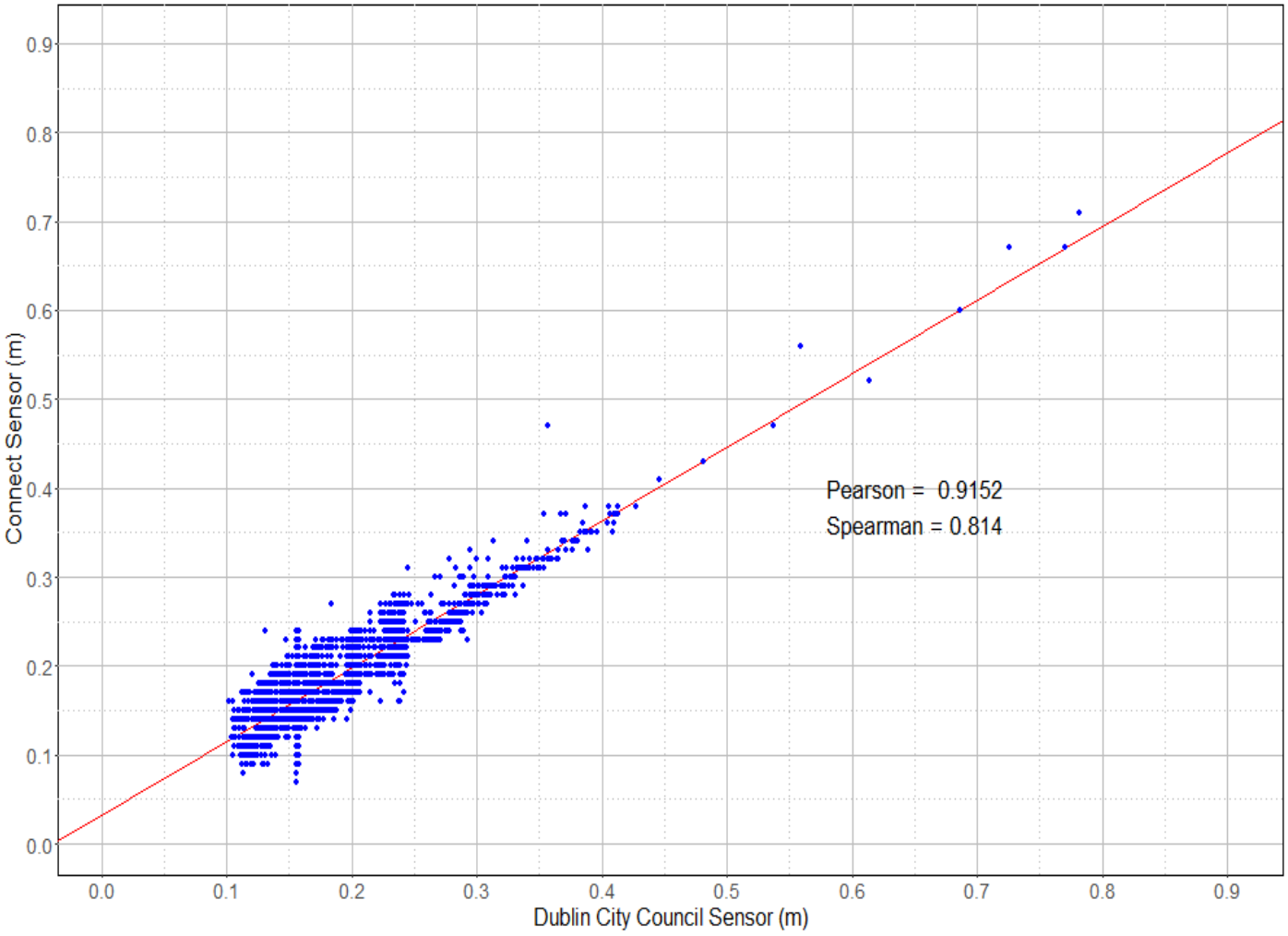

Figure 2.7: Correlation between the Connect Sensor and the reference sensor, provided by DCC, at Bohernabreena when the data collection logs were set to record once every hour.

Correlation was categorised as follows using Evans (1996) guide: 0.00 - 0.19 “very weak”, 0.20-0.39 “weak”, 0.40-0.59 “moderate”, 0.60-0.79 “strong”, 0.80-1.0 “very strong” (Evans, 1996). In Figure 2.7 a very strong correlation result was obtained for both the Pearson and Spearman correlation coefficients of 0.9 and 0.8 respectively. When graphed it was noted that the clear majority of readings fell below 0.4 m of river level and an accurate indication of correlation for higher water level was not obtained when hourly readings where used. A lower Spearman correlation result than Person informs us that the ranked correlation is not as strong as the true to value correlation. This means that the readings taken by each sensor were not correlating strongly at each ranked water level.

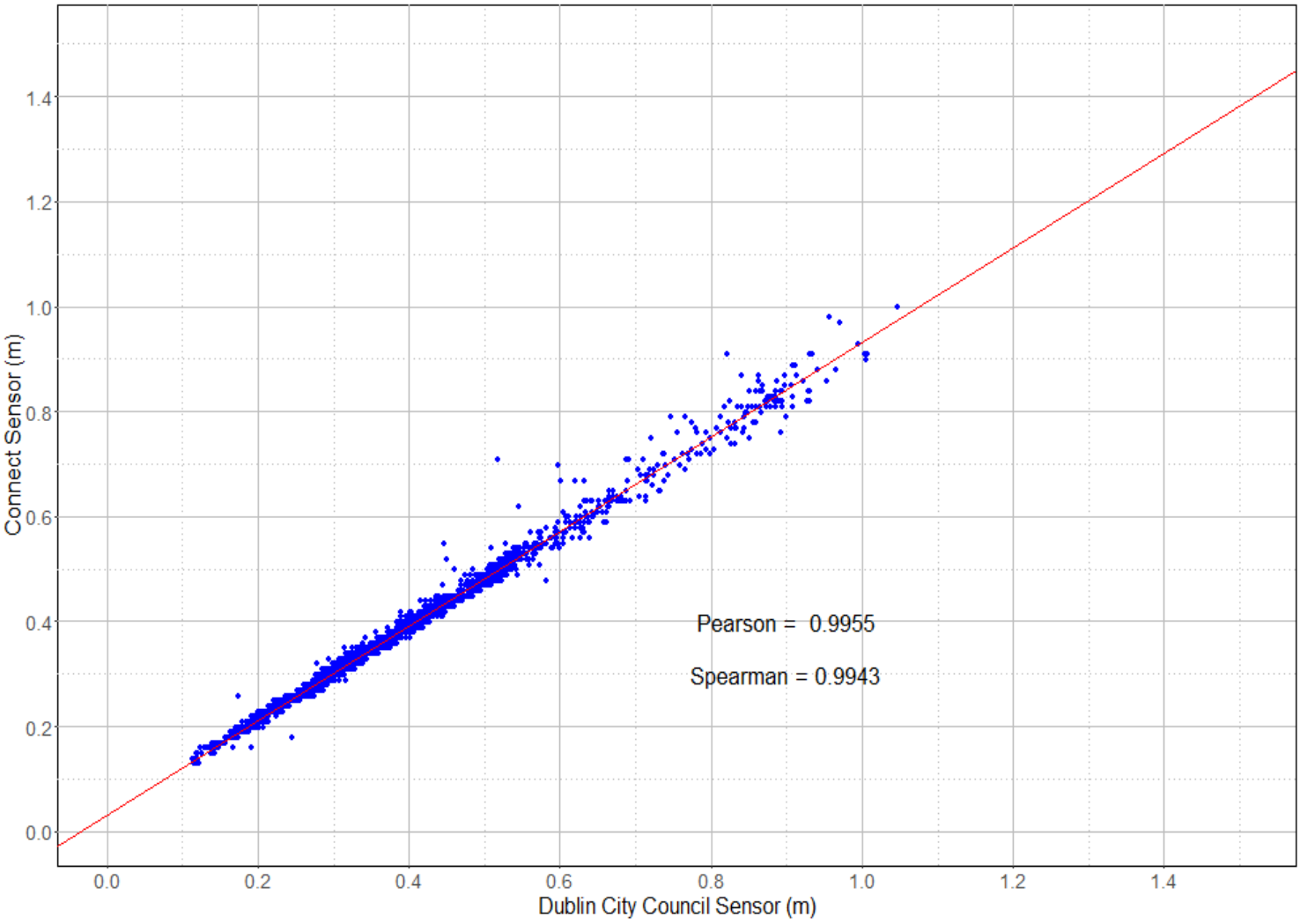

Figure 2.8: Correlation between the Connect Sensor and the reference sensor, provided by DCC, at Bohernabreena when the data collection logs were set to record every 15 min.

In Figure 2.8 there is a rise in Pearson and Spearman correlation coefficient after switching to 15 min data logs. True correlation and ranked correlation were both very strong between the connect sensor and the DCC reference sensor. There was also an increase in the amount of readings taken after 0.4 m. A Spearman coefficient result of 0.9943 tells us that the sensor was correlating very well at both the lower and higher ranked data logs.

Increasing the data logs improved the correlation results for all sensors (Table 2-2). All sensors recorded the max water level when taking readings every 15 min.

Table 2.2: Correlation results between the reference sensors, provided by DCC, and the Connect Sensor.

	Waldron's Bridge	Bohernabreena	Lady's Lane	Gandon
Initial Pearson	0.9619	0.9152	0.9465	0.8749
Pearson after change to 15 min data logs	0.9925	0.9955	0.9965	0.9857
Initial Spearman	0.8375	0.814	0.914	0.7969
Spearman after change to 15 min data logs	0.9751	0.9943	0.9936	0.9796

The initial sampling rate was set to an hour as fewer samples would consume less battery power. However, this sampling rate proved unrealistic for obtaining real-time water level measurements and the increase in water level occurs much quicker than the hourly readings were able to monitor. In Table 2 the results for Spearman's and Pearson's correlation coefficients before and after the sampling frequency was changed to 15 min showed an improvement in all sites. An improvement in Pearson's correlation coefficient means the overall correlation between the sensors improved at the true value of each water level reading. An improved Spearman's correlation coefficient means that the correlation has improved for each ranked variable. The results obtained, after the change of sampling rate, showed that the highest and lowest ranked water levels had a strong correlation so there was an improvement of correlation when the water level was low as well as when it was high.

2.3.2 Correlation between sites on the Dodder

To determine what influence downstream water levels had on upstream water levels a Spearman's ranked coefficient was performed for each of the sites on the Dodder. The downstream sites were correlated against the upstream sites from lag 0 to lag 1440 min. Spearman's ranked coefficient was chosen to determine if there was an

equal ranked influence of river height at high and low river levels irrespective if they rose and fell at the same rate. A Person's coefficient gives a correlation coefficient result for the true values and will not give a strong result if the variables increase at different rates, so it was not used.

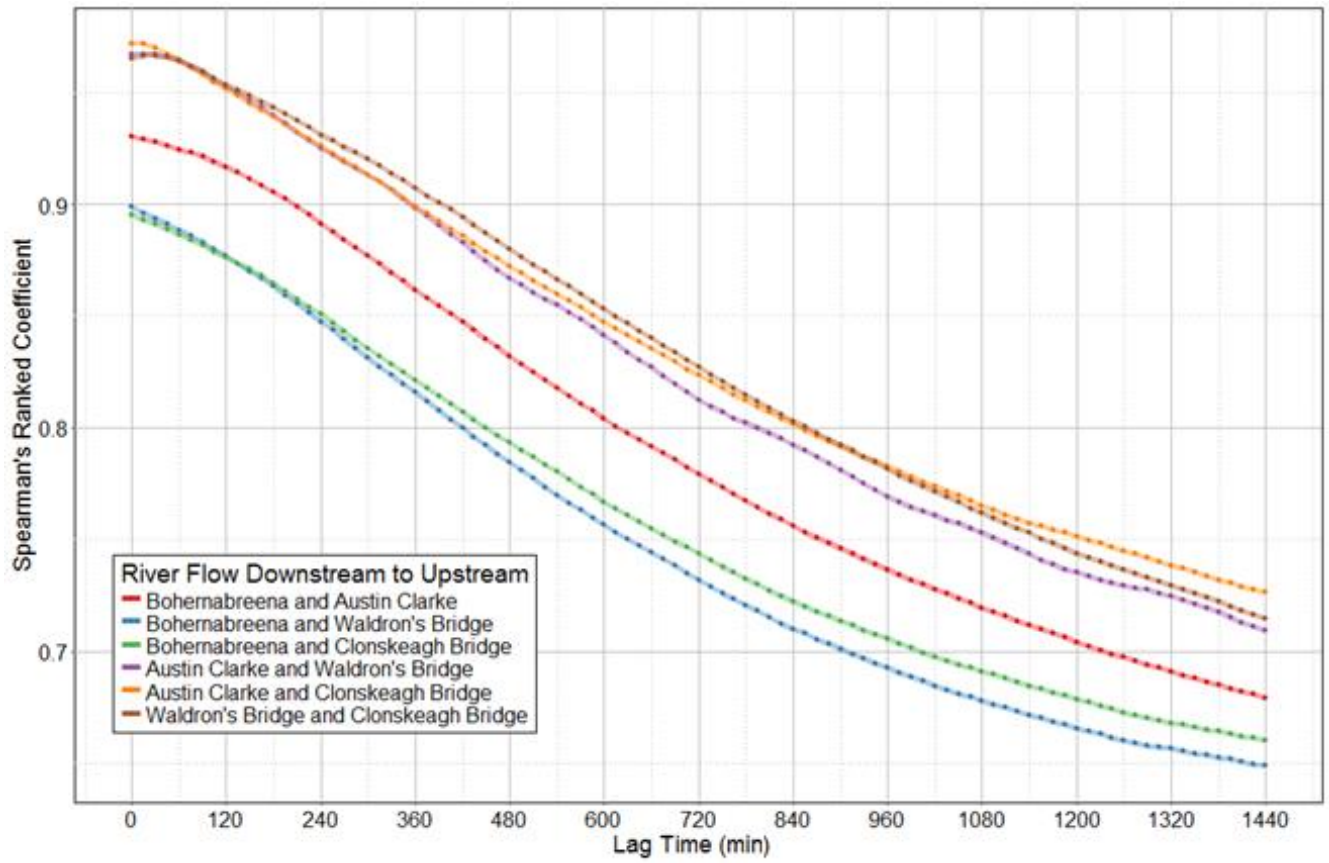

Figure 2.9: Spearman's ranked coefficient results between sites located on the Dodder from lag 0 to 1440 min.

In Figure 2.9 a downward trend is seen for all the sites in the Dodder. Overall correlation is decreasing between the sites as lag time increases. The maximum lag times can be seen in Figure 2.11.

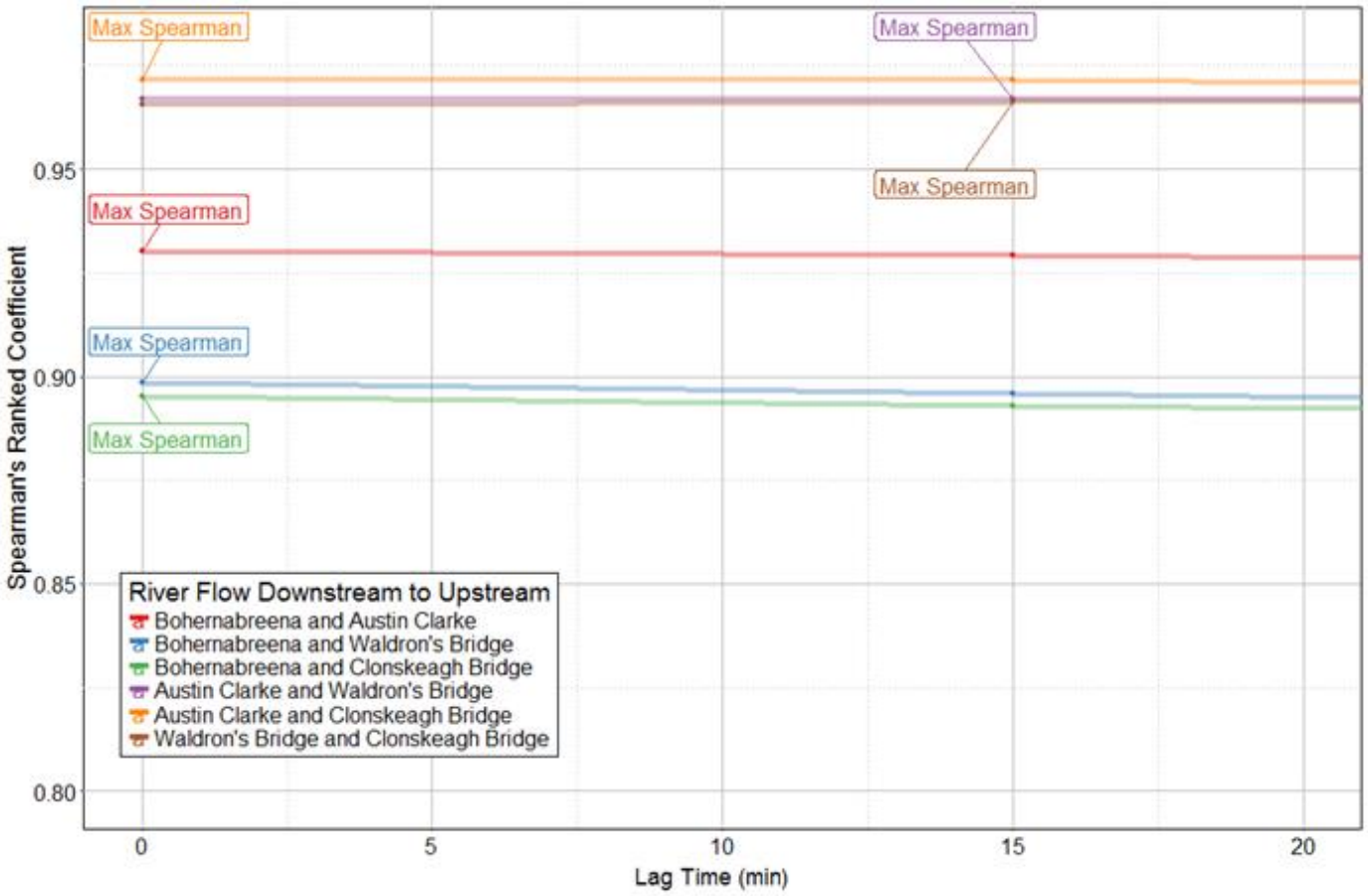

Figure 2.10: Close-up graph of the maximum Spearman coefficients found for sites on the Dodder river.

In Figure 2.10 the maximum spearman coefficient for most sites had a lag time of zero. This means that the influence downstream is happening at a faster rate than the sensors can measure or that the river levels are rising at the same time and there is no influence on the upstream river levels from the downstream levels. A 15 min delay was found between Austin Clarke and Waldron's Bridge, and Waldron's Bridge and Clonskeagh Bridge. These sites are also the closest in relation to each other in distance. Austin Clarke is approximately 3 km from Waldron's Bridge and Waldron's Bridge is approximately 3 km from Clonskeagh bridge. Austin Clarke and Clonskeagh bridge had a max Spearman result at lag 0 min. It was expected that if Austin Clarke and Waldron's bridge have a 15 min lag and Waldron's Bridge and Clonskeagh Bridge also have a 15 min lag the lag between the two sites, then Austin Clarke and Clonskeagh Bridge, located upstream and downstream from Austin Clarke

respectively, would have a greater than 15 min lag time between them. Distance may be a factor in why these sensors appeared to have no lag time; the distance between Austin Clarke and Clonskeagh Bridge is 6 km.

2.3.3 Correlation between sites on the Dodder tributaries

Spearman ranked coefficients were calculated between sites on the Dodder tributaries and the sites upstream from where the tributaries joined the Dodder to determine if there was any influence from the tributaries.

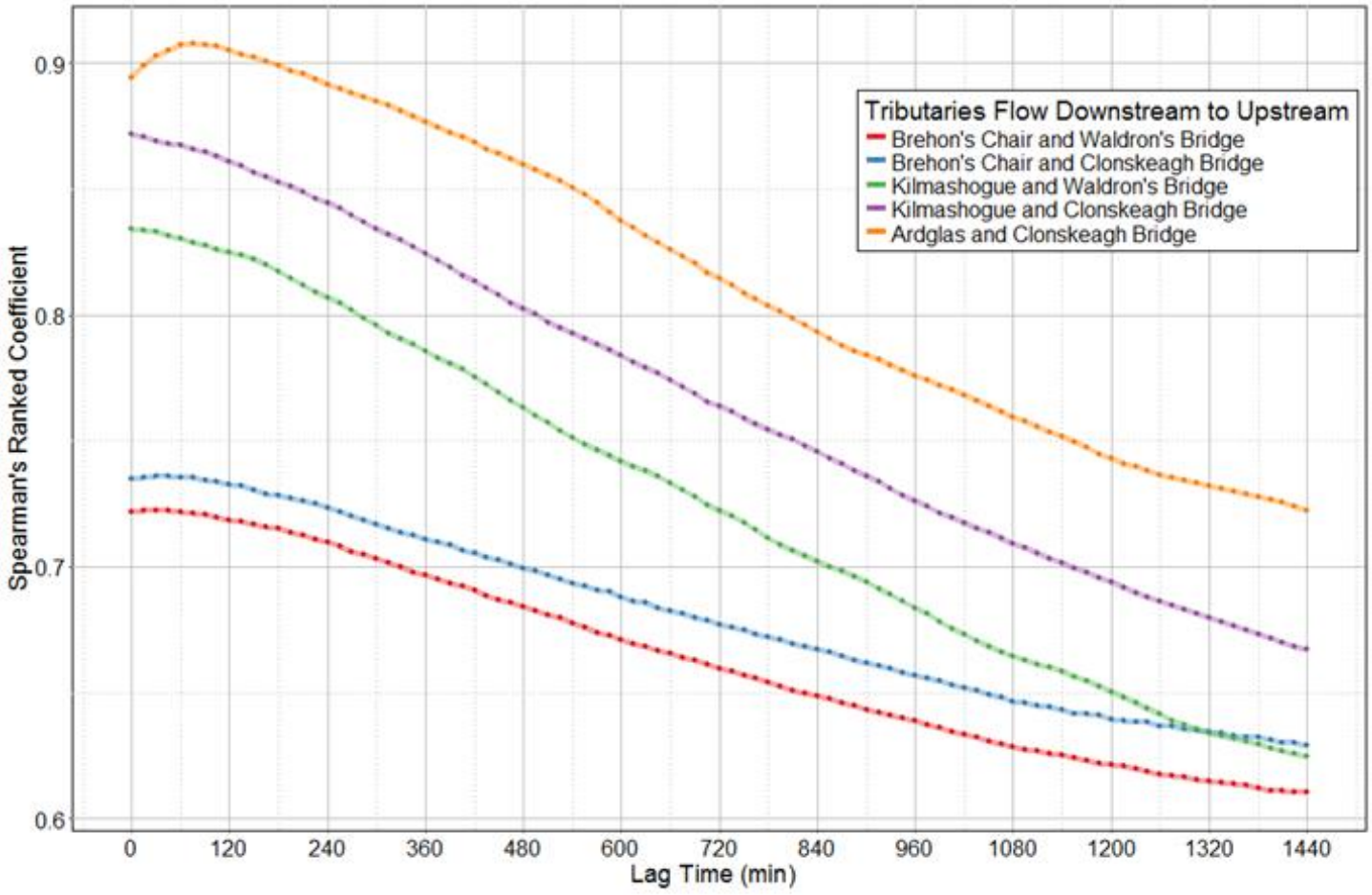

Figure 2.11: Spearman's ranked coefficient results between sites located on the Dodder tributaries from lag 0 to 1440 min.

In Figure 2.11 an overall decrease in Spearman coefficient was seen for each of the tributaries and the sites upstream over the course of 1440 min.

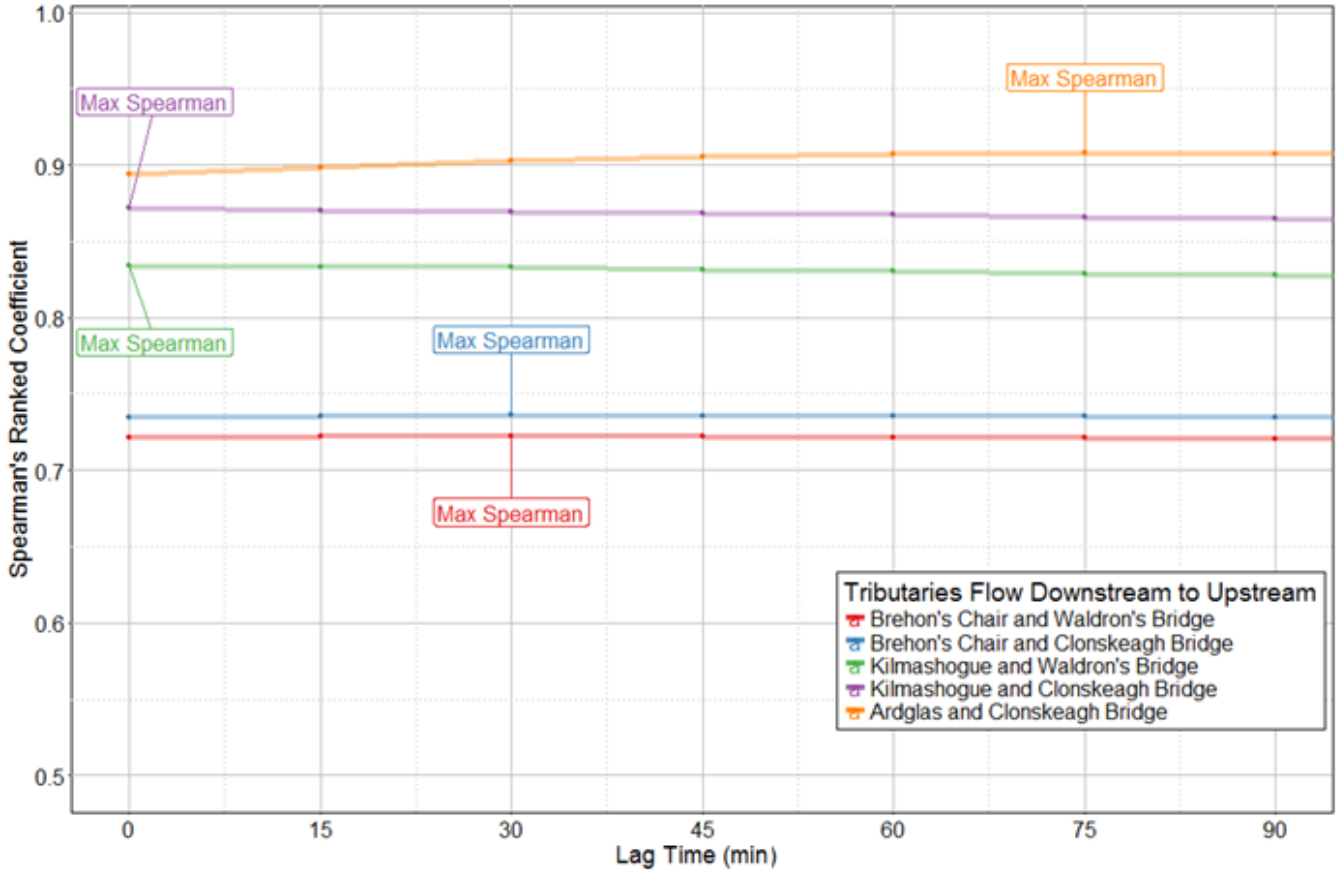

Figure 2.12: Close-up graph of the maximum spearman coefficient found for the Dodder tributaries.

The greatest lag time at maximum Spearman coefficient found between Ardglas and Clonskeagh Bridge, which are also the closest in location having a distance of 4 km between them, was 75 min (Figure 2.12). Brehon's Chair had a maximum Spearman coefficient at lag 30 min for both Waldron's Bridge and Clonskeagh Bridge, which are a distance of 5 km and 8 km between them and Brehon's Chair respectively. A strong correlation was seen between these sites and it affects the two sites upstream at the same rate. Kilmashogue has the furthest distance to travel to the sites on the Dodder these are 6 km to Waldron's Bridge and 9 km to Clonskeagh Bridge; the max Spearman coefficient was found at lag 0 for Kilmashogue. The site may be too far away for an accurate assessment of river height influence to be made. It appears from examining the tributaries and the Dodder that the strongest correlation with a resulting lag time resulted from sensors located 4 km or less from each other.

2.3.4 Correlation between rainfall and all sites

Rainfall was taken from 5 min data logs taken from Bohernabereena and Dublin City Council. When a Pearson correlation was performed between these rainfall data sets a result of 0.34 was obtained. This correlation was weak, and it was decided that an average of these sites would not be used; instead the rain gauge closest to the river site was used to determine the correlation between rainfall and site location. A list of sites and the closet rain gauge in available in Table 2.3.

Table 2.3: Distance to the closest available rain gauge for each site monitored.

Site	Closest Rain Gauge	Distance (km)
Lady's Lane	Dublin City Council	2
Bohernabreena	Bohernabreena	0
Gandon	Dublin City Council	2
Waldron's Bridge	Dublin City Council	4
Clonskeagh Bridge	Dublin City Council	4
Austin Clarke	Dublin City Council	6
Ardglas	Dublin City Council	8
Brehon's Chair	Bohernabreena	7
Kilmashogue	Bohernabreena	7

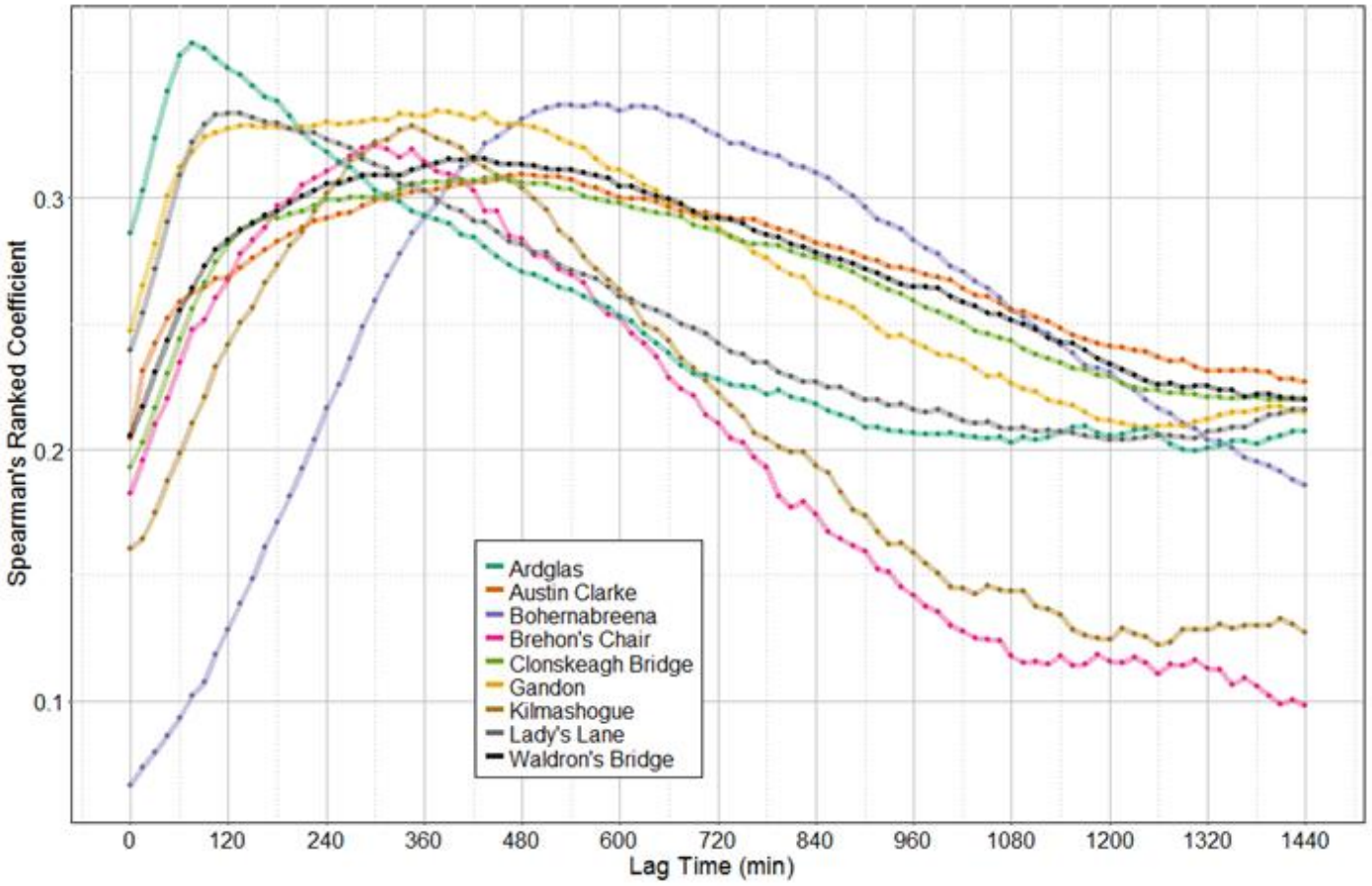

Figure 2.13: Spearman's ranked coefficient results between each site and the closest located rainfall gauge from lag 0 to 1440 min.

In Figure 2.13 each site has been correlated to its closest rain gauge. None of the sites reach a greater Spearman coefficient than 0.4, therefore all results are classified as weak. All sites follow different trends rising and falling at different lag times. Bohernabreena has the latest rise of lag time which may be due to the fact that is a reservoir and is expected to need a greater amount of rainfall to influence water level height.

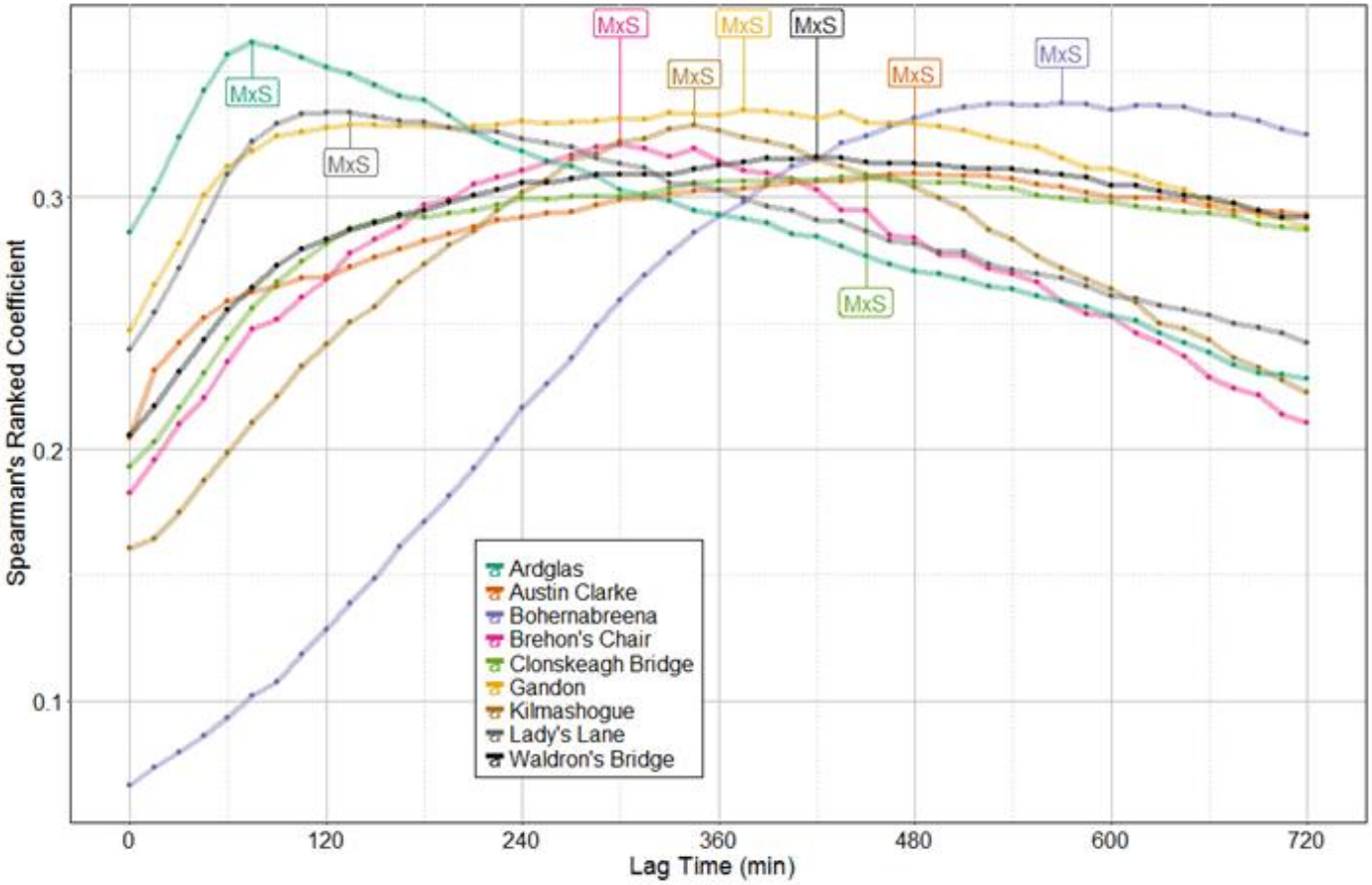

Figure 2.14: Close-up graph of the maximum spearman (MxS) coefficient found for each site and the nearest rainfall gauge.

Table 2.4: Maximum recorded Spearman coefficient and the corresponding lag time for each monitoring site.

Site	Maximum Spearman	Lag Time Hours:Min
Ardglas	0.36	00:75
Lady's Lane	0.33	02:15
Brehon's Chair	0.32	05:00
Kilmashogue	0.33	05:45
Gandon	0.33	06:15
Waldron's Bridge	0.32	07:00
Clonskeagh Bridge	0.31	07:30
Austin Clarke	0.31	08:00
Bohernabreena	0.34	09:30

In Figure 2.14 and Table 2.4 the highest Spearman coefficient, 0.36, was found for Ardglas at a lag time of 75 min. This was also the shortest lag time recorded. Spearman coefficient decreased with lag time increase, except for Bohernabreena. The max lag time indicates when a rainfall reaches its max ranked height and the river is reaching its corresponding max ranked height. This is not an indication of when rainfall begins to influence river level height; it is an indication of the amount of time for rainfall to reach its greatest influence over river height. There was no relationship seen between location of river sites and influence of rainfall. The rainfall influence is likely due to the shape of the riverbed as well as drainage in the area. These times are also a representation of average rainfall events occurring over the period of the 17-11-2015 to the 02-02-2016 and do not account for different responses in river height due to rainfall intensity. It was decided to examine a period of intense rainfall to determine the influence on river height when rainfall is expected to have the greatest influence on river height.

2.3.3 Case study Storm Desmond

On the 03-12-2015 Storm Desmond brought intense rainfall to Ireland. A daily total of 27.4 mm of rainfall was recorded at the Dublin City Council rainfall gauge and 32.4 mm was recorded at Bohernabreena rainfall gauge. The cross-correlation function (CCF) estimation was performed in R using the 'stats' core R package (R Core Team, 2016), between each site and the closest rain gauge for the 03-12-2015 to determine how much influence the storm had on river height.

2.3.3.1 Results at Clonskeagh Bridge

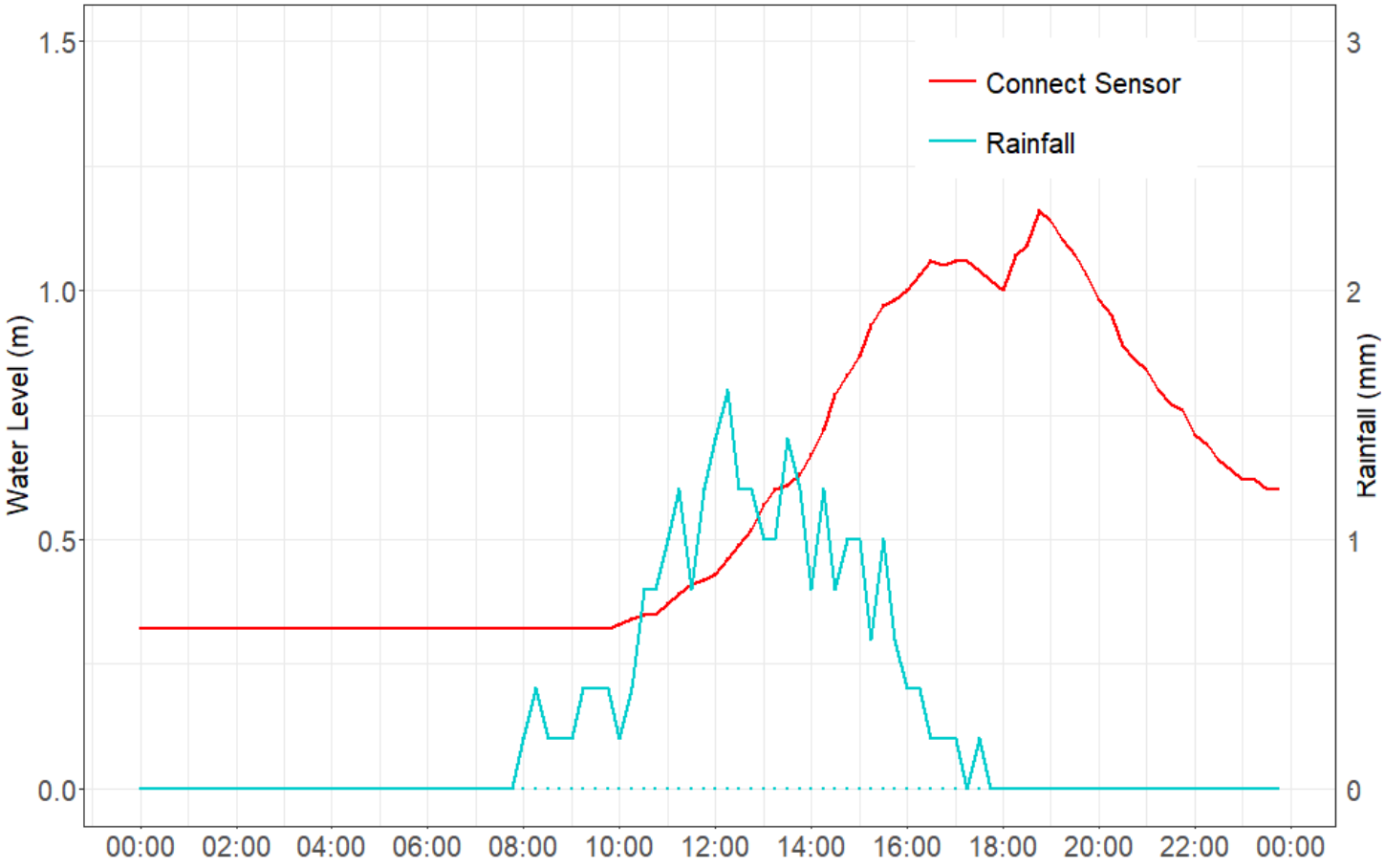

Figure 2.15: Overlay of rainfall and river height at Clonskeagh Bridge on the 03-12-2015 during Storm Desmond.

In Figure 2.15 rainfall begins at 08:00 and continues until 17:30. A cross-correlation function (CCF) plot was used to determine at what lag time does rainfall begin to impact water height and at what lag time does it reach peak impact and then begin to taper off and lose influence on the river height.

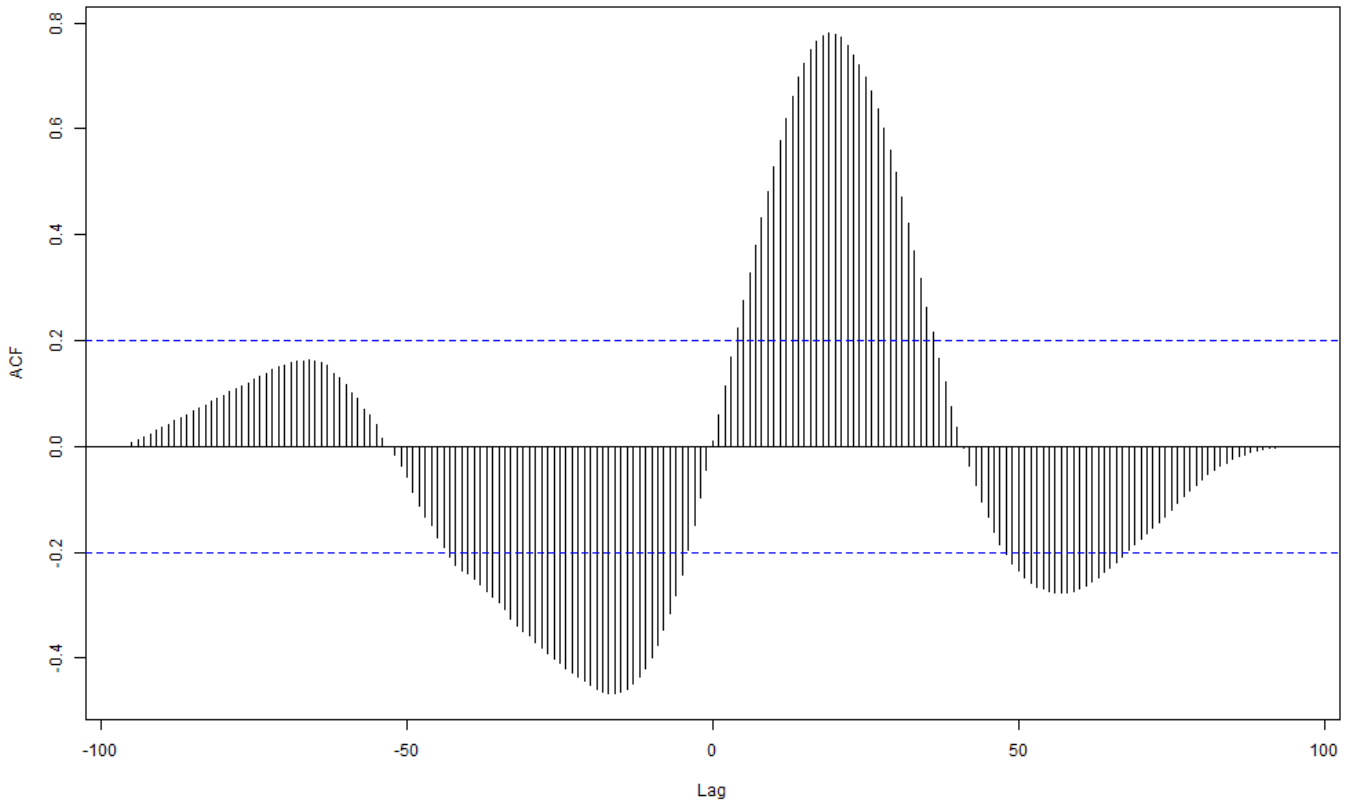

Figure 2.16: CCF plot for rainfall and water level at Clonskeagh Bridge on the 03-12-2015 during Storm Desmond.

In Figure 2.16 the blue line indicates the point of statistical significance over 0.5 alpha. When the correlation exceeds this line this correlation value can be said to be true for the population within a 95% confidence level. The negative lag influence is of no interest as it describes the relationship between rainfall and river height negatively against time. The correlation of interest begins at lag 0 and continues to lag 40. A positive correlation at lag 0 indicates that there is no delay between the beginning of the rainfall event and its influence on river height. As the rainfall data is taken at 15 min intervals lag 40, where the positive correlation ends, corresponds to 10 hours after

the beginning of the rainfall, 16:45. From Figure 2.16 at 16.45 the river level begins to drop off here and rainfall level has dropped intensity. The drop in intensity of rainfall here will not significantly influence the river level height as the river is now lowering too quickly for it to be affected by a weak rainfall intensity. There is a slight rise at 18:45 but the CCF tells us that this is not due to rainfall and may be from another source. The peak correlation of 0.782 occurs at lag 19 or 4 hours and 45 min after the rainfall event begins. This corresponds to 11.30 on Figure 2.16. This is the time where the river level is most influenced by rainfall and begins to rise rapidly.

2.3.3.2 Results for all other sites.

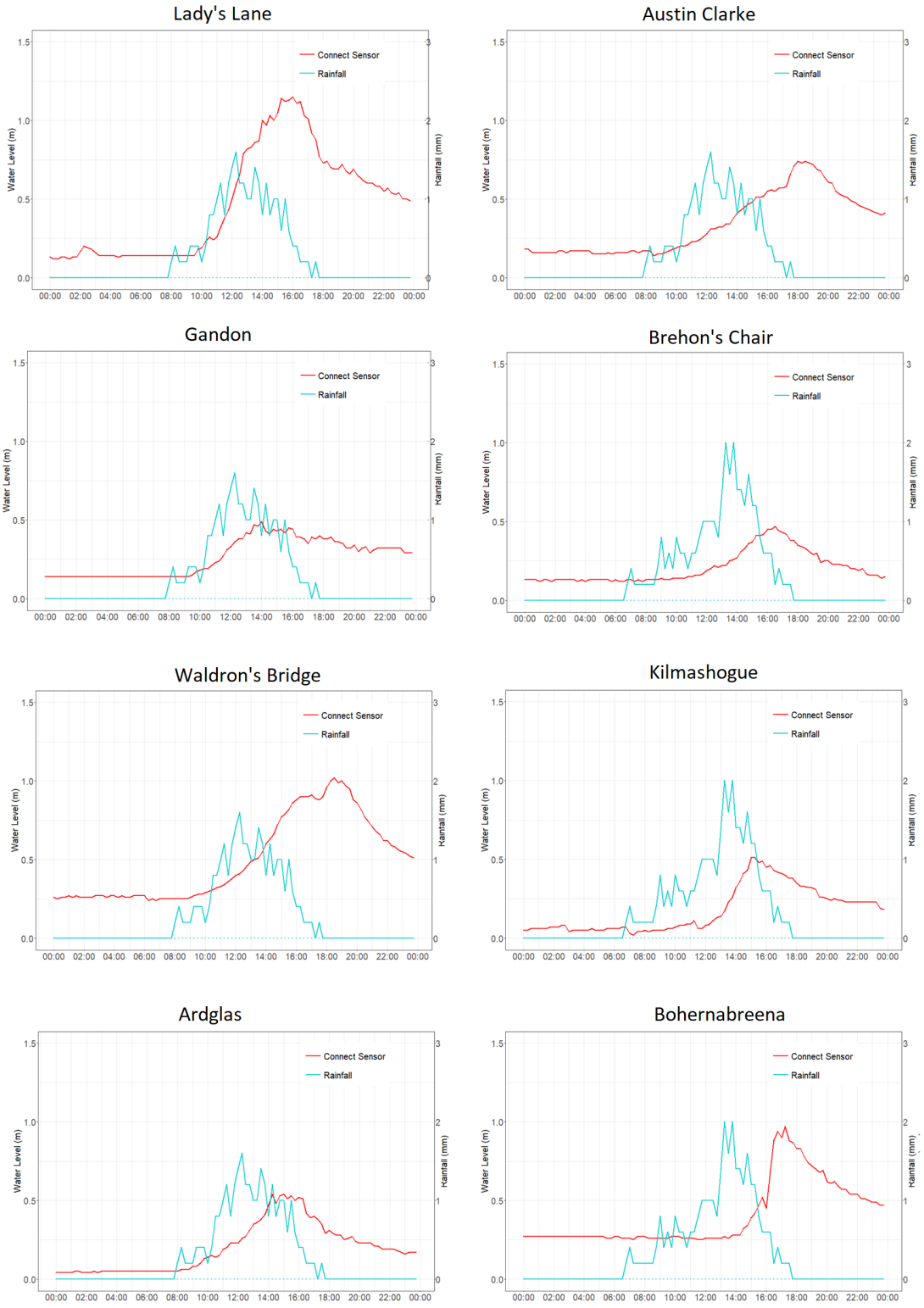

Figure 2.17: Overlay of rainfall and river height at all other sites on the 03-12-2015 during Storm Desmond.

Table 2.5: CCF results for rainfall and water level at all other sites on the 03-12-2015 during Storm Desmond.

Lady's Lane			
	Lag (min)	**Corresponding Real Time**	**Correlation Result**
Start of positive correlation	0	08:00	0.40
Peak of positive correlation	11	10:45	0.78
End of positive correlation	35	16:45	0.01
Gandon			
	Lag (min)	**Corresponding Real Time**	**Correlation Result**
Start of positive correlation	0	08:00	0.49
Peak of positive correlation	9	10:15	0.71
End of positive correlation	36	17:00	0.02
Waldron's Bridge			
	Lag (min)	**Corresponding Real Time**	**Correlation Result**
Start of positive correlation	0	08:00	0.02
Peak of positive correlation	19	12:45	0.78
End of positive correlation	40	18:00	0.03
Ardglas			
	Lag (min)	**Corresponding Real Time**	**Correlation Result**
Start of positive correlation	0	08:00	0.55
Peak of positive correlation	10	10:30	0.83
End of positive correlation	30	15:30	0.01

Table 2.5 contd.

Austin Clarke			
	Lag (min)	**Corresponding Real Time**	**Correlation Result**
Start of positive correlation	1	08:15	0.03
Peak of positive correlation	22	13:30	0.75
End of positive correlation	42	18:30	0.02
Brehon's Chair			
	Lag (min)	**Corresponding Real Time**	**Correlation Result**
Start of positive correlation	0	06:45	0.30
Peak of positive correlation	12	09:45	0.85
End of positive correlation	33	15:00	0.01
Kilmashogue			
	Lag (min)	**Corresponding Real Time**	**Correlation Result**
Start of positive correlation	0	06:45	0.26
Peak of positive correlation	12	09:45	0.72
End of positive correlation	38	16:15	0.03
Bohernabreena			
	Lag (min)	**Corresponding Real Time**	**Correlation Result**
Start of positive correlation	5	08:00	0.04
Peak of positive correlation	19	11:30	0.72
End of positive correlation	44	17:45	0.01

For most river sections there was no delay between the beginning of the rainfall and the beginning of an increase in river height which was determined by the presence of a positive correlation at lag 0 (Table 2.5). There were two exceptions the positive correlation began at lag 1 at Austin Clarke and lag 5 at Bohernabreena, this corresponds to fifteen min and an hour and fifteen min after the beginning of the rainfall event respectively. It is hypothesised that the depth and width of the river in these areas may increase the amount of time rainfall needs for the river height to impact. Bohernabreena is a reservoir and therefore has a large capacity, this could explain the hour and a half of rainfall that fell before the water level was impacted. The peak correlation time indicates the time at which the river is rising at its fastest in response to the rainfall. The peak river time was varied. This case study allowed us to show how a sensor network can determine how each river section reacted to the same storm event. Determining what areas of the river may be at greater risk of flooding closer to the beginning of a storm event can allow people to prioritise storm mitigation in these areas in the event of a storm.

2.4 Conclusion

2.4.1 Correlation between the connect sensor and the reference sensor provided by DCC

When testing the connect sensor against the reference sensors already in use by Dublin City Council a very strong correlation was achieved when 15 min data logs were used. The affordable cost of the connect sensor means that a network of sensors may be deployed through a catchment without being hugely expensive to the consumer and can still provide precise results.

2.4.2 Optimal location of network deployment

It was noted from the network deployed in the Dodder catchment that when sensors were located closer together a higher max spearman correlation was seen between them at a lag time of greater than zero. It would be more beneficial in this case to have a network of sensors located no more than 8 km from each other.

2.4.3 Rainfall and site location

When examining the Spearman's ranked correlation between each site and the closest available rain gauge no strong correlation was seen at any site. Each site reacted differently in response to rainfall and a different max Spearman was achieved at each site. This can be used as an indication of the average lag time between a rainfall event and water level response. To determine how each site responded in the most adverse rainfall conditions a case study of Storm Desmond on the 03-12-2015 was examined.

2.4.4 Storm Desmond case study

The case study showed that the river level was affected from the beginning of the rainfall event but was not affected by the end stages of the rainfall event. The rainfall event started gradually building to intensity and dropped intensity quickly. The river level also increased in height faster than it reduced in height and may not have been affected by the reduced level in rain as it was lowering in level quicker than the rainfall could affect the rise of the river level. The case study also showed how most areas of the catchment were affected immediately by the storm event however, each part of the catchment reached its peak correlation at slightly different times.

References

Arnbjerg-Nielsen, K. and Fleischer, H. S. (2009) ‘Feasible adaptation strategies for increased risk of flooding in cities due to climate change’, *Water Science and Technology*. IWA Publishing, 60(2), pp. 273–281.

Bonnet, P., Gehrke, J. and Seshadri, P. (2000) ‘Querying the physical world’, *IEEE personal Communications*. IEEE, 7(5), pp. 10–15.

De Bruijn, E. I. F. and Brandsma, T. (2000) ‘Rainfall prediction for a flooding event in Ireland caused by the remnants of Hurricane Charley’, *Journal of Hydrology*. Elsevier, 239(1–4), pp. 148–161.

Carsell, K. M., Pingel, N. D. and Ford, D. T. (2004) ‘Quantifying the benefit of a flood warning system’, *Natural Hazards Review*. American Society of Civil Engineers, 5(3), pp. 131–140.

Christensen, J. H. and Christensen, O. B. (2003) ‘Severe summertime flooding in Europe’, *Nature*. Nature Publishing Group, 421(6925), p. 805. Available at: https://doi.org/10.1038/421805a.

Evans, J. D. (1996) *Straightforward statistics for the behavioral sciences.* Pacific Grove, CA: Brooks/Cole Publishing.

Faisal, I. M., Kabir, M. R. and Nishat, A. (1999) ‘Non-structural flood mitigation measures for Dhaka City’, *Urban Water*. Elsevier, 1(2), pp. 145–153.

Gervais-Ducouret, S. (2011) ‘Next smart sensors generation’, in *Sensors Applications Symposium (SAS), 2011 IEEE*. IEEE, pp. 193–196.

Gourbesville, P. *et al.* (2012) ‘Flood warning systems and ubiquitous computing’, *La Houille Blanche*. Société Hydrotechnique de France, (6), pp. 11–16.

He, X. *et al.* (2014) ‘The construction of monitoring and warning system for flash flood defense in China’.

Jonkman, S. N. *et al.* (2009) ‘Loss of life caused by the flooding of New Orleans after Hurricane Katrina: analysis of the relationship between flood characteristics and mortality’, *Risk Analysis: An International Journal*. Wiley Online Library, 29(5), pp. 676–698.

Keoduangsine, S. and Goodwin, R. (2012) ‘A GPRS-based data collection and transmission for flood warning system: The case of the Lower Mekong River Basin’, *International journal of innovation, management and technology*. IACSIT Press, 3(3), p. 217.

R Core Team (2016) ‘R: A Language and Environment for Statistical Computing’. Vienna, Austria. Available at: https://www.r-project.org/.

Ridoutt, B., Sanguansri, P., Bonney, L., Crimp, S., Lewis, G., Lim-Camacho, L. (2013) ‘Climate Change Adaptation Strategy in the Food Industry — Insights from Product Carbon and Water Footprints’, *Climate Change*, 4(2), p. 26. doi: 10.3390/cli4020026.

Tekle, T. T. (2014) ‘Ultrasonic stream bridge sensors (USBS) error in water level estimation’. University of Iowa.

Wadey, M. P. *et al.* (2015) ‘A comparison of the 31 January–1 February 1953 and 5–6 December 2013 coastal flood events around the UK’, *Frontiers in Marine Science*, 2, p. 84. doi: 10.3389/fmars.2015.00084.

Chapter 3: Evaluation of the Occurrence of Bacterial Contamination in Grand Canal Basin Dublin

Abstract

The Grand Canal Basin is a popular recreational freshwater dock located in central Dublin City. Due to a storm water outfall located directly in the basin there is a concern of raw sewage entering the basin after heavy rainfall events. To prevent exposure of harmful bacterial concentrations to the local population, who use the basin for bathing activities, the water in the basin has been routinely sampled from 2004 until present by Waterways Ireland. This study takes historical data and examines trends in response to site location, seasonal trend, daily rainfall, and changing rainfall characteristics and intensity to determine what conditions are more likely to cause a pollution event and how long until the water is likely to be safe for use by bathers again. It also examines the ColiSense as a method for fast analysis for bacterial concentration for the Grand Canal Basin.

3.1 Introduction

3.1.1 Sewage in fresh water

Surface water is highly susceptible to bacterial pollution from sources such as run off (Mekonnen and Hoekstra, 2018), animal faeces (Nguyen *et al.,* 2018) and sewage discharges (Olds *et al.,* 2018). As many fresh water sources are used by humans for fishing, recreation, and drinking water it is important that these places are routinely monitored. However due to the standard method used for monitoring bacterial concentrations, grab sampling, results are not obtained in a timely manner. Grab sampling is also an expensive method of sampling and therefore is usually used sparingly depending on the budget of the implementing organization (Namieśnik *et al.*, 2005). There are EPA directives for bathing water thresholds in place to determine whether water bodies are safe to drink from or swim in, however there are no strict guidelines to determine the length of time, after a pollution event occurs, a waterbody requires to be off-limits before it is safe for use by the public once more (The Environmental and Protection Agency, 2001). Being able to estimate the amount of time required for a water body to reach safe concentrations for use by people again can help determine the best time to take samples after a pollution event has occurred (Boehm, Graham and Jennings, 2018). Once a microbial pollutant has entered a water body they are dispersed, and the number of pollutant microbes begin to reduce due to inactivation. Inactivation can occur due to stress, change in temperature, and directly by being photochemically damaged by the sun (Dick *et al.,* 2010).

3.1.2 The Grand Canal Basin

The Grand Canal Basin is being increasingly used as a water body for recreation in the inner-city area of Dublin (53.3424° N, 6.2413° W) (Moore, 2008). Monitoring of bacterial concentrations has taken place from 2004 to present to ensure the safety of the people who bathe in the basin. Currently a storm water outfall is located directly in the basin which is suspected to contribute majorly to the bacterial concentration in the basin. The storm water outfall carries excess sewage into the basin during high flow events, which usually occur during periods of high rainfall. This report aims to determine if there is any link between bacterial concentration and location of the storm water outfall as well as incidences of high rainfall.

3.1.3 Effects of storm water outfalls

Storm water outfalls carrying sewage can dramatically increase the number of pathogens, and faecal bacteria entering a water body. This can have very harmful effects on the people who use these water bodies recreationally. Studies show that many storm water overflow systems suffer greatly from the first flush phenomenon, where a high load of pollutants enter a water body through the system as a result of a large rainfall event (Gupta and Saul, 1996; Peng et al., 2016). Faecal bacterial concentrations obtain the highest levels early in storms, before peak flow is achieved, and this affect can also be seen early in storm seasons (Tiefenthaler, Stein and Schiff, 2011), and that these concentrations decline in the final stages of storms (Krometis *et al.*, 2007). Rainfall intensity has also been shown to impact bacterial concentration in urban settings (Hathaway, Hunt and McCarthy, 2015).

3.1.4 Bathing water thresholds

The Grand Canal Basin is not an official designated bathing area, but fresh water bathing standards are used to monitor the area due to its popularity for recreational use. The bathing water thresholds used in this project were set by the Environmental Protection Agency (EPA). The 76/160/EEC directive threshold values were taken for total coliforms, faecal coliforms, and E. coli (Directive, 1976). The 2006/7/EC directive threshold values were used for Enterococci (Directive, 2006). The fresh water bathing standards are as follows: total coliform > 5000 Most Probable Number (MPN/100 mL, faecal coliform > 1000 MPN/ 100 mL monitored in the years 2004 to 2014 and E. coli > 900 MPN/ 100 mL; Enterococci > 330 MPN/100 mL monitored in the years 2015 to present. This study examined the effects on these thresholds in response to site location, seasonal trend, daily rainfall, and changing rainfall characteristics and intensity. Also examined was how the site reacts under different conditions to help determine when is the best time to take future samples and have the most likely chance of detecting a pollution event. Informed sampling allows protection zlorganisations to reduce the cost of sampling as well as increasing the likelihood major pollution events won't be missed. This type of informed sampling is referred to as 'ecological forecasting' (Dietze *et al.,* 2018).

3.1.5 The ColiSense for sampling in the Grand Canal Basin

The ColiSense was designed to prove rapid analysis and potential for on-site application of bacteria (Heery *et al.,* 2016). The ColiSense was examined as an alternative to grab sampling for future testing. The technology is relatively new, and this study was the first time its effectiveness was tested in a canal ecosystem. The ColiSense consists of a sensitive portable fluorimeter with incubating capability and triplicate sample chambers. The target analyte was β-D-Glucuronidase (GUS) hydrolyses a synthetic substrate 6-Chloro-4-Methyl-Umbelliferyl-β-D-Glucuronide (6-CMUG) to release the fluorescent molecule 6-Chloro-4-Methyl-Umbelliferyl (6-CMU). The breakdown of 6-CMUG is measured over time to determine GUS activity.

3.2 Methods

3.2.1 Catchment description

The Grand Canal Basin is located near the city centre of Dublin 53.3424° N, 6.2413° W. It is enclosed by a harbour between the River Liffey and the Grand Canal. It consists of an inner and outer basin where the Grand Canal ends before it meets the River Liffey. Since 2000 the surrounding area has undergone significant redevelopment as part of the Dublin Docklands area redevelopment project.

3.2.2 Sampling sites

Figure 3.1 shows the Grand Canal Basin and lists the sampling sites used in this project. There is a storm water outfall located directly in the Grand Canal Basin. Sites 3, 4, and 5 are contained within the inner basin while sites 6, and 7 are within the outer basin. Site 3 is the closest in location to the storm water outfall. Site 1 is located upstream from the basin and was not sampled in the years pertaining to this study. Site 2 was originally where site 3 is now located and was renamed when the site locations decreased. All site 2 data was renamed to site 3.

3.2.3 Sample analysis using ColiSense (Heery et al., 2016)

Syringe filters are used for recovering the bacteria from the samples through filtration of fixed volumes using 50 mL syringes. The volume used in this experiment was 100 mL for each sample. The pore size on the filters is 0.45 µm. Once filtered the bacteria are washed to remove any residual compounds. Then 100 µL of lytic agent was added to the syringe filter, the filter is capped and in incubated for 30 min at 37 ^{o}C. The E. coli bacteria are lysed by the lytic agent and the marker enzyme Beta-D-Glucuronidase (GUS) is released. The enzyme is then recovered in 1.9 mL of buffer in glass vials. A fluorogenic substrate is added to the glass vials and 3 vials from each site are then placed in ColiSense for detection in triplicate. GUS breaks down the substrate and releases the fluorescent molecule which is detected over time.

Ten sample sites were used for the ColiSense analysis (Figure 3.2). Total coliform, E. coli, and enterococci were measured using Colilert 18 and Enterolert E. For total coliform and E. coli determination 100 mL of each water sample was placed into a sterile bottle. Colilert 18 was then added to the sample and inoculated into a Quanti-Trays and sealed. For enterococci detection Enterolert E was added to the 100 mL of

each sample and then inoculated into a Quanti-Trays and sealed. For E. coli and coliform enumeration, samples were incubated at 37 °C for 18 to 20 h. For enterococci enumeration, samples were incubated at 25 °C for 24 to 28 h. Following incubation, the Quanti-Tray using Colilert 18, wells were read visually for yellow colour indicating the presence of coliforms and for blue fluorescence indicating the presence of E. coli. The Quanti-Tray, using Enterolert E, wells were read visually for blue fluorescence indicating the presence of enterococci.

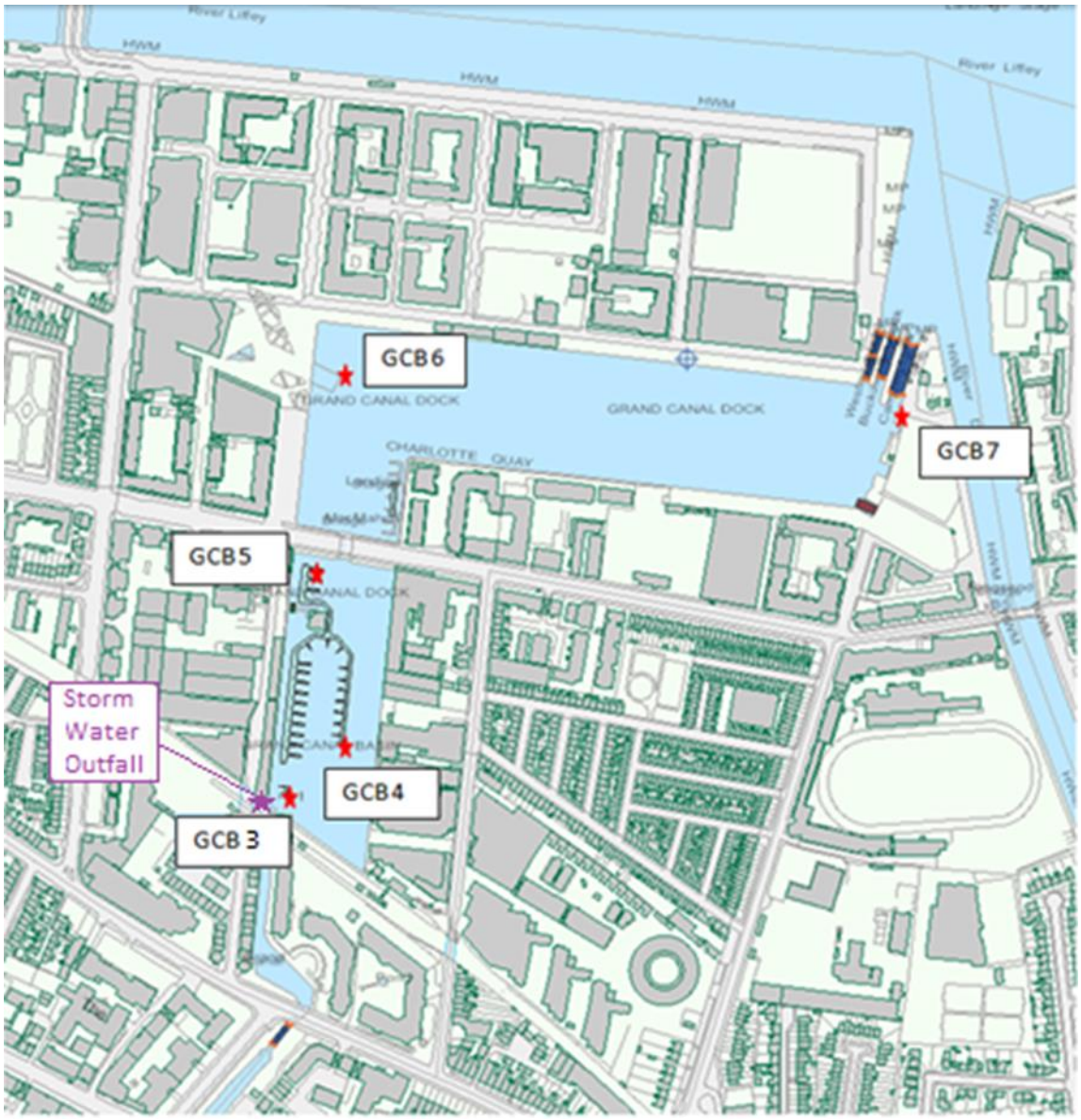

Figure 3.1: Historical sampling sites used by Waterways Ireland at the Grand Canal Basin from 2004 to present. Sites 3, 4, and 5 are located closest to the storm water outflow in the upper level of the basin. Sites 6 and 7 are located in the lower level.

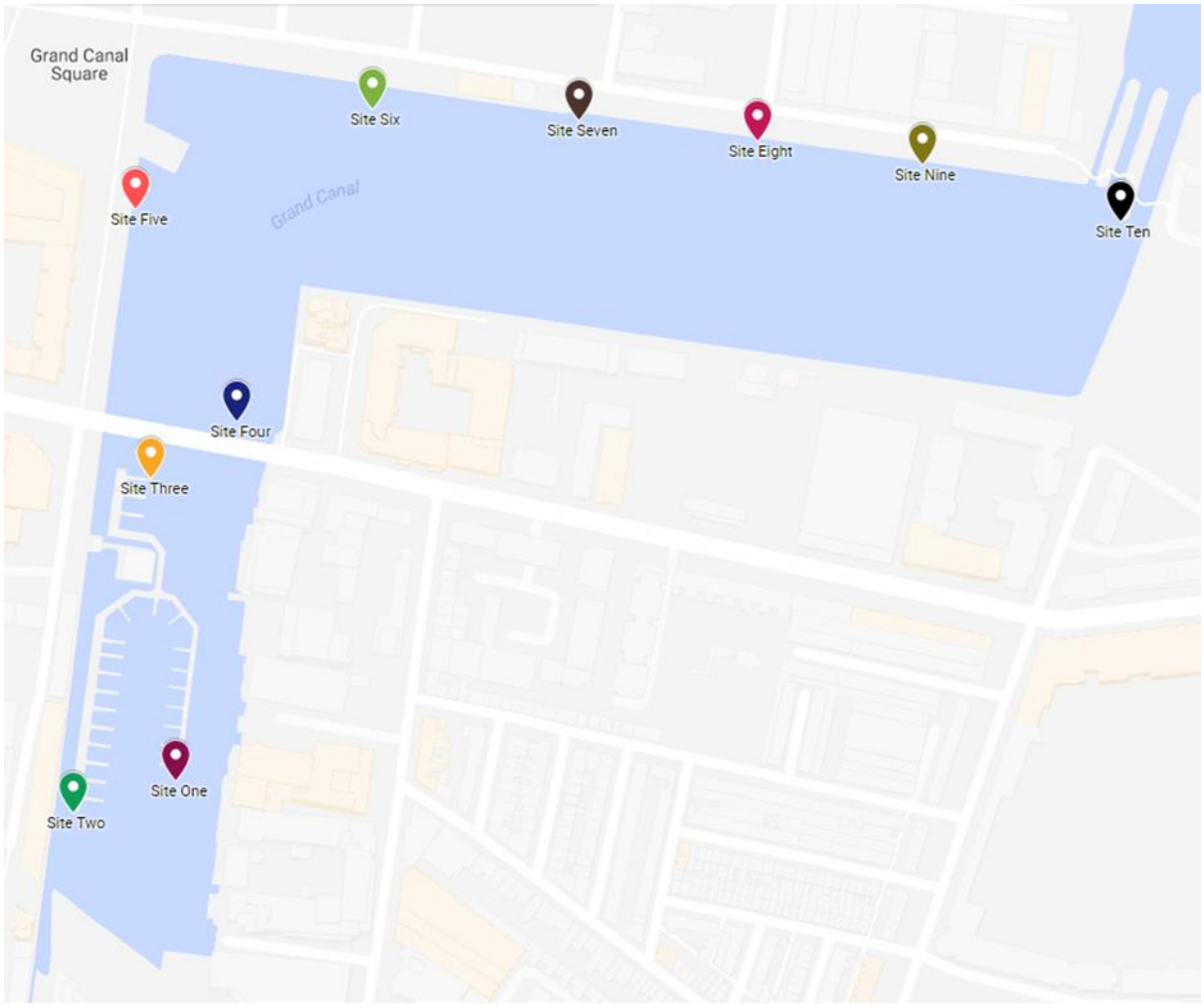

Figure 3.2: Locations were grab samples were taken, for analysis using the ColiSense, performed on the 09-02-17 and the 23-02-17.

3.3.4 Data and statistical analysis

The data used was sourced from 13 years of spot samples acquired from Waterways Ireland. Table 3.1 outlines the number of samples taken of each bacterium monitored in each year. A total of 1104 samples of total coliforms, 1012 samples of faecal coliforms, 276 samples of E. coli, and 184 samples of Enterococci were taken. Initially samples of total coliform and total faecal coliform where taken from November 2004 until January 2014 when sampling was changed to total coliform and E. coli. From July 2015 to 2016 all samples taken were of E. coli and Enterococci. The data contained information from five different sample sites located in the Grand Canal Basin; sites 3, 4, 5, 6, and 7. Samples were collected in different frequencies each year depending on what resources were available. There was not data collected for Site 3 for E. coli

or Enterococci. Upon receiving the data, it was changed to .csv files and analysed in R. Data analysis was split into two sections: 1) Analysis without rainfall and 2) Analysis with rainfall.

Correlation analysis was tested under Pearson and Spearman correlation factors. Spearman is computed on ranks and measures a monotonic relationship; while Pearson is on true values and measures a linear relationship. Both were measured to determine if the data had a linear or monatomic relationship. Correlation was categorised as follows using Evans (1996) guide: 0.00 - 0.19 “very weak”, 0.20-0.39 “weak”, 0.40-0.59 “moderate”, 0.60-0.79 “strong”, 0.80-1.0 “very strong” (Evans, 1996).

For analysis without rainfall each site was examined to determine the percentage breaches of all bacterium monitored. Fresh water bathing water limits of: total coliform > 5000 MPN/100 mL; faecal coliform > 1000 MPN/ 100 mL; E. coli > 900 MPN/ 100 mL; Enterococci > 330 MPN/ 100 mL. Then an average was determined for each bacterial concentration result obtained in each month for all sampling years. Lastly correlation was performed between each site location to determine if there was a delay between incidents of high and low bacteria samples.

For analysis with rainfall it was first necessary to determine the optimum rain gauges for analysis. A total of 8 rain gauges ranging from 1 to 34 km distance from sample site were compared to find the optimum rain gauge. When found these results were then correlated against each site to determine individual sites response to rainfall. The rainfall data was expressed in total mm per day.

Table 3.1: Total observations recorded in each year for each monitored bacterium.

Years	Total Coliform Observations	Faecal Coliform Observations	E. Coli Observations	Enterococci Observations
2004	59	59	0	0
2005	253	253	0	0
2006	231	231	0	0
2007	158	158	0	0
2008	129	129	0	0
2009	62	62	0	0
2010	42	42	0	0
2011	8	8	0	0
2012	14	14	0	0
2013	56	56	0	0
2014	88	0	88	0
2015	4	0	48	44
2016	0	0	140	140
Total	1104	1012	276	184

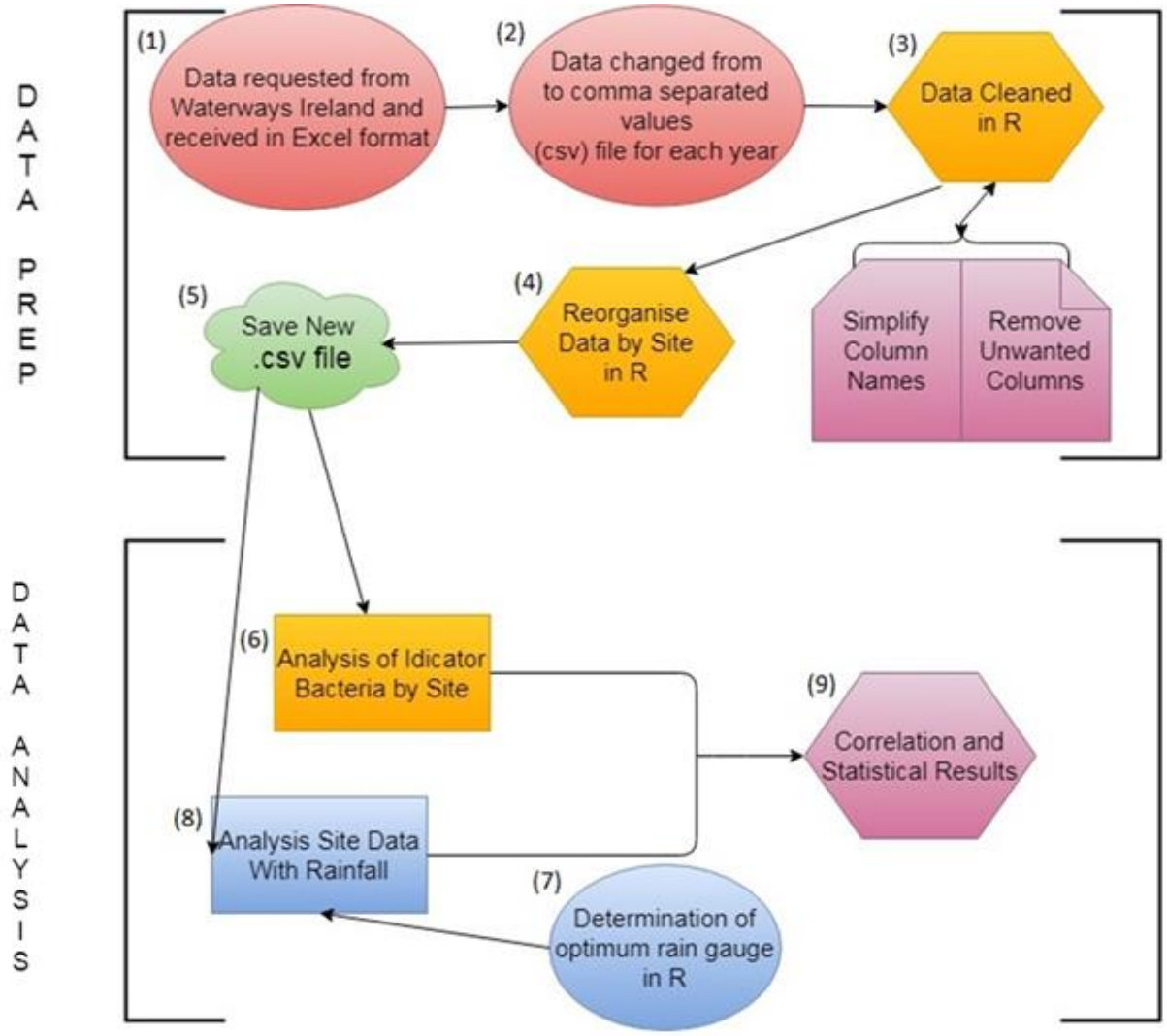

Figure 3.3: Experimental schematic outlining the order analysis was taken from data collection and preparation to data analysis.

Figure 3.3 describes the routes taken in order to source, compile, and analyse the data for the experiment. Data was sourced from Waterways Ireland in Excel format. (2) This was converted to a comma separated values (csv) file for work in R (R Core Team, 2016). (3) Data was cleaned in R: column names were simplified, missing data was removed, and unwanted columns were removed. (4) Initially data was organised by year. This data was then reorganised by site so that each data set contained all the recorded indicator bacteria measurements for one site. (5) These data sets were saved to new csv files. (6) The site data was first analysed against each other to determine if the bacterial concentrations in each site was related. (7) Then the optimum rain gauge was determined in order to be used for the second part of analysis. (8) The data was analysed against the indicator bacteria at each site to

determine what impact rainfall had on each site. (9) Correlation and statistical results were achieved.

Table 3.2: Data type and method performed using R programming (R Core Team, 2016).

Data Type	Analysis Method
Site Only	Percentage Breach, Year and Month levels (Total and Faecal Coliform only) Pearson and Spearman Correlation, Kolmogrov-Smirnove test, Colisense analysis.
Site and Rainfall	Pearson and Spearman Correlation, Rainfall Characteristic Analysis, Rainfall Threshold Analysis, Student T-Test. Percentage Breach.

Each method used for analysis on site only and for analysis on site only and analysis on site and rainfall is listed in Table 3.2.

3.3 Results and discussion

Deterring a pollution event should be in theory a simple process. A sample is taken, it is tested, and if the results fall above the threshold for safe water guidelines then the waterbody is determined to be polluted. However due to the costly nature of sampling it is impossible to take a sample every time the river is to be used by the population. To overcome this a ‘best time to sample’ practice is usually put in place. Historical data can be used to examine trends which can be used as indicators when trying to determine when a pollution event is most likely to have occurred. Prioritising sampling to occasions where there is a higher chance of a pollution event occurring can save cost on routine sampling and insure that routine sampling is not missing pollution events by only sampling on previously set dates.

The Grand Canal Basin has become increasingly important as a site for recreational water use. To protect the health of people who use the basin water quality must be monitored accurately. Untreated sewage entering the basin through the storm water outfall, located near site 3 (Figure 3.1), is of particular concern to the health of people who use the basin. This study aims to show if there is a relationship between location of the storm water outfall and the sites in the basin, as well as to determine if there is a relationship between each site. The impact rainfall has on coliform concentration at each site was also investigated. Determining how rainfall affects bacterial concentrations can help make informed decisions about when the basin may be at higher risk to breaching the recommended bathing levels. ColiSense analysis was performed in order to monitor the levels in the basin ourselves to get a sense of current levels.

3.3.1 Analysis of bacterial concentrations and site location

3.3.1.1 Percentage breaches of EPA bathing water thresholds.

Initially (Table 3.4) the overall percentage of breaches of bathing water levels were found for each site to give an indication of which sites had a higher frequency of pollution incidents

Table 3.3: Percentage breaches of EPA recommended bathing water thresholds at each site.

Microbe sample taken	Site 3	Site 4	Site 5	Site 6	Site 7
E. coli 900 MPN/100 mL		24.6	23.2	14.5	10.1
Enterococci 330 MPN/100 mL		15.2	13	6.5	6.5
Total Coliform 5000 MPN/100 mL	34.8	29.2	24.7	16.4	5.4
Faecal Coliform 1000 MPN/100 mL	27.5	21.9	16.14	7.9	2.5

The percentage breaches of, the daily guideline levels, was calculated at each sample site (Table 3.3). The highest percentage of breaches was at Site 3 which is in closest proximity to the storm water outfall. The breaches decrease in percentage as the sites get further away from the outfall.

3.3.1.2 Average monthly coliform concentrations

As total coliform and faecal coliform samples were taken over the longest amount of years, 12 and 10 years respectively, these measurements were averaged for each month to determine what months in what years had the highest average microbe levels. Table 3.4 displays the average recorded level for total coliforms in the sampling years 2004 to 2015. The highest average concentrations occurred in the second half of the year from June to November with November having the highest average concentrations in all years. Similar results were seen for average faecal coliform (Table 3.5) with higher averages occurring in June to December and November being the month with the highest averages. The lowest concentrations were predominately found in April and May.

Table 3.4: Average total coliform concentrations by month (results displayed in MPN/100 mL).

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2004											11377	8926
2005	8132	8864	1262	1038	5701	454	592	2135	6746	19499	20393	18818
2006	2304	2805	1459	469	534	3247	770	1730	6648	21061	4130	2571
2007	3482	7280	1434	254	1864	22158	14160	1776	2473	824	8055	1734
2008	1452	483	8820	793	1813	16037	1676	20906	10233	4516	3476	1157
2009	555	14974		3242	436	1340	10128		971	325		9403
2010	823	3642	1076	114	399	408	800		3795	497	440	
2011			346		495			41			31768	
2012		41			274			2798	1119	849	915	
2013		79			1128	210	130	489	2437	10938	37367	
2014	3430	5277	924	4215	3074	1965	766	2888		712		
2015		675										
Total Average	2883	4412	2189	1447	1572	5727	3628	4095	4303	6580	13102	7101

Table 3.5: Average faecal coliform concentrations by month (results displayed in MPN/100 mL).

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2004		463		244							8449	1833
2005	1806	277	783	36	878	79	65	225	3919	2020	1877	6621
2006	401	3578	283	23	53	132	78	181	1248	4396	616	386
2007	1067	53	190	121	28	4570	2713	117	94	115	1125	253
2008	397	3535	720	533	376	1435	183	1716	3109	473		67
2009	66	804		24	99	166	1730		186	117		2353
2010	135		466		94	58	296		595	66	313	
2011			41		105			20			104	
2012		20			178			319	148	71	5733	
2013		41			125	60	46	161		4017	229	
Total Average	645	1096	414	163	215	929	730	391	1328	1410	2306	1919

3.3.1.3 Site correlation.

Correlation between each site was analysed to determine if there was any trend between site location and bacterial concentration or if the sites behave independently of each other. The migration of bacterium, between the sites was determined by examining the correlation and slope between the sites. A high correlation signified that the bacterium level at one site had a strong relationship with the next site, a low correlation signified that the two sites were not strongly related and what is happening at one site does not give any indication to what is happening at the corresponding site. A slope close to 1 indicates that the bacterial concentration levels between the sites are changing at similar rates. A slope of exactly 1 indicates that changes in bacterial concentration happens at the exact same rate. A slope much lower than 1 indicates that the rate of change of bacterial concentration is faster for the site on the x-axis than the site on the y-axis. A slope much higher than 1 indicated the rate of change of bacterial concentration is faster for the site on the y-axis than the site on x-axis.

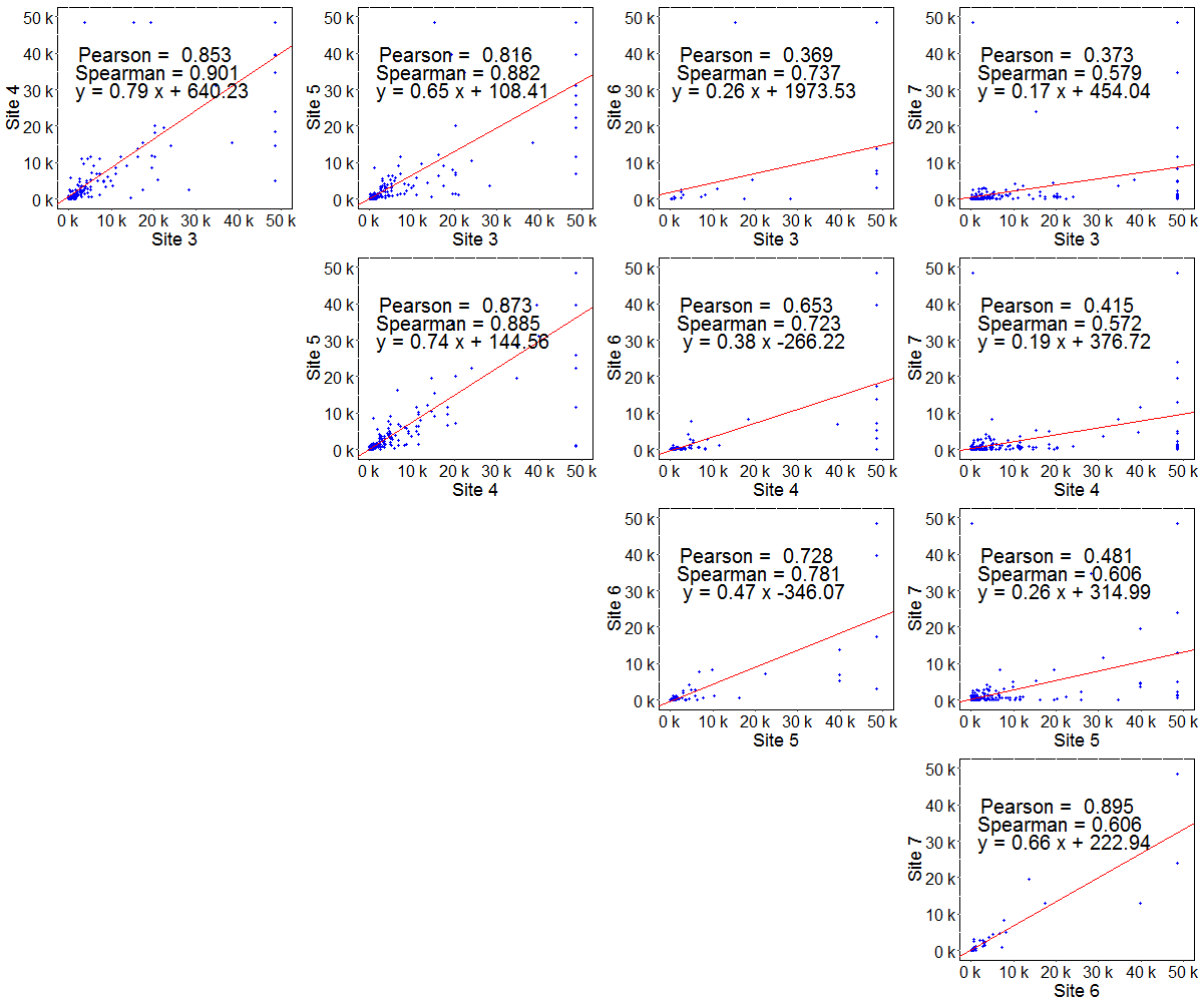

Figure 3.4: Correlation results for total coliform concentrations at sites 3 to 7 all results displayed are in MPN/100 mL.

In Figure 3.4 the sites closer in location to each other had the greatest Pearson and Spearman correlation results. A greater Spearman correlation than Pearson suggests the relationship between the data is positively monotonic meaning there is an increase at one site where the corresponding site increases at a different rate. A fall in slope occurs relative to increase in distance between the sites suggesting that it takes some time for the bacteria to reach the same concentration in the sites furthest from the outfall.

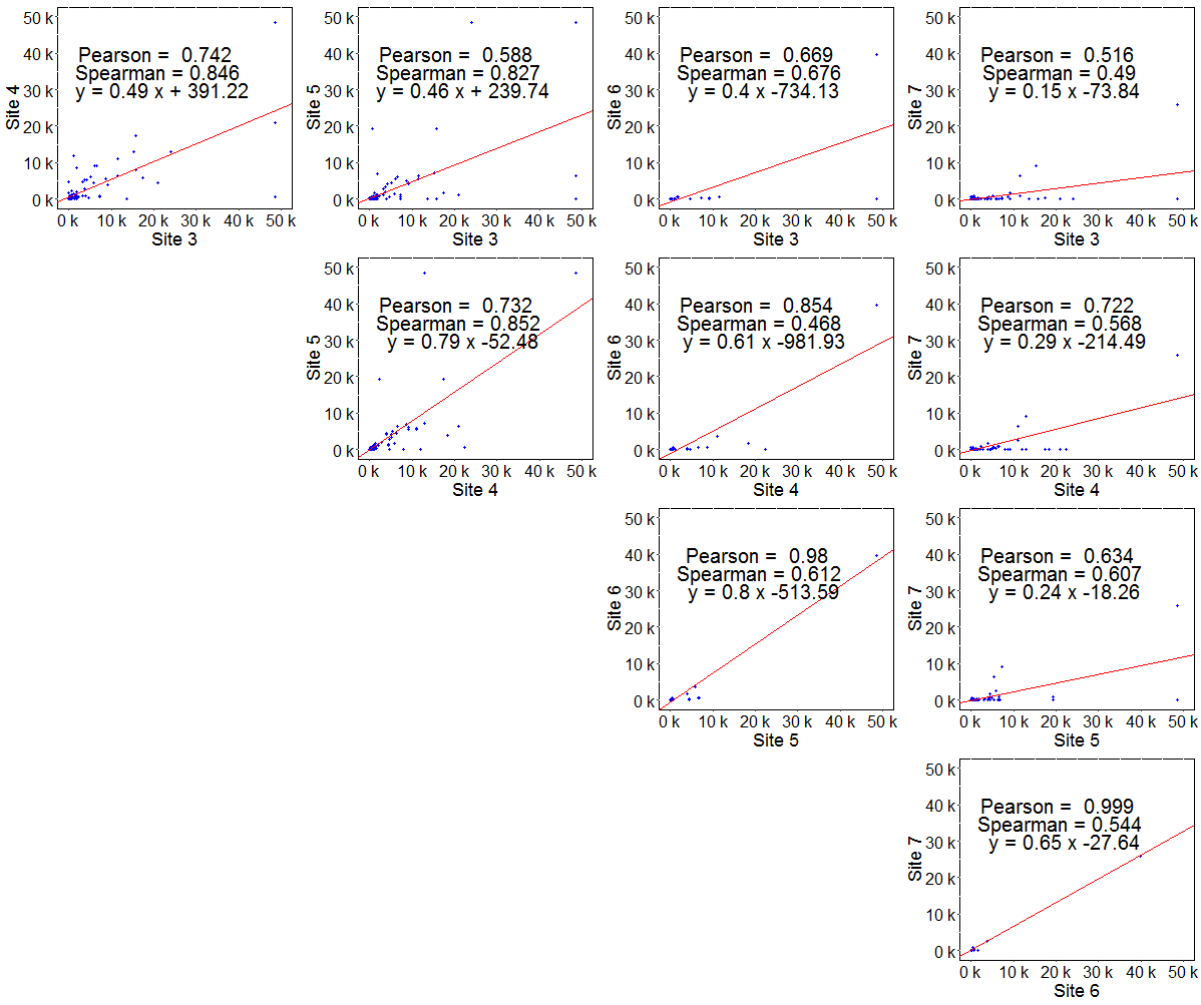

Figure 3.5: Correlation results for faecal coliform level at sites 3 to 7 all results displayed are in MPN/100 mL.

The correlation results between faecal coliform level at each site show similar trends to the results for total coliform levels (Figure 3.5). Here the Pearson and Spearman correlation results are still greatest for sites in closer proximity. Again, the slope of each line decreases as the sites get further apart but not at as high a rate as for total coliform. This means that the levels of faecal coliform at each site are changing at a lower rate than for total coliform levels.

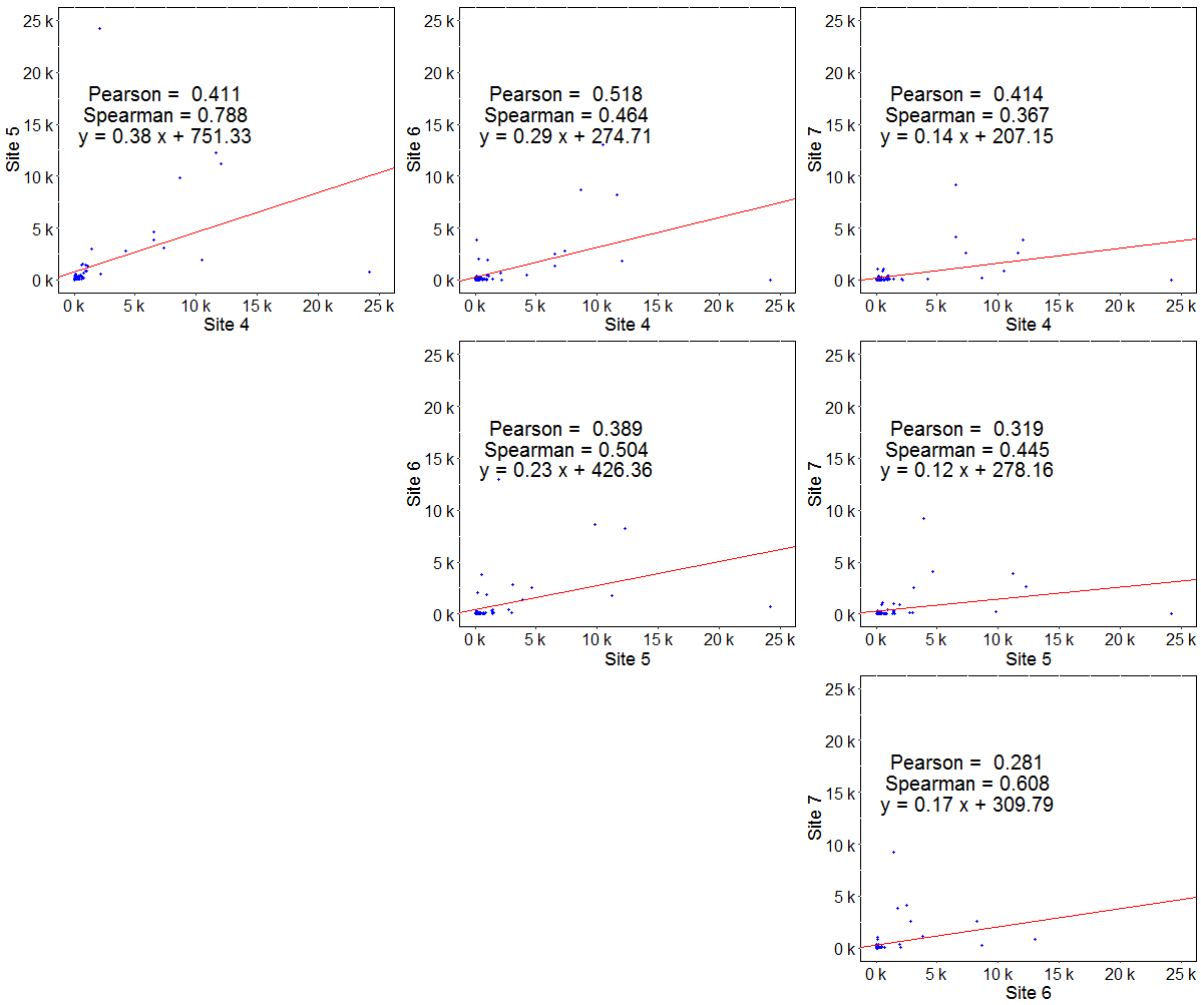

Figure 3.6: Correlation results for E. coli levels at sites 4 to 7 all results displayed are in MPN/100 mL.

The correlation between sites results for E. coli levels are overall weaker than for faecal coliform or total coliform (Figure 3.6). The correlation results here are moderate meaning more outliers are present for E. coli levels and there is no clear relationship between the sites. However higher levels were found to occur at sites closest to the storm water overflow as shown in Figure 3.1.

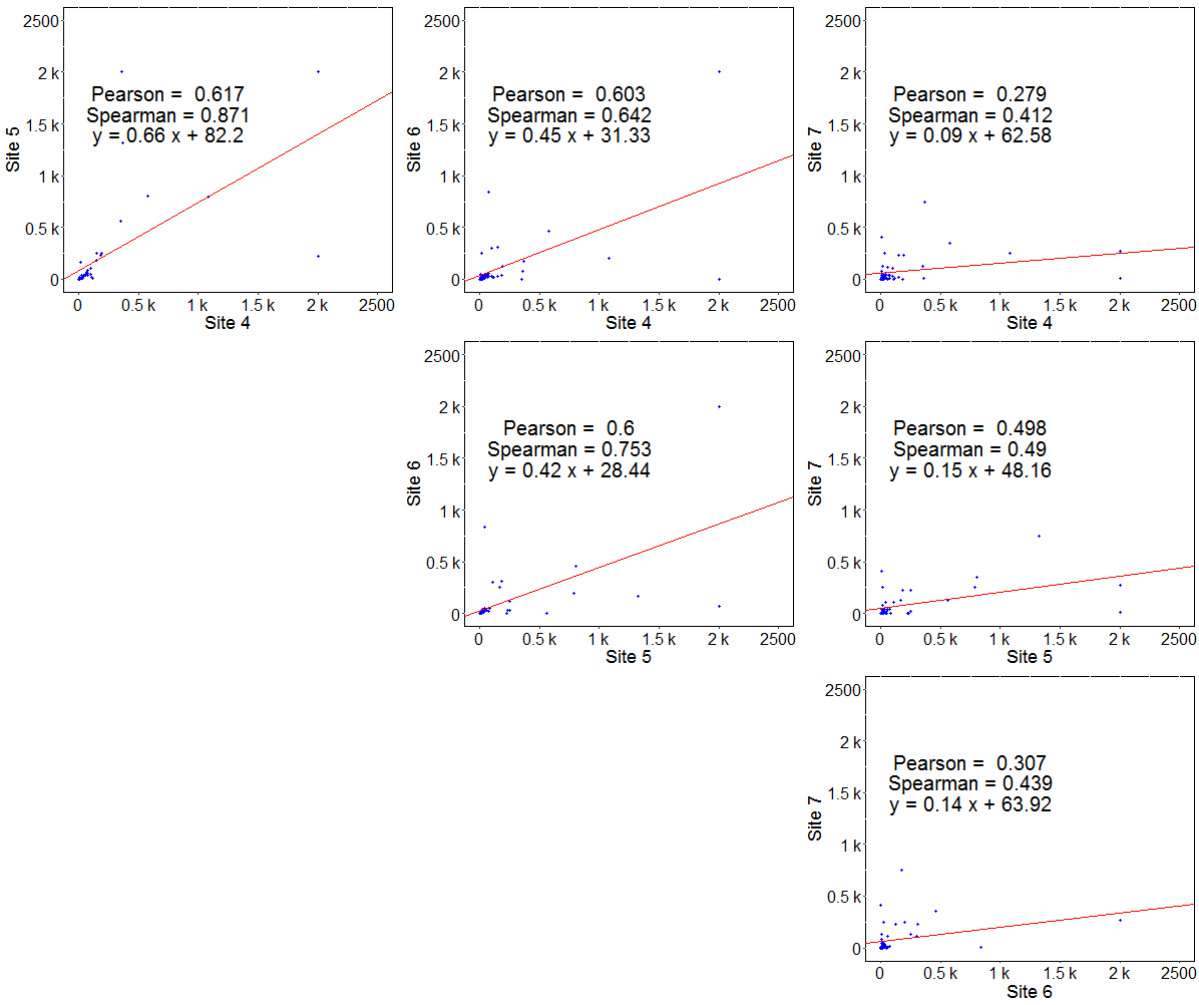

Figure 3.7: Correlation results for Enterococci levels sites 4 to 7 all results displayed are in MPN/100 mL.

It was found that enterococci had similar correlation strengths to that found for E. coli (Figure 3.7). It was found that site 7 had a greater correlation with site 5 than site 6. This illustrated that enterococci do not follow the same pattern of sites with closest distance having the greatest correlation. The slope decreases for sites that are further away from each other meaning rate of change between the sites become less as the sites furthest from the storm water outfall do not reach the same heights as the sites closest on a given sampling day.

3.3.1.4 Statistical tests

To determine if the samples from the sites were from the same population density a two sample Kolmogorov-Smirnov test (KS-test) was performed (Lopes, 2011). The KS-test is a non-parametric test that makes no assumption about the distribution of

the data, so it can be used for not normally distributed data. It is a robust test that can handle highly non-normal data. It compares the two empirical distribution functions. That is,

$$D=\frac{\sup}{i}|E_1(i)-E_2(i)|$$

where E_1 and E_2 are the empirical distribution functions for the two samples and 'sup' is the supremum function. Both E_1 and E_2 are computed at each point in each sample.

The null and accepted hypothesis are as follows:

H_0: The two samples come from the same population density.

H_a: The two samples do not come from the same population density.

The KS-test uses the maximum vertical deviation between the two curves as the statistic D. The null hypothesis is rejected at a level of α if D > p where p = $(\alpha)\sqrt{\frac{n+m}{nm}}$, n and m are the sizes of the first and second sample respectively and c(α) = $\sqrt{-\frac{1}{2}\ln\frac{\propto}{2}}$. The value of alpha uses on this experiment was 0.05 for 95% confidence level.

Table 3.6: Kolmogorov-Smirnov test results for total coliform at sites 3 to 7.

	Site 4	**Site 5**	**Site 6**	Site 7
Site 3	D=0.066313 p=0.7809	D=0.16567 p=0.009543	D=0.35798 p=1.754e-05	D=0.4176 p=2.22e-15
Site 4		D=0.12021 p=0.08696	D=0.32697 p=7.423e-05	D=0.37721 p=4.741e-14
Site 5			D=0.29295 p=0.0005741	D=0.33991 p=2.187e-11
Site 6				D=0.14968 p=0.2339

p values greater than D (in green) are where the null hypothesis has been accepted and this shows that the two samples come from the same population density.

The results for total coliform determined that site 3 and 4 were from the same population (Table 3.6). The only other two sites determined to be from the same population were sites 6 and 7. Sites 3 and 4 are the sites located closest to the storm water outfall, so it would suggest that they are being affected by sewage from the outfall on a similar scale. Site 6 and 7 are furthest from the outfall. They have been shown to have lower total coliform levels and fewer breaches. If coliforms being brought along by the water's current are reaching site 6 and 7 at a similar time this may be the reason for their levels to be from the same population.

Table 3.7: Kolmogorov-Smirnov test results for faecal coliform at sites 3 to 7.

	Site 4	**Site 5**	**Site 6**	**Site 7**
Site 3	D=0.095792 p=0.3589	D=0.18773 p=0.002975	D=0.38291 p=0.0002056	D=0.48114 p=2.2e-16
Site 4		D=0.13255 p=0.06622	D=0.32626 p=0.002281	D=0.44818 p=2.2e-16
Site 5			D=0.21519 p=0.106	D=0.33619 p=5.106e-10
Site 6				D=0.33986 p=0.001259

p values greater than D (in green) are where the null hypothesis has been accepted and this shows that the two samples come from the same population density.

For faecal coliform only site 3 and 4 were found to be from the same population (Table 3.7). These are the two sites closest to the storm water outfall and therefore may be influenced by the outfall at a similar enough scale to be counted as being from the same population.

Table 3.8: Kolmogorov-Smirnov test results for E. coli at sites 4 to 7.

	Site 5	**Site 6**	**Site 7**
Site 4	D=0.33986 p=0.9566	D=0.4058 $p=2.325e^{-05}$	D=0.47826 $p=2.797e^{-07}$
Site 5		D = 0.37681 p = 0.0001112	D=0.52174 $P=1.393e^{-08}$
Site 6			D=0.21739 p=0.07671

p values greater than D (in green) are where the null hypothesis has been accepted and this shows that the two samples come from the same population density.

For E. coli sites 4 and 5 were the only sites found to be from the same population. These sites are the only sites in the upper basin with site 4 being the closest to the storm water outfall (Table 3.8). Sites in the lower basin were not found to be from the same population as sites in the upper basin. Each site in the lower basin was not found to be from the same population. This suggests E. coli levels are more diversely spread in the lower basin.

Table 3.9: Kolmogorov-Smirnov test results for Enterococci at sites 4 to 7.

	Site 5	**Site 6**	**Site 7**
Site 4	D=0.1087 p=0.9487	D=0.26087 p=0.08739	D=0.23913 p=0.144
Site 5		D=0.15217 p=0.6612	D=0.21739 p=0.2271
Site 6			D=0.17391 p=0.4899

p values greater than D (in green) are where the null hypothesis has been accepted and this shows that the two samples come from the same population density.

For Enterococci all sites apart from site 4 were found to be from the same population (Table 3.9). Site four was only found to be from the same population as site 5; all sites further down the basin were found not to be from the same population as site 4. Site 4 is closest to the storm water outfall and it may be the case that the influence of bacterial influx here is too strong to be from the same population distribution than sites further downstream.

3.3.2 Analysis of site data with rainfall

3.3.2.1 Determination of optimal rain gauge

Eight rain gauges ranging from 1 to 34 km distance from sample site were compared to find the optimum rain gauge. Rainfall measurements were taken in total mm per day from 2004 to 2016. A Pearson's correlation test was performed to determine if any rain gauges were significantly different from one another. All rain gauges with a 0.8 or higher Pearson's correlation were averaged, and the result was correlated with the microbiological data from the sample sites. A level of 0.8 or higher Pearson's correlation was chosen as it shows a very strong correlation between the results. The results of the rain gauges Pearson's test can be seen in Figure 3.8.

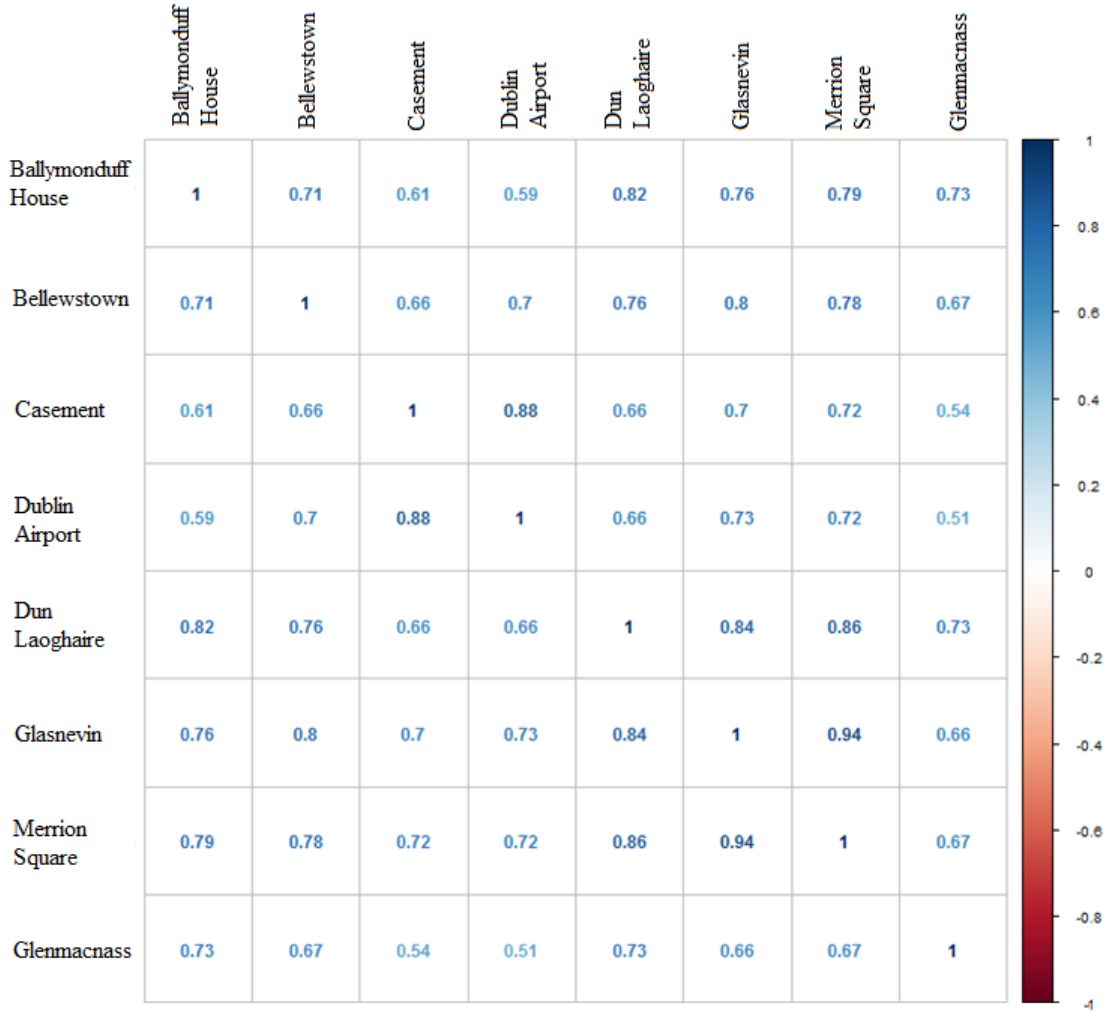

Figure 3.8: Pearson results for eight rain gauges surrounding the Grand Canal Basin.

Each rain gauge was correlated against site 4 individually. The averages of the rain gauges with Pearson's correlation of above 0.8 were also correlated against each site.

The correlation results for site 4 and total coliform samples can be seen in Table 3.10. Site 4 was chosen as it had the greatest amount of observations over the sampling period.

Table 3.10: Correlations calculated between total coliform concentrations, individual rain gauges and average rain gauges at site 4.

Rainfall Site	Distance from Site 4 (km)	Slope	Pearson	Spearman
Bellewstown (BN)	34	835.01	0.81	0.58
Dublin Airport (DA)	9.16	480.1	0.49	0.62
Glasnevin (GV)	3.8	707.23	0.78	0.65
Merrion Square (MS)	1	774.38	0.78	0.72
Casement (CT)	13.9	363.87	0.37	0.64
Ballyedmonduff House (BF)	12	505.06	0.78	0.68
Dun Laoghaire (DL)	9.7	781.42	0.77	0.59
Glenmacnass	32.2	235.87	0.49	0.51
MS DL GV	_	799.99	0.8	0.66
MS DL	_	808.67	0.79	0.67
MS GV	_	759.21	0.79	0.7
DL GV	_	796.83	0.8	0.66
BN GV	_	825.33	0.82	0.66
CT DA	_	447.36	0.44	0.65
DL BF	_	378.34	0.36	0.64

Merrion Square showed very strong Pearson and Spearman results. However, the combination of Glasnevin, Merrion Square, and Dun Laoghaire was chosen as is

showed a very strong Pearson's result and a strong Spearman's result whilst giving a larger rainfall span that may affect the catchment (Table 3.10). It was decided that it was best to take the largest rainfall span so all rainfall that may affect the catchment would be included. Dun Laoghaire and Glasnevin showed very strong correlations with the basin when tested as individual sites.

3.3.2.2 Rainfall correlation results for individual sites.

To determine how many days of rainfall had the most influence on the indicator bacteria in the basin a correlation between different rainfall days prior to sampling was analysed. Finding the greatest correlation can be beneficial in choosing how many days after rainfall occurrences should the basin be monitored in the future.

The optimum rain gauge chosen, the average of Merrion Square, Glasnevin, and Dun Laoghaire, was the correlated against each individual Site. To determine how many days prior to sampling gave the best correlation site 4 was correlated against different days and different averages of days, these included rainfall taken: 1 day prior to sampling; two days prior to sampling; 3 days prior to sampling; 4 days prior to sampling; the average of 2 days prior to sampling; the average of 3 days prior to sampling; the average of 4 days prior to sampling. Site 4 was chosen as it was used for all four microbiological measurements and it has higher incidences of pollution. A student's t-test was performed to test if the result was statistically valid within in a 95% confidence rate (Kim, 2015). The null hypothesis of no correlation in the population was tested. A p-value greater than the alpha value of 0.05 accepts the null hypothesis.

Table 3.11: Correlation results between days of rainfall and total coliform concentrations at site 4.

Rainfall Prior to Sample Collection	Pearson	p	Spearman	p
One Day	0.495	$6.17E^{-15}$	0.479	$5.71E^{-14}$
Two Days	0.353	$8.12E^{-08}$	0.375	$9.79E^{-09}$
Three Days	0.193	0.004174	0.326	$8.38E^{-07}$
Four Days	-0.006	0.9338	0.081	0.231
Average of Two Days	0.551	$< 2.2E^{-16}$	0.532	$< 2.2E^{-16}$
Average of Three Days	0.513	$4.52E^{-16}$	0.522	$< 2.2E^{-16}$
Average of Four Days	0.452	$1.91E^{-12}$	0.505	$1.50E^{-15}$

The optimal rainfall chosen is in yellow. T-test that gave a result of no correlation within a 95% confidence level is in red.

For total coliform the greatest correlation was found to be the average of two days prior to sample collection (Table 3.11). The null hypothesis for these correlation result was rejected, and the resulting correlation coefficient can be said to be statistically valid up to a 95% confidence level. The rainfall was averaged against all sites for total coliform.

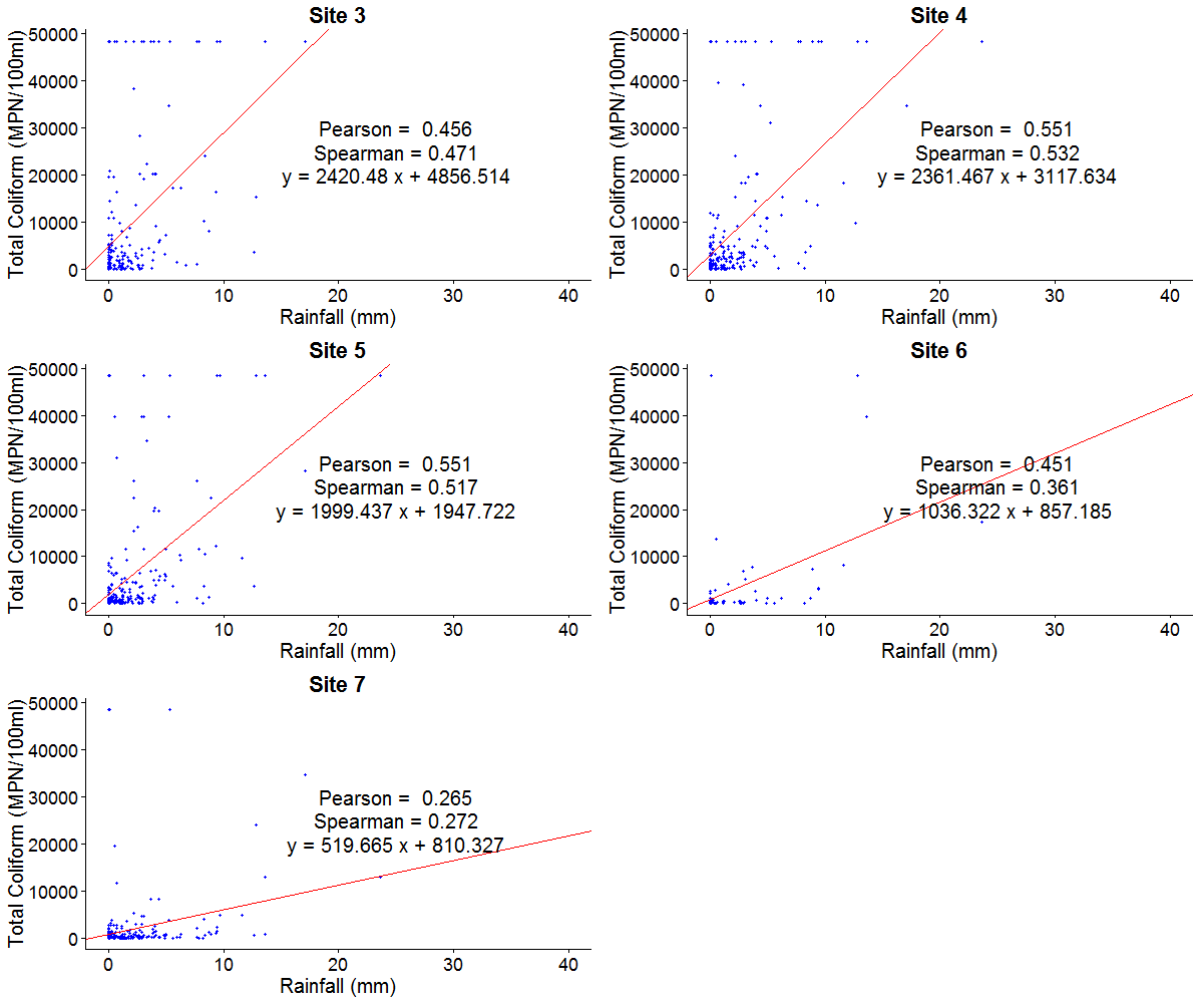

Figure 3.9: Correlation of total coliform levels at each site with rainfall from the average of the two days prior to sampling.

A moderate Pearson and Spearman correlation was found for total coliform at sites 3, 4, and 5 (Figure 3.9). Site 6 had moderate Pearson correlation but a weak Spearman correlation suggesting that relationship is not monotonic. Site 7 has both a weak Pearson and Spearman correlation. A stronger correlation for sites 3, 4, and 5, support the: Kolmogorov-Smirnov test results that these sites are closely related (Table 3.6). Sites 3, 4, and 5 also show similar slopes in response to the rainfall. Site 6 and 7 have lower slopes in response to rainfall. The slopes for each site decreases from Site 3 to Site 7 as they became further away from the storm water outfall.

Table 3.12 Correlation results between days of rainfall and faecal coliform concentrations at Site 4.

Rainfall Prior to Sample Collection	Pearson	p	Spearman	p
One Day	0.29	3.81E-05	0.46	1.21E-11
Two Days	0.202	0.004493	0.348	6.02E-07
Three Days	0.087	0.2266	0.277	8.53E-05
Four Days	-0.087	0.2252	0.08	0.267
Average of Two Days	0.321	4.36E-06	0.507	3.32E-14
Average of Three Days	0.294	2.90E-05	0.494	1.81E-13
Average of Four Days	0.218	0.002142	0.487	4.31E-13

The optimal rainfall chosen is in yellow. T-test that gave a result of no correlation within a 95% confidence level is in red.

The greatest rainfall correlation result for faecal coliform was with the average of two days prior to sampling. It achieved a weak Pearson correlation and a moderate Spearman correlation (Table 3.12). The null hypothesis for these correlation result was rejected, and the resulting correlation coefficient can be said to be statistically valid up to a 95% confidence level. Faecal coliform and rainfall display a more monotonic relationship than direct correlation with rainfall meaning it needs a very high rainfall volume to display an increase in faecal coliform levels.

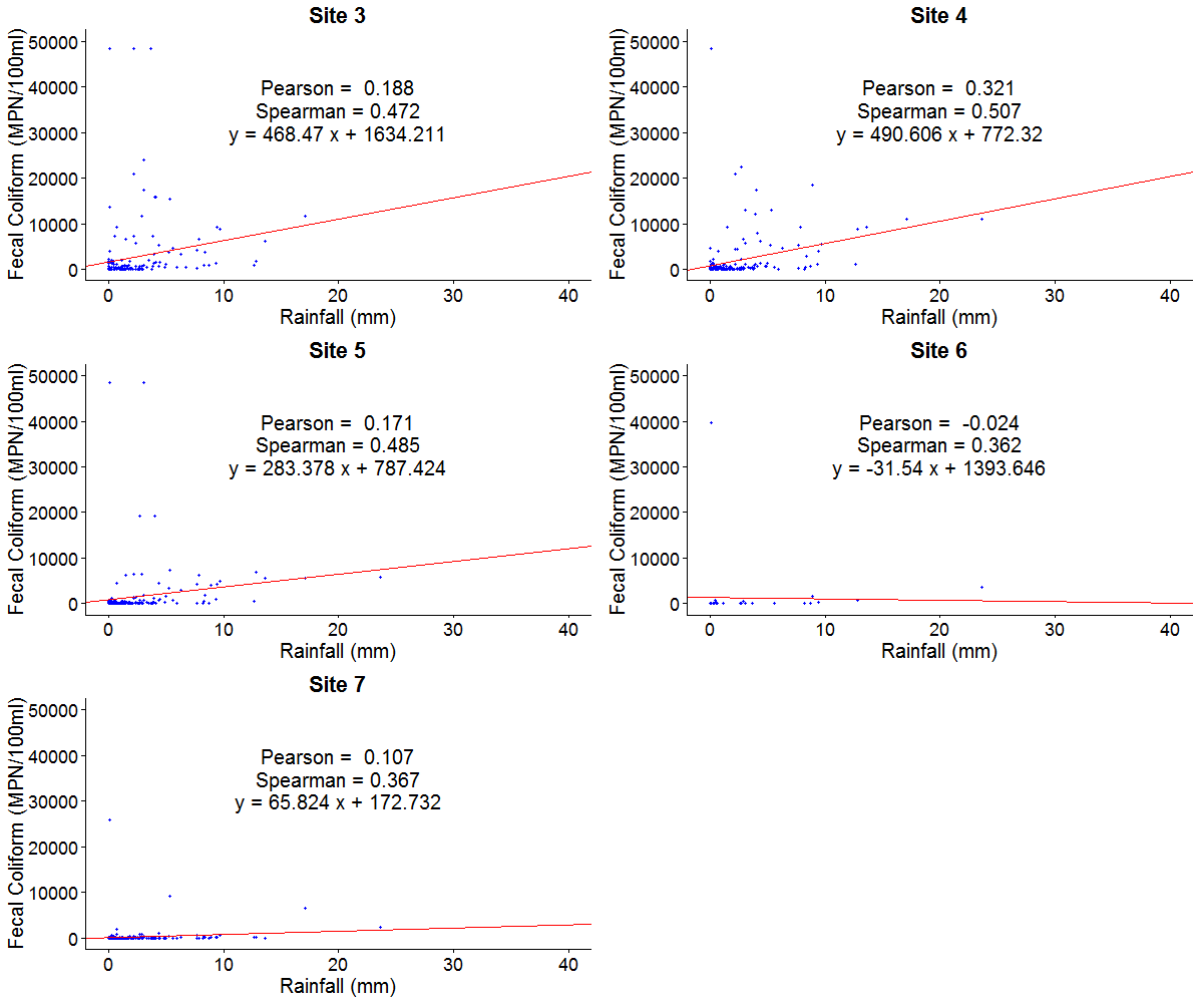

Figure 3.10: Correlation of faecal coliform levels at each site with rainfall from the average of the two days prior to sampling.

A moderate Spearman correlation was found at sites 3, 4, and 5 meaning these sites have a monotonic relationship with rainfall and therefore there is a ranked level relationship but not a direct increase relationship (Figure 3.10). The Pearson correlation at sites 3, 4, and 5 were weak and thus the results at these sites show no direct correlation with rainfall. Sites 6 and 7 show a weak Spearman correlation and a very weak Pearson correlation with rainfall and therefore it can be said these sites are not impacted by the average of two days rainfall prior to sampling.

Table 3.13: Correlation results between days of rainfall and E. Coli concentrations at Site 4.

Rainfall Prior to Sample Collection	Pearson	p	Spearman	p
One Day	0.436	0.0001808	0.504	1.02E-05
Two Days	0.285	0.01752	0.287	0.01684
Three Days	0.022	0.8555	0.201	0.09793
Four Days	0.441	0.0001517	0.195	0.1092
Average of Two Days	0.513	6.43E-06	0.502	1.12E-05
Average of Three Days	0.378	0.001345	0.382	0.001192
Average of Four Days	0.46	6.91E-05	0.402	0.000613

The optimal rainfall chosen is in yellow. T-test that gave a result of no correlation within a 95% confidence level is in red.

The greatest Pearson and Spearman correlations were found from the average of two days prior to sampling for E. coli levels at site 4 (Table 3.12). The null hypothesis for this correlation result was rejected, and the resulting correlation coefficient can be said to be statistically valid up to a 95% confidence level. A moderate Pearson and Spearman was achieved for this type of rainfall.

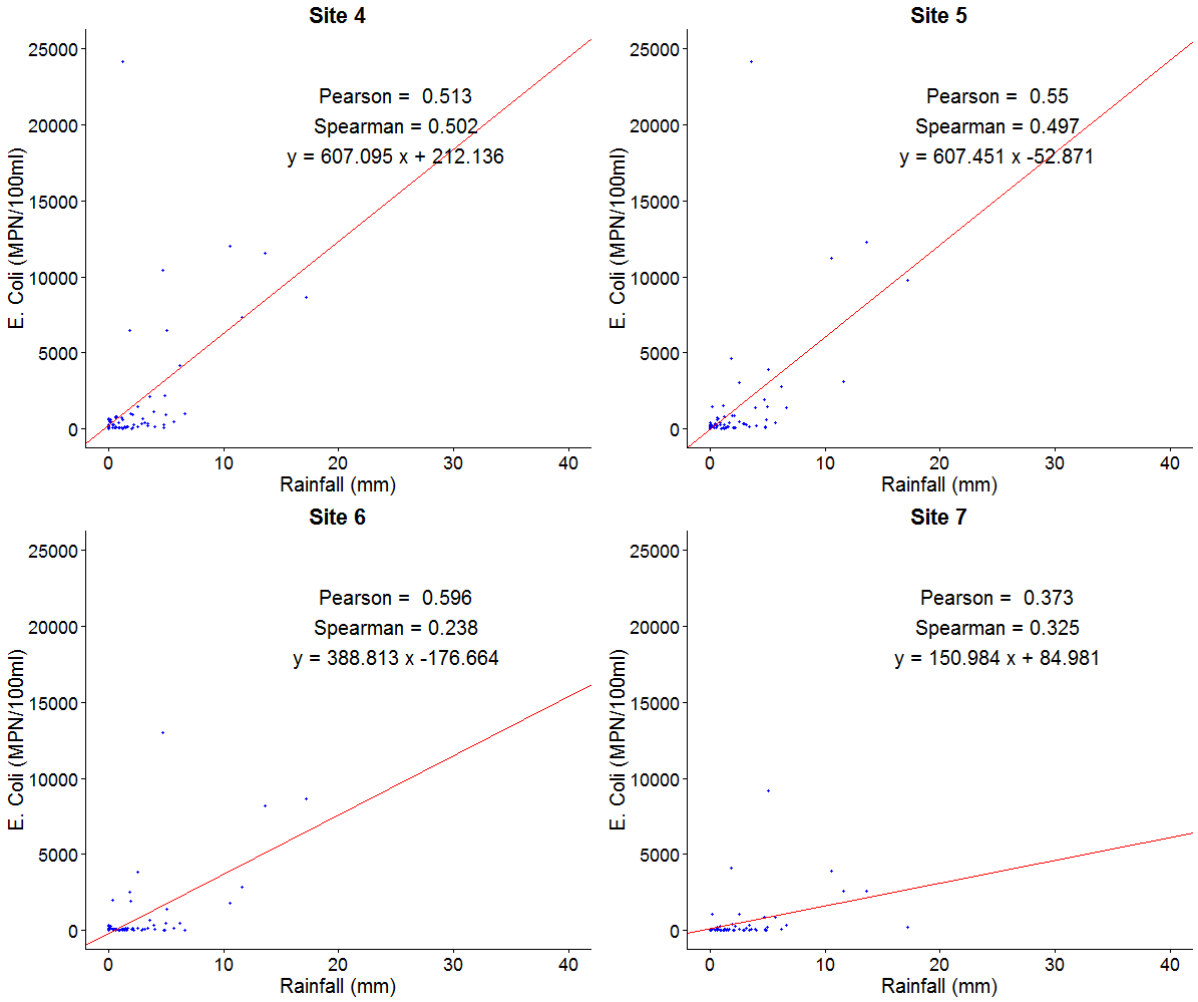

Figure 3.11: Correlation of E. coli levels at each site with rainfall from the average of the two days prior to sampling.

The greatest Pearson and Spearman result was found and sites 4 and 5 for E. coli; a moderate result was achieved for both (Figure 3.11). The slopes for these two sites were very similar suggesting the rate of increase in E. coli at site 4 is very close to the rate of increase of E. coli at site 5. Sites 4 and 5 were also found to be from the same population from the Kolmogorov–Smirnov test results suggesting the impact of E. coli levels in the upper part of the basin is happening rapidly (Table 3.8). Site 6 gave a moderate Pearson result and a weak Spearman result.

Table 3.14: Correlation results between days of rainfall and Enterococci concentrations at Site 4.

Rainfall Prior to Sample Collection	Pearson	p	Spearman	p
One Day	0.615	$5.34E^{-06}$	0.465	0.001142
Two Days	-0.027	0.8601	0.217	0.14700
Three Days	-0.109	0.4719	0.283	0.05678
Four Days	0.212	0.1578	0.332	0.02415
Average of Two Days	0.572	$3.23E^{-05}$	0.449	0.00175
Average of Three Days	0.376	0.009962	0.404	0.00534
Average of Four Days	0.375	0.01029	0.465	0.00113

The optimal rainfall chosen is in yellow. T-test that gave a result of no correlation within a 95% confidence level is in red.

The greatest correlation result for Enterococci was with rainfall taken one day prior to sampling (Table 3.14). The null hypothesis for this correlation result was rejected, and the resulting correlation coefficient can be said to be statistically valid up to a 95% confidence level. A strong Pearson and moderate Spearman correlation was achieved for this type of rainfall.

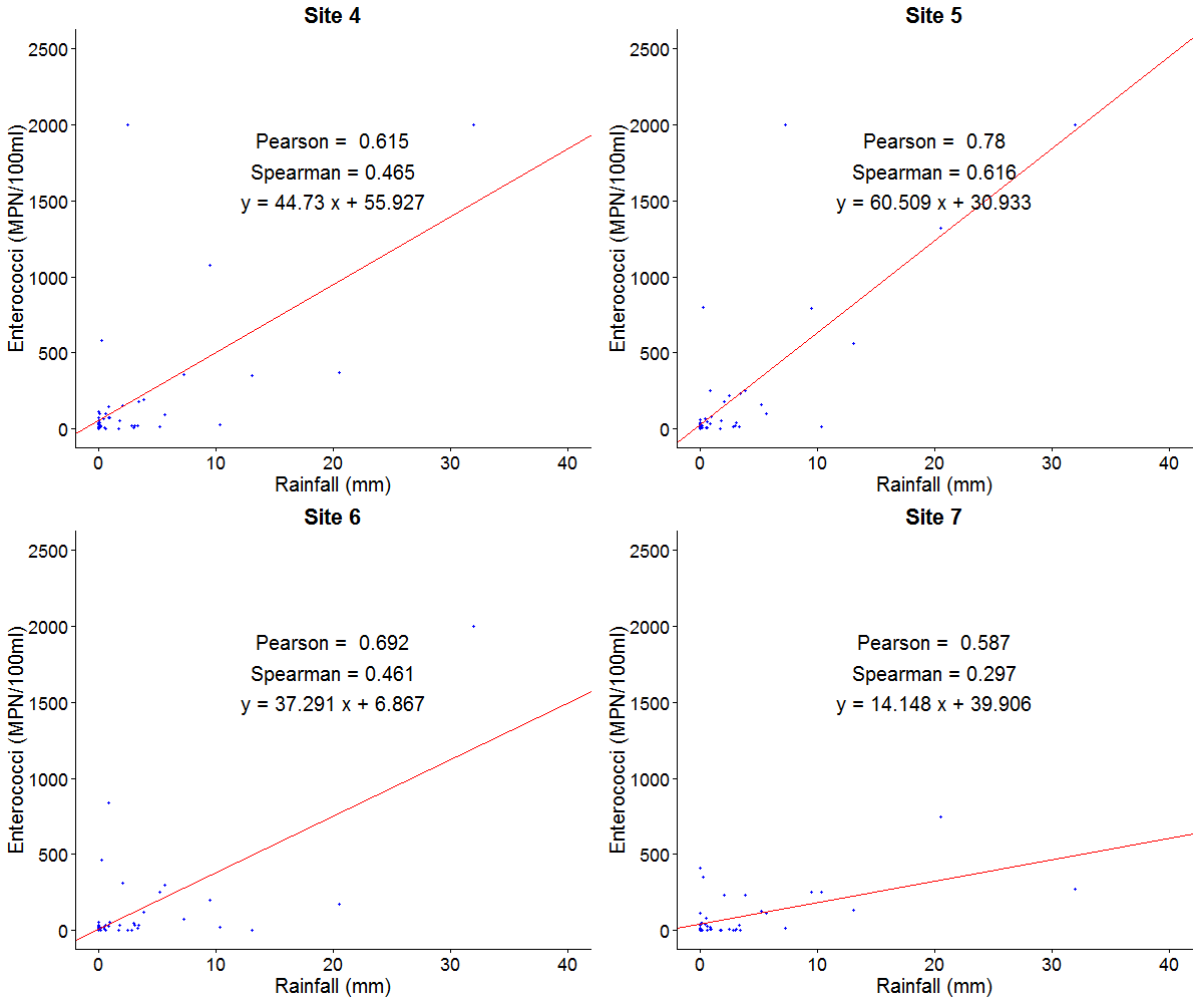

Figure 3.12: Correlation of Enterococci levels at each site with rainfall from one day prior to sampling

A strong Pearson correlation between rainfall and Enterococci was achieved at site 4, 5, and 6 (Figure 3.12). A moderate Spearman correlation was achieved for sites 4, and 6, with a strong Spearman correlation found at Sit 5. Site 7 had a moderate Pearson and weak Spearman result and shows the poorest correlation to rainfall. Site 5 was found to have the highest slope and correlation results which means it reacts to rainfall quicker than that of Site 4, which has a lower slope, this differs from other bacterial results where the sites closer to the storm water outfall were rising quicker. The fact that no statistical differences were found between any of the Enterococci results could mean that diffusion of Enterococci happens at a greater rate than other microorganisms tested in the basin.

3.3.2.3 Defining rainfall conditions

Different rainfall categories were used to assess how different lengths of rainfall affected mean bacterial levels. The rainfall levels used were $x < 2.5$ mm, $2.5 \text{ mm} > x < 6.4$ mm, $6.4 \text{ mm} > x < 25$ mm, and $x > 25$ mm in a 24-hour period. The rainfall categories were taken from a similar study on beach bacterial concentrations on Santa Monica Bay beaches (Ackerman and Weisberg, 2003). All sites do not contain results for each rainfall condition due to the different frequencies of samples taken.

3.3.2.4 Average bacterial levels under different rainfall conditions with respect to days from rainfall

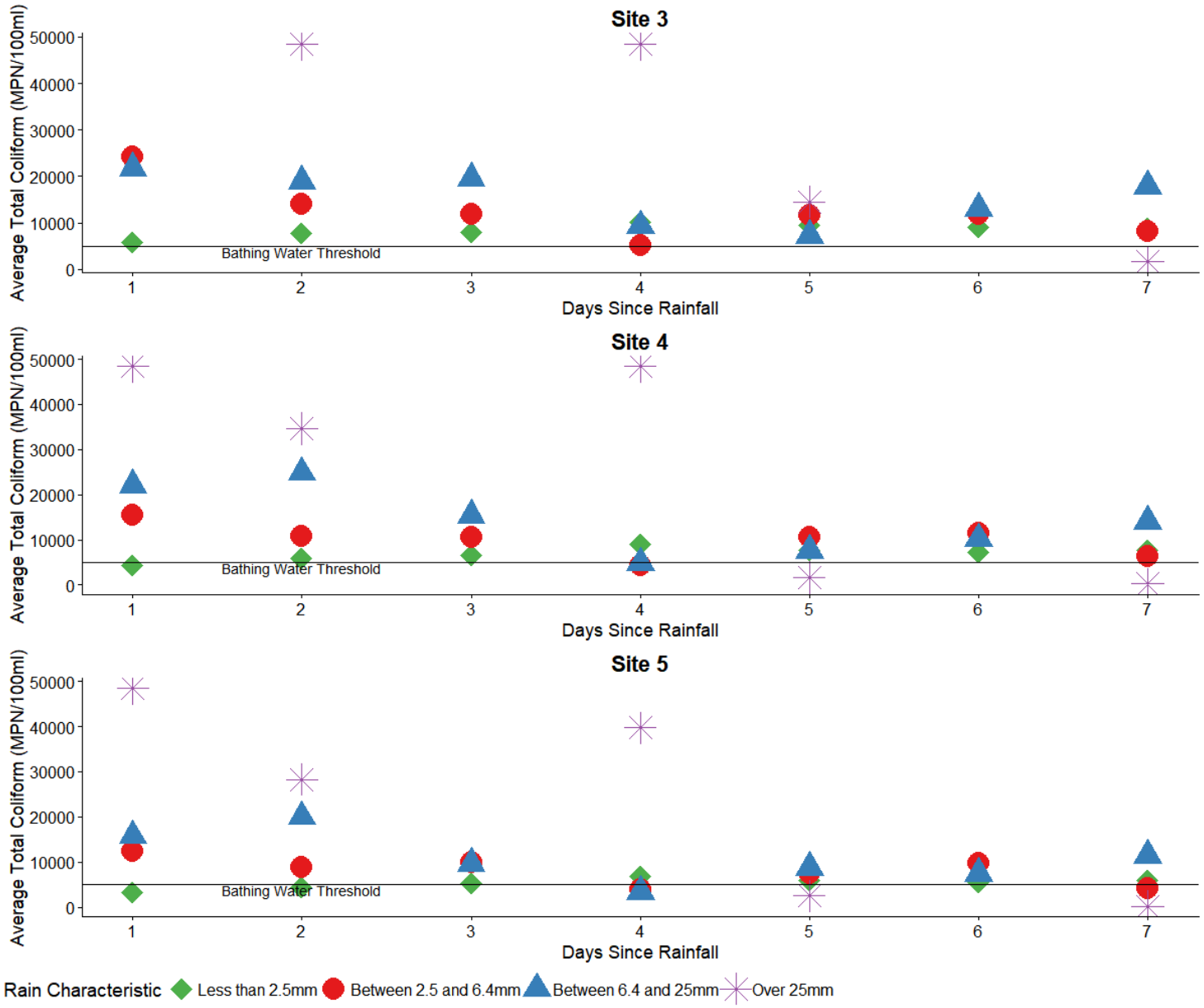

Figure 3.13: Average total coliform levels under different rainfall conditions with respect to days since rainfall at sites 3, 4, and 5 located in the upper level of the Grand Canal Basin.

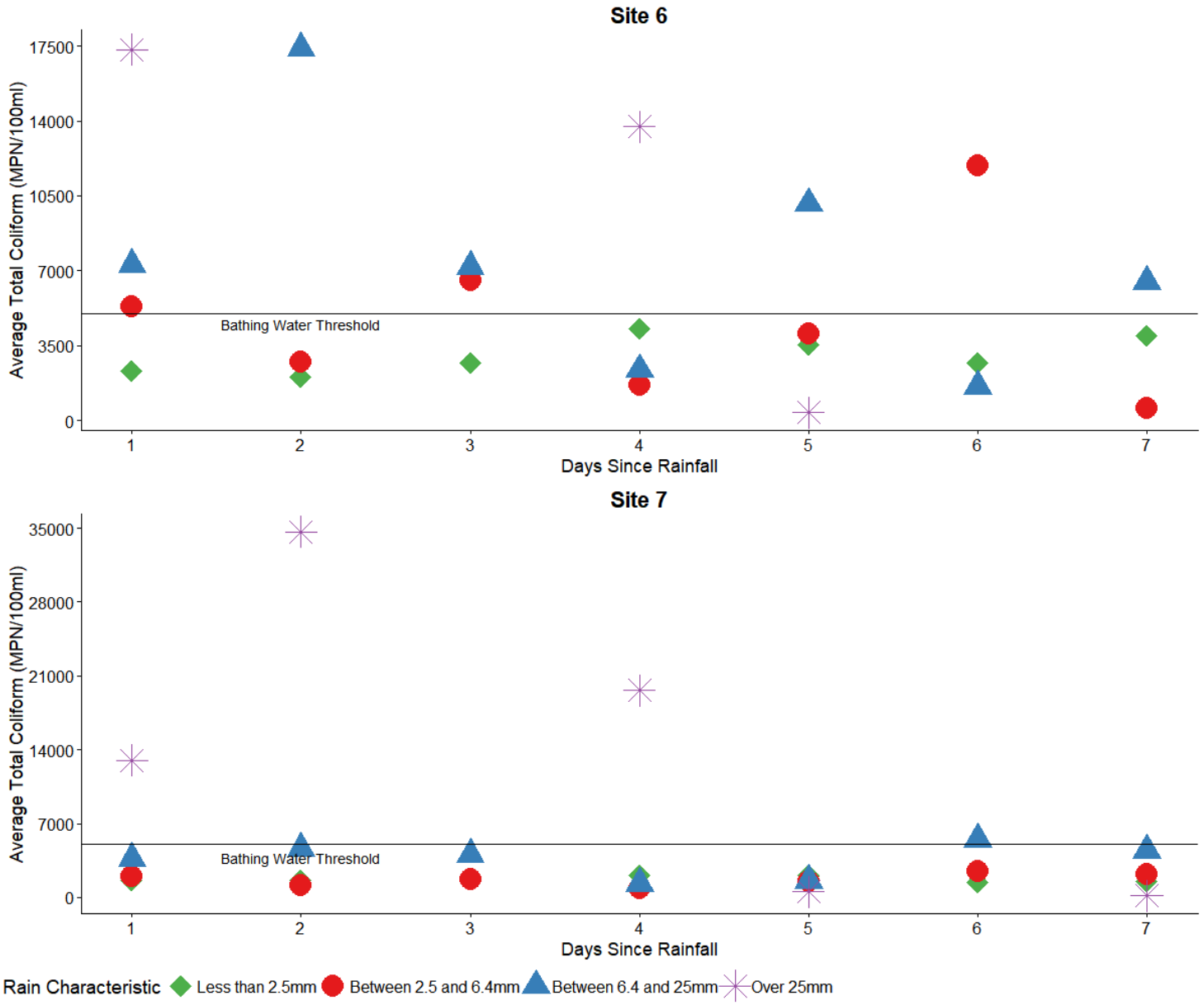

Figure 3.14: Average total coliform levels under different rainfall conditions with respect to days since rainfall at sites 6 and 7 located in the lower level of the Grand Canal Basin.

For large rainfall events (x > 25 mm) the total coliform is highest after 1 day in sites 4, and 5 (Figure 3.13). There is no first day data for Site 3 however there is a clear difference between results for sites 4 and 5, in the upper basin, and sites 6 and 7 (Figure 3.14), in the lower basin which have a much lower response in the first day. It is still however the highest level seen at sites 6 and 7 for a 1st day response after a rainfall event. A larger second day response is seen at site 7 which could be due to the flow of water through as high levels at site 4 and 5 are carried towards site 7. For

every site, except site 3, total coliform levels stay above the bathing water threshold for up to 5 days after the heavy (x > 25 mm) rainfall event. Site 3 which is closest to the storm water outfall does not have a recorded drop till 7 d after the event. For rainfall events that are between 6.4 mm > x < 25mm there is a slightly higher total coliform response on the second day for site 4, 5, and 6. Site 7 which is the furthest from the storm water outfall is only considerably affected by heavy rainfall (x > 25 mm). The second day response may be due to the lower flow rate trough the storm water outfall taken longer to disperse through the basin. Site 3 has the highest reaction on the 1st day after rainfall events that were between 6.4 mm > x < 25 mm and between 2.5 mm > x < 6.4 mm. These results show that when the rainfall is greater than 2.5 mm there is an large impact on the water quality at site 3. Site 3 is the site which is closest to the storm water outflow and was expected to have the greatest impact from low rainfall amounts. Site 4 and 5 also had the highest response for rainfall events between 2.5 mm > x < 6.4 mm on the first day but this result was lower than the impact of rainfall between 6.4 mm > x < 25 mm. Site 6 was not impacted hugely by rainfall events between 2.5mm > x < 6.4mm and results tended to stay below or close to the bathing water threshold. Some late day rises, day 5 to 7, at site 6 and 7 may also be due to the diffusion of coliforms through the water from the higher impacted sites near the storm water outfall. No sites were greatly impacted by rainfall events below 2.5 mm.

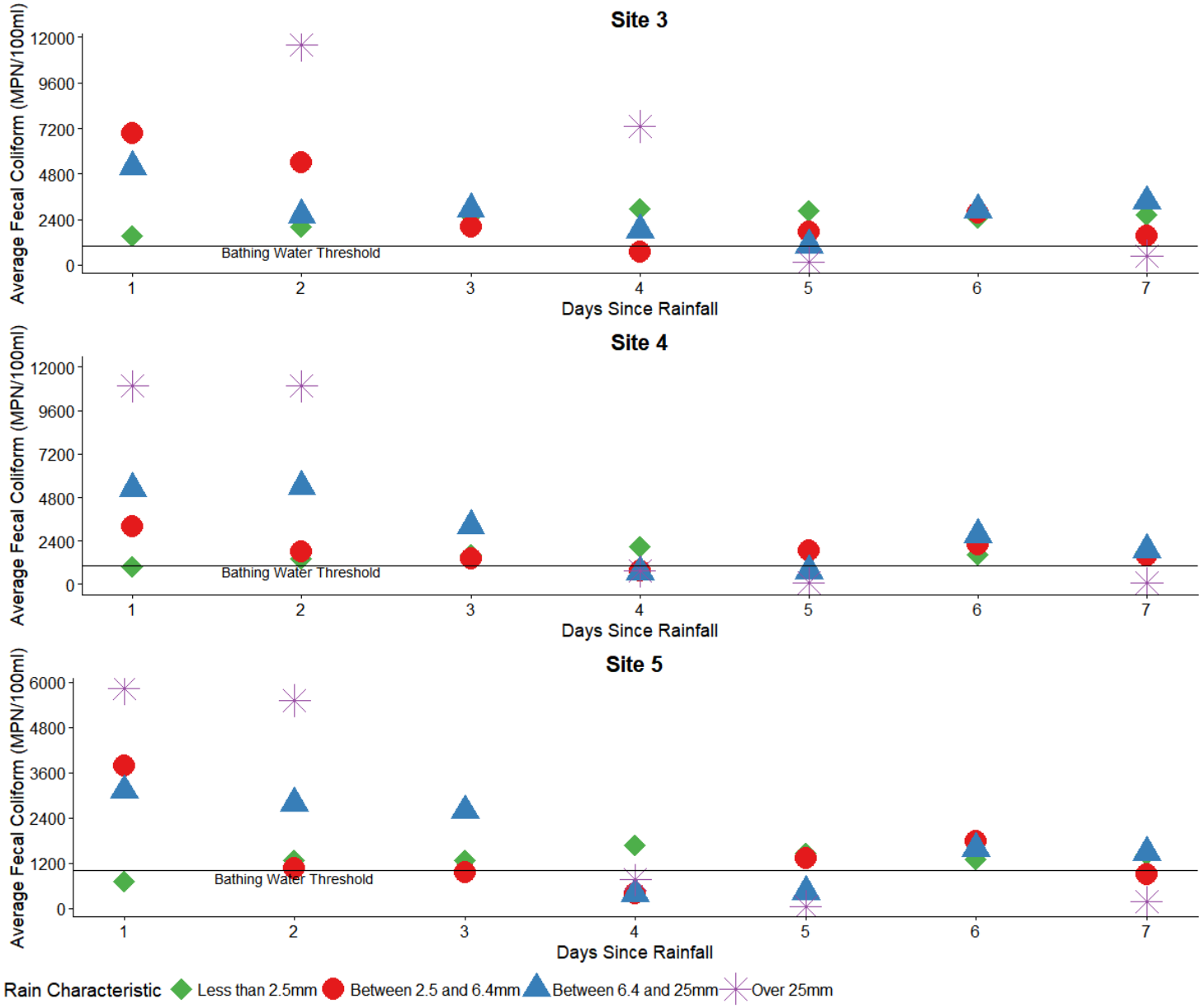

Figure 3.15: Average faecal coliform levels under different rainfall conditions with respect to days since rainfall at sites 3, 4, and 5 in the upper level of the Grand Canal Basin.

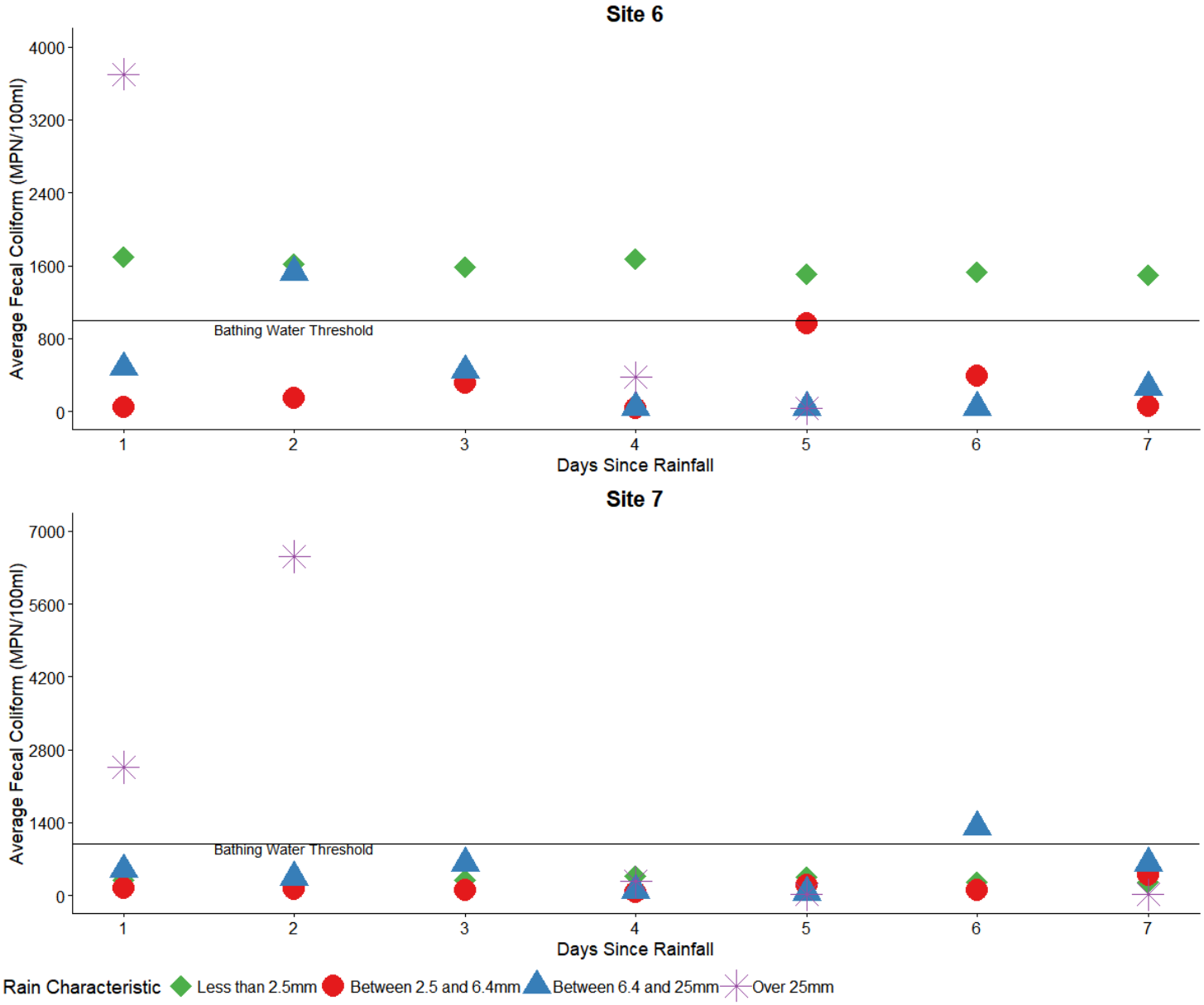

Figure 3.16: Average faecal coliform levels under different rainfall conditions with respect to days since rainfall at sites 6 and 7 in the lower level of the Grand Canal Basin.

Similar to total coliform level faecal coliform had the highest faecal coliform level for rainfall greater than 25 mm was on the first day for sites 4 and 5 in the upper basin (Figure 3.15) and on the second day for site 7 in the lower basin (Figure 3.16). Unlike total coliform levels both site 6 and site 7 are only greatly affected by the highest rainfall level ($x > 25$ mm). Moderate storm levels between $6.4 \text{ mm} > x < 25$ mm and between $2.5 \text{ mm} > x < 6.4$ mm do not have any or as large a second day effect in any of the sites. All sites in the upper basin have the most days with results being over the bathing water threshold. Site 5 does not reach the same height levels as site 4, as it did with total coliform levels, indicating that faecal coliform levels may take longer to disperse along the basin. Again, no sites were greatly impacted by rainfall events below 2.5 mm.

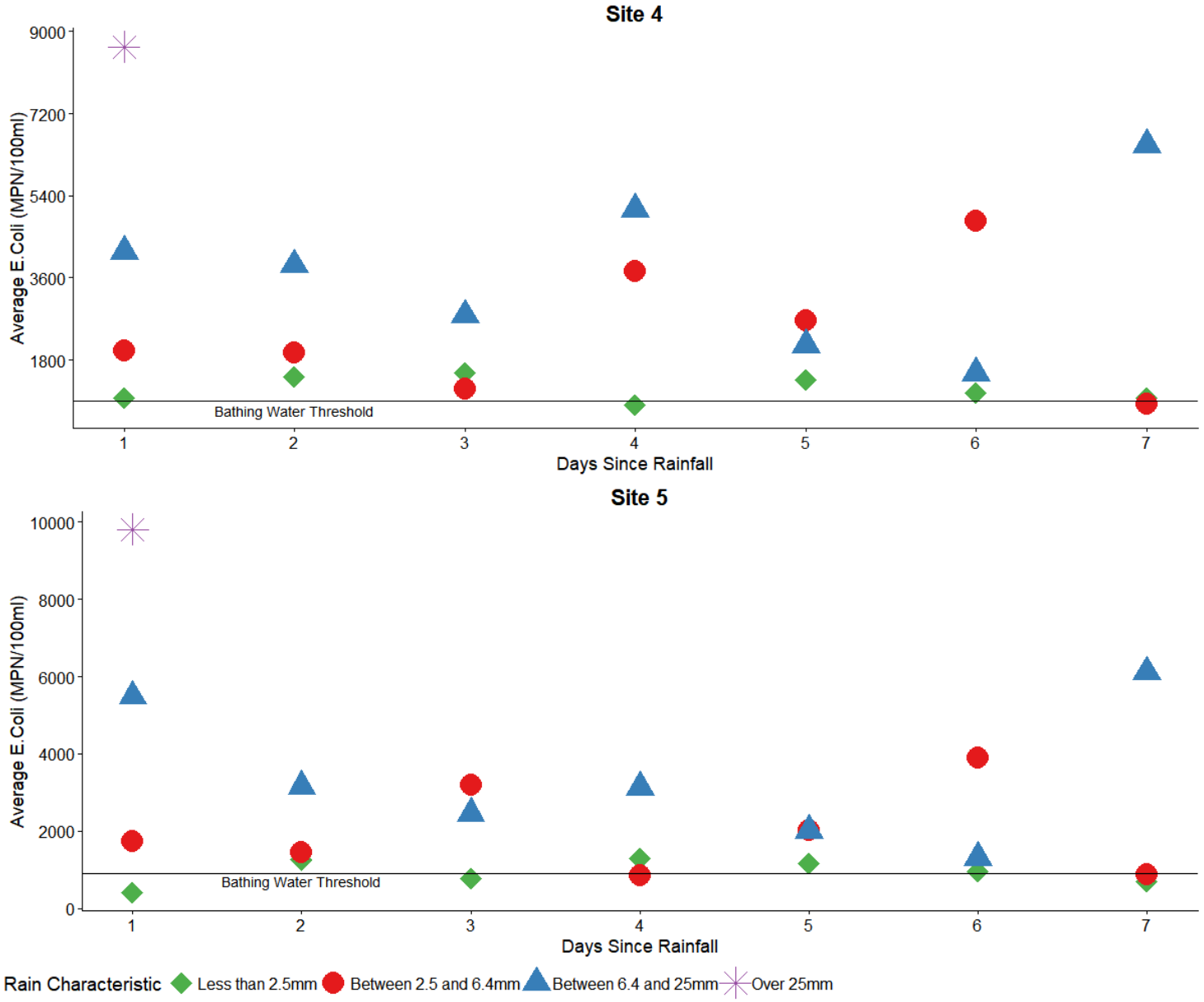

Figure 3.17: Average E. coli levels under different rainfall conditions with respect to days since rainfall at sites 4 and 5 in the upper level of the Grand Canal Basin.

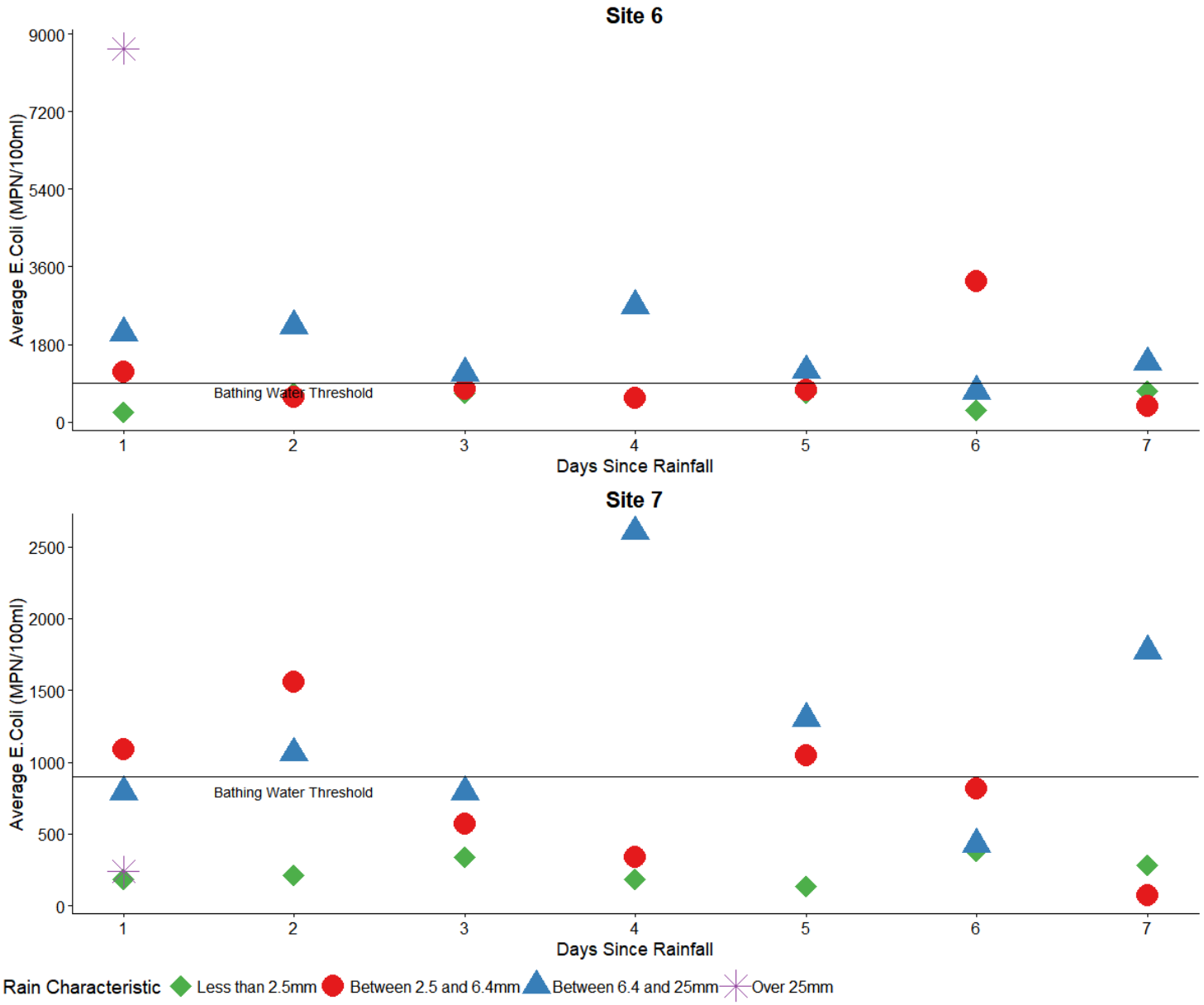

Figure 3.18: Average E. coli levels under different rainfall conditions with respect to days since rainfall sites 6 and 7 in the lower level of the Grand Canal Basin.

The highest rainfall (x > 25 mm) conditions give the greatest E. coli levels in sites 4, 5, and 6 (Figure 3.17 and 3.18). Site 7 does not breach the bathing water threshold for the highest rainfall condition. Site 6 and site 7 have more breaches for samples taken after rainfall events between 6.4 mm > x < 25 mm than between 2.5 mm > x < 6.4 mm and none for below 2.5mm. Site 4 and 5 in the upper basin have no visible pattern for events between 6.4 mm > x < 25 mm and between 2.5 mm > x < 6.4 mm but levels appear to be high and stay relatively high. Site 4 and 5 are also subject to more E. Coli breaches for rainfall events less than 2.5 mm.

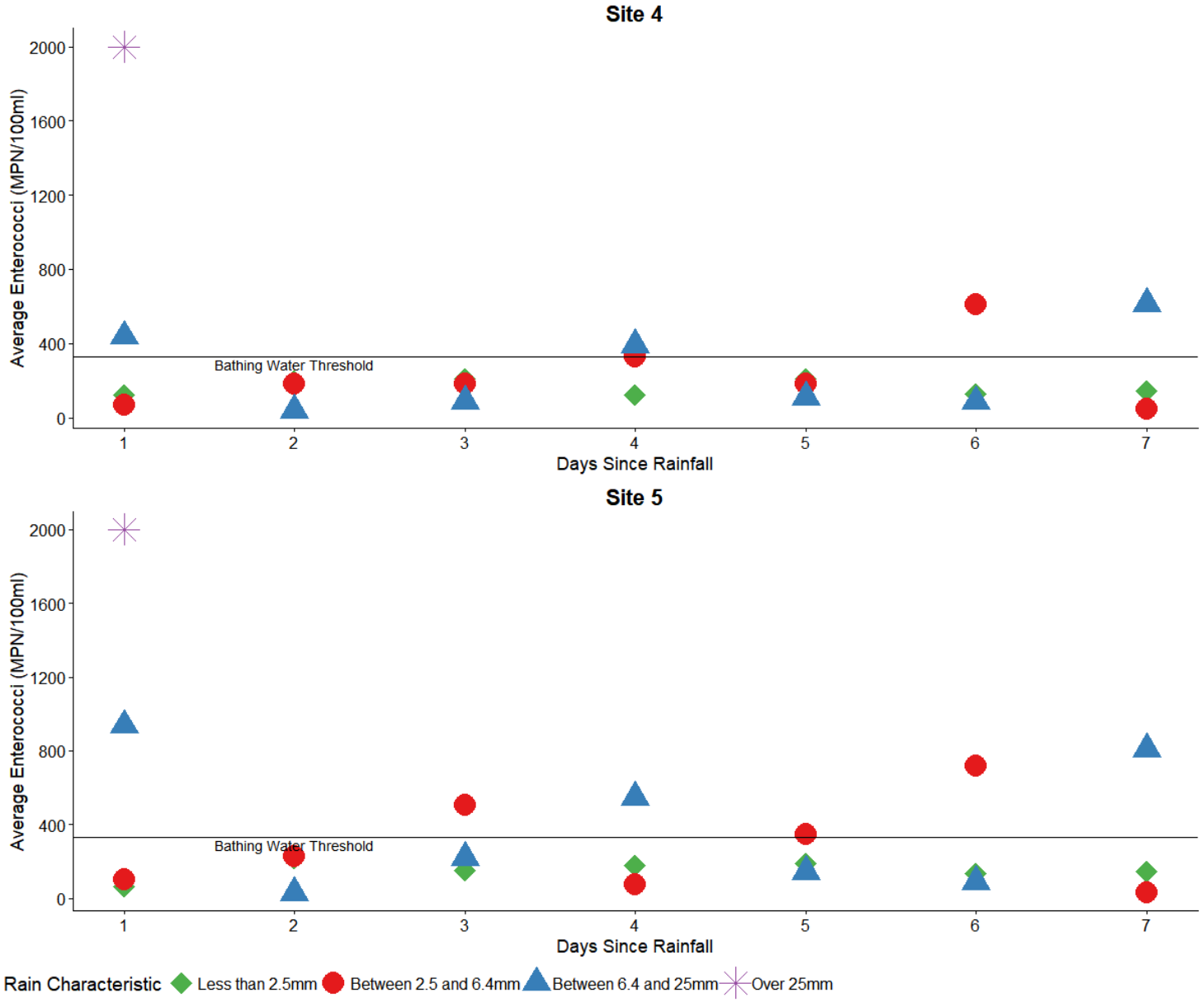

Figure 3.19: Average Enterococci levels under different rainfall conditions with respect to days since rainfall at sites 4 and 5 in the upper level of the Grand Canal Basin.

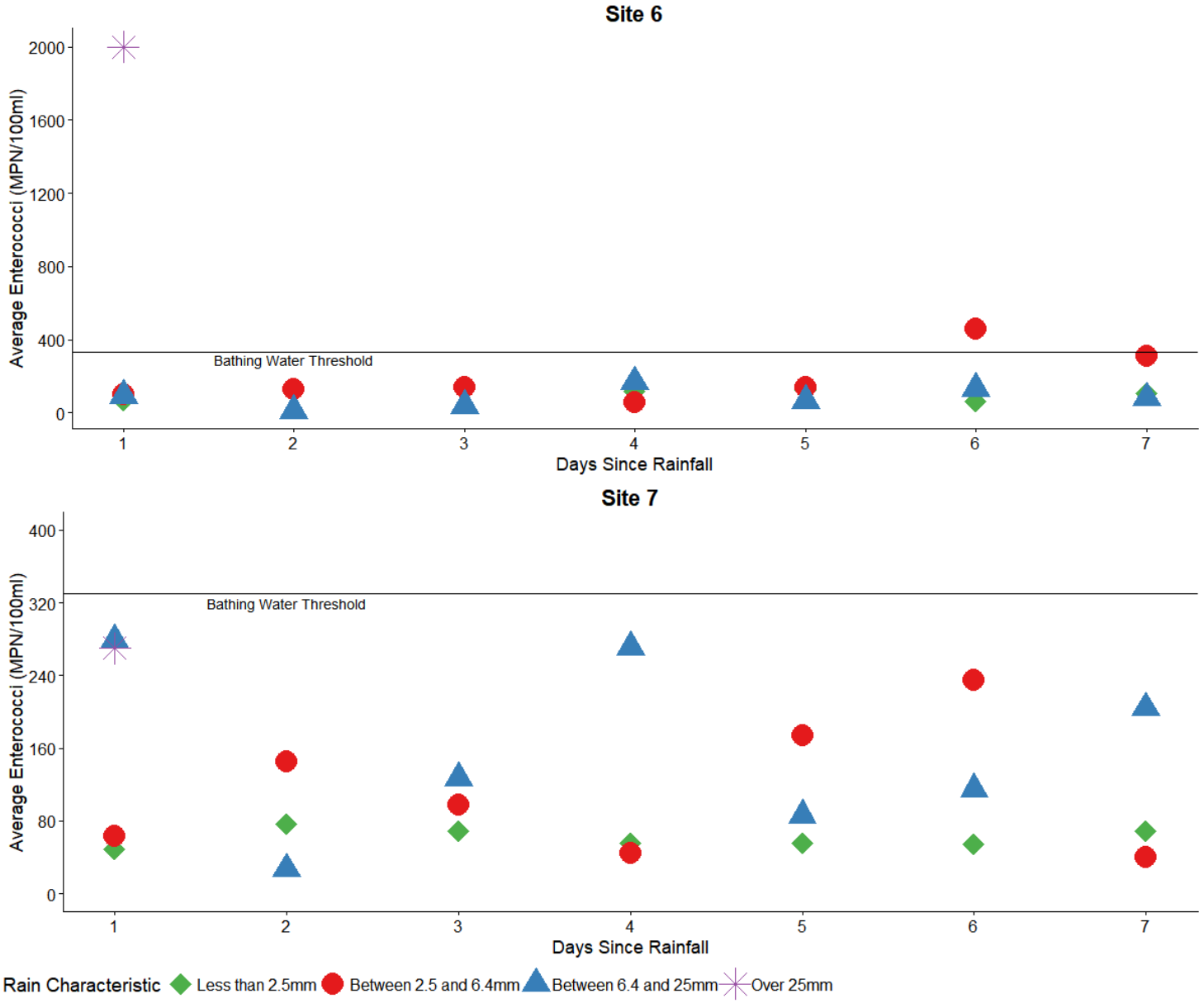

Figure 3.20: Average Enterococci levels under different rainfall conditions with respect to days since rainfall sites 6 and 7 in the lower level of the Grand Canal Basin.

Enterococci as with E. coli levels are impacted greatly by rainfall events over 25 mm in sites 4, 5, and 6 on the first day with no impact being seen in site 7 on the first day (Figure 3.19 and 3.20). No great impact in seen in site 6 or 7 for any other rainfall type. No clear patterns are seen in site 4 and 5 for rainfall between 6.4 mm > x < 25 mm and between 2.5 mm > x < 6.4 mm but again as with E. coli levels most breaches are occurring for these rainfall levels. No breaches occur for rainfall levels below 2.5 mm for any site.

3.4.2.5 Percentage of bathing water threshold breaches as a function of rainfall in the upper basin

It was noted from examining each site in terms of rainfall conditions that most breaches were occurring in the upper basin. To examine what amount of rainfall would cause the most amount of breaches in the upper basin, rainfall one day prior to samples being taken was examined.

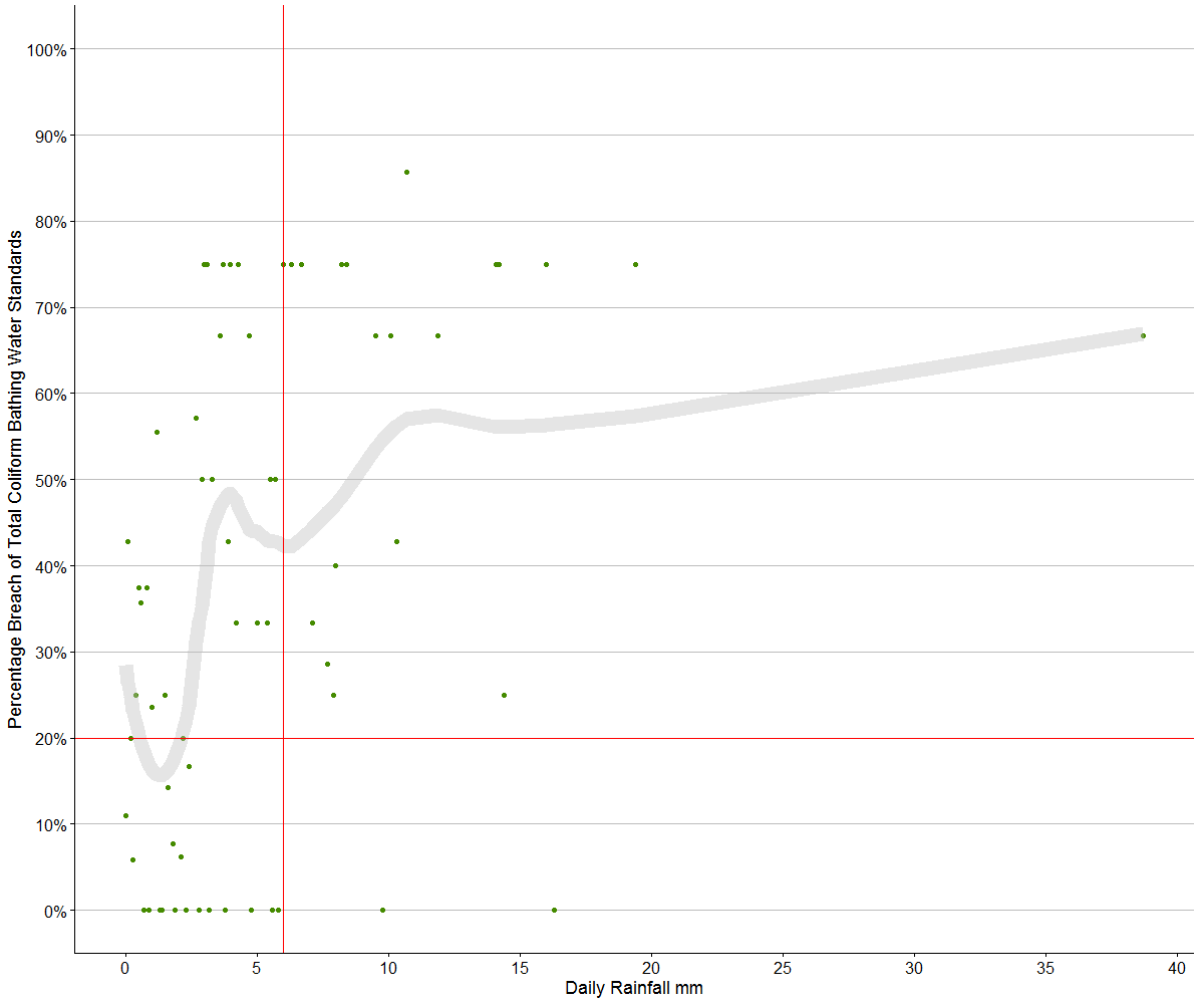

Figure 3.21: Percentage of upper basin breaches of total coliform bathing water quality standards as a function of rainfall.

Total coliform levels had the majority of samples with over 20% of breaches, in the upper basin, when over 6mm of rainfall occurred (Figure 3.21).

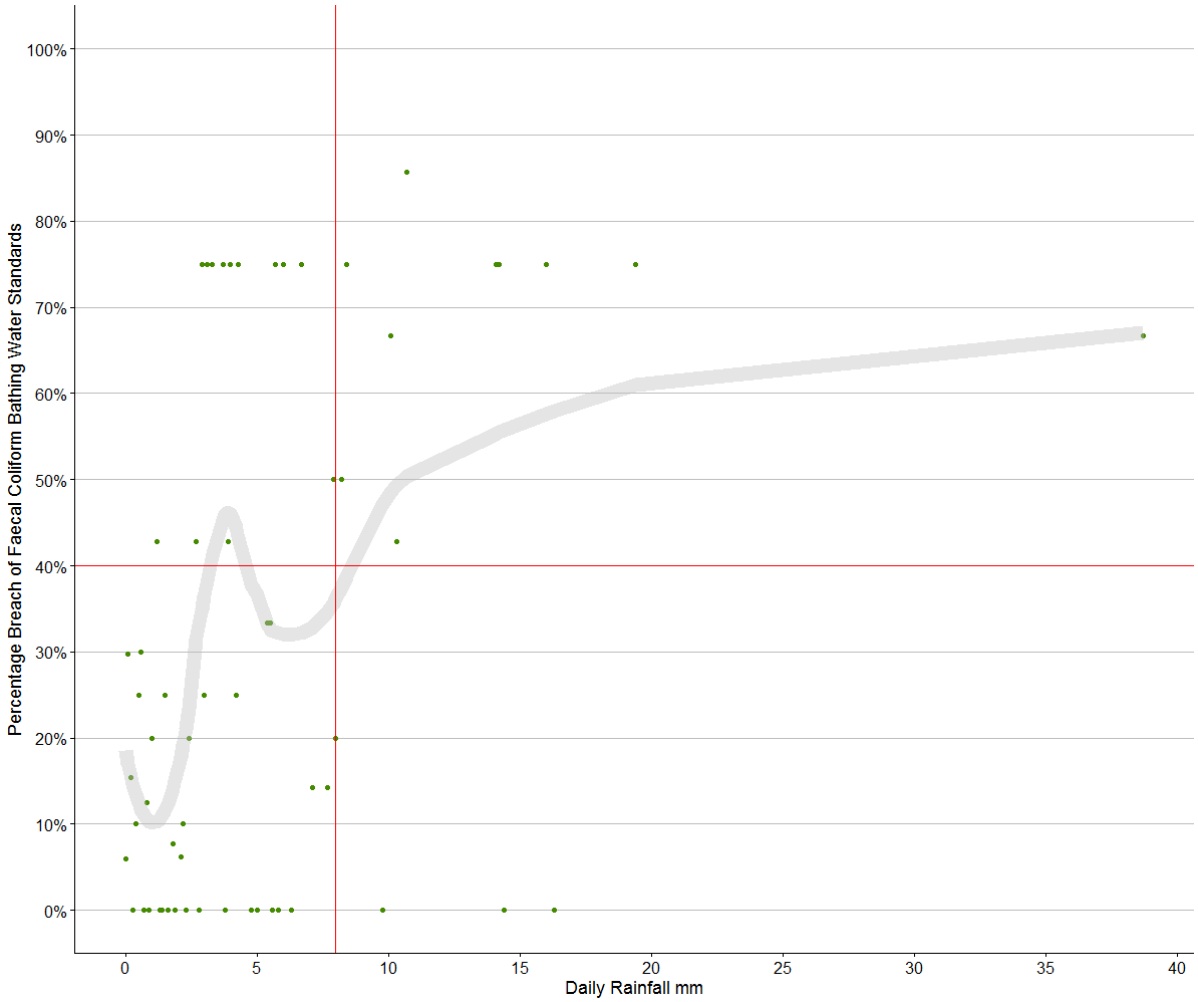

Figure 3.22: Percentage of upper basin breaches of faecal coliform bathing water quality standards as a function of rainfall.

Faecal coliform results had most of the samples with over 40% of breaches, in the upper basin, when over 7 mm of rainfall occurred (Figure 3.22).

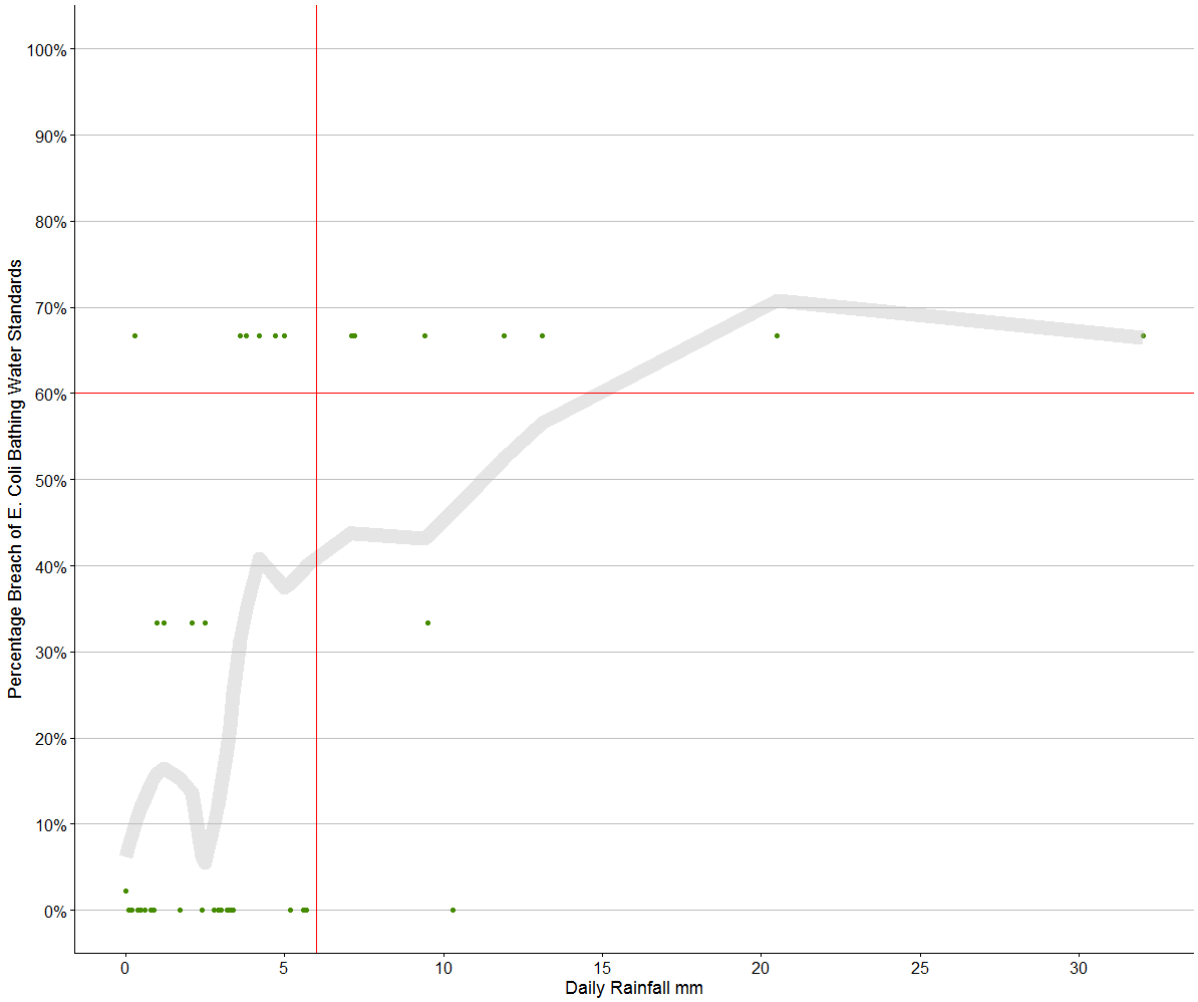

Figure 3.23: Percentage of upper basin breaches of E. coli bathing water quality standards as a function of rainfall.

E. coli results had most of the samples with over 60% of breaches, in the upper basin, when over 6mm of rainfall occurred (Figure 3.23).

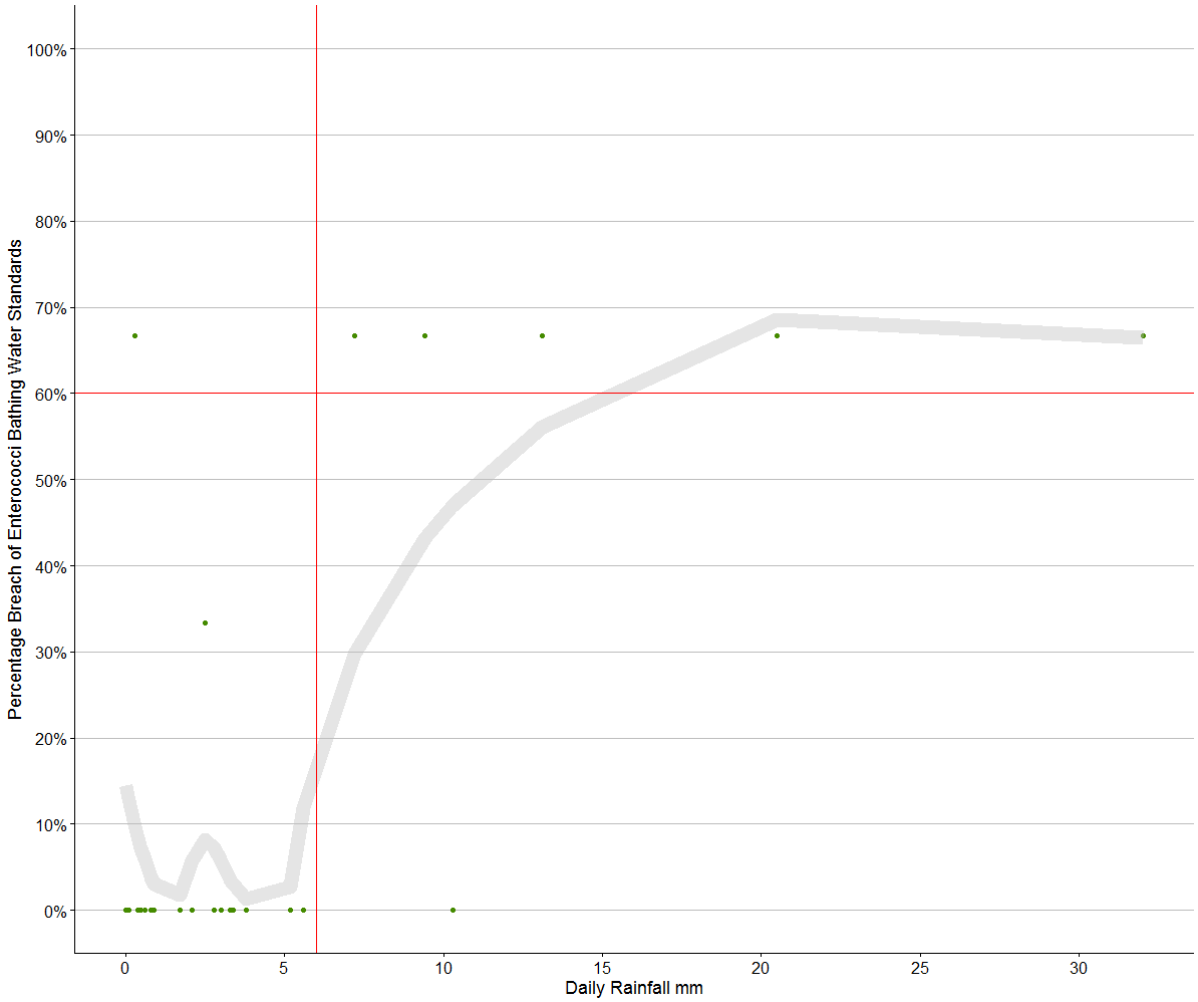

Figure 3.24: Percentage of upper basin breaches of enterococci bathing water quality standards as a function of rainfall.

Enterococci results had most of the samples with over 60% of breaches, in the upper basin, when over 6mm of rainfall occurred (Figure 3.24).

3.4.2.6 ColiSense Analysis

ColiSense analysis was performed to determine total coliform, E. Coli, and enterococci concentrations at 10 sites in the basin (Figure 3.2). This was done in order to examine the use of a novel technology, capable of rapid results, to determine current indicator bacteria levels in the basin. Samples were taken on two separate days 09-02-17 where 0.2 mm of rainfall fell in the average of 2 days before sampling and the 23-02-17 where 5.1 mm of rainfall fell in the average of 2 days before sampling. The ColiSense results for GUS activity were compared to Colilert 18 and Entorolert E results from the same spot sample.

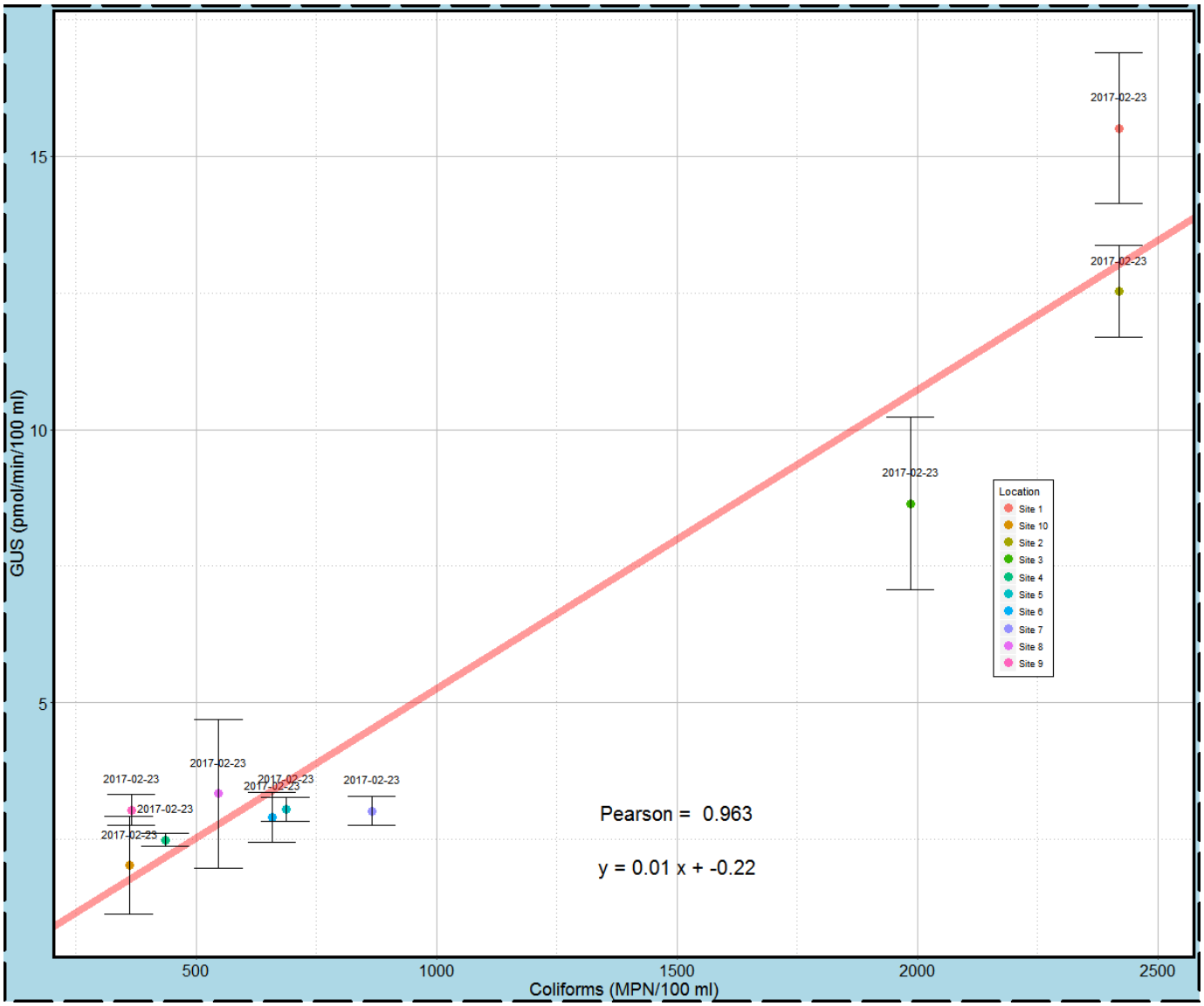

Figure 3.25: The relationship between ColiSense GUS activity results and concentration of total coliforms for each site monitored.

Total coliform had a very strong Pearson result with GUS activity (Figure 3.25). Gus activity was highest at sites 1, 2, and 3, located in the upper basin, and this correlated with equally high total coliform levels for these sites.

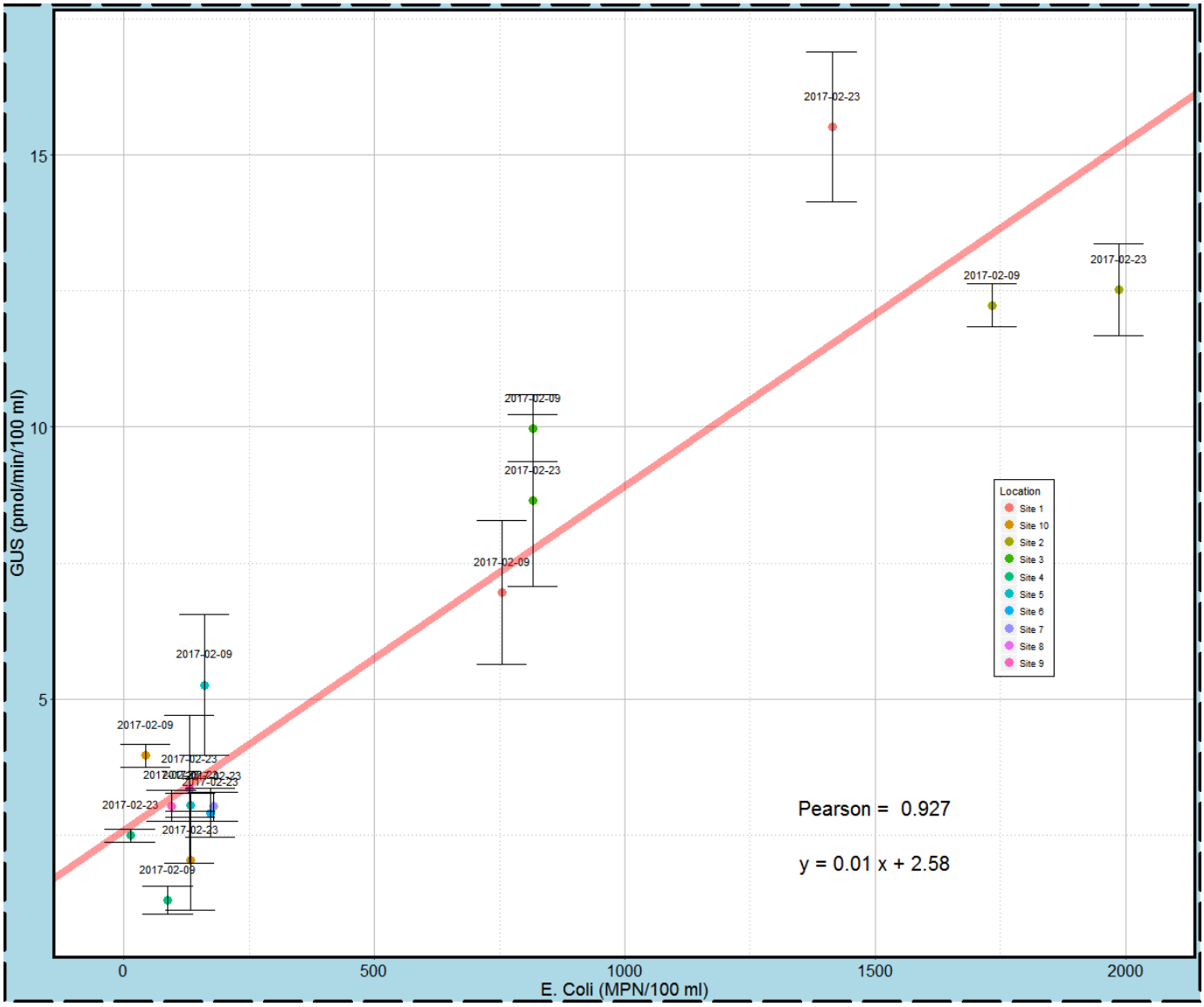

Figure 3.26: The relationship between ColiSense GUS activity results and concentration of E. coli for each site monitored.

E. coli showed the highest GUS activity in sites 1, 2, and 3 in the upper basin (Figure 3.26). The correlation between GUS activity and E. Coli was very strong with a Pearson of 0.927. Higher GUS activity was recorded on the 23-02-17 than the 09-02-17 which was expected due to the higher level of rainfall.

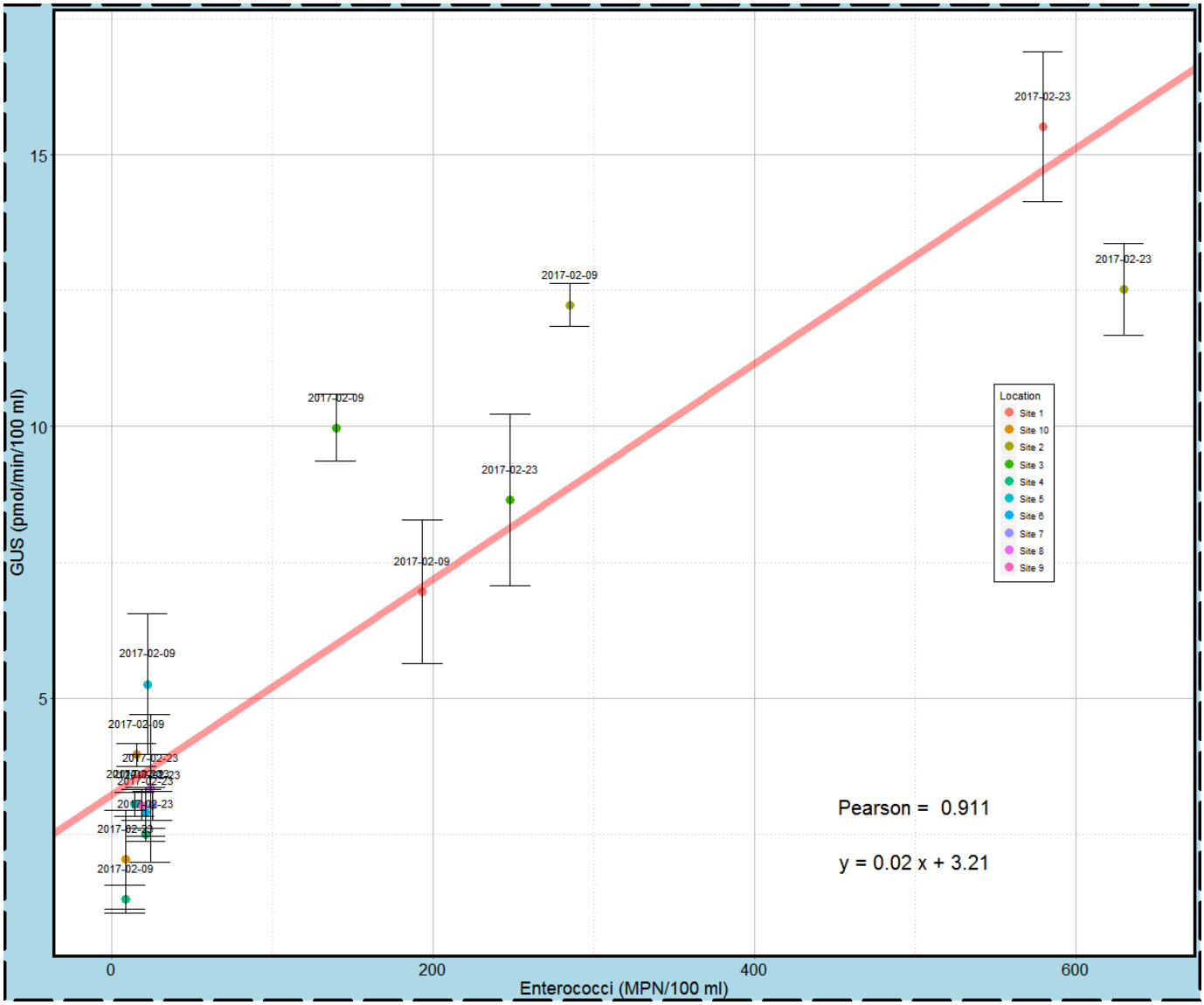

Figure 3.27: The relationship between ColiSense GUS activity results and concentration of enterococci for each site monitored.

Enterococci results also has a very strong correlation with GUS activity (Figure 3.27). The highest GUS activity was seen 1, 2, and 3 in the upper basin which also had the corresponding highest enterococci levels.

These correlation results may be used for future testing in the basin as rating curves by which future total coliform, E. coli, and Enterococci levels can be determined when GUS activity is monitored.

3.4 Conclusion

To establish a safe environment for the public to use as a recreational water body efficient assessment of indicator bacteria levels must occur. There was evidence found from the results of what sites contained the highest levels, how they reacted in relation to each other and to rainfall. This information may be used in combination with future tests to aid with decisions about bathing water safety.

3.4.1 Observations in relation to site location

Initial examination of the entire data set, from 2004 to 2016, showed that the highest percentage of breaches of net samples were taking place in site 3. Sites closer to the outfall had a higher percentage breach than sites further from the outfall. Sites closest in location had the greatest correlation suggesting site location and indicator bacteria level are closely related. It was also noted that sites closer in location had a slope that was close to one meaning the rate of change of bacterial level is similar at these sites. Sites further away from each other had a much lower slope than one. This suggests that the rate of change of bacteria level at sites further away from each other are happening at much different rates. It is apparent from the results that the rate of change of indicator happens quicker at sites closer to the storm water outfall that at sites further away from the storm water outfall.

3.4.2 Effect of rainfall on levels of bacteria in the Grand Canal Basin

The majority of indicator bacteria correlated best with the average of two days rainfall before sampling. Enterococci correlated best with samples taken one day before sampling. Sites in the upper basin showed better correlation with rainfall than sites in the lower basin which may be due to the proximity to the storm water outfall (Figure 1). When rainfall categories were examined every site showed high levels of indicator bacteria levels on the first day after a rainfall event of < 25 mm. This level usually subsided by 5 days after the event for all sites. When rainfall fell between 6.4 mm > x < 25 mm and between 2.5 mm > x < 6.4 mm sites in the upper basin, 3, 4, and 5, had a high result on the first day and then sometimes would present a rise in the second day. Rainfall events of this magnitude would usually take 4 days before bacterial concentration levels lessen and samples show as being below the bathing water threshold. Site 6 and 7 in the lower basin were not affected immediately after a rainfall event of this magnitude; however, they sometimes showed a rise in day 6 or 7 which could be due to bacteria being carried through the water from the sites in the upper

basin which are affected on day one and two. No sites were greatly affected by rainfall events under 2.5 mm. When percentage breaches were examined, for sites in the upper basin, against rainfall intensity when over 6 mm of rainfall fell the majority of total coliform breaches were found above 20%. When above 7 mm of rainfall the majority of faecal coliform breaches were found above 40%. For both E. coli and enterococci when over 6 mm the majority of breaches were above 60%. For the majority of indicator bacteria rainfall events larger than 6 mm had a visible increase in the amount of breaches being recorded in the upper basin.

3.4.3 Recommendations on a monitoring programme

From the results recorded it is recommended after rainfall events of 6 mm in magnitude on there is a risk to bathers on the first and second day after the storm. For rainfall events larger than 25 mm there is an immediate risk to bathers in the entire basin and sampling should take place as soon as possible. There was no great risk to bathers from rainfall events less than 6 mm.

3.4.4 Sensing technologies that can be incorporated for improved monitoring

The ColiSense was tested as a method for improved monitoring in the basin as it has a rapid result time. Very strong correlation was found for the GUS activity monitored from the ColiSense with the, coliforms, E. coli, and enterococci monitored in the basin. These correlation results may be used for future testing in the basin as rating curves by which future total coliform, E. coli, and Enterococci levels can be determined when GUS activity is monitored. It is recommended to use ColiSense so as to obtain results in a much shorter time. This can enable scientists to build larger more accurate data sets for future analysis.

References

Ackerman, D. and Weisberg, S. B. (2003) 'Relationship between rainfall and beach bacterial concentrations on Santa Monica Bay beaches', *Journal of Water and Health*, 1(2), pp. 85–87.

Boehm, A. B., Graham, K. E. and Jennings, W. (2018) 'Can we swim yet? Systematic review, meta-analysis, and risk assessment of aging sewage in surface waters', *Environmental Science & Technology*, p. acs.est.8b01948. doi: 10.1021/acs.est.8b01948.

Chen, E. H. (1971) 'The Power of the Shapiro-Wilk W Test for Normality in Samples from Contaminated Normal Distributions', *Journal of the American Statistical Association*. Taylor & Francis, 66(336), pp. 760–762. doi: 10.1080/01621459.1971.10482342.

Dick, L. K. *et al.* (2010) 'Relative decay of bacteroidales microbial source tracking markers and cultivated escherichia coli in freshwater microcosms', *Applied and Environmental Microbiology*, 76(10), pp. 3255–3262. doi: 10.1128/AEM.02636-09.

Dietze, M. C. *et al.* (2018) 'Iterative near-term ecological forecasting: Needs, opportunities, and challenges.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(7), pp. 1424–1432. doi: 10.1073/pnas.1710231115.

Directive, C. (1976) '76/160/EEC of 8 December 1975 concerning the quality of bathing water', *OJ L*, 31(5.2).

Directive, E. U. B. W. (2006) 'Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160', *EEC*.

Evans, J. D. (1996) *Straightforward statistics for the behavioral sciences.* Pacific Grove, CA: Brooks/Cole Publishing.

Gupta, K. and Saul, A. J. (1996) 'Specific relationships for the first flush load in combined sewer flows', *Water Research*, 30(5), pp. 1244–1252. doi: 10.1016/0043-1354(95)00282-0.

Hathaway, J. M., Hunt, W. F. and McCarthy, D. T. (2015) 'Variability of Intra-event Statistics for Multiple Fecal Indicator Bacteria in Urban Stormwater', *Water Resources Management*, 29(10), pp. 3635–3649. doi: 10.1007/s11269-015-1020-0.

Heery, B. *et al.* (2016) 'ColiSense, today's sample today: A rapid on-site detection of β-d-Glucuronidase activity in surface water as a surrogate for E. coli', *Talanta*, 148, pp. 75–83. doi: 10.1016/j.talanta.2015.10.035.

Kim, T. K. (2015) 'T test as a parametric statistic', *Korean Journal of Anesthesiology*. The Korean Society of Anesthesiologists, 68(6), pp. 540–546. doi: 10.4097/kjae.2015.68.6.540.

Krometis, L.-A. H. *et al.* (2007) 'Intra-storm variability in microbial partitioning and microbial loading rates', *Water Research*, 41(2), pp. 506–516. doi: 10.1016/j.watres.2006.09.029.

Lopes, R. H. C. (2011) 'Kolmogorov-Smirnov Test BT - International Encyclopedia of Statistical Science', in Lovric, M. (ed.). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 718–720. doi: 10.1007/978-3-642-04898-2_326.

Mekonnen, M. M. and Hoekstra, A. Y. (2018) 'Global Anthropogenic Phosphorus Loads to Freshwater and Associated Grey Water Footprints and Water Pollution Levels: A High-Resolution Global Study', *Water Resources Research*, 54(1), pp. 345–358. doi: 10.1002/2017WR020448.

Moore, N. (2008) *Dublin docklands reinvented: the post-industrial regeneration of a European city quarter*. Four Courts Pr Ltd.

Namieśnik, J. *et al.* (2005) 'Passive sampling and/or extraction techniques in environmental analysis: A review', *Analytical and Bioanalytical Chemistry*, 381(2), pp. 279–301. doi: 10.1007/s00216-004-2830-8.

Nguyen, K. H. *et al.* (2018) 'Determination of wild animal sources of fecal indicator bacteria by microbial source tracking (MST) influences regulatory decisions', *Water Research*. Elsevier Ltd, 144, pp. 424–434. doi: 10.1016/j.watres.2018.07.034.

Olds, H. T. *et al.* (2018) 'High levels of sewage contamination released from urban areas after storm events: A quantitative survey with sewage specific bacterial indicators', *PLOS Medicine*, 15(7), p. e1002614. doi: 10.1371/journal.pmed.1002614.

Peng, H.-Q. *et al.* (2016) ‘Event mean concentration and first flush effect from different drainage systems and functional areas during storms’, *Environmental Science and Pollution Research*, 23(6), pp. 5390–5398. doi: 10.1007/s11356-015-5657-2.

R Core Team (2016) ‘R: A Language and Environment for Statistical Computing’. Vienna, Austria. Available at: https://www.r-project.org/.

The Environmental and Protection Agency (2001) ‘Parameters of water quality’, *Environmental Protection*, p. 133. doi: 10.1017/CBO9781107415324.004.

Tiefenthaler, L., Stein, E. D. and Schiff, K. C. (2011) ‘Levels and patterns of fecal indicator bacteria in stormwater runoff from homogenous land use sites and urban watersheds.’, *Journal of water and health*. England, 9(2), pp. 279–290.

Chapter 4: A Real-Time Monitoring Approach for Examining Water Quality Changes Upstream and Downstream From a Cattle Access Point

Abstract

Unrestricted cattle access to waterbodies is known to introduce nutrients into aquatic systems. The exact extent and impact of how cattle's behaviour when entering these water bodies has not been studied in great detail. It may not be possible in all areas to exclude cattle from waterbodies. This study uses real-time turbidity measurements and real-time motion detecting camera data to examine the influence of cattle on water quality when they enter a stream. Two sensors were used; one placed upstream from a cattle access point, and one placed downstream. The difference in upstream and downstream turbidity during a cattle access event was examined to determine what factors impacted the turbidity of the stream most. A cattle access event was described as any continuous entry of cattle into the stream and was captured by motion detecting cameras. The amount of entries, made by each cow, to the stream was recorded as the number of cows. When more than 5 min elapsed between cattle exit from the stream and the entry of more cattle this was counted as the start of a new event. The length of each event from first entry to the last exit of the cattle was also recorded. Results found that when there was greater than 8 cows or an event lasted longer than 14 min in the stream there will be a higher impact on downstream turbidity.

4.1 Introduction

The introduction of nutrients into rivers can have a serious negative effect on water quality (Mekonnen and Hoekstra, 2018). Nutrient enrichment of river bodies can result in eutrophication, which in turn causes great damage to freshwater ecosystems. The greatest effect of eutrophication is algal bloom, which are frequently toxic, and cause a great risk to fisheries, human health, and livestock (Backer *et al*., 2015; Paerl, Otten and Kudela, 2018).

Agricultural activities have been shown to be a large contributor of nutrients to freshwater (Lizotte and Locke, 2018). This is largely from run off due to the spreading of fertilizer, however studies have shown that direct pollution from livestock is also a significant contributing factor (Reis *et al.,* 2010; Oliver *et al.,* 2018). Livestock left to graze for long periods of time remove vegetation, expose soil, and increases the amount of runoff of sediment, nutrients, and pathogens found in waterbodies (Hughes *et al.,* 2016).

The impact of the contribution of cattle to the turbidity levels of a stream was examined in this study. Turbidity is a measure of the optical properties of water and is frequently used as an indicator of sediment levels (Riley, 1998; Sun, Cornish and Daniell, 2001; Arismendi *et al.,* 2017). Turbidity sensors can be deployed for long periods and take real-time data points. They are advantageous over grab samples, used to measure sediment, as they can take far more data points (Voigt *et al.,* 2007; O'Flynn *et al.,* 2010). Cattle with direct access to streams contribute to the increased sediment and nutrient level by directly urinating and defecating in the river, as well as by loosening the soil of the river bank located at the access point which can alter the hydrology and the drainage pathways of the site (Vidon, Campbell and Gray, 2008; Hughes *et al.,* 2016; Carter *et al.,* 2017). Studies have shown that cattle preferentially defecate in streams which elevates nitrogen and phosphorus levels (Bond, Sear and Edwards, 2012).

Ireland has a large agricultural land use comprising of 60% of the total land use of rural areas. This study takes place in July and August which is during the Irish grazing season.

Sensors capable of real time continuous monitoring, were used to examine the difference in river water quality upstream and downstream of the cattle access point.

Real time monitoring can be used to examine the long-term effects of cattle access as well as examining the direct effects under different weather and flow conditions.

This study uses YSI sensors to measure the difference in turbidity upstream and downstream of the cattle access point was compared to determine the influence of cattle on suspended sediment. It also allowed for examination of the background level of turbidity when no cattle have entered the stream.

This study is part of an EPA funded research project which examines cattle exclusion from watercourses and its environmental and socio-economic implications in partnership with Teagasc and Dundalk Institute of Technology.

4.2 Method

4.2.1 Catchment description

The site was located in Dunleer, County Louth (53°50'01.7"N 6°25'04.9"W). The land is used as a grazing area for cattle. It contains one stream which is a tributary of the White river. This field was chosen as it contained an accessible cattle access point with sufficient water levels, deep enough, for the sensors to be completely submerged when placed in the respective upstream and downstream monitoring points. The access point is only accessible from one side of the bank. The cattle only enter the stream for access to water and not as a crossing point which is common in other cattle access cases. Sensors were placed 3 m upstream from the cattle access point and 1 m downstream from the access point insuring they were at the point of deepest water. The study took place between the 12th June 2017 and the 3rd of August 2017.

Figure 4.1: Location of cattle access point, upstream and downstream sensors. The cattle access point was only accessible from one point. Upstream, downstream, and the north bank from this access point was fenced off.

4.2.2 Instrumentation and structure for deployment

4.2.2.1 Multi-parameter sondes

Two multi-parameter sondes were deployed: one upstream and one downstream from the cattle access point. The model of sonde used was YSI 6600 EDS V2-2, manufactured by YSI Environmental, and each was deployed with a probe to measure temperature (Celsius), conductivity (millisiemens per centimetre), turbidity (nephelometric turbidity units (NTU)), optical dissolved oxygen (milligram per litre), and pH. However, turbidity is the only parameter examined in this study.

Figure 4.2: Sensor deployment in stream. Sensors were deployed in a steel cage and laid horizontally along the river bed.

The sondes were first calibrated in a laboratory, by adjusting the reading output to comply with YSI standards for pH, turbidity, and conductivity; before being transferred to the site. As the sondes were being deployed in a shallow fresh water stream a cage was constructed because the sondes had to be laid flat in the water for the probes to be completely submerged. Steel cages were constructed to protect the sondes from being dragged along the riverbed as well as to protect them from theft as they would be visible in times of shallow water level (Figure 4.2). The cages were constructed from steel and the sensor was secured with hinged steel covers that were bolted. The steal securing the sonde was lined with rubber to insure the sondes were immovable within the cage.

Visits to the site were made every 4 weeks to inspect for damage and clean the sonde. The sensor was cleaned to reduce fouling of the instrument which leads to decreased sensitivity of the paramotors. At each site visit the data recorded up until the site visit was transferred to the laptop brought on site. Then the sonde was removed from its cage, cleaned, inspected for damage and redeployed. The data collected was corrected for value drift. Sensor fouling occurs gradually over time resulting in reduced sensitivity of the sensors and the data recorded slight deviation in the recorded data. After cleaning data correction for fouling drift was applied between two servicing dates as described by Wagner et al. (Wagner *et al.,* 2006). Sensor drift begins as soon as the sensor is deployed and is assumed to occur at a constant rate. Zero correction is applied at the start of the interval, the full correction at the end of it and between these dates data is linearly interpolated. The following equation (Wagner *et al.,* 2006) was used in this case for linear drift correction,

$$V_C = V + \left(V_f - V_f\right)\left(\frac{T_t - T}{T_t}\right)$$

V_c is the drift corrected value, V is the original measured value, V_f is the response of the sensor immediately before cleaning and validation at the end of the correction interval; V_s is the response of the sensor after cleaning and calibration; T_s is the total time interval for which the correction is applied and T is the time between the end of deployment and the time when the value is measured.

4.2.2.2 Motion detecting camera data

The image data taken by two Bushnell Trophy HD motion-activated camera, model 119676 located set to capture two different angles of the cattle access point. The camera worked by detecting the motion of cows entering the stream and leaving the stream. The camera included night vision which captured any events taking place in darkness. Site visits were used to inspect the camera for damage and ensure they were still in the correct line of sight. The photos were stored on a memory card where they were then transferred from the site to a laptop for examination. Some examples of images taken can be seen in Figure 4.3.

Figure 4.3: A) Displays the cattle access point taken from camera angle 1 when no cows were in the stream. B) Displays two cows in the stream taken from camera angle one. C) Displays four cows in the stream taken from camera angle 2. D) Displays three cows taken in night vision at camera angle 2.

4.2.2.3 River discharge monitoring

The daily river discharge data was provided by Teagasc, a semi-state authority in the Republic of Ireland responsible for research and development, training and advisory services in the agri-food sector. It consisted of an Orpheus mini, manufactured by OTT HydroMet, which measures water level, pressure, and temperature of the stream. The level data is used in the calculation of discharge when combined with a rating curve for the outlet. River discharge is reported in cubic meter per second (m^3/s).

4.2.3 Data and statistical analysis

4.2.3.1 Number of data points recorded

Table 4.1: Number of recorded data points for the period of monitoring which took place from between the 12th June 2017 and the 3rd of August 2017.

Number of turbidity data points collected	Number of images taken on Camera	Number of flow data points	Number of events were cows entered the stream
13887	41689	365	68

The camera data was recorded by noting the time the cows entered a stream (Table 4.1), how long the cows stayed in the stream, and the number of cows in the stream. As well as this the time of the next turbidity reading after the cows entered the stream was recorded. Events where people, and other animals (dogs, foxes, etc.) entered the water were taken out of the study as they were able to venture further upstream and downstream than the cows meaning they were walking too close to the placement of the sensors for accurate readings of their influence on turbidity levels at the cattle access point. Turbidity data was reported as difference in turbidity for each event.

4.2.3.1 Analysis performed

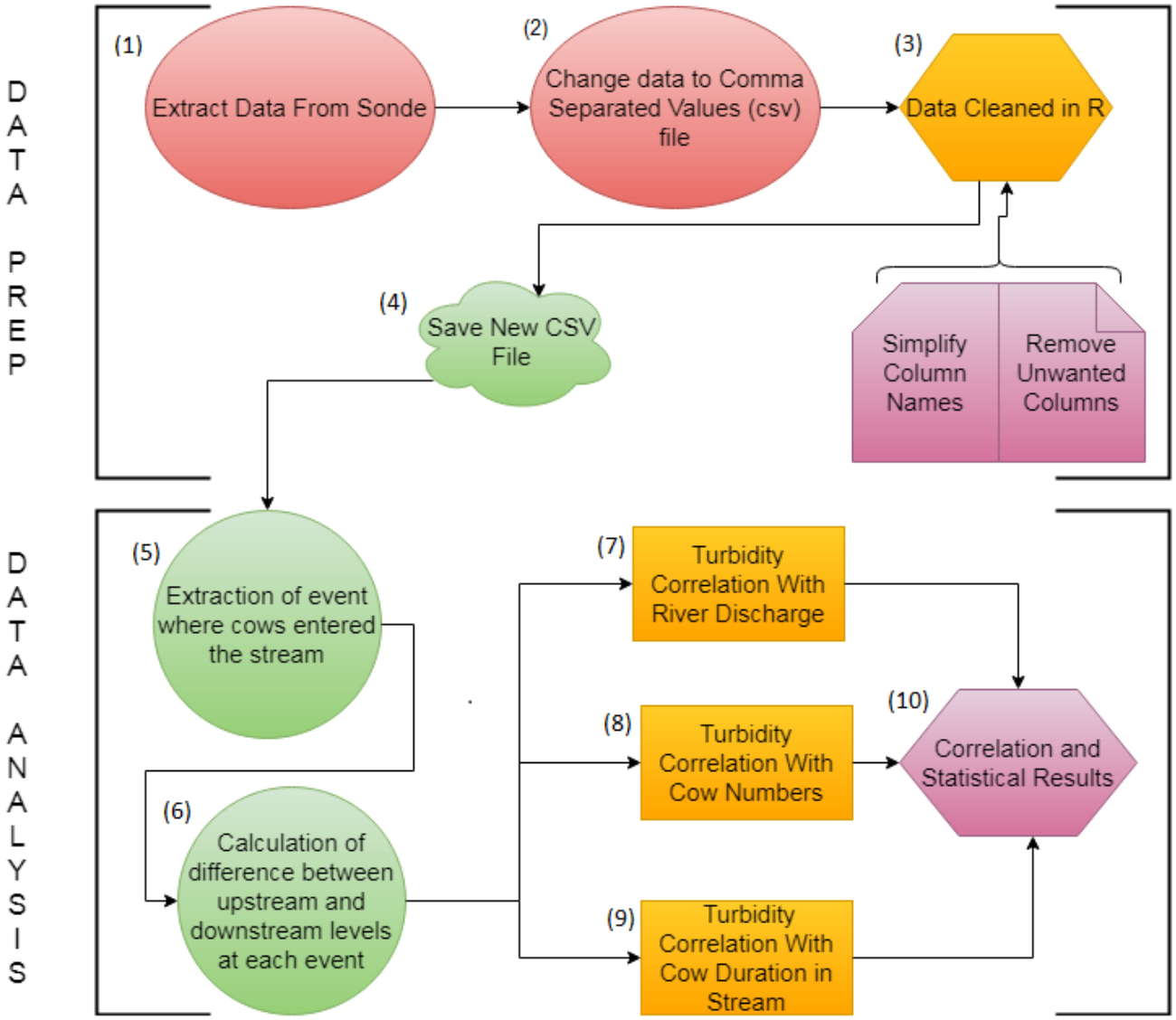

Figure 4.4: Experimental schematic outlining the order analysis was taken from data collection and preparation to data analysis.

In Figure 4.4 data was collected from the sondes on site by using a usb connection connected to a laptop. (2) The files were then taken from this laptop and loaded into a working computer where they were converted into csv files for use in R. (3) Data was cleaned in R: column names were simplified, missing data was removed, data that may have been interfered with by any known events of humans entering the water was removed, and any unwanted columns were removed. (4) The cleaned data was saved as a .csv file; these files were used for all analysis. (5) From examination of the camera footage, events where cows entered the stream were noted and the time of the event,

how many cows were involved, and the length of time the cows spent in the water was noted. As turbidity reading were only recorded every 15 min the first upstream and downstream reading taken at the beginning of an event was taken as the event turbidity. (6) To calculate the difference in turbidity at each event the value of upstream turbidity was taken from the value of downstream turbidity. A large positive result indicates downstream turbidity levels are much higher than upstream turbidity results taken at the same time. (7) Difference in turbidity was correlated against the river discharge obtained from Teagasc. (8) Difference in turbidity was correlated against the number of cows entering the stream in one event obtained from camera data. (9) Difference in turbidity was correlated against the length of time of an event where cows have entered the stream. (9) Correlation and statistical results were tabulated and graphed.

4.3 Results and discussion

The aim of this study is to determine the conditions in which the cattle's effect on turbidity is greatest and at which levels they cause minimal effect. Correlation results for Pearson, linear correlation, and Spearman, ranked monotonic correlation, were categorised using the Evans (1996) guide where a result of 0.00 – 0.19 is very weak, of 0.20 – 0.39 is weak, of 0.40 – 0.59 is moderate, of 0.60 - 0.79 is strong, and of 0.80 – 1.0 is very strong.

4.3.1 Event characteristics

An event was described as any continues entry of cattle into the stream. The amount of individual entries to the stream was recorded as the cow numbers, meaning if one cow entered exited and then entered the stream again the second entry, as well as the first, of this cow, was counted as an individual cow number. When more than 5 min elapsed between cattle exit from the stream and the entry of more cattle this was counted as the start of a new event. The length of each event from first entry to the last exit of the cattle was also recorded.

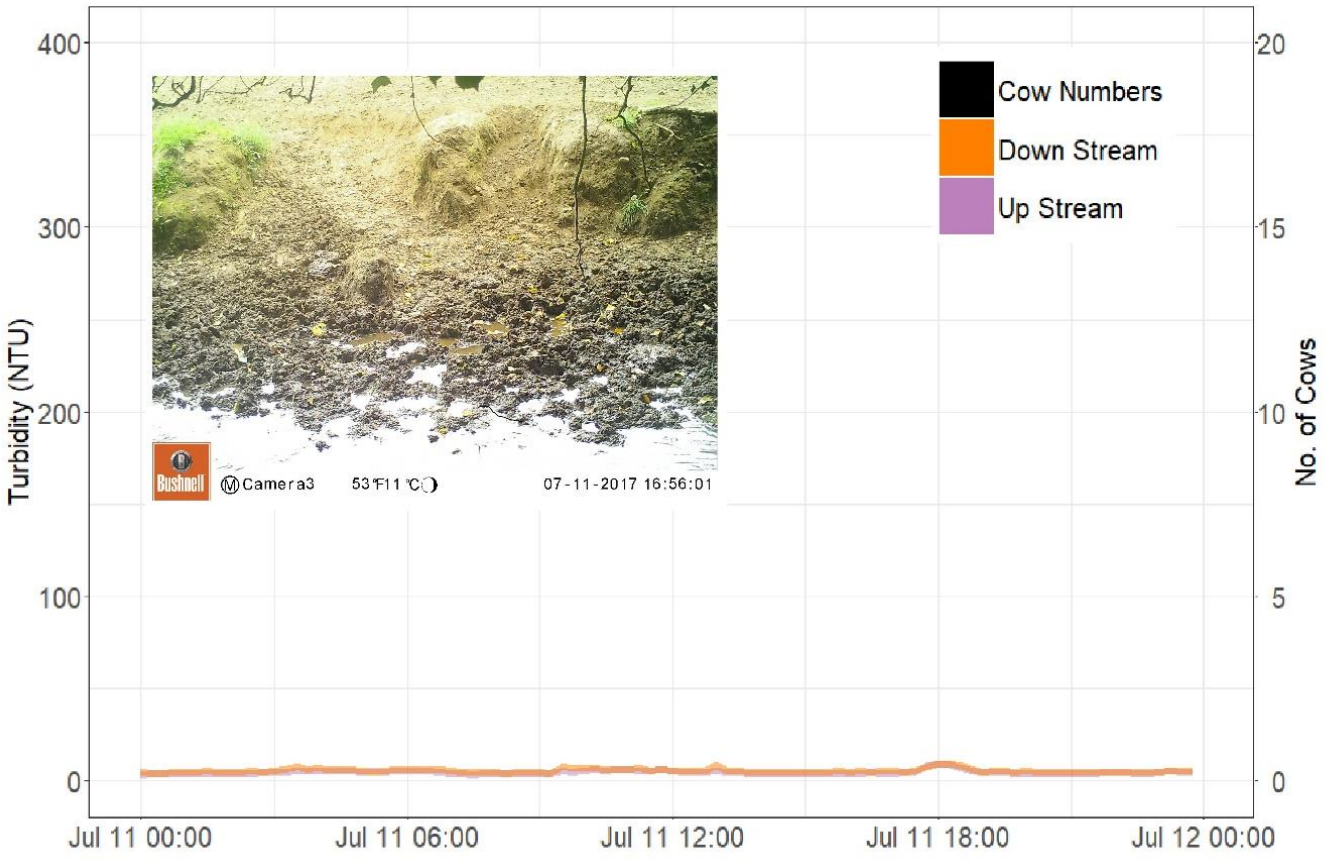

Figure 4.5: Camera images and turbidity levels taken on the 11th July 2017.

On the 11th of July 2017 no cattle access events took place (Figure 4.5). River discharge for this day was recorded to be 0.07 m^3/s and no differences in upstream and downstream turbidity were noted.

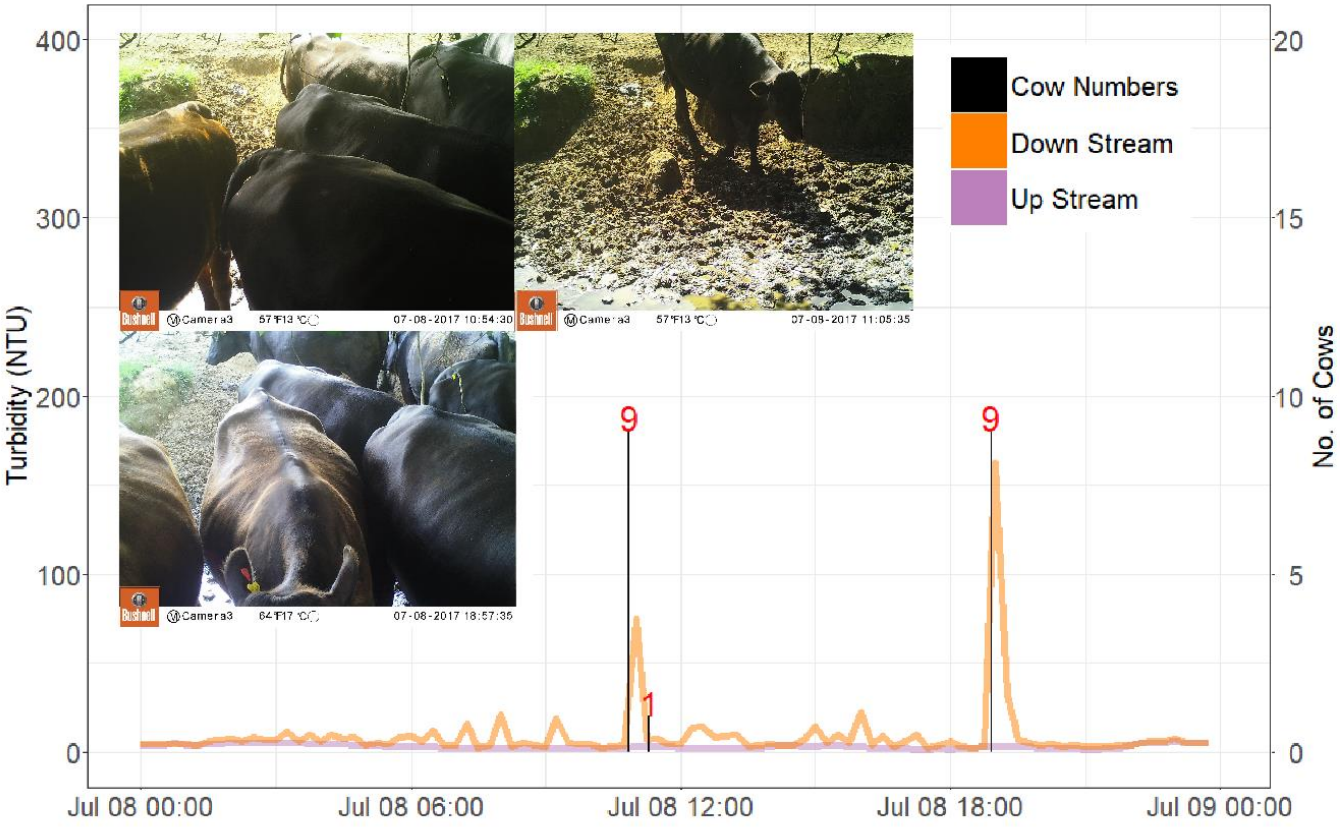

Figure 4.6: Camera images and turbidity levels taken on the 8th July 2017.

On the 8th July 2017 3 cattle access events took place (Figure 4.6). River discharge for this day was recorded to be 0.08 m^3/s and downstream turbidity levels were overall higher than upstream. The first event started at 10:52, a total of 9 cows entered the stream, and the event lasted 17 min. A large spike in downstream turbidity was recorded after this event. The second event took place at 11:19, one cow entered the stream, and the event lasted 5 min. No large difference in upstream and downstream turbidity levels was recorded for this event. The third and final event took place at 18:55, a total of 9 cows entered the stream, this event lasted 34 min. An extremely large spike in downstream turbidity was seen for this event much larger than the earlier event with equal amounts of cattle entries. It was determined from this to investigate the difference in turbidity of each event in relation to river discharge level, number of cows entering the stream, and length of the event.

4.3.2 Turbidity results for each event.

There were 69 recorded cattle access events. To determine under what conditions cows impacted the downstream turbidity of a stream 3 areas were examined:

(i) The difference in upstream and downstream turbidity in relation do different river discharge levels.

(ii) The difference in upstream and downstream turbidity in relation to different numbers of cows entering the stream for one event.

(iii) The difference in upstream and downstream turbidity in relation to the length of each cattle access event.

Table 4.2: Distribution of difference in turbidity recorded for each cattle access event.

Minimum (NTU)	**Maximum (NTU)**	**Median (NTU)**	**Average (NTU)**	**Standard Deviation (NTU)**
-8.6	159.9	2.8	16.2	± 30.7

There were 5 events out of 36 where upstream turbidity levels were higher than downstream and resulted in a negative difference in turbidity (Table 4.2). There was a large deviation, ± 30.7 NTU, in the difference between upstream and downstream turbidity for each event. The reason for this disparity was investigated by correlating the difference in upstream and downstream turbidity against river discharge, the number of cows entering the stream, the length of the event.

4.3.3 Analysis of the impact of river discharge on turbidity during a cattle access event.

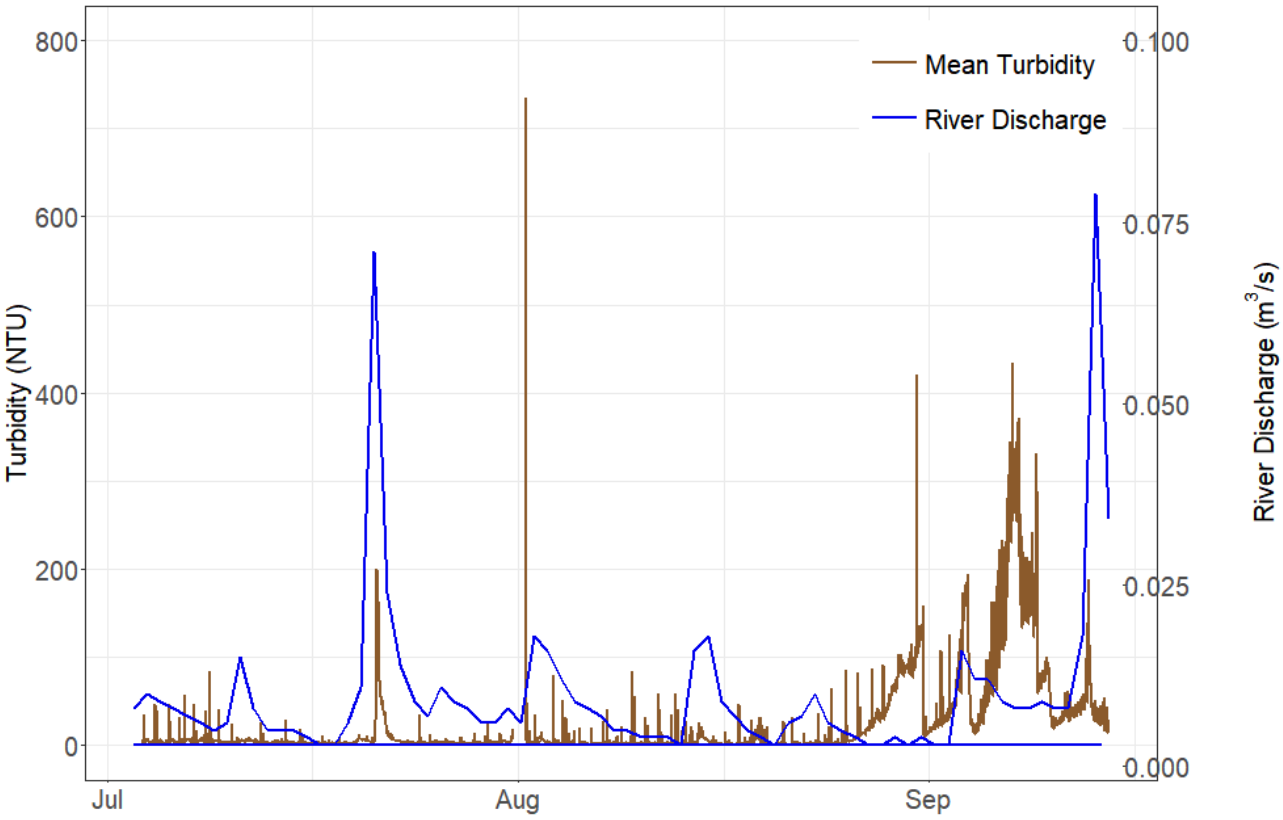

Figure 4.7: Mean upstream and downstream turbidity and river discharge levels for the duration of sensor deployment.

Overall river discharge level for the period remained low with only two occasions where levels went above 0.025 m^3/s (Figure 4.7). Average turbidity levels varied considerably more for the same period.

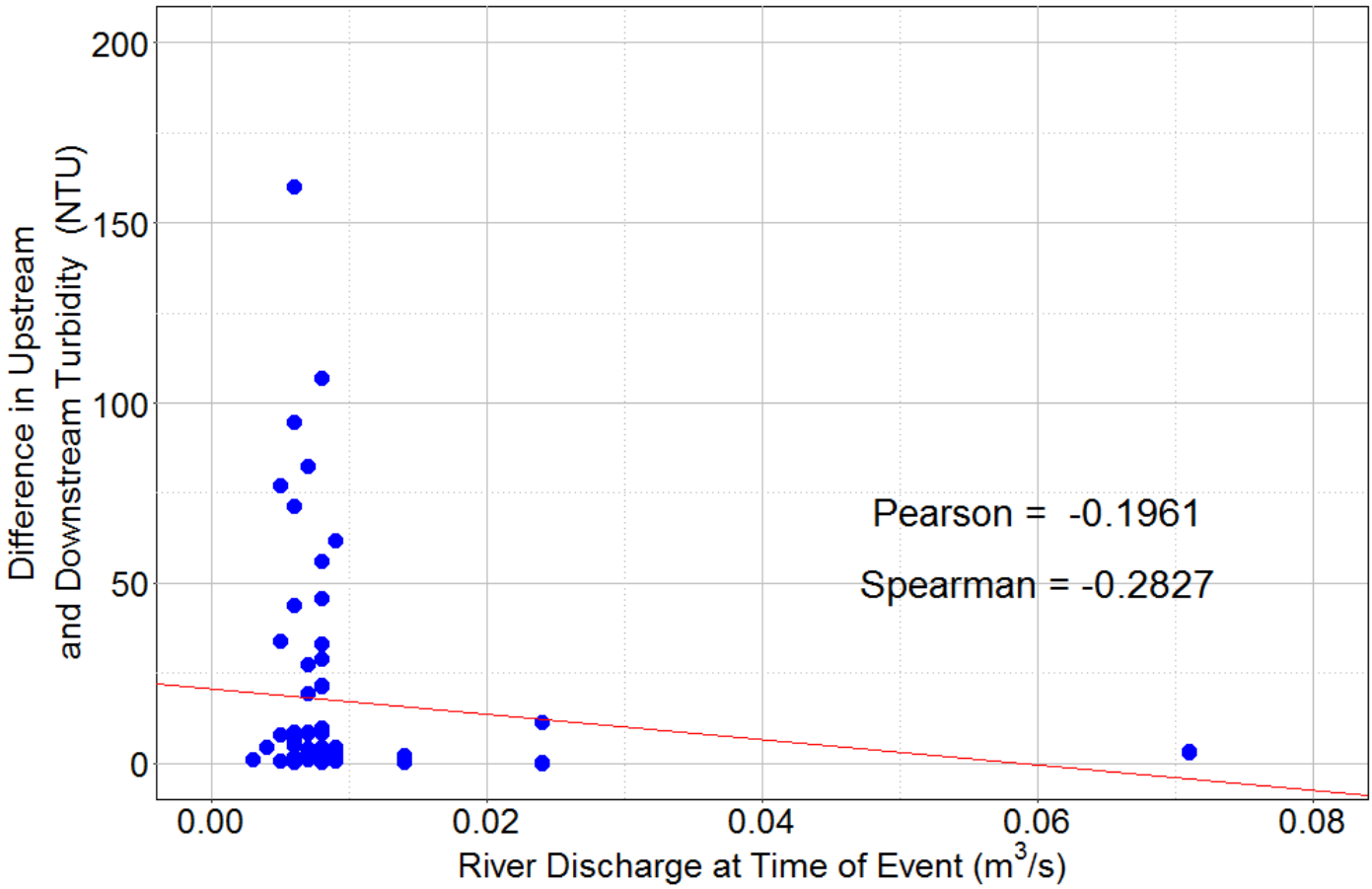

Figure 4.8: Correlation results for difference in turbidity and river discharge at each cattle access event

There was very weak Pearson and Spearman correlation seen between river discharge and the difference in turbidity in each event (Figure 4.8). The speed of the river discharge appeared to have little influence over the difference in upstream and downstream turbidity levels. Most events occurred at low discharge speeds, less than 0.03 m^3/s.

Table 4.3: Distribution of river discharge levels for each cattle access event.

Minimum (m^3/s)	**Maximum (m^3/s)**	**Median (m^3/s)**	**Average (m^3/s)**	**Standard Deviation (m^3/s)**
0.003	0.071	0.008	0.013	± 0.017

The maximum river discharge for a cattle access event was 0.071 m^3/s (Table 4.3). There was not a huge disparity between the river discharge levels and the recorded standard deviation was a low ± 0.017 m^3/s. The results from the river discharge levels suggest that cows prefer to enter the water at low levels.

4.3.2 Analysis of the impact of the number of cows entering the stream in each event on turbidity levels

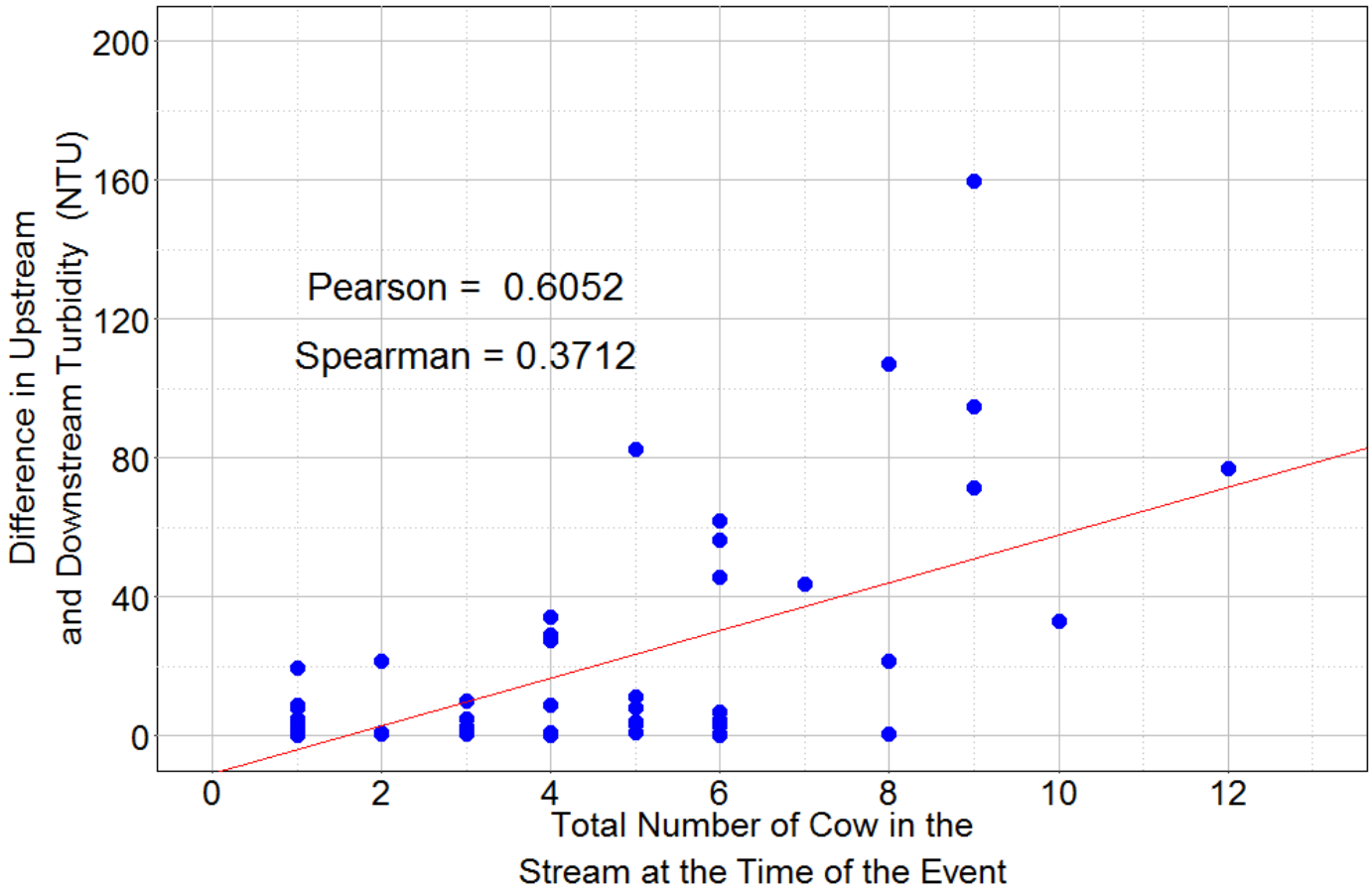

Figure 4.9: Correlation results between the total number of cows entering the stream and the difference in upstream and downstream turbidity at each event.

A strong Pearson correlation result was seen between the number of cows entering the stream and the resulting difference in upstream and downstream turbidity (Figure 4.9). A weak Spearman result was seen which informs us that a monotonic relationship is unlikely and an increase in cow numbers in a stream will not necessarily result in an increased level of turbidity downstream. Results indicate that when above 8 cows enter the water downstream turbidity is always largely affected, although as there were only 5 events where 9 cows or over entered the stream further study is recommended to confirm this.

Table 4.4: Distribution of the number of cows entering the stream during each cattle access event.

Minimum	Maximum	Median	Average	Standard Deviation
1.0	12.0	4.0	3.9	± 2.7

The minimum number of cows entering the water for one event was 1 and the maximum recorded was 12 (Table 4.4). The median number of cows seen entering was 4. The standard deviation was ± 2.7 suggesting cows in this study preferred to enter the stream in groups of 2 to 6. Events of 7 and above cows entering the stream were a lot less frequent only happening 6 times over the course of the project in comparison to events with below 7 cows entering the stream happening 60 times.

4.3.3 Analysis of the impact of the length of time of each event on turbidity levels

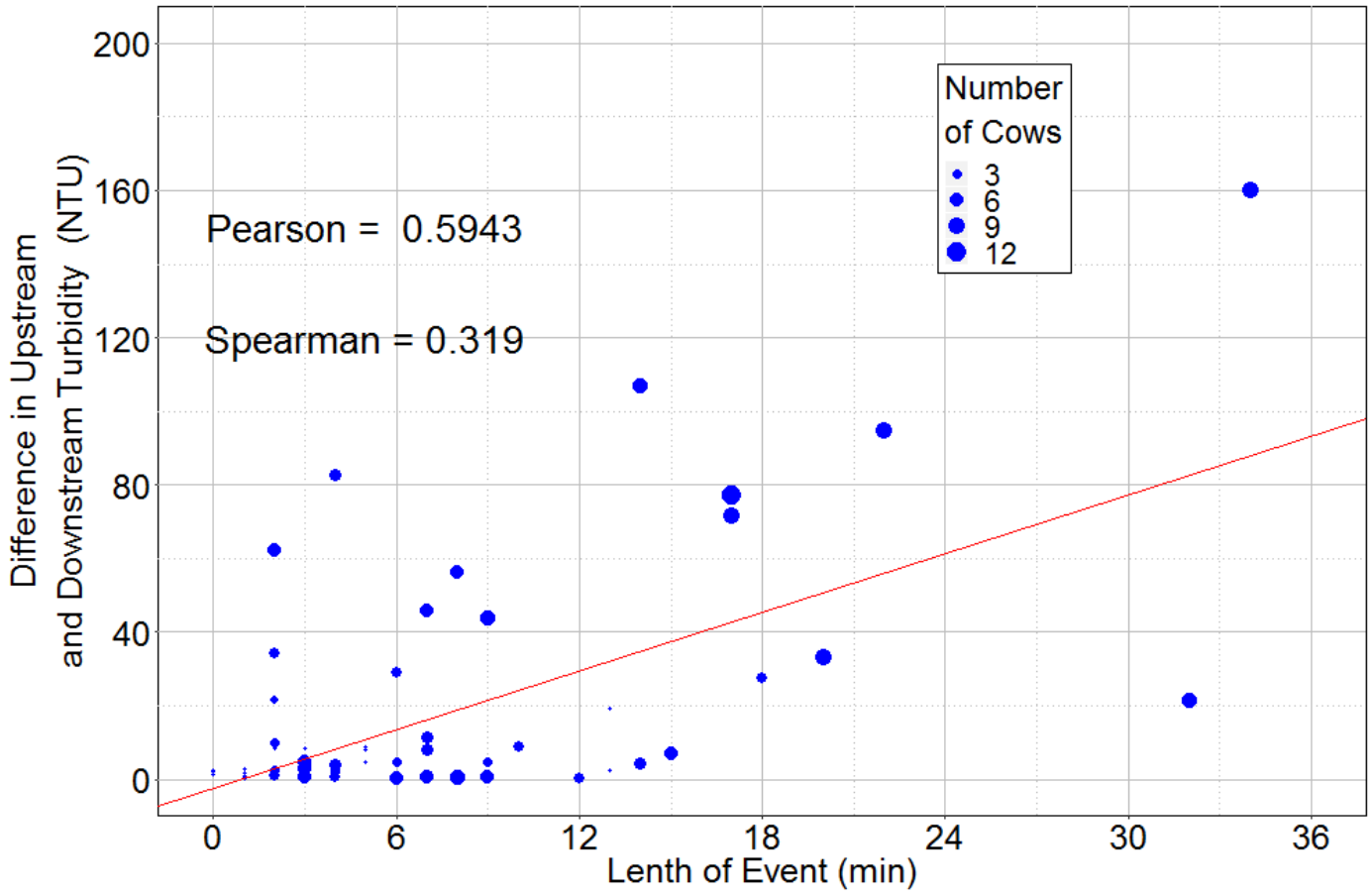

Figure 4.10: Correlation results between the length in min of a cattle access event and the difference between upstream and downstream turbidity. The number of cows at each event is represented by point size.

There was a moderate Pearson relationship between the length of time of an event and the difference in upstream and downstream turbidity (Figure 4.10). A weak Spearman correlation result here indicates there is not a monatomic relationship indicating that if an event increases in time it will not always result in a higher turbidity level downstream. The longest recorded event lasted 34 min and had the largest recorded difference in upstream and downstream turbidity. The general distribution of the number of cows suggest that smaller number of cows have a greater impact on difference in turbidity the longer they are in the stream. When cows were in the stream for over 14 min a greater rise in turbidity level downstream was seen. It was noted that cows spending longer than 13 min were always in groups of 2 or more.

Table 4.5: Distribution of the length of time spent by each group of cattle spent in stream during of each event.

Minimum (min)	Maximum (min)	Median (min)	Average (min)	Standard Deviation (min)
<1	34	5	7	± 7

There was a large standard deviation of ± 7 and cows appear to largely vary the amount of time spent in the stream (Table 4.5).

4.0 Conclusion

A cattle access event was described as any continues of entry of cattle into the stream where less than 5 min elapsed between each cow entering. The amount of entries to the stream was recorded as the number of cows. When more than 5 min elapsed between cattle exit from the stream and the entry of more cattle this was counted as the start of a new event. The length of each event from first entry to the last exit of the cattle was also recorded. It was observed that cattle access events with the same amount of entries by cows to the stream but of different lengths resulted in different levels of downstream turbidity. To determine what causes the disparity in these results the difference in downstream and upstream turbidity levels of each event was compared to daily river discharge levels, number of cows entering the stream, and length of the cattle access event. It was observed that cows mostly entered the stream in times of low river discharge this was not seen as a having a large impact on the change in turbidity levels. A strong Pearson correlation was seen between cow numbers in stream and difference in upstream and downstream turbidity and a moderate Pearson correlation was seen between the length of time of each event and the difference in upstream and downstream turbidity. Spearman correlation results were weak for both cow numbers and length event and no monotonic relationship was seen. The results concluded that when there is greater than 8 cows, or any number of cows spend over 14 min in the stream, there will be a higher result in downstream turbidity in comparison to upstream turbidity levels. Limiting the number of cows in stream to less than 8 and the reducing amount of time spent by each cow instream to less than 14 min has the potential to reduce impact on downstream turbidity levels.

References

Arismendi, I. *et al.* (2017) ‘Suspended sediment and turbidity after road construction/improvement and forest harvest in streams of the Trask River Watershed Study, Oregon’, *Water Resources Research*, 53(8), pp. 6763–6783. doi: 10.1002/2016WR020198.

Backer, L. C. *et al.* (2015) ‘Cyanobacteria and Algae Blooms: Review of Health and Environmental Data from the Harmful Algal Bloom-Related Illness Surveillance System (HABISS) 2007–2011’, pp. 1048–1064. doi: 10.3390/toxins7041048.

Bond, T. A., Sear, D. and Edwards, M. (2012) ‘Temperature-driven river utilisation and preferential defecation by cattle in an English chalk stream’, *Livestock Science*. Elsevier B.V., 146(1), pp. 59–66. doi: 10.1016/j.livsci.2012.02.022.

Carter, J. *et al.* (2017) ‘Upland Water and Deferred Rotation Effects on Cattle Use in Riparian and Upland Areas’, *Rangelands*. The Authors, 39(3–4), pp. 112–118. doi: 10.1016/j.rala.2017.06.003.

Hughes, A. O. *et al.* (2016) ‘Unrestricted dairy cattle grazing of a pastoral headwater wetland and its effect on water quality’, *Agricultural Water Management*. Elsevier B.V., 165, pp. 72–81. doi: 10.1016/j.agwat.2015.11.015.

Lizotte, R. E. and Locke, M. A. (2018) ‘Assessment of runoff water quality for an integrated best management practice system in an agricultural watershed’, 73(3), pp. 247–256. doi: 10.2489/jswc.73.3.247.

Mekonnen, M. M. and Hoekstra, A. Y. (2018) ‘Global Anthropogenic Phosphorus Loads to Freshwater and Associated Grey Water Footprints and Water Pollution Levels: A High-Resolution Global Study’, *Water Resources Research*, 54(1), pp. 345–358. doi: 10.1002/2017WR020448.

O’Flynn, B. *et al.* (2010) ‘Experiences and recommendations in deploying a real-time, water quality monitoring system’, *Measurement Science and Technology*, 21(12). doi: 10.1088/0957-0233/21/12/124004.

Oliver, D. M. *et al.* (2018) ‘Science of the Total Environment A catchment-scale model to predict spatial and temporal burden of E . coli on pasture from grazing livestock’,

Science of the Total Environment. The Author(s), 616–617, pp. 678–687. doi: 10.1016/j.scitotenv.2017.10.263.

Paerl, H. W., Otten, T. G. and Kudela, R. (2018) 'Mitigating the Expansion of Harmful Algal Blooms Across the Freshwater-to-Marine Continuum', *Environmental Science & Technology*. American Chemical Society, 52, pp. 5519–5529. doi: 10.1021/acs.est.7b05950.

Reis, S. *et al.* (2010) 'Impacts of European livestock production : nitrogen , sulphur , phosphorus and greenhouse gas emissions , land-use , water eutrophication and biodiversity Impacts of European livestock production : nitrogen , sulphur , phosphorus and greenhouse gas emissions , land-use , water eutrophication and biodiversity'. IOP Publishing.

Riley, S. J. (1998) 'The sediment concentration-turbidity relation: Its value in monitoring at Ranger Uranium Mine, Northern Territory, Australia', *Catena*, 32(1), pp. 1–14. doi: 10.1016/S0341-8162(97)00054-4.

Sun, H., Cornish, P. S. and Daniell, T. M. (2001) 'Turbidity-based erosion estimation in a catchment in South Australia', *Journal of Hydrology*, 253(1–4), pp. 227–238. doi: 10.1016/S0022-1694(01)00475-9.

Vidon, P., Campbell, M. A. and Gray, M. (2008) 'Unrestricted cattle access to streams and water quality in till landscape of the Midwest', *Agricultural Water Management*, 95(3), pp. 322–330. doi: 10.1016/j.agwat.2007.10.017.

Voigt, T. *et al.* (2007) 'Sensor Networking in Aquatic Environments - Experiences and New Challenges', in *32nd IEEE Conference on Local Computer Networks (LCN 2007)*, pp. 793–798. doi: 10.1109/LCN.2007.23.

Wagner, R. J. *et al.* (2006) *Guidelines and standard procedures for continuous water-quality monitors: station operation, record computation, and data reporting*.

Chapter 5: Conclusions

Overall summary and conclusion

The aim of this research is to use R statistical programming as an effective tool for trend analysis of sensor and historical data in order to determine the optimum conditions for deployment, future monitoring, and determine what precautions if any are to be taken. R statistical programming was chosen as a method of analysis as it is capable of dealing with large data sets whilst making analysis easily reproducible by other scientists. Water monitoring requires analysis over many different areas with biological and physical parameters required to be measured to insure the safety and health of all people. This study touched on three main areas of water analysis in three different types of waterbodies:

Ch2: Water level analysis in an urban river catchment.

Ch3: Bacterial analysis in a canal basin.

Ch4: Turbidity analysis at a cattle access point in a rural stream.

It also used three different types of water sensor deployment and data:

Ch2: Network sensor deployment.

Ch3: Historical grab sample analysis and single use rapid result sensor monitoring.

Ch4: Upstream and downstream multi- parameter sonde deployment for monitoring of a specific event.

This study tested the use of the Connect Sensor as a method for affordable water level network deployment in an urban catchment. It found that when the Connect Sensor used for affordable water level analysis it is most beneficial when each sensor is located no more than 8 km from each other. A case study of Storm Desmond showed how analysis of the data using the cross-correlation function used in R was able to provide information about how the different parts of the catchment behaved in an extreme rainfall event. Examining how each section of the river behaves in a storm invent can enable people to prioritise areas of the river that are more at risk earlier on in an event.

The examination of historical grab sample data in a canal basin was used to determine if rainfall had an impact on bacterial level due to the location of a storm water outflow on the edge of the basin. It was determined by the use of correlation

and statistical analysis performed in R that after rainfall events of 6 mm in magnitude there is a risk to bathers using the basin on the first and second day after the storm. For rainfall events larger than 25 mm there is an immediate risk to bathers in the entire basin and sampling should take place as soon as possible. There was no great risk to bathers from rainfall events less than 6 mm.

The ColiSense was tested as a method for improved monitoring in the basin as it has a rapid result time. Very strong correlation was found for the GUS activity monitored from the ColiSense with the, coliforms, E. coli, and enterococci monitored in the basin. It is recommended to use ColiSense so as to obtain results in a much shorter time. This can enable scientists to build larger more accurate data sets for future analysis.

Correlation and statistical analysis of real time turbidity data was performed in R to examine the impact of cows entering a stream. Results found that when there is greater than 8 cows, or any number of cows spend over 14 min in the stream, there will be a higher level of downstream turbidity in comparison to upstream levels. Limiting the number of cows in stream to less than 8 and the reducing amount of time spent by each cow instream to less than 14 min has the potential to reduce impact on downstream turbidity levels.

It is concluded that R statistical programming is an efficient tool for the trend analysis of large homogenous data sets relating to water bodies. A great deal of information can be obtained using relatively easy to use software packages provided by the programme. R programming can be used to cover a wide range of different methods of water analysis using sensor and historical monitoring data. It is recommended that R programming be used for future analysis of water monitoring.

Many advantages and disadvantages were noted between the different approaches of monitoring used in this thesis. A network of sensors gives the greatest picture of the water quality changes in an entire catchment. However the cost of each sensor makes network deployment expensive. While a low cost sensor like the Sonic Signalman, for measuring water level, is easily deployed amongst a catchment, an expensive multiparameter sonde such as the YSI would have a huge cost attached to network deployment. The use of sensors for rapid sampling is highly advantageous for bacterial monitoring where traditional grab sampling techniques

are timely and costly. It would not be practical to measure parameters used in the other studies taken place in the same way. The change in water level happened at very high rates, which would not be seen, if a real time sensor had not been used. The turbidity sensing was needed for the measurement of an uncontrolled event. Sensors here were required to be in the water measuring constantly in order to capture the exact moment cattle entered the stream.

Each approach to water monitoring and analysis was tailored to the needs of the type of water body in question, the events being examined, and the water parameter being measured. Water quality management will only be effective when the correct methods and analysis is employed. Sensor cost and available technology are the biggest factors in water quality monitoring but the use of historical monitoring and event monitoring can give valuable insight into how water quality changes. Historical monitoring gives an accurate glimpse of overall change of water quality over time. Event monitoring can give a valuable insight into what is specifically causing changes in water parameters. It is therefore desirable that sensors regardless of type can be used for both types of analysis, of historical data and event data.

Chapter 6: Appendices

6.1 R code used in Chapter 2: Affordable Water Level Monitoring Sensors for Network Deployment

6.1.1 Sensor validation overlay graph

```
######################
#Programmes needed
library(ggplot2)
library(gridExtra)
library(scales)
library(corrplot)
######################
######################
g1<-ggplot(Ddata, aes(x=dt)) +
  geom_line(aes(y=volume, color="Connect Sensor"))+
  geom_point(aes(y=volume, color="Connect Sensor"))
g1
g2<-g1+ geom_line(data=Ddata,
aes(y=volume, color="Dublin City Sensor"))+
                        geom_point(data=Ddata, aes(y=volume, color="Dublin City
Sensor"))+
  scale_color_manual(name = "Sensors",
                     breaks = c("Connect Sensor","Dublin City Sensor"),
                     values = c("#FF0003", "#00CCCC"))+
  theme(
panel.background = element_rect(fill = 'white', colour = 'grey'),
```

```
                legend.position = c(0.86,0.9), legend.box = "horizontal",
                plot.title = element_text(size = rel(1.5)))+
                theme_bw()+
 scale_y_continuous(
                limits=c(0,0.4),
               breaks = seq(-0.2,1.2, 0.05))+
  scale_x_datetime(
breaks = date_breaks("12 hours"),
labels=date_format("%d %b %H:%M"),
xlab("Date")+
ylab("Water Level (m)")

g2
#########################
#########################
```

6.1.2 Sensor validation correlation graph

```
######################
#Programmes needed
require(ggplot2)
require(gridExtra)
require(scales)
require(corrplot)
######################
######################
Merged<-merge(Ddata, Mdata, by=intersect("dt","dt"), all=T)
Merged<-Merged[,c("volume.x","volume.y")]
Merged<-na.omit(Merged)
Merged<-Merged[order(Merged$volume.x),]
nrow(Merged)
#######
max=max(Merged$volume.x)
min=min(Merged$volume.x)
x=0 #0 to 10
y=x+1
data <- subset(Merged, Merged$volume.x >= min + (x * (max/3)))
data<- subset(data, data$volume.x <= min + (y * (max/3)))
nrow(data)
ccf(Mdata$volume, rain$mm, lag.max=1000, plot=TRUE, main="River Level and Rain Lady's Lane", cex=66)
pear<-cor(data$volume.x, data$volume.y, method="pearson")
```

```
line<-coef(lm(volume.y ~ volume.x, data = data))

#################

#Plot Merged Data

ggplot(data, aes(volume.x,volume.y))+

geom_point(col="blue")+

geom_abline(intercept=line[c(1)] ,  slope=line[c(2)] , col="red" )+

annotate("text", x = 0.25, y = 0.08, label = "Pearson Correlation Coefficient=")+

annotate("text", x = 0.25, y = 0.07, label = round(pear, digits = 4))+

xlab("Dublin City Sensor (m)")+ #x axis label

ylab("Connect Sensor (m)") +  #y axis lable

 theme(axis.text.x  = element_text(angle=90, vjust=0.5))
```

6.1.3 Graph of lag time between sites

```
######################
#Programmes needed
require(ggplot2)
require(ggrepel)
require(RColorBrewer)
######################
######################


theme_next <-  theme(
                        panel.background = element_rect(fill = 'white', colour = 'black'),
                        legend.position = c(0.2,0.2),
                        legend.box = "horizontal",
                        legend.title = element_text(size = 19),
                        plot.title = element_text(size = 22),
                        axis.text = element_text(size = 16, colour = "black"),
                        axis.title = element_text(size = 18),
                        legend.text = element_text(size= 16),
                        legend.key = element_rect(fill = "white"),
                        legend.background = element_rect(colour = "black"),
                        panel.grid.major = element_line(colour = "grey"),
                         panel.grid.minor = element_line(colour = "grey", linetype =
"dotted"))


col <-brewer.pal(7, "Set1")
```

```
####################

#####################1

ggplot(BAC,

                aes(colour="Bohernabreena and Austin Clarke", x =time, y=pear )) +

geom_line( alpha=0.4, size=1.5) +

geom_point()+

geom_label_repel(data=subset(BAC, pear==max(pear)),

aes(time,pear,label="Max Spearman", colour="Bohernabreena and Austin Clarke"),

                nudge_y=0.01,size = 6)+

geom_label_repel(data=subset(BWB, pear==max(pear)),

aes(time,pear,label="Max Spearman", colour="Bohernabreena and Waldron's Bridge"),

                nudge_y=0.01,size = 6)+

geom_label_repel(data=subset(BCB, pear==max(pear)),

aes(time,pear,label="Max Spearman", colour="Bohernabreena and Clonskeagh Bridge"),

               nudge_y=-0.01,size = 6)+

geom_label_repel(data=subset(ACWB, pear==max(pear)),

aes(time,pear,label="Max Spearman", colour="Austin Clarke and Waldron's Bridge"),

                nudge_y=2, nudge_x =-1,size = 6)+

geom_label_repel(data=subset(ACCB, pear==max(pear)),

aes(time,pear,label="Max Spearman", colour="Austin Clarke and Clonskeagh Bridge"),

                 nudge_y=1,size = 6)+

geom_label_repel(data=subset(WBCB, pear==max(pear)),
```

```
aes(time,pear,label="Max Spearman", colour="Waldron's Bridge and Clonskeagh Bridge"),
                nudge_y=-0.02,nudge_x=-1,size = 6)+
geom_line(data=BWB,
aes(x=time,y=pear,colour="Bohernabreena and Waldron's Bridge"),alpha=0.4, size=1.5)+
geom_point(data=BWB,
                aes(x=time,y=pear,colour="Bohernabreena and Waldron's Bridge"))+
geom_line(data=BCB,
aes(x=time,y=pear,colour="Bohernabreena and Clonskeagh Bridge"),alpha=0.4, size=1.5)+
geom_point(data=BCB,
                aes(x=time,y=pear,colour="Bohernabreena and Clonskeagh Bridge"))+
geom_line(data=ACWB,
aes(x=time,y=pear,colour="Austin Clarke and Waldron's Bridge"),alpha=0.4, size=1.5)+
geom_point(data=ACWB,
                aes(x=time,y=pear,colour="Austin Clarke and Waldron's Bridge"))+
geom_line(data=ACCB,
aes(x=time,y=pear,colour="Austin Clarke and Clonskeagh Bridge"),alpha=0.4, size=1.5)+
geom_point(data=ACCB,
                aes(x=time,y=pear,colour="Austin Clarke and Clonskeagh Bridge"))+
geom_line(data=WBCB,
aes(x=time,y=pear,colour="Waldron's Bridge and Clonskeagh Bridge"),alpha=0.4, size=1.5)+
```

```
geom_point(data=WBCB,
aes(x=time,y=pear,colour="Waldron's Bridge and Clonskeagh Bridge"))+
ggtitle(element_blank()) +
coord_cartesian(xlim =c(0,20), ylim =c(0.8,0.98))+
scale_x_continuous( breaks = seq(0, 1440, by = 5))+
xlab("Lag Time (min)")+
ylab("Spearman's Ranked Coefficient") +
scale_colour_manual( name = "River Flow Downstream to Upstream",
                     values = c("Bohernabreena and Austin Clarke"=col[1],
                                "Bohernabreena and Waldron's Bridge"=col[2],
                                "Bohernabreena and Clonskeagh Bridge"=col[3],
                                "Austin Clarke and Waldron's Bridge"=col[4],
                                "Austin Clarke and Clonskeagh Bridge"=col[5],
                                "Waldron's Bridge and Clonskeagh Bridge"=col[7]),
                     breaks=c("Bohernabreena and Austin Clarke",
                              "Bohernabreena and Waldron's Bridge",
                              "Bohernabreena and Clonskeagh Bridge",
                              "Austin Clarke and Waldron's Bridge",
                              "Austin Clarke and Clonskeagh Bridge",
                              "Waldron's Bridge and Clonskeagh Bridge"))+
theme_next
```

6.1.4 Graph of lag time between sites and the closest rainfall gauge

```
######################
#Programmes needed
require(ggplot2)
require(ggrepel)
require(RColorBrewer)
######################
######################
theme_next <-  theme(
  `       panel.background = element_rect(fill = 'white', colour = 'black'),
          legend.position = c(0.4,0.2),
          legend.box = "horizontal",
          legend.title = element_blank(),
          plot.title = element_text(size = 22),
          axis.text = element_text(size = 16, colour = "black"),
          axis.title = element_text(size = 18),
          legend.text = element_text(size= 16),
          legend.key = element_rect(fill = "white"),
          legend.background = element_rect(colour = "black"),
          panel.grid.major = element_line(colour = "grey"),
          panel.grid.minor = element_line(colour = "grey", linetype = "dotted"))
cole = c(brewer.pal(8, "Dark2")[c(1,2,3,4,5,6,7,8)], brewer.pal(11, "RdGy")[c(11)])


ggplot(data=BB, aes(x=time))+
geom_label_repel(data=subset(DCL, pear==max(pear)),
```

```
                aes(time,pear,label="MxS", colour="Lady's Lane"),
                nudge_y=-0.02,size = 6)+
geom_label_repel(data=subset(DCG, pear==max(pear)),
                aes(time,pear,label="MxS", colour="Gandon"),
                nudge_y=0.04, size = 6)+
geom_label_repel(data=subset(BK, pear==max(pear)),
aes(time,pear,label="MxS", colour="Kilmashogue"),
nudge_y=0.02,size = 6)+
geom_label_repel(data=subset(DCAG, pear==max(pear)),
aes(time,pear,label="MxS", colour="Ardglas"),
nudge_y=-0.02,size = 6)+
geom_label_repel(data=subset(BBC, pear==max(pear)),
aes(time,pear,label="MxS", colour="Brehon's Chair"),
nudge_y=0.05,size = 6)+
geom_label_repel(data=subset(BB, pear==max(pear)),
aes(time,pear,label="MxS", colour="Bohernabreena"),
nudge_y=0.02,size = 6)+
geom_label_repel(data=subset(DCCB, pear==max(pear)),
aes(time,pear,label="MxS", colour="Clonskeagh Bridge"),
nudge_y=-0.05,size = 6)+
geom_label_repel(data=subset(DCAC, pear==max(pear)),
aes(time,pear,label="MxS", colour="Austin Clarke"),
nudge_y=0.04,size = 6)+
geom_label_repel(data=subset(DCWB, pear==max(pear)),
```

```
                aes(time,pear,label="MxS", colour="Waldron's Bridge"),
nudge_y=1,size = 6)+
geom_line(aes(y=pear, color="Bohernabreena"), alpha=0.4, size=1.5)+
geom_point(aes(y=pear, color="Bohernabreena"))+
geom_line(data=BBC, aes(x=time, y=pear, color="Brehon's Chair"), alpha=0.4,
size=1.5)+
geom_point(data=BBC, aes(x=time, y=pear, color="Brehon's Chair"))+
geom_line(data=DCG, aes(x=time, y=pear, color="Gandon"), alpha=0.4, size=1.5)+
geom_point(data=DCG, aes(x=time, y=pear, color="Gandon"))+
geom_point(data=BK, aes(x=time, y=pear, color="Kilmashogue"))+
geom_line(data=BK, aes(x=time, y=pear, color="Kilmashogue"),alpha=0.4,
                size=1.5)+
geom_point(data=DCAC, aes(x=time, y=pear, color="Austin Clarke"))+
geom_line(data=DCAC, aes(x=time, y=pear, color="Austin Clarke"), alpha=0.4,
                size=1.5)+
geom_point(data=DCAG, aes(x=time, y=pear, color="Ardglas"))+
geom_line(data=DCAG, aes(x=time, y=pear, color="Ardglas"), alpha=0.4, size=1.5)+
geom_point(data=DCCB, aes(x=time, y=pear, color="Clonskeagh Bridge"))+
geom_line(data=DCCB, aes(x=time, y=pear, color="Clonskeagh Bridge"), alpha=0.4,
size=1.5)+
geom_point(data=DCL, aes(x=time, y=pear, color="Lady's Lane"))+
geom_line(data=DCL, aes(x=time, y=pear, color="Lady's Lane"),alpha=0.4,
                size=1.5)+
geom_point(data=DCWB, aes(x=time, y=pear, color="Waldron's Bridge"))+
```

```
geom_line(data=DCWB, aes(x=time, y=pear, color="Waldron's Bridge"),alpha=0.4,
                size=1.5)+
scale_x_continuous(
                limits=c(0,720),
                breaks = seq(0, 1440, by = 120) )+
labs(x="Lag Time (min)", y="Spearman's Ranked
Coefficient",title=element_blank())+
theme_next+
scale_color_manual(values=cole)
```

6.1.5 CCF and overlay graph of connect sensor and rainfall during Storm Desmond

```
####################
#Programes needed
library(ggplot2)
library(scales)
library(gtable)
library(cowplot)
library(grid)
library(gridExtra)
library(plotrix)
library(dplyr)
######################
####################
Merged<-merge(rdataB, dataB, by=intersect("datetime","datetime"), all=T)
Merged<-na.omit(Merged)
nrow(Merged)
Merged<- subset(Merged,
                datetime >= as.POSIXct('2015-12-03 00:00:00') &
                  datetime <= as.POSIXct('2015-12-03 23:55:00'))
```

```
#########################
########Graph 1 CCF#####
#########################
lagcorr<-ccf(Merged$volume, Merged$mm, lag.max=1000, plot=F, main="")
attributes(lagcorr)
write.csv(lagcorr$acf,"severe.csv")
ccf(Merged$volume, Merged$mm, lag.max=1000, plot=T, main="")
#########################
########Graph 2 Overlay###
#########################
p1<-ggplot(data=Merged, aes(x=datetime))+
geom_line(aes(y=volume, color="Connect Sensor"), size=0.8)+
geom_point(aes(y=volume, color="Connect Sensor"), size=0.8)+
geom_point(data=Merged, aes(x=datetime, y=0, color="Rainfall"), size=0.8)+
scale_color_manual(name = element_blank(),
                   breaks = c("Connect Sensor","Rainfall"),
                   values = c("#FF0003", "#00CCCC"))+
scale_y_continuous(limits=c(0,1.5))+
scale_x_datetime(date_labels = "%H:%M",date_breaks = "2 hour")+
labs(x="", y="Water Level (m)",title=element_blank())+
theme_bw()+
theme(legend.justification=c(0.6,0.95),
          legend.position=c(0.8,0.95),
          legend.text = element_text(size= 20),
```

```
        legend.key.size = unit(1.5, "cm"),
        plot.title=element_text(size=30,vjust=1),
        axis.text.x=element_text(size=20),
        axis.text.y=element_text(size=20),
        axis.title.x=element_text(size=20),
        axis.title.y=element_text(size=20))
p2<-ggplot() + geom_line(data=Merged, aes(x=datetime, y=mm,
        color="Rainfall"), size=0.8)+
geom_point(data=Merged, aes(x=datetime, y=mm, color="Rainfall"), size=0.8)+
scale_color_manual(name = element_blank(),
                   breaks = c("Rainfall"),
                   values = c("#00CCCC"))+
scale_y_continuous(limits=c(0,3))+
scale_x_datetime(date_labels = "%H:%M",date_breaks = "2 hour")+
theme_bw()+
theme(
        panel.background = element_rect(fill = NA),
        panel.grid.major.x=element_blank(),
        panel.grid.minor.x=element_blank(),
        panel.grid.major.y=element_blank(),
        panel.grid.minor.y=element_blank(),
        axis.text.y=element_text(size=20),
        axis.title.x=element_text(size=20),
        axis.title.y=element_text(size=20))+
```

```
xlab("")+
ylab("Rainfall (mm)")
g1 <- ggplot_gtable(ggplot_build(p1))
g2 <- ggplot_gtable(ggplot_build(p2))
pp <- c(subset(g1$layout, name == "panel", se = t:r))
g <- gtable_add_grob(g1,
                     g2$grobs[[which(g2$layout$name == "panel")]],
                     pp$t, pp$l, pp$b, pp$l)
alab <- g2$grobs[[which(g2$layout$name=="ylab-l")]]
ia <- which(g2$layout$name == "axis-l")
ga <- g2$grobs[[ia]]
ax <- ga$children[[2]]
ax$widths <- rev(ax$widths)
ax$grobs <- rev(ax$grobs)
ax$grobs[[1]]$x <- ax$grobs[[1]]$x - unit(1, "npc") +
unit(0.15, "cm")
g <- gtable_add_cols(g, g2$widths[g2$layout[ia, ]$l],
                     length(g$widths) - 1 )
g <- gtable_add_cols(g, g2$widths[g2$layout[ia, ]$l],
                     length(g$widths) - 1 )
g <- gtable_add_grob(g, ax, pp$t, length(g$widths) - 2, pp$b)
g <- gtable_add_grob(g, alab, pp$t, length(g$widths) - 1, pp$b)
grid.draw(g)
```

6.2 R code used in Chapter 3: Evaluation of the Occurrence of Bacterial Contamination in Grand Canal Basin Dublin

6.2.1 Graph of bacterial concentration correlation between sites

```
##################
#Programmes needed
library(ggplot2)
library(gridExtra)
library(scales)
library(corrplot)
library(grid)
library(sitools)
######################
######################
theme_other <-  theme(
        panel.background = element_rect(fill = 'white', colour = 'black'),
        legend.position = c(0.85,0.9), legend.box = "horizontal",
        plot.title = element_text(size = 22),
        axis.text = element_text(size = 16, colour = "black"),
        axis.title = element_text(size = 18),
        legend.background = element_rect(colour = "black"))
Merged<-merge(data6, data7, by=intersect("date","date"), all=T)
Merged<-Merged[,c("ecoli.x","ecoli.y")]
Merged<-na.omit(Merged)
Merged<-Merged[order(Merged$volume.x),]
```

```
nrow(Merged)
pear<- cor(Merged$ecoli.x, y= Merged$ecoli.y, method="pearson")
spear <- cor(Merged$ecoli.x, y= Merged$ecoli.y, method="spearman")
 line<-coef(lm(ecoli.y ~ ecoli.x, data = Merged))
label1 <- paste("Pearson = ", round(pear,3))
label2 <- paste("Spearman =", round(spear,3))
label3 <- paste('y =',round(line[[2]],2),'x', '+', round(line[[1]],2))
ggplot(Merged, aes(ecoli.x, ecoli.y))+
geom_point(col="blue", size=0.9)+
geom_abline(intercept=line[c(1)] ,  slope=line[c(2)] , col="red" )+
ylim(0,25000)+
 xlim(0,25000)+
 scale_y_continuous(labels=f2si, limits = c(0, 25000))+
 scale_x_continuous(labels=f2si, limits = c(0, 25000))+
 annotate("text", x = 7500, y = 18000, label = label1,  size = 7)+
 annotate("text", x = 7500, y = 16000, label = label2, size = 7)+
 annotate("text", x = 7500, y = 14000, label = label3,  size = 7)+
 xlab("Site 6")+ #x axis label
 ylab("Site 7") +  #y axis lable
 theme_other+
 ggtitle(element_blank())
```

6.2.2 Pearson results for rain gauges

```
##################

#Programmes needed

library(corrplot)

######################

######################

corrplot(datar, method = "number")
```

6.2.3 Correlation results between days if rainfall and bacterial concentration

```
###################

#Programmes needed

library(ggplot2)

library(scales)

library(gtable)

library(cowplot)

library(grid)

library(gridExtra)

library(plotrix)

###################

###################

#Change reading to average of 2 days previous

rain2<-rdata

rain2$rain<-filter(x=rain2$rain, filter=rep(1/2, 2), sides=1)



#################################
```

```
##########Site minus Day########
#################################
data<-site7.csv
data<-data[,c(1,21)]
data$date<- as.Date(data$date, format="%d/%m/%Y")
data<-na.omit(data)
day1<-datamo
day1$date<-data$date -1
########################
########################
theme_other <-  theme(
  panel.background = element_rect(fill = 'white', colour = 'black'),
  legend.position = c(0.85,0.9), legend.box = "horizontal",
  plot.title = element_text(size = 22),
  axis.text = element_text(size = 18, colour = "black"),
  axis.title = element_text(size = 20),
  legend.background = element_rect(colour = "black"))
Merged<-merge(rain2, day1, by=intersect("date","date"), all=T)
names(Merged)<-c("date","rain","reading")
Merged<-Merged[,c("rain","reading")]
Merged<-na.omit(Merged)
#Merged$reading<-log10(Merged$reading)
nrow(Merged)
summary(Merged)
```

```
pear<- cor(y=Merged$reading, x= Merged$rain, method="pearson")
spear <- cor(y=Merged$reading, x= Merged$rain, method="spearman")
line<-coef(lm(reading ~ rain, data = Merged))
label1 <- paste("Pearson = ", round(pear,3))
label2 <- paste("Spearman =", round(spear,3))
label3 <- paste('y =',round(line[[2]],3),'x', '+', round(line[[1]],3))
ggplot(Merged, aes(rain, reading))+
geom_point(col="blue", size=0.9)+
geom_abline(intercept=line[c(1)] ,  slope=line[c(2)] , col="red" )+
annotate("text", y = 22000, x =  25, label = label1,  size = 7)+
annotate("text", y = 20000, x =  25, label = label2,  size = 7)+
annotate("text", y = 18000, x =  25, label = label3,  size = 7)+
xlim(0,40)+
ylim(0,25000)+
ylab("E. Coli (MPN/100ml)")+#x axis label
ylab(expression(paste(Log[10], " E. Coli (MPN/100ml)")))+
xlab("Rainfall (mm)") +  #y axis lable
ggtitle("Site 7")+
theme_other
```

6.2.4 Rainfall condition graph

```
##################
#Programmes needed
library(ggplot2)
library(scales)
library(gtable)
library(cowplot)
library(grid)
library(gridExtra)
library(plotrix)
library(RColorBrewer)
#################
#################
theme_other <-  theme(
        panel.background = element_rect(fill = 'white', colour = 'black'),
        legend.position = c(0.85,0.8), legend.box = "horizontal",
        legend.title=element_text(size=19),
        legend.text=element_text(size=16),
        plot.title = element_text(size = 22),
        axis.text = element_text(size = 16, colour = "black"),
        axis.title = element_text(size = 18),
        legend.background = element_rect(colour = "black"))
cols <- brewer_pal(palette = "Set1")(4)
ggplot(storm2.5, aes(day, reading,color="Less than 2.5mm",shape="Less than
2.5mm"))+
```

```
geom_point( size=10)+
geom_point(data=storm6.4,aes(y=reading, color="Between 2.5 and 6.4mm",
        shape="Between 2.5 and 6.4mm"), size=10)+
geom_point(data=storm25,aes(y=reading,color="Between 6.4 and
        25mm",shape="Between 6.4 and 25mm"),  size=10)+
geom_point(data=stormB,aes(y=reading,  color="Over 25mm", shape="Over
        25mm"), size=10)+
 geom_abline(intercept=900 ,  slope=0 , col="black" )+
 annotate("text", y = 800, x =  2,col="black", label = "Bathing Water Threshold",  size
        = 5)+
ylim(0,3000)+
ylab("Average E.Coli (MPN/100ml)")+#x axis label
xlab("Days Since Rainfall") +  #y axis lable
ggtitle("Site 7")+
theme_other+
scale_y_continuous(breaks = seq(0,2500, 500))+
scale_x_continuous(breaks = seq(0,100, 1))+
scale_colour_manual(name = "Rain Characteristic",
                    breaks=c("Less than 2.5mm","Between 2.5 and 6.4mm","Between 6.4
                    and 25mm","Over 25mm"),
                    labels = c("Less than 2.5mm","Between 2.5 and 6.4mm","Between 6.4
                    and 25mm","Over 25mm"),
                    values = (cols)) +
scale_shape_manual(name = "Rain Characteristic",
```

```
        breaks=c("Less than 2.5mm","Between 2.5 and 6.4mm","Between 6.4
        and 25mm","Over 25mm"),
        labels = c("Less than 2.5mm","Between 2.5 and 6.4mm","Between 6.4
        and 25mm","Over 25mm"),
        values = c(16, 17, 18, 8))
```

6.2.5 Graph of percentage bathing water breaches

```
##################
#Programmes needed
library(ggplot2)
library(data.table)
library(reshape)
###################
###################
theme_next <-  theme(
        panel.background = element_rect(fill = 'white', colour = 'black'),
        legend.position = c(0.85,0.9), legend.box = "horizontal",
        plot.title = element_text(size = 22),
        axis.text = element_text(size = 16, colour = "black"),
        axis.title = element_text(size = 18),
        legend.background = element_rect(colour = "black"),
        panel.grid.major = element_line(colour = "grey"),
        panel.grid.major.x = element_blank(),
        panel.grid.minor = element_blank())
```

```
rain2<-rdata
rain2$rain<-filter(x=rain2$rain, filter=rep(1/2, 2), sides=1)
rain3<-rdata
rain3$rain<-filter(x=rain3$rain, filter=rep(1/3, 3), sides=1)
rain4<-rdata
rain4$rain<-filter(x=rain4$rain, filter=rep(1/4, 4), sides=1)
day1<-data
day1$date<-data$date -1
day2<-data
day2$date<-data$date -2
summary(day1)
Merged<-merge(rdata, day1, by=intersect("date","date"), all=T)
names(Merged)<-c("date","rain","reading")
Merged<-Merged[,c("rain","reading")]
Merged<-na.omit(Merged
All<-rbind(Merged4,Merged5)
read<-dcast(setDT(All), rain~rowid(rain, prefix="reading"), value.var="reading")
read$percent<-rowSums(read[,-1]> 330,na.rm=TRUE)/rowSums(!is.na(read))*100
per<-read[,c(1,36)]
per$percent <- round(per$percent,2)
average<-loess(percent ~ rain, data=per, span=0.6)
smoothed<-predict(average)
ggplot(per, aes(x=rain, y=(percent / 100)))+
geom_bar(aes(x=rain,y=(percent / 100)), stat="identity",col="chartreuse4")+
```

```
geom_point(col="chartreuse4", size=2)+
geom_line(aes(x=per$rain,y=(smoothed/ 100)), col="grey", size=5, alpha=0.4)+
geom_abline(intercept=0.6 ,  slope=0 , col="red" )+
geom_vline(xintercept=6, col="red" )+
scale_y_continuous(breaks = seq(0,1, 0.1), labels = percent, limits=c(0,1))+
scale_x_continuous(breaks = seq(0,100, 5))+
                    limits=c(0,20))+
ylab("Percentage Breach of Enterococci Bathing Water Standards")+#x axis label
ylab(expression(paste(Log[10], " E. Coli (MPN/100ml)")))+
xlab("Daily Rainfall mm") +
theme_next
```

6.2.6 Graph for ColiSense analysis

```
##################
#Programmes needed
library(ggplot2)
library(scales)
library(gtable)
library(grid)
###################
###################
theme_next <-  theme(
        panel.background = element_rect(fill = 'white', colour = NA),
        panel.border = element_rect(fill = NA, colour = "black", size = 2),
        legend.position = c(0.85,0.4), legend.box = "horizontal",
```

```
        plot.title = element_text(size = 22),
        axis.text = element_text(size = 16, colour = "black"),
        axis.title = element_text(size = 18),
        legend.background = element_rect(colour = "black"),
        panel.grid.major = element_line(colour = "grey"),
        panel.grid.minor = element_line(colour = "grey", linetype = "dotted"),
        plot.background = element_rect(fill = "lightblue", colour = "black",
        size = 2, linetype = "longdash"))
limits <- aes(ymax = GUS + STD, ymin=GUS - STD)
pear<- cor(data$GUS, y= data$reading, method="pearson")
spear <- cor(data$GUS.average, y= data$E.Coli, method="spearman")
line<-coef(lm( GUS~ reading, data = data))
label1 <- paste("Pearson = ", round(pear,3))
label2 <- paste('y =',round(line[[2]],2),'x', '+', round(line[[1]],2))
label2 <- paste('y =',round(line[[2]],2),'x', round(line[[1]],2))
ggplot(data, aes(x=reading, y=GUS, label=Date))+
geom_point(aes(col=Location), size=4)+
geom_text(vjust = 0, nudge_y = 0.5)+
geom_abline(intercept=line[c(1)] ,  slope=line[c(2)] , col="red" , size=3, alpha=0.4)+
annotate("text", x = 1500, y = 3, label = label1,  size = 7)+
annotate("text", x = 1500, y = 2, label = label2,  size = 7)+
geom_errorbar(limits, width=100)+
scale_y_continuous(breaks = seq(0,1, 0.1), labels = percent,
                   limits=c(0,1))+
```

```
scale_x_continuous(breaks = seq(0,100, 5))+
        limits=c(0,20))+
 xlab("E. Coli (MPN/100 ml)") +
ylab("GUS (pmol/min/100 ml)") +
theme_next
```

6.3 R code used in Chapter 4: A Real-Time Monitoring Approach for Examining Water Quality Changes Upstream and Downstream From a Cattle Access Point

6.3.1 Event characteristic graph

```
##################
#Programmes needed
library(ggplot2)
library(scales)
library(gtable)
library(cowplot)
library(grid)
library(gridExtra)
library(plotrix)
library(RColorBrewer)
###################
###################
col <-brewer.pal(9, "Set1")
col2 <-brewer.pal(12, "Set3")
```

```
date<-as.POSIXct(c("2017-07-06 00:00:00","2017-07-06 23:55:00"))
p1<- ggplot(datau, #makes sure ggplot knows what type of data is has to plot
        aes(x = datetime)) +
 geom_line(size=2, aes(y= turbid, colour="Up Stream", fill="Up Stream")) +
geom_line(data=datad, aes(x=datetime,y=turbid,colour="Down Stream", fill="Down
                Stream"),size=2)+
geom_bar(data=datac, aes(x=dt, y=0, fill="Cow Numbers"),stat = 'identity', position
                'dodge')+
scale_colour_manual( name = element_blank(), guide=FALSE,
                values = c(
                                alpha(("Down Stream"=col[5]),0.5),
                                alpha(("Up Stream"=col2[10]),0.5),
                                alpha(("Cow Numbers"=NA),1)))+
scale_fill_manual(name = element_blank(),
                values = c(
                                alpha(("Cow Numbers"="black"),1),
                                alpha(("Down Stream"=col[5]),1),
                                alpha(("Up Stream"=col2[10]),1)))+
scale_x_datetime( limits = date)+
scale_y_continuous(limits=c(0,400))+
labs(x="", y="Turbidity (NTU)",title=element_blank())+
theme_bw()+
theme(legend.justification=c(0.6,1),
        legend.position=c(0.85,0.95),
```

```
        legend.text = element_text(size= 20),
        legend.key.size = unit(1.5, "cm"),
        plot.title=element_text(size=30,vjust=1),
        axis.text.x=element_text(size=20),
        axis.text.y=element_text(size=20),
        axis.title.x=element_text(size=20),
        axis.title.y=element_text(size=20))
p2<-ggplot() + geom_bar(data=datac, aes(x=dt, y=cows, fill="Cow
Numbers"),size=0.5,stat =
                'identity', position = 'dodge', width=60, alpha=1)+
scale_fill_manual(name = element_blank(),
                breaks = c("Cow Numbers"),
                values = c("black"))+
scale_y_continuous(limits=c(0,20))+
scale_x_datetime( limits = date)+
geom_text(data=datac,aes(label=cows,y=cows, x=dt), vjust=0, colour="red",
                size=9)+
theme_bw()+
theme(panel.background = element_rect(fill = NA),
        panel.grid.major.x=element_blank(),
        panel.grid.minor.x=element_blank(),
        panel.grid.major.y=element_blank(),
        panel.grid.minor.y=element_blank(),
        axis.text.y=element_text(size=20),
        axis.title.y=element_text(size=20),
```

```
        axis.title.x=element_text(size=20))+
xlab("")
ylab("No. of Cows")
g1 <- ggplot_gtable(ggplot_build(p1))
g2 <- ggplot_gtable(ggplot_build(p2))
pp <- c(subset(g1$layout, name == "panel", se = t:r))
g <- gtable_add_grob(g1,
                     g2$grobs[[which(g2$layout$name == "panel")]],
                     pp$t, pp$l, pp$b, pp$l)
alab <- g2$grobs[[which(g2$layout$name=="ylab-l")]]
ia <- which(g2$layout$name == "axis-l")
ga <- g2$grobs[[ia]]
ax <- ga$children[[2]]
ax$widths <- rev(ax$widths)
ax$grobs <- rev(ax$grobs)
ax$grobs[[1]]$x <- ax$grobs[[1]]$x - unit(1, "npc") +
                     unit(0.15, "cm")
g <- gtable_add_cols(g, g2$widths[g2$layout[ia, ]$l],
                     length(g$widths) - 1 )
g <- gtable_add_cols(g, g2$widths[g2$layout[ia, ]$l],
                     length(g$widths) - 1 )
g <- gtable_add_grob(g, ax, pp$t, length(g$widths) - 2, pp$b)
g <- gtable_add_grob(g, alab, pp$t, length(g$widths) - 1, pp$b)
grid.draw(g)
```

6.3.2 Mean turbidity and river discharge graph

```
##################
#Programmes needed
library(ggplot2)
library(scales)
library(gtable)
library(cowplot)
library(grid)
library(gridExtra)
library(plotrix)
library(RColorBrewer)
###################
###################
Merged<-merge(datad, datau, by=intersect("datetime","datetime"), all=T)
Merged$avg<-rowMeans(Merged[c('turbid.y','turbid.x')])
Merged<-Merged[,c("datetime","diff")]
Merged<-na.omit(Merged)
nrow(Merged)
col <-brewer.pal(9, "Set1")
col2 <-brewer.pal(12, "Set3")
p1<- ggplot(Merged, #makes sure ggplot knows what type of data is has to plot
                    aes(x = datetime)) + #Index and variable to be plotted
geom_line(size=1, aes(y= avg, colour="Mean Turbidity")) +
geom_line(data=dataf, aes(x=date,y=flow,colour="River Discharge"),size=1)+
scale_colour_manual( name = element_blank(),
```

```
                    values = c("Mean Turbidity"="tan4",
                     "River Discharge"="blue2")
                          )+
scale_y_continuous(limits=c(0,800))+
labs(x="", y="Turbidity (NTU)",title=element_blank())+
theme_bw()+
theme(legend.justification=c(0.6,1),
          legend.position=c(0.87,0.98),
          legend.text = element_text(size= 20),
          legend.key.size = unit(1.5, "cm"),
          plot.title=element_text(size=30,vjust=1),
          axis.text.x=element_text(size=20),
          axis.text.y=element_text(size=20),
          axis.title.x=element_text(size=20),
          axis.title.y=element_text(size=20))
p2<-ggplot() +
geom_line(data=dataf, aes(x=date, y=flow, colour="River Discharge"),size=1)+
scale_colour_manual(name = element_blank(),
                    breaks = c("River Discharge"),
                    values = c("River Discharge"="blue2"))+
 scale_y_continuous(limits=c(0.003,0.1))+
  theme_bw()+
 theme(panel.background = element_rect(fill = NA),
```

```
        panel.grid.major.x=element_blank(),
        panel.grid.minor.x=element_blank(),
        panel.grid.major.y=element_blank(),
        panel.grid.minor.y=element_blank(),
        axis.text.y=element_text(size=20),
        axis.title.y=element_text(size=20),
        axis.title.x=element_text(size=20))+
xlab("")+
ylab(expression(paste("River Discharge (",m^3, "/s)", sep="")))
g1 <- ggplot_gtable(ggplot_build(p1))
g2 <- ggplot_gtable(ggplot_build(p2))
pp <- c(subset(g1$layout, name == "panel", se = t:r))
g <- gtable_add_grob(g1,
                     g2$grobs[[which(g2$layout$name == "panel")]],
                     pp$t, pp$l, pp$b, pp$l)
alab <- g2$grobs[[which(g2$layout$name=="ylab-l")]]
ia <- which(g2$layout$name == "axis-l")
ga <- g2$grobs[[ia]]
ax <- ga$children[[2]]
ax$widths <- rev(ax$widths)
ax$grobs <- rev(ax$grobs)
ax$grobs[[1]]$x <- ax$grobs[[1]]$x - unit(1, "npc") +
unit(0.15, "cm")
g <- gtable_add_cols(g, g2$widths[g2$layout[ia, ]$l],
```

```
                length(g$widths) - 1 )
g <- gtable_add_cols(g, g2$widths[g2$layout[ia, ]$l],
                length(g$widths) - 1 )
g <- gtable_add_grob(g, ax, pp$t, length(g$widths) - 2, pp$b)
g <- gtable_add_grob(g, alab, pp$t, length(g$widths) - 1, pp$b)
grid.draw(g)
```

6.3.3 Turbidity correlation graphs

```
##################
#Programmes needed
require(ggplot2)
#require(gridExtra)
#require(scales)
require(corrplot)
###################
###################
Merged<-merge(datau, datac, by=intersect("dt","dt"), all=T)
Merged<-Merged[,c(1,6:9)]
Merged<-na.omit(Merged)
nrow(Merged)
colnames(Merged)[2]<-"turbidu"
Merged<-na.omit(Merged)
Final<-merge(Merged, datad, by=intersect("dt","dt"), all=T)
Final<-Final[,c(1:5,10)]
colnames(Final)[6]<-"turbidd"
```

```
Final<-na.omit(Final)
ready<-subset(Final, comment %in% keep)
ready$tdiff<-ready$turbidd - ready$turbidu
ready$length<-format(ready$length, format="%H:%M:%S")
ready$length<-sapply(strsplit(ready$length,":"),
      function(x) {
        x <- as.numeric(x)
        x[1]+x[2]/60
      }
)
theme_next <-  theme(
          panel.background = element_rect(fill = 'white', colour = 'black'),
          plot.title = element_text(size = 22),
          axis.text = element_text(size = 16, colour = "black"),
          axis.title = element_text(size = 18),
          legend.background = element_rect(colour = "black"),
          panel.grid.major = element_line(colour = "grey"),
          panel.grid.minor = element_line(colour = "grey", linetype = "dotted"))
graph<-ready
pear<- cor(graph$, y= graph$tdiff, method="pearson")
spear <- cor(graph$length, y= graph$tdiff, method="spearman")
line<-coef(lm(tdiff ~ length, data = graph))
label1 <- paste("Pearson = ", round(pear,4))
label2 <- paste("Spearman =", round(spear,4))
```

```
ggplot(graph, aes(length,tdiff))+
geom_point(aes(size=graph$cows), fill="#56B4E9", color="darkred", shape=22)+
geom_abline(intercept=line[c(1)] ,  slope=line[c(2)] , col="red" )+
annotate("text", x = 0.5, y = 70, label = label1,  size = 6)+
annotate("text", x = 0.5, y = 50, label = label2, size = 6)+
xlab("Length of Event (min)")+ #x axis label
ylab("Difference in Upstream and Downstream Turbidity (NTU)") +
theme_next
```