Analysis of Fluvial Dissolved Organic Carbon using High Resolution UV-Visible Spectroscopy and Raman Spectroscopy

Martin Coleman

MEng Electronic and Electrical Engineering

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

School of Geographical and Earth Sciences

College of Science and Engineering

University of Glasgow

August 2017

Abstract

This dissertation focusses on some advancements in methodology for measuring and analysing dissolved organic carbon (DOC): analysing data from a high resolution sensor generating DOC concentrations, [DOC] and secondly the use of Raman spectroscopy to analyse the composition of DOC. Recent advances in sensor technology have enabled the collection of DOC data with greater frequency over extended time periods than was previously possible through manually collecting water samples. In this research a time series of 30 minute [DOC] data for 2.5 years from Drumtee water, a peaty catchment in Scotland, was generated and analysed using a Spectro::lyser^{TM} from S::Can^{TM}, with a customised algorithm for calculating [DOC]. The time series revealed details of events and strong seasonal variation in the [DOC], with a range of 8.0 mg/l to 55.7 mg/l. During the same time period measurements made using manual sampling of river water were very similar, ranging from 10.2 mg/l to 81.1 mg/l (with the second largest value at 64.1 mg/l).

Similar DOC export budgets were calculated from Spectro::lyser^{TM} measurements and from the laboratory-analysed samples for both the hydrological year 2012/13 (HY 2012/13) and hydrological year 2013/14 (HY 2013/14). For the HY 2012/13 year the DOC budgets using the field collected data and the laboratory collected data were 16.6 $gCm^2.yr^{-1}$ and 19.8 $gCm^2.yr^{-1}$ respectively. For the HY 2013/14 year the DOC budgets using the field collected data and the laboratory collected data were 18.1 $gCm^2.yr^{-1}$ and 19.5 $gCm^2.yr^{-1}$ respectively. The similarity between the budgets calculated using the high-resolution [DOC] sensor and the budget calculated using laboratory measured [DOC] samples indicated that seasonal variation had a greater influence on export budgets than short term events had. GAMs were used to model the high resolution [DOC] data, and the model generated an R^2 value of 0.75 and a p-value of $< 2.2 \times 10^{-16}$. It was also identified statistically that there were regular [DOC] dilutions during events and that these dilutions tended to coincide with the time period when discharge was increasing most rapidly.

To identify relationships and periodicities in the high resolution [DOC] time series that would otherwise be challenging to identify three forms of wavelet analysis were used. These were continuous wavelet transforms (CWTs), maximal overlap discrete wavelet transforms (MODWTs) and wavelet coherence transforms (WTCs). Using the WTCs, it was determined that there were short term correlations between the [DOC] and pH between 25 June 2013 and 17 July 2013, between [DOC] and SC during 7 August 2013 and 7 October 2013 and between [DOC] and water temperature during 19 June 2013 and 30 June 2013. Although the although the relationship between [DOC] and temperature is

strong over a full year it was over these shorter time periods the weakest of the three relationships established. Identifying this coherence was not possible using bivariate analysis and the long periods of no coherence obscured these responses when analysing the data on scatter plots. Although wavelet analysis has been used in other applications this is one of the first instances in which this technique has been applied to [DOC] time series.

Raman spectroscopy, conducted using a 785 nm laser, was explored as an analytical tool that could enable a better understanding of DOC composition, as an alternative to the use of fluorescence spectroscopy. Tests were conducted using both Stokes and anti-Stokes Raman spectroscopy measurements with the best results obtained using anti-Stokes Raman spectroscopy. Solid phase measurements were made of glucose, fructose, sucrose, glycine, tyrosine, tryptophan and phenylalanine, but only the glucose produced a measurable spectrum of these substances. Measurements (powders and solutions) were made of humic and fulvic acids and these produced spectra that were measurably different from the background signals. The limit of detection was measured to be approximately 500 mg/l for both the humic acid and fulvic acid. It was identified that comparing the sections of the measured spectra between wavenumbers -1100 cm^{-1} to -1400 cm^{-1} to -1800 cm^{-1} to -2000 cm^{-1} could be used to differentiate between humic and fulvic acids.

In summary, this research has focussed on the use of use high resolution sensor technology to generate and then analyse a long time series in a fluvial system with a particularly high [DOC], and made advances in being able to model the [DOC] using a GAM model, despite the complex relationship measured between discharge and [DOC]. Additionally, wavelet analysis has been applied to a [DOC] data set to identify trends in the [DOC] time series that would otherwise be hard to identify. Wavelet analysis has been applied to other geophysical time series such as those found in atmospheric research, but this appears to be the first time it has been applied to [DOC]. Additionally, the use of the anti-Stokes region of the Raman spectra has allowed identification of humic and fulvic acids, and established a limit of detection. Furthermore, an absorbance ratio was identified that can be used to determine whether a solution of humic substances is dominated primarily by humic acid or fulvic acid. This research appears to be the first study to explore this.

Table of Contents

Table of Figures

Table of tables

Acknowledgements

Over the last five years I have spent much of my time working on the research that is presented in this thesis. However, this has not been an individual endeavour and there have been many significant roles of those involved.

Firstly, I must thank my funders for providing me with financial support and the University for allowing me to complete the thesis part-time, when I had to start working after my grant came to an end. Both the Natural Environmental Research Council (NERC) and the Analytical Chemistry Trust Fund have funded this research and have supported me for years. Without this backing the project would not have been possible.

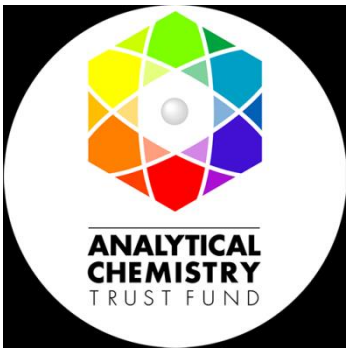
I am also immensely grateful to various individuals who have had a massive role in helping me with this project especially since my previous academic background, which was in electronic and electrical engineering. I am extremely thankful to Prof Susan Waldron for lending me her expertise and experience in working with fluvial systems in both an academic and fieldwork environment and also for countless hours working with me and enabling me to develop knowledge and skills required to complete the research I carried out. Thank you to Prof Marian Scott OBE who introduced me to many of the statistical tools that I used in this project and helped me to expand my knowledge in this area.

I would also like to thank Kenny Roberts for his frequent help on a regular basis for five years for both helping set up experiments in the laboratory and assistance on fieldwork. I must also thank Prof Jon Cooper and Peter Chung for introducing me to Raman spectroscopy and directing me in being able to carry out the experimental work discussed in this thesis. Thank you to Prof Hugh Flowers and Michael Beglan for helping me set up and conduct experimental work in the chemistry department.

I also wish to thank Dr Alona Armstrong, Dr Simon Drew and Dr Leena Vihermaa for their help and advice, at various points during my PhD as sources of knowledge and for helping to train me to work outdoors in the field.

To many others, a massive thank you for the various roles at the university. You have made my time doing my PhD infinitely more enjoyable. Thank you to Roger Grau Andres, Amy Childs, Dr Hu Ding, Sofia Daouadji Marie Denis, Lidia Escudero, Reine Blair Espinoza, Hazel Long, Hemanth Pasumarthi, Antony Phin, Simon Quemin, Dr Andy Singleton, Dr Ben Smith, Leena Vihermaa, Ying Zheng and Zhiwei Zhou.

I could not complete my acknowledgements without offering the biggest thank you to my family and in particular my parents, who have provided me with support for 28 years and enabled me to complete the education I have.



Author's Declaration

I declare that the work outlined and described in this thesis has been carried out by myself unless otherwise acknowledged. This thesis is completely my own composition and has not, in whole or part, been submitted for any other degree at this or any other university.

Martin Coleman

August 2017

Thesis Citation: Coleman, M (2017) *Analysis of Fluvial Dissolved Organic Carbon using High Resolution UV-visible Spectroscopy and Raman Spectroscopy*. Ph.D. thesis, University of Glasgow, School of Geographical and Earth Sciences, Glasgow, U

1.0 Introduction

This PhD research focusses on finding improved methods of measuring the dissolved organic C pool in catchment drainage systems. This objective was met, firstly, by testing an existing market product, establishing its accuracy and then determining what can be learned from this new technology. The second focus of the project was to test the effectiveness of Raman spectroscopy as a tool for analysing the composition and potentially the concentration of the dissolved organic carbon (DOC).

Measuring concentration is important because of the need to understand the drivers behind the transport of organic C from a terrestrial landscape into a fluvial environment and ultimately being processed through the carbon cycle. Being able to determine the composition of the exported C is also important because different components of organic C will break down at different rates, which can vary from hours to millennia (Buffle, 1990, Moran and Hodson, 1990, Mopper et al., 1991, Carlson and Ducklow, 1996) and therefore the time it takes to incorporate river C into other pools in the C cycle, such a conversion to CO_2 and efflux to the atmosphere, will vary. Therefore, in this project I have generated understanding pertaining to both the concentration and composition of dissolved organic carbon.

1.1 – Capturing the detail of C export in drainage systems

The central concept of the carbon cycle is that carbon is recycled through the Earth's different natural environmental systems (Cole et al., 2007). This means that the same molecules are used in multiple different processes again and again. As a simple example the same carbon atom can be part of a CO_2 molecule in the atmosphere that is taken in by a plant during photosynthesis and converted into a sugar; which may then be consumed by an animal and then during respiration converted back into CO_2 and emitted into the atmosphere. The different components of the carbon cycle, sometimes referred to as pools or reservoirs include the atmosphere, hydrosphere and biosphere, and carbon can be transferred between these different reservoirs (Detwiler and Hall, 1986). It is an important concept in fields such as climate change, ecology, oceanography and geochemistry (Schimel, 1995). In recent years concern has been raised over the contribution of terrestrial carbon to atmospheric greenhouse gases such as CO_2 and CH_4 (Gorham, 1991, Limpens et al., 2008, Billett et al., 2010), especially as soil organic carbon contains more carbon than is present in the atmosphere (Davidson and Janssens, 2006). There is also the possibility that increased CO_2 concentrations in the atmosphere can accelerate the rate of peatland degradation (Freeman et al., 2004, Worrall et al., 2004a), indicating that it is important to also consider feedback loops in environmental systems.

Sensor technology has been adopted in greater quantities in fluvial systems in recent years with sensors designed to measure water temperature, pH, conductivity, dissolved oxygen and now dissolved organic carbon concentrations, [DOC]. The deployment of such sensors in the field for extended periods of time has enabled the collection of large amounts of data and generated detailed data sets that have the potential to increase our knowledge and understanding of hydrological systems. The creation of high resolution time series (Kirchner et al., 2004, Jeong et al., 2012, Neal et al., 2012, Strohmeier et al., 2013) reveal just how much detail can be missed through intermittent sampling rates and how much exciting fine detail can be extracted from the sensor time series.

Using sensors to measure dissolved organic carbon (DOC) is an intrinsically more difficult task than measuring variables like pH or specific conductivity (SC). The first problem is that there is no common chemical formula for DOC so identifying a single test for measurement of this mixture is challenging. In recent years commercial solutions have emerged for measuring DOC concentration, [DOC], using UV-visible absorbance technology, including systems such as the S::can^{TM} Spectro::lyser^{TM}, which was used in this project. However, lagging behind this are sensors to consider the composition of the DOC pool.

The adoption of automated sensors to measure [DOC] is dependent upon acceptable reliability and accuracy. Reliability refers both to physical robustness and an ability to make accurate measurements under non-ideal circumstances. For example, can a field-deployed sensor accurately measure [DOC] when there is biofouling on the sensor or sediment washed downstream has accumulated on the detection point? Accuracy refers to how close the sensor estimated [DOC] value is to the true or standard value, and this is important in a field environment where samples are not pre-treated and the sensing conditions change rapidly and over longer timescales, such as annually. If there is a loss of accuracy in a sensor-generated estimate (compared to the laboratory measured composition), is this compensated for by the increased data resolution? Whilst deployment of sensors for [DOC] is becoming more commonplace, there is still a lack of data sets of length and from systems with higher concentration of DOC.

Hydrological systems can be very responsive to events (Clark et al., 2007a, Bass et al., 2011) and field-deployed sensor technology allows us to easily measure short timescale responses to hydrological variation (commonly termed event flow). Seasonal variation is also a driver of change, and therefore, to take full advantage of this technology we need to measure both seasonal long trends and short term responses. As a result, with high

frequency measurements, data sets of considerable size can be generated, which in itself presents challenges in how best to analyse the data.

One of the exploratory tools used to analyse the data in this project was wavelet analysis. Wavelet analysis is a statistical tool that can be used to analyse data in the frequency domain and extract data at multiple periodicities simultaneously. The ability of wavelets to analyse data at multiple periodicities simultaneously makes it ideal as an exploratory tool in natural environments where data sets are usually non-stationary, where properties such as mean, variance and auto-correlation change over time (Nason, 2006a). The wavelet analysis allowed for the exploratory analysis of the high resolution time series across a range of different time periods (periodicities) simultaneously.

The monitoring of [DOC] at high resolutions and statistical tools to analyse them over a range of time scales is important because it encapsulates hydrological responses occurring at a range of different time periods. For example, [DOC] can be influenced by: events (Clark et al., 2007a); a series of events, where [DOC] can decrease over successive events (Worrall et al., 2002, Jennings et al., 2009) and by seasonal variations such as the decomposition of organic matter (Kaiser et al., 2001), (Hongve, 1999).The type of high resolution analysis presented in this project will not necessarily reveal what is causing variations at a specific time scale but it can identify that they are occurring for other related research.

1.2 Determining the composition of DOC

DOC has been a widely studied source of natural carbon with particular interest emerging in recent decades due to its role in the carbon cycle (Mopper et al., 1991, Monteith et al., 2007). In particular the export of DOC from peatland soils has raised concerns related to both the direct and indirect contributions it is making to atmospheric CO_2 (Limpens et al., 2008, Billett et al., 2010) and the issue of water colouration (Wallage et al., 2006). This is especially pertinent because land use changes have resulted in increased exports of DOC in recent years (Evans and Monteith, 2001, Worrall et al., 2004b, Evans et al., 2006).

The composition of DOC is not uniform and can vary from site to site and also over time (Ågren et al., 2013). This is because DOC is not a specific substance but an umbrella term that incorporates a large number of different substances (Lutz et al., 2012). This includes sugars, amino acids and humic substances (which itself includes humus, humic acids and fulvic acids). Consideration of the composition of humic substances reveals further complexity: unlike many other acids, such as hydrochloric or sulphuric acid, there is not a single chemical formula for humic acid or fulvic acid (McKnight et al., 1992).

Humic and fulvic acids are some of the most abundant chemicals in peat soils but they are some of the least understood substances (Piccolo, 2002, Rodríguez et al., 2016). As the breakdown of different DOC molecules can vary widely depending on their composition (Sondergaard and Middelboe, 1995, Carlson and Ducklow, 1996), the carbon pool composition is worthy of consideration.

There are many options available for determining the composition of a particular DOC sample. The traditional methods have been the use of resins to chemically separate different components of DOC from one another (Aiken et al., 1979, Chow, 2006, Esteves et al., 2009) and the use of fluorescence spectroscopy (Senesi et al., 1991, Frimmel, 1998, Herzsprung et al., 2012). There are, however, some alternative methods. Using absorbance spectroscopy to compare different wavelengths has been used by some researchers to determine information about the relative proportion of humic acid to fulvic acid through the E4:E6 ratio, which compares the absorbance at 465 nm with the absorbance at 665 nm (Tadini et al., 2014) This approach originated in the 1960s (Kononova, 1966) and is still used today (Tadini et al., 2014, Jorge-Araújo et al., 2015, Zalba et al., 2016). Another method of determining DOC composition is to use NMR spectroscopy, which is debatably the best way to probe the structure of humic substances (Al-Faiyz, 2012, Rodríguez et al., 2016). However, if the long-term goal of this research is to develop a field-deployable sensor, then the NMR approach is unlikely to be the selected choice of technology as it is one of the more expensive spectroscopic techniques and it also has a relatively poor sensitivity (Grosch et al., 2013, Sundekilde et al., 2013, Kumar et al., 2015).

One potential method for investigating changes in DOC composition is Raman spectroscopy. Raman spectroscopy is an analytical tool that can be used to determine detailed features about a substances molecular structure (Vankeirsbilck et al., 2002a). For example, Raman spectroscopy has been shown to be a useful tool for studying graphene (another C compound) and identifying sample purity or the number of layers in a sample (Frank et al., 2011, Ferrari and Basko, 2013, Nanda et al., 2016). It is also, relatively to other types of spectrometer, simple to produce portable Raman spectrometers, with some becoming available on the market, through companies such as Rigaku corporation, SciAps Inc and B&W Tek.

While fluorescence spectroscopy is already widely used to analyse DOC (Coble et al., 1990, Coble, 1996, Chen et al., 2003, Johnson et al., 2011) there is justification for experimenting with Raman spectroscopy to analyse its composition. In some cases Raman spectroscopy has the potential to reveal more detailed structural information about particular substances than does fluorescence spectroscopy (Kneipp et al., 1999,

Vankeirsbilck et al., 2002a). Consequently, it is worthwhile investigating the effectiveness of Raman spectroscopy for analysing DOC. If the Raman spectroscopy were able to provide greater structural detail of the DOC than fluorescence spectroscopy without sacrificing sensitivity then this may indicate that Raman spectroscopy would be a good basis for a field deployable sensor.

1.3 Challenges producing a field deployable sensor

Before a field based sensor can be used for effective measurement of any environmental variable, whether that be hydrological or some other type of sensor it is necessary for other steps to have been taken first. The consideration below outlines the sequence of logic associated with this:

1) Confirm that measurements can be made with the intended technology in a controlled laboratory setting. Relevant to this particular project was whether Raman spectroscopy can actually measure DOC and could it also differentiate between different types of DOC? Laboratory environments make controlling for variables such as temperature, pressure, contaminants etc easier. The influence of these external influences can be minimised more easily as testing can be conducted in more ideal settings. It is not only important to establish whether the technique works but whether the technique can produce replicable results when experiments are repeated.
2) If measurements can be made, can limits of detection be established for DOC or its various components. Is this limit of detection below the concentrations that would be measured in the field? If not is there a method for concentrating the solution? If so is this realistic within a time frame – e.g., if the rate of measurement is every 30 minutes, then the concentration process would have to take place in less than 30 minutes.
3) Can the technology be miniaturised to the point where it is portable, deployable and able to be powered with batteries? If so, does this cause a loss in performance?
4) Making the miniaturised sensors durable enough to withstand the outdoor environment. For example, can it function under a wide range of temperatures? This refers not only to the accuracy and precision of the instruments, but also the durability of the casing, and making sure it does not get damaged by changes in temperature, or when it might be impacted with large force either by water, objects flowing in the water such as rocks, or withstand an impact from wildlife.
5) The next question is, are the measurements in the field actually reflecting what was tested in the laboratory? In many cases this can be achieved by collecting

samples (in this case water samples) and returning these to the laboratory to allow for the comparison between measurements made in the field and the laboratory. In doing this the accuracy and precision of the new sensor can be evaluated.

6) Finally, when the above points have been satisfied, then the sensor can be deployed to measure [DOC] with confidence.

The above points are required to produce a working field deployable sensor but getting from point one to point six can take a long time - potentially years. In this project the use of the Spectro::lyser™, a purchased sensor was tested against points 5 and 6. Other experimentation, that dealt with Raman spectroscopy, was concerned with points 1 and 2.

1.4 Research objectives and aims

This PhD research was initially conceived to develop a sensor for measuring DOC concentrations. However, just before starting, such sensor technology emerged more routinely on the market and so it was decided to use this existing technology and assess its 'fit-to-purpose' and instead explore if laboratory approaches to characterise DOC composition may offer a suitable framework for translation to a field-based sensor. Within these broad areas, the following specific objectives were identified:

1) to generate continuous time series of [DOC] and associated hydrochemistry that capture annual, seasonal and event-based variation at a field site.
2) to assess the quality of the field-based sensor [DOC] measurements by comparison with laboratory generated data.
3) to use the data to better understand the hydrological drivers that influence [DOC] and magnitude of the DOC exported.
4) to explore the advantages conferred by the sensor technology (or not) in developing this understanding and the challenges that need to be overcome in processing such large volumes of data.
5) to explore if other hydrochemistry measurements that may be measured more simply, can be used as proxy measurements to generate this understanding of [DOC] via statistical modelling. This would be achieved by comparing [DOC] measurements against [DOC] that is calculated using a statistical model, based on easier to measure hydrological variables.
6) to determine the suitability of Raman spectroscopy for studying the composition of DOC, which could offer a framework for an emerging sensor and explore its viability.

Additionally, it was important to ensure the data sets generated were of value in their own right and so the field-research was based in Drumtee catchment, a 9.4 km^2 catchment draining part of Europe's largest wind farm, Whitelee and monitored already since 2006, by both manual field-sampling and with the installation of some hydrochemistry sondes. This was also a research site for fellow PhD students on DOC degradation (Zheng, submission due 2017) and dissolved inorganic C cycling and efflux (Long, 2016), and so the understanding generated here contributes importantly to a larger scale approach to describe C cycling.

1.5 Thesis Structure

There are eight chapters in this thesis, including this chapter, and they focus in the following areas.

1) Chapter 2 reviews the literature focussing on the importance of the fluvial carbon cycle in peatlands and why the DOC pool is of particular interest. The evolution of measurement processes for quantifying DOC concentration and how sensor technology is changing the analysis of hydrology data, are explored. This chapter finishes by consideration of Raman spectroscopy as a framework for a DOC quality sensor, considering particularly where it has been applied in research cognate to the analysis of DOC that indicates this approach could be viable
2) Chapter 3 details the methods used in the field and laboratory, although some of these are further considered in later chapters where more relevant. In particular, this chapter discusses assessing the accuracy of the field-based sensor.
3) Chapter 4 presents and interprets the time series data generated by the almost continuous time series generated by the sensors at Drumtee focussing on inter-annual variation, event flow responses, and the effectiveness of simple and more complex statistical approaches to modelling [DOC]. This chapter also constructs export budgets and compares these to other sites.
4) Chapter 5 focussed on a statistical approach to identify signals within and between non-stationary time series that have multiple periodicities and to examine the series for coherence. The approach employed, wavelet analysis, is less common place in DOC analysis, and was chosen since it provides a powerful tool, to visually examine the modes of variation and thus support interpretation of complex time series.
5) Chapter 6 explores the use of Raman spectroscopy in the analysis of DOC and particularly humic substances, with key components of this chapter being the exploration of how best to use this approach to provide internationally reproducibility in spectral responses.

6) Chapter 7 explores use of anti-Stokes Raman spectroscopy as an alternative method to the techniques for the analysis of humic substances. There is very little prior research to inform these approaches and so this chapter is experimental in nature.
7) Chapter 8 concludes the research, considering progress against the five specific objectives and identifying research advances before concluding with exploration of future research directions.

2.0 Literature Review

This chapter provides background on DOC and how it can be measured with a view to setting up the later chapters dealing with advancing analytical techniques for measuring [DOC] and analysing the composition of DOC. The first section establishes what DOC is in terms of chemistry (including a description of humic and fulvic acids) and discusses how DOC fits into the carbon cycle. The focus of the chapter then moves to discussing existing methods for measuring DOC, a discussion about the benefits and potential drawbacks of high resolution sensors for collecting hydrology data, a discussion about why wavelet analysis was selected as a tool for analysing data and what this can achieve and finally there is a discussion about Raman spectroscopy. The discussion about Raman spectroscopy incorporates the physics of how the Raman phenomena occurs, discusses other applications for which it has been applied to and why it might be a useful tool for analysing DOC.

2.1 DOC introduction

Dissolved Organic Carbon (DOC) describes a mixture of organic carbon-based molecules that can pass through a membrane filter (Sondergaard and Middelboe, 1995, Evans et al., 2005). The components of DOC include sugars, amino acids, peptides, proteins, carboxylic acids, nucleic acids and humic substances (Thomas, 1997). The loose definition of DOC produces some potentially challenging issues from an analytical standpoint as there are potentially thousands or millions of organic carbon molecules that could fit this description. Some of the substances within DOC are easy to categorise and are understood very well – this includes sugars and amino acids, which have single easy to classify molecular structures but there are also more complex components and most of these complex DOC molecules are classified as humic substances (Moran and Hodson, 1990, De Lange et al., 2003).

Collectively DOC is an important component of the natural carbon cycle (Mopper et al., 1991, Anesio and Graneli, 2003). The carbon cycle describes the processes by which carbon is transported through the Earth's different systems including the atmosphere, biosphere, hydrosphere and lithosphere (Detwiler and Hall, 1986, Limpens et al., 2008). In recent decades, there has been an increased interest that the role that terrestrial carbon can contribute to the global carbon cycle (Gorham, 1991, Anesio and Graneli, 2003, Cole et al., 2007, Limpens et al., 2008).

Increased awareness and interest related to global warming is also contributing to an interest in better understanding the potential contributors to the carbon cycle, which may

be incorporated into feedback loops within the carbon cycle that may exacerbate global warming effects (Cox et al., 2000). DOC may also be an important regulator within fluvial systems through its influence on light penetration of the water column (Siebeck et al., 1994, Morris et al., 1995, Williamson et al., 1999a, Williamson et al., 1999b, De Lange et al., 2003).

DOC is also important because not only can it be mobilised from a terrestrial to an aquatic system but it can also influence the mobilization of metals, nutrients and contaminants within soils (Munch et al., 2002). Interestingly DOC can result in either decrease or increase in the mobilization of these other substances depending on the particular composition of DOC present (Totsche et al., 1997, McCarthy et al., 1998).

As DOC is a mixture of different molecules it follows that the chemical reactivity of these molecules will not be completely uniform (Thomas, 1997, De Lange et al., 2003). One of the chemical characteristics that varies between different DOC molecules is their susceptibility to chemical breakdown. Those that breakdown quickly are termed labile and those that break down over longer periods of time are termed non-labile or refectory (Sondergaard and Middelboe, 1995).

DOC is important in aquatic ecosystems because it is the primary substrate for bacterial growth (Williamson et al., 1999b, De Lange et al., 2003). Interesting it is not only the labile component of DOC that can be acted upon by bacteria and although non-labile DOC tends to be resistant to breakdown there is evidence to suggest that these materials can be acted upon by micro-organisms in the water (Moran and Hodson, 1990), particularly if the more resistant fractions such as humic substances are degraded by UV light (De Lange et al., 2003, Fasching and Battin, 2012). The presence of coloured DOC in the water column is however, not always a bad thing. In addition to be a substrate for bacterial growth it can also shield organisms living in the water (Allen, 1973, Sondergaard and Middelboe, 1995, Williamson et al., 1999a).

Although this project is primarily concerned with the methods of measurement of DOC rather than its functioning in terms of catchment dynamics, ecosystems, climate change or the carbon cycle as a whole it is important to consider the background because it is the motivation behind wanting to advance measurement techniques. Better understanding of the [DOC] such as variations during events (as presented here) can further understanding about how DOC exports from terrestrial landscapes. Similarly, better understanding about variations in the composition of DOC over time could reveal more details about catchment functioning.

2.1.1 Differentiating between dissolved and particulate organic C

The complex composition of organic carbon found in soils necessitates that an operational definition is required when attempting to classify it in a workable manner. One of the simplest and most common ways of doing this is to classify organic carbon into dissolved and particulate fractions. Therefore, DOC is operationally defined as an organic carbon molecule that can pass through a membrane filter (Sondergaard and Middelboe, 1995, Roulet and Moore, 2006), although there isn't a universally agreed upon filter size for this distinction. Typically filters vary in pore size from 0.2 µm (Sondergaard et al., 1995), 0.22 µm (Morris et al., 1995) 0.45 µm (Evans et al., 2005) and up to 0.7 µm (Fellman et al., 2008, Waldron et al., 2009). Anything within this range should be suitable and enable comparisons with other published studies.

It has been argued by some researchers that the use of filter sizes to distinguish DOC from POC was an arbitrary decision and that a filter size of 0.1 µm would produce a more accurate representation of true dissolved material (Chow et al., 2005). Chow discussed the characterisation of organic carbon in four categories rather than the usual two of DOC and POC. Chow categorised organic carbon as POC, DOC, colloidal organic carbon (COC) and fine colloidal organic carbon (FCOC). In his paper, he states that a filter paper with a pore size of 0.1 µm would remove the POC and COC and that a paper of 0.025 µm would remove the FCOC, leaving only the DOC component of the organic carbon. This was not the first published work to discuss the challenges of producing a true extract of DOC from a soil or water sample, with a paper from Sharp in 1973 discussing many of the same issues (Sharp, 1973). Chow's paper discussed this topic with respect to soil samples and Sharp's paper is dealing with seawater but the principle point of the discussion is the same – that the filter sizes used in most laboratory settings do not necessarily produce a true representation of DOC.

However, there isn't really a pragmatic alternative to using filter papers at present. In principle separating the hydrophilic and hydrophobic components might provide us with a more accurate distinction (Provenzano et al., 2006) but there is not really a workable method just now to allow us to do this due to the extremely high heterogeneity of DOC. The use of filter sizes of arbitrary size is not an ideal solution but it enables a categorisation of sorts in as much as DOC and POC will be segregated between larger humic substances and everything else. As discussed many of these larger molecules will be non-labile and will therefore be processed and broken down at a slower rate than the molecules classified as dissolved. Using this definition, it is estimated that the POC pool is ~20% or less of the DOC pool (Spyres et al., 2000).

The issue of filtration is important when considering field deployable sensors. Filtering the water can enable a simpler analysis of a substance such as DOC as some other substances (such as particulate material) can be removed and therefore it is not necessary to consider differentiating between different sources of carbon, some of which may not be DOC. However, filters can become quickly blocked and keeping a filter sufficiently clean to allow a water sample to be taken can be challenging.

2.1.2 Peatlands

In the context of this project the terrestrial landscape contribution of carbon from a temperate peatland system was used as an environment to test the effectiveness of the S::can™ Spectro::lyser™ as a sensor for measuring [DOC] at high resolutions. One of the reasons that peatlands are of interest to different researchers is because of their high carbon content (Gorham, 1991, Page et al., 2002). Here the importance of peatlands to wider research on the carbon cycle and catchment functioning is discussed. This particular project is focussed on advancing the methods used to measure DOC rather than advancing specific knowledge of the carbon cycle or the catchment functioning. However, this focus on advancing sensor technology could ultimately enable a greater understanding of catchment functioning by allowing the export of [DOC] to be better understood.

Globally peatlands are estimated to contain large stores of carbon although it is difficult to calculate a precise figure but a 2012 review of various literature resources concluded that an estimate for global carbon stocks was about 500 Gigatonnes of carbon (GtC) ± 100 GtC (Yu, 2012). However, within this study there is reference to various approaches by many different authors using different approaches. For example, one approach calculated the volume of peat and produced a range of 180 g C m^{-2} yr^{-1} (Gorham, 1990) and 455 g C m^{-2} yr^{-1} (Gorham, 1991) and a second approach that calculated based primarily on the density of the peat, which generated estimates between 137 g C m^{-2} yr^{-1} (Schlesinger, 1977) and 466 g C m^{-2} yr^{-1} (Adams and Faure, 1998). For instance sphagnum peatlands are estimated to contain one third of the world's soil organic carbon (Gorham, 1991, Fenner et al., 2007). In particular, changes to terrestrial landscape usage (such as construction, ditch cutting etc) have potentially resulted in a net increase of carbon from soils (Page et al., 2002, Fiedler et al., 2008, Armstrong et al., 2010). Additionally, due to feedback mechanisms increased CO_2 in our atmosphere may be contributing to the faster degradation of soil stocks of organic carbon (Tipping et al., 1999, Freeman et al., 2004, Monteith et al., 2007, Limpens et al., 2008). However, whether they are a net source or sink the contributions they could provide to the atmosphere need to be considered (Billett et al., 2004).

When a peatland is in good condition it can be a net sink of carbon (Turetsky et al., 2002, Dawson et al., 2004, Kayranli et al., 2010) – whereby they draw in more carbon from the natural environment than they release. Peatlands can draw down CO_2 from the atmosphere but they also emit CO_2 and CH_4, both of which are greenhouse gases (Gorham, 1991) but they can also export, which can ultimately be converted to greenhouse gases (Limpens et al., 2008, Fenner et al., 2011). However when the peatland becomes damaged either by natural or anthropogenic changes it can become a net source of carbon to the carbon cycle, which can have both short and long term consequences (Dawson et al., 2004). Peatland degradation and increased C exports can lead to increased water colouration, which is undesirable in drinking water and necessitates the increased application of water cleaning procedures, which could potentially result in an increased likelihood of producing disinfectant by-products (DBPs) which can be harmful for humans (Li et al., 1998, Kitis et al., 2001, Chow, 2006, Armstrong et al., 2010). Longer term effects might be a contribution to atmospheric climate change as processes such as microbial activity breakdown the influx of DOC into CO_2 which can be released and contribute to atmospheric CO_2 (Billett et al., 2004, Limpens et al., 2008).

In recent years there has been increased interest in Scotland on the potential impacts that the extensive construction of windfarms on peatlands are having in terms of impact on local wildlife such as birds (Bright et al., 2008, Pearce-Higgins et al., 2008). Additionally, there has been increased interest in analysing whether the disturbance to the peat caused by windfarm construction is resulting in increased C exports and long term damage (Grieve and Gilvear, 2008, Nayak et al., 2008, Waldron et al., 2009, Nayak et al., 2010, Murray, 2012, Smith et al., 2012, Smith, 2016).

Carbon exports have been monitored around Whitelee windfarm over recent years to determine what, if any, impact the windfarm development has had on carbon exports (Waldron et al., 2009, Murray, 2012). These studies do not appear to indicate a significant impact on the DOC exports as a result of the construction work at Whitelee, although this may not be reflected at other sites. Some potential impact may also have been mitigated by the developers Scottish Power carrying out peatland restoration works during the construction work (ScottishPower, 2008, ScottishPower, 2009).

2.1.3 Humification categorisation

One of the key components of DOC to consider, particularly with regards to compositional analysis, are humic substances. One of the key goals of this project is to evaluate how

effective Raman spectroscopy is for analysing DOC and identifying key components of DOC. In organic soils such as peatlands humic substances represent a very important source of carbon and it is therefore critical to consider the humic and fulvic acids.

Humic substances are a complex mixture of aromatic and aliphatic hydrocarbon structures with attached amide, carboxyl, ketone and other functional groups (Leenheer and Croué, 2003, Evans et al., 2005). They are one of the most complex components of DOC (Kuwatsuka et al., 1978) but also one of the most important as they can account for 70-80% of organic carbon in mineral soils (Piccolo, 2002); 60-70% of the total carbon found in soils and they also account for 40-60% of DOC found in natural waters (Mobed et al., 1996). A more recent estimate is that humic substances account for approximately half of the DOM within natural surface water runoff supplies (Beggs and Summers, 2011). They are found in peats, sediments, lignites, brown coal, sewage and other deposits (Grinhut et al., 2007). Unlike other components of organic carbon that breakdown more readily humic substances can reside in their current state for potentially thousands of years before they start to be processed through the carbon cycle (Grinhut et al., 2007). Research from sea water studies suggests that the DOC found in the deep sea (>500 m) is refectory, containing many substances such as humic materials, with an average age of about 6,240 years (Williams and Druffel, 1987), indicating that the turnaround for some of this material is very lengthy although there is also research to suggest that some fractions of this DOC can breakdown via photochemical degradation (Mopper et al., 1991).

Humic substances originate from degraded organic matter, particularly plants, although it also includes material from other dead organisms (Grinhut et al., 2007, Cao et al., 2016). Humic substances play a large role in the colouration of water and also are often highly fluorescent (Stevenson, 1994, Sutton and Sposito, 2005). The fluorescence property has led to many studies to measure and analyse humic substances using fluorescence spectroscopy (Coble et al., 1990, Mopper and Schultz, 1993, Chen et al., 2003, Herzsprung et al., 2012).

Humic substances are usually classified as either humic acids, fulvic acids and humins (Chen and Wang, 2007). However, humic acids, fulvic acids and humins are themselves umbrella terms that fit three broad categories. Humins are classified by being insoluble under any pH conditions, fulvic acids are soluble in all pH conditions and humic acids are generally insoluble at pH's lower than 2 (Sutton and Sposito, 2005). The proportion of humic substances present in a DOC sample will vary between different sources and methods of extraction (Mobed et al., 1996). This near limitless variation in the composition of humic substances contributes significantly to the challenges faced when attempting to analyse it.

There are a number of studies being produced that are starting to look at humic substances using nuclear magnetic resonance (NMR) spectroscopy, which is perhaps the most promising method of unravelling whether there is an identifiable structure within humic or fulvic acids (Francioso et al., 2001, Šmejkalová and Piccolo, 2007, Al-Faiyz, 2012). However, the use of NMR to study humic acid is time consuming and is at present not a tool for performing fast analyses with a high throughput.

2.2 Methods of DOC measurement and analysis

There are two approaches that can be applied to DOC measurements and analysis. The first approach that can be taken is a quantitative analysis, which can be used to determine the quantity of [DOC] present in a sample. The second approach is a qualitative analysis, which can be used to determine the composition of a particular DOC mixture.

Historically the quantitative approach has been more common due to the highly heterogeneous composition of DOC. Quantitative approaches are performed by performing a bulk measurement of carbon. Approaches that have been used over the years include wet chemical oxidation (WCO) (Ogawa and Ogura, 1992), high temperature combustion (HTC) (Jonathan H, 1973), high temperature catalytic oxidation (HTCO) (Sugimura and Suzuki, 1988, Spyres et al., 2000) and more recently UV absorbance spectroscopy (Müller et al., 2014).

Qualitative analysis has been less common but there are still examples of it over several decades despite the difficultly of analysing such a heterogeneous mixture of molecules. The two most common approaches up until the start of the 21st century were the use of resins to chemically fractionate different types of DOC (Chow, 2006) and fluorescence spectroscopy (Herzsprung et al., 2012) which was used to examine variations in the fluorescence response of different substances.

Within this project I have made use of different analyses to both measure the concentration of DOC and through the use of Raman spectroscopy some compositional analysis. The theory and application of both [DOC] measurements and compositional analyses are discussed in this sub-chapter.

2.2.1 Quantitative approaches

Quantitative approaches are the simpler and more common types of DOC analyses. The quantitative analysis was used to determine how much organic carbon was present in a water sample by both mass and concentration. The advantage of this type of analysis is

that it does not matter if the heterogeneity of the original sample is destroyed, provided that all the carbon is measured. To do this most approaches use oxidation to convert the various different carbon molecules into CO_2. The two most common methods for doing this have been wet chemical oxidation (WCO) and high temperature catalytic oxidation (HTCO). WCO was used as a technique to convert organic carbon from its various components into CO_2 using a chemical oxidising agent (Cauwet, 1994) at a moderate temperature (e.g. 98°C) (Osburn and St-Jean, 2007). The WCO method was the traditional older form of analysing DOC (Krogh and Keys, 1934, Williams, 1971, Strickland, 1972, Jorgensen, 1976) although the WCO has also been used more recently (Osburn and St-Jean, 2007). Experiments had been carried out during the 1960's and 1970's on developing a high temperature oxidation approach (Menzel and Vaccaro, 1964, Jonathan H, 1973) but there were problems making this approach work for salty seawater (Cauwet, 1994). However, in 1988 Sugimura and Suzuki published a paper that significantly influenced the analysis of DOC by describing an effective method of converting organic carbon to CO_2 using high temperature catalytic oxidation (Sugimura and Suzuki, 1988). For HTCO a furnace with a catalyst was employed to convert the mixture of carbon to CO_2 with a common temperature of about 680°C has been widely adopted (Cauwet, 1994). During the 1990s the HTCO method became the most common method of measuring [DOC] (Chen and Wangersky, 1993, Skoog et al., 1997, Wiebinga and de Baar, 1998, Spyres et al., 2000, Cherukuru et al., 2016).

Once all of the carbon has been converted into CO_2 the gas can be measured and the number of carbon atoms in the sample can be measured. The important thing regardless of whether WCO or HTCO is used is that all other forms of carbon are removed from the sample besides the DOC as otherwise the value of the [DOC] will be overstated (Spyres et al., 2000). In this sense WCO and HTCO are very similar techniques – they are converting a complex mixture into a single easy to quantify substance so that the total quantity of DOC can be calculated.

When HTCO was first introduced some papers were suggesting significant differences in the measurements of WCO and HTCO (Sugimura and Suzuki, 1988, Ogawa and Ogura, 1992, Chen and Wangersky, 1993). However, a 1993 paper by Suzuki identified that the 1988 paper with Sugimura may have contained some unreliable measurements and that although the technique was valid there was a lack of certainty with regards to some of the results presented in that original study (Suzuki, 1993).

In the 1990s there seems to have been an overhaul of the approach to monitoring DOC as the interest in monitoring DOC increased. From a 1993 report resulting from a workshop chaired by Jonathan Sharp it was stated that studies into DOM had received

"'Poor stepsister' treatment" in that there had been insufficient attention to detail with regards to producing adequate procedural methods (Sharp et al., 1993). Increased interest in the monitoring of DOC coincided with discussions about whether WCO or HTCO should be the preferred method for measuring DOC. It can be inferred that the introduction of the HTCO technique had led to a rethinking of how the entire set of procedures relating to DOM were being managed. Within this report a series of recommendations were described and this includes methods for the collection, processing and storing of samples and these are essentially the same procedural steps that are carried out to this date.

Papers by Skoog(Skoog et al., 1997) and Spyres (Spyres et al., 2000)stated that it was this overhauling of the procedural approach and the advancing of the technological protocol that was the real benefit from the discussion surrounding the development of the HTCO technique (Skoog et al., 1997, Spyres et al., 2000). HTCO is probably the most common method of measuring DOC today and this approach was used in this project.

2.2.2 Compositional analysis

Being able to analyse the constituent parts of DOC has represented a greater challenge largely because there is not a specific chemical or chemical signature that can be targeted for analysis in the same way that CO_2 is targeted for analysis by experiments that rely on oxidising organic carbon either by chemical processes or through the application of heat via combustion. It was noted in 2.1 that the heterogeneity of DOC (Ågren et al., 2013) makes qualitative approaches difficult because the composition is variable. The two most common methods for conducting quantitative analysis regarding DOC are the use of resins and fluorescence spectroscopy.

Resins used for segregating components of DOC are porous materials produced with high surface areas that have high adsorption properties that enabled substances such as organic carbon to adsorb to its surface at different sections depending on the absorbance properties of the particular organic carbon molecules (Aiken et al., 1992). These resins such as XAD-8 and later DAX-8 have been used to separate out components such as humic and fulvic acids (Aiken et al., 1979, Malcolm and Maccarthy, 1992, Tomasbarberan et al., 1992, Kalbitz et al., 2003, Chow, 2006).

One of the long-term goals of this line of research would be the development of a field based sensor that could be used to determine the composition of organic carbon at high frequencies. The use of resins would likely necessitate the collection of samples,

collected in the field and returned to the laboratory for analysis, which would reduce the sample rate throughput.

A different approach that has been used since at least the early 1990s is the use of fluorescence spectroscopy which enables the identification of several different components of organic carbon (Coble et al., 1990). One of the benefits of using fluorescence spectroscopy with regards to DOC has been the ability to identify differences between humic acids and fulvic acids, which fluoresce at different wavelengths (Senesi et al., 1991, Matthews et al., 1996, Chen et al., 2003, Baker and Spencer, 2004, Esteves et al., 2009).

Fluorescence spectroscopy can also be used to differentiate between different amino acids – for example tryptophan ($C_{11}H_{12}N_2O_2$) and tyrosine ($C_9H_{11}NO_3$) as both fluoresce in similar regions but there is nevertheless some overlap, making this challenging. For example tryptophan fluoresces when the excitation wavelength is between 295 nm and 305 nm but it can be preferable to use 295 nm because this avoids the excitation of tyrosine (Lakowicz, 2013).

Fluorescence describes the physical process within molecules of absorbing light at a particular wavelength and emitting at a wavelength at a longer wavelength than was absorbed at – this process is called a Stokes shift (Lakowicz, 2013). The Stokes shifts are different for different substances and this can enable the identification of particular materials via there spectral signature (Mopper and Schultz, 1993). Spectral signatures are very useful because they enable the identification of particular substances through the analysis of its light. Fluorescence spectroscopy also benefits from generating relatively powerful spectral peaks, which is why it is used over Raman in some instances (Farquharson et al., 1999, McCreery, 2005).

Modern fluorescence analysis of DOC utilises excitation emission matrices (EEMs) which are produced by scanning across the full range of excitation (absorption) and emission ranges simultaneously (Matthews et al., 1996, Baker, 2001). EEMs are a powerful method for determining different components of DOC by producing a surface plot (usually rendered in either 3D or colour coded) to show the excitations and emissions on an area plot where certain regions have been linked to different substances such as amino acids, humic acids and fulvic acids (Baker, 2001, Zepp et al., 2004, Carstea et al., 2010, Herzsprung et al., 2012).

The disadvantage of fluorescence and the motivation behind trying to determine if Raman spectroscopy could be a useful instrument for analysing DOC is that it is in theory possible

to obtain more detailed structural information about a material using Raman spectroscopy than it is with fluorescence spectroscopy (Kneipp et al., 1999, Vankeirsbilck et al., 2002b)

2.3 Benefits of High-Resolution Sensors

A goal of this research was to test the robustness of a high resolution sensor that can be used for measuring [DOC]. Measuring anything at a higher temporal frequency will generate more data, which can enable a greater insight into the process or processes controlling a particular variable. There are many ways in which larger datasets can be beneficial to a research project such as this one although the volume of data can present a couple of challenges as discussed below.

There are several papers that discuss the benefits and drawbacks of in-situ measurements (Taillefert et al., 2000, Walker et al., 2004, Liu et al., 2006, Winkler et al., 2008). Taillefert et al (Taillefert et al., 2000) discusses the advantages and disadvantages of electromagnetic instruments in aquatic systems in general terms. Some potential benefits include not requiring to collect samples and high data acquisition in a relatively short period of time, whereas some of the drawbacks are that the limits of detection are often poorer and that measured responses can be more open to interpretation (Taillefert et al., 2000). Some studies view in-situ measurements as critical to the research being carried out such as for measuring soil moisture (Walker et al., 2004). There is published data showing that tests on an in-situ spectrometer showed greater uncertainties were associated with the in-situ measurements when compared to the laboratory measurements (Winkler et al., 2008). Although in general laboratory based sensors are more accurate than field based in-situ sensors, improvements in technology (e.g. pH sensors) mean that the accuracy of field based sensors is improving and in doing so reducing the gap between the accuracy of field based sensors and laboratory based sensors (Liu et al., 2006). This means that what might be considered not suitable variables for field measurements just now based on current technology might be in a few years' time.

Traditional approaches to measuring [DOC] have largely been restricted to the laboratory. A sample would be obtained from the field and returned to the laboratory for analysis. However, this approach is limited by the sampling frequency where rapid changes in DOC are unlikely to be recorded. Intensive fieldwork by an individual at more regular intervals is possible (Smith, 2016) but is physically demanding and not a viable long term solution.

In recent years sensors measuring different hydrological variables have come onto the market and have been used by different research groups (Kirchner et al., 2004, Neal et

al., 2012). For example, the In-Situ Inc TROLL® was used in this project to measure water temperature, pressure, pH and SC. Other research has shown that sampling this type of variable at a frequency of weekly or monthly is insufficient to capture the full picture of what is happening in the time series. Kirchner has compared low frequency analysis of water hydrochemistry to trying to listen to a symphony where only a small fraction of the notes are actually played (Kirchner et al., 2004).

Traditional [DOC] measurements tend to rely on sampling rates such as weekly (Eckhardt and Moore, 1990, Frank et al., 2000) bi-weekly (Schiff et al., 1990, Hope et al., 1996) or monthly (Worrall et al., 2004b, Evans et al., 2005). The use of automated samplers to collect event data do exist and have been used to collect hourly samples of water (Clark et al., 2007b) although these are limited to the number of samples that can be stored in the analyser. While sensors such as the Spectro::lyser™ are also limited by memory limits the amount of data storable on that exceeds the number of samples which could be held in an auto-sampler.

The development of field deployable UV-vis spectrophotometers and an effective method for using this spectral data to calculate [DOC] provides an opportunity to collect [DOC] data in the same manner that other hydrochemistry data can be collected. The two main products that have been deployed in scientific studies in recent years are the Spectro::lyser™ from S::can™ and the TriOS ProPS-Kits. There have been several published Earth sciences studies where data was collected using Spectro::lysers™ – these include studies of the Amazonian rainforest (Waterloo et al., 2006); analysing drinking water (Appels et al., 2007); storm event responses (Bass et al., 2011, Jeong et al., 2012, Jones et al., 2014); peatland research (Koehler et al., 2009, Grayson and Holden, 2012) and the analysis of [DOC] in lakes (Müller et al., 2014). Analysis conducted using Spectro::lyser™ data has been used for waste water quality studies (Langergraber et al., 2003, Rieger et al., 2004). Studies that have used data collected using the Trios sensor are less numerous but do exist for Earth sciences studies (Sandford et al., 2010, Rinke et al., 2013).

The Spectro::lyser™ comes inbuilt with its own proprietary algorithm. However, other methods of interpreting the absorbance data have been tested. A promising method of calculating [DOC] from absorbance data is to use two wavelengths of light that are weighted and summed together to produce a [DOC] value has been developed and tested by Ed Tipping's group at the Centre for Ecology and Hydrology (CEH) at Lancaster University (Tipping et al., 2009). Tipping indicates that so long as the two wavelengths of light are within a reasonable boundary of suggested values then the wavelengths selected are not set in stone. In this 2009 paper Tipping used a 254 nm and a 340 nm wavelength

of light. As will be discussed in greater detail in the methods the Spectro::lyser™ did not have a 254 nm wavelength and so the nearest value of 255 nm was used instead, in conjunction with the 340 nm wavelength of light.

At present, there is no study that directly compares the data collected from a Spectro::lyser™ and a TriOS ProPS-Kit. There are two reasons for this. The studies conducted using these sensors are all compared against laboratory data. The second reason is cost – at present UV-vis absorbance spectrometers designed to measure [DOC] are expensive and so it is often beyond the budget of most studies to be able to compare the two sensors in a cost-effective manner.

Another benefit of using field based sensors is that natural processes can sometimes be taking place in a field that either don't occur when samples are returned to a laboratory setting or those processes are taking place at a different rate (Herrero-Hernández et al., 2015). Therefore, differences in the field and laboratory measurements do not necessarily always mean that the laboratory experiments are more accurate. However, in this particular study the DOC largely consists of materials that are resistant to degradation and breakdown (Thomas, 1997) so it is likely that the experimental results would be more accurate in the laboratory in this instance, although some DOC molecules that are usually resistant to breakdown can degrade when exposed to sunlight (Moran and Zepp, 1997).

2.4 Operational and modelling challenges with high resolution environmental data

While high resolution datasets can provide significant insights into understanding environmental systems they also present a number of challenges. In addition to the issues discussed on page 19 and 20, as discussed in papers by (Taillefert et al., 2000, Walker et al., 2004, Liu et al., 2006, Winkler et al., 2008), there are some issues relating to field collected high resolution data that may be particularly pertinent to this project.

The first potential issue is that by working with sensors directly in the field the weather and environmental influence can result in greater errors (Hart and Martinez, 2006, Wark et al., 2008). In the case of hydrology projects influences include sensors being buried in mud or sediment, materials which can also accumulate on the surface of the sensors, distorting measured results. Bio-fouling growth also presents a common source of error in hydrology systems as this can grow on the surface of sensors or around the casing for sensors and distort results (Lehaitre et al., 2008, Tercier-Waeber et al., 2009). Metals found in streams such as manganese or iron, which have been identified as materials found in peatlands – particularly those that have experienced soil disturbance (Muller and

Tankéré-Muller, 2012), can also represent a potential source of error. To combat these issues it was necessary keep the Spectro::lyser™ as clean as possible.

The inability of regular calibration has also been identified a major issue to the deployment of environmental sensors (Barrenetxea et al., 2008). In a laboratory setting analytical equipment can be calibrated as regularly as is deemed necessary by the researcher. In a field environment calibration cannot generally be carried out as often as sensors are often left unattended for long periods unattended.

Although the volume of data obtainable from high resolution sensors represents an opportunity to improve understand the environmental system being studied, the volume of data collected by high resolution sensors can present a major analytical challenge by making the analysis more complex compared to the analysis that would be carried out on a smaller dataset. If appropriate methods are not identified for analysing the data then there is the risk that the volume of information could become overwhelming and the database becomes a "data dump" (Keim, 2002, Keim et al., 2006, Muller et al., 2012). As opposed to fragmentary signals a full dataset can be generated from automated sensors and this can require more complex statistical tools as time series may now contain variation at a range of different frequencies (Neal et al., 2012) especially when the data is being collected faster than it is analysed. The data in this type of project does not fall into a technical category of big data which would contain datasets of 10^{18} or bigger (Kaisler et al., 2013) but these datasets are significantly larger than those produced by manual sampling and appropriate statistical techniques need to be identified and applied. For this reason, scripts in R were written to aid in the analysis of the data.

2.4.1 Wavelet analysis

Frequency domain analysis can provide a complementary analysis to time domain analysis (Rua, 2012) and can determine how much of a signal lies within a given frequency range by allowing different frequencies to be identified and sampled independently (Anemüller et al., 2003). Two common and powerful approaches that were considered for use in this project were Fourier analysis and wavelet analysis. Both techniques can enable the identification and extraction of different frequency components within a complex signal (Rua, 2012). However as has been noted Fourier analysis is not as adaptable or suitable for analysing non-stationary time series (Merry and Steinbuch, 2005). A stationary time series refers to a time series in which properties such as mean, variance and auto-correlation do not change over time (Nason, 2006a). As the variables of a natural fluvial system are unlikely to be stationary, wavelet analysis was selected as a more appropriate tool over Fourier analysis. Indeed, most geophysical systems can be described as nonstationary (Stockwell et al., 1996).

Wavelet analysis is a powerful statistical tool that has been implemented in different fields such as economics (Gençay et al., 2001, Schleicher, 2002, Crowley, 2007, Rua, 2012); communications and signal processing (Rioul and Duhamel, 1992, Lakshmanan and Nikookar, 2006); electronics (Wernekinck et al., 1993) and soil analysis (Lark, 2006). There are a lack of studies in hydrology using wavelet analysis but with the increasing abundance of large hydrological data sets techniques such as wavelet analysis may provide a useful method of analysing and interpreting data.

There are several types of wavelet analysis but a general description of wavelet analysis is that it enables the various frequency components of a signal to be identified individually (Graps, 1995). For example, the temperature of an outdoor environment on most regions of the Earth's surface have both seasonal and diurnal patterns and these can be identified and analysed separately. However, unlike temperature, in many cases the frequency components of a signal will not be known when analysing a signal. One advantage of the wavelet approach is that they provide exact scale-based decomposition results (Serroukh et al., 2000).

One type of wavelet analysis is the maximal overlap discrete wavelet transform (MODWT) and this enables a visualisation of the decomposition of a time series with different frequency components extracted and presented individually (Percival and Walden, 2006). There is precedent for using MODWTs in geographical studies such as such as hydrological studies (Percival and Mofjeld, 1997, Kallache et al., 2005, Bogner and Kalas, 2008) as well as research conducted in other areas such as on atmospheric turbulence (Cornish et al., 2006).

Continuous wavelet transforms (CWTs) are another method of wavelet analysis that have a less coarse breakdown of different frequency components than do discrete techniques like MODWT (Percival and Walden, 2006). One example of CWTs used in a geosciences setting is their use in the analysis of the temporal variability of rainfall and runoff, which showed that in a catchment in eastern Australia that since the early 20th century the frequency of extreme events had been increasing (Nakken, 1999).

Wavelet coherence transforms (WTCs) are used for the comparison of two variables and identifying the strength of relationship that exists at a variety of temporal time scales between these two time series and has been used in the context of a number of geographical studies (Torrence and Compo, 1998, Grinsted et al., 2004). For example, the Grinstead paper used the example of analysing the Arctic oscillation and the Baltic sea ice extent record to demonstrate the usefulness of WTCs to geographical problems. WTCs are

similar to Fourier analysis but can analyse coherence as a function of time (Lachaux et al., 2002). Wavelet coherence phase analysis is used here to respectively identify the strength of relationship between different variables such as [DOC] and temperature at a range of temporal time scales.

When conducting wavelet analysis it has been noted that it is important to perform a Monte-Carlo analysis to assess the statistical significance against red noise background (Torrence and Compo, 1998, Grinsted et al., 2004). A Monte-Carlo analysis is achieved by comparing the regions of coherence with the regions of coherence with random spectra (RIVERA et al., Grinsted et al., 2004). The greater the number of times that this is done, then the more precise the regions of coherence are constrained.

One of the challenges of using wavelet analysis was discussed by Terrance and Compo in that they are all too often used for qualitative rather than quantitative purposes as they are used to produce “Colourful pictures” (Torrence and Compo, 1998). In part, this is brought about because it is not easy to translate the output of a wavelet analysis into a numerical output.

In this project I have used the wavelet functions CWT, MODWT and WTC as exploratory tools in attempting to extract data that may be correlated for short timescales. Different hydrological relationships can be present at different (potentially unknown) timescales, and wavelet analysis is a useful method for determining if transient patterns or relationships exist at an unknown periodicity without having to test at those different timescales individually.

2.4.2 Alternative approaches

Some other methods of analysing the data besides wavelets were considered for this project. Other methods considered included de-trending the data before analysis, hysteresis, analysing event spacing and dynamic harmonic regression.

When compared to de-trending the data, using hysteresis or event spacing wavelets have the advantage of being able to analyse a wide range of periodicities simultaneously. The other key benefit of wavelet analysis is that it is capable of identifying transient relationships – those relationships that exist only over part of the time series, such as a relationship that exists during the summer months but not during the winter months.

An alternative type of frequency domain analysis considered to wavelets was dynamic harmonic regression. In standard harmonic regression a periodicity is selected and this is used to build a regression model and generate a regression equation. In dynamic

harmonic regression the periodicity is varied dynamically, enabling an estimate of the periodicity. Therefore, it provides a very similar output to wavelet analysis. Dynamic harmonic regression has shown to be useful for tasks such as adaptive seasonal adjustment, signal extraction, forecasting and backcasting (Bujosa et al., 2007), solar irradiation forecasts (Trapero et al., 2015), and the diurnal oscillations of groundwater-surface water fluxes (Keery et al., 2007). However, in this project the primary function of the frequency domain analysis was as an exploratory analytical tool rather than as a forecasting tool. Rather than using frequency domain analysis to directly construct a forecast it was used to inform the relevant periodicities within the time series that might be relevant to consider when constructing a model.

One other type of analysis that was considered as an alternative to using wavelets but that was ultimately not used was to de-trend the time series and then analysing the remainder of the signal could be analysed. The simplest approach to detrending the data of certain known trends would be to model the known trend (e.g. seasonality) and then analysing the residual of the remaining signal (Hanson et al., 2004). However, in this instance, the main dataset to be analysed is [DOC] and although this approach could be used to remove the seasonal variation from the signal there are still several overlapping frequency components within the time series. Some of these may be overlapping one another making them more difficult to identify, especially when it is not known which frequencies exist within the time series.

Another approach that could have been used as exploratory analysis would have been hysteresis. Hysteresis has been used for processes exploration e.g., the reasons behind differences between laboratory and field measured hydraulic properties (Basile et al., 2003); the hydraulics as a consequence of slope failure (Ebel et al., 2010); analysing the influence of soil moisture and rainfall on discharge and suspended sediments (Seeger et al., 2004). While hysteresis consideration could have potentially proven a useful method of analysing events due to the abrupt shifts in measured hydrology, it would not necessarily provide a summary over long time periods. The main advantage of wavelet analysis is that it enables both short and long-frequency components to be analysed simultaneously.

The primary intention of using wavelet analysis was to identify variation at specific periodicities that would otherwise be difficult to observe in the time series measured, rather than necessarily being used as a method in and of itself for modelling the data. If desired thereafter, approaches such as hysteresis could subsequently be used to analyse the data after the periodicity components could be identified. Another reason for not using hysteresis in favour of wavelets was an analysis of the data in the time domain had already been conducted (presented in chapter four). The major motivation behind conducting a wavelet

analysis was to determine if information could be extracted in the frequency domain that might not be apparent in a time domain analysis.

In contrast wavelet analysis was deemed to be the most appropriate form of frequency domain analysis for this project because it allowed for a wide range of frequency components to be analysed simultaneously in non-stationary time series without having to manipulate the signal. The other significant advantage of wavelet analysis is that it allows for transient relationships to be identified.

2.4.3 Generalised Additive Models (GAMs)

Hydrological systems often produce complex responses to changes such as events or variations in seasonality. Therefore, it is often insufficient to model the response of a system on a single variable. It can therefore be advantageous to make use of a technique that can generate a predictive model using multiple variables. One of the tools that can be used for these purposes are generalised additive models (GAMs) ,which have been used for purposes such as the study of soil moisture (García-Baquero et al., 2016). GAMs are a form of multi-variate analysis that are capable of producing statistical models that are dependent on more than one variable.

GAMs are semi-parametric extensions of generalised linear models (Guisan et al., 2002), which are also capable of multi-variate analyses. GAMs use link functions to establish the relationships between the mean of the response variable and a smoothed function of the explanatory variables (Guisan et al., 2002). GAMs are very useful methods for identifying nonlinear covariate effects in data (Hastie and Tibshirani, 1990, Guisan et al., 2002). As some of the responses of the hydrological variables to events, seasonality etc, may not necessarily be linear GAMs have been selected as an appropriate tool for analysing and modelling [DOC].

2.5 Raman Spectroscopy

When considering how to advance methods of DOC measurements and analysis one method that seemed to merit consideration was Raman spectroscopy. Raman spectroscopy has been developed as part of portable systems and has shown capability about identifying structural information about different substances (Vankeirsbilck et al., 2002b). Both the capacity to miniaturise Raman and its known capability of being able to identify structural information about substances would seem to make it a potentially ideal for future sensor development. However, before it is used as the basis of a field deployable DOC sensor it is prudent to establish its effectiveness in the laboratory.

2.5.1 What is Raman and how does it work?

Raman spectroscopy is a physical phenomenon that is also sometimes known as the Smekal Raman effect owing to the two men to which its discovery can be primarily attributed in published works (Smekal, 1923, Raman and Krishnan, 1928). Raman spectroscopy is the result of inelastic scattering of light that is absorbed and then emitted (Raman and Krishnan, 1928, Collette and Williams, 2002). Electromagnetic radiation is absorbed by a molecule and is then emitted after a very short time interval, with the process being specific to the molecule in question. The Raman process of absorption and emission is superficially similar to fluorescence spectroscopy but whereas fluorescence is a photoluminescence process that depends on electron transitions between real energy levels, Raman is a molecular scattering process that depends on temporarily existing virtual states (Collette and Williams, 2002). The 2002 paper by Collette and Williams illustrates the difference between Raman and fluorescence (Figure 2.1). The time difference between Raman and fluorescence processes is that Raman absorption and emission processes can occur on a time scale of (<1 picosecond (ps)) whereas fluorescence spectra are typically measured in hundreds or thousands of ps (Everall et al., 2001).

One of the characteristics of Raman spectroscopy (and one of the potential downsides of using it) is that it is a rather weak phenomenon, especially when compared with fluorescence spectra (McCreery, 2005), which can be four to eight orders of magnitude greater and completely obscure the Raman signals (Farquharson et al., 1999). Weak signal strength is particularly problematic for substances that exhibit strong fluorescence signals because these can obscure the weaker Raman signals which they will often overlap with. One method of reducing the influence of fluorescence in Raman measurements is to select a longer wavelength laser, which produces less fluorescence (Viskari and Landers, 2006, Qin et al., 2010, Vítek et al., 2012).

The major potential advantage of Raman spectroscopy compared to fluorescence spectroscopy is that it can be used to collect far more structural information about the molecular structure (Kneipp et al., 1999, Vankeirsbilck et al., 2002a). One other advantage of Raman spectroscopy over fluorescence is that fluorescence can cause photo-bleaching, which may potentially damage target materials such as proteins (Han et al., 2009).

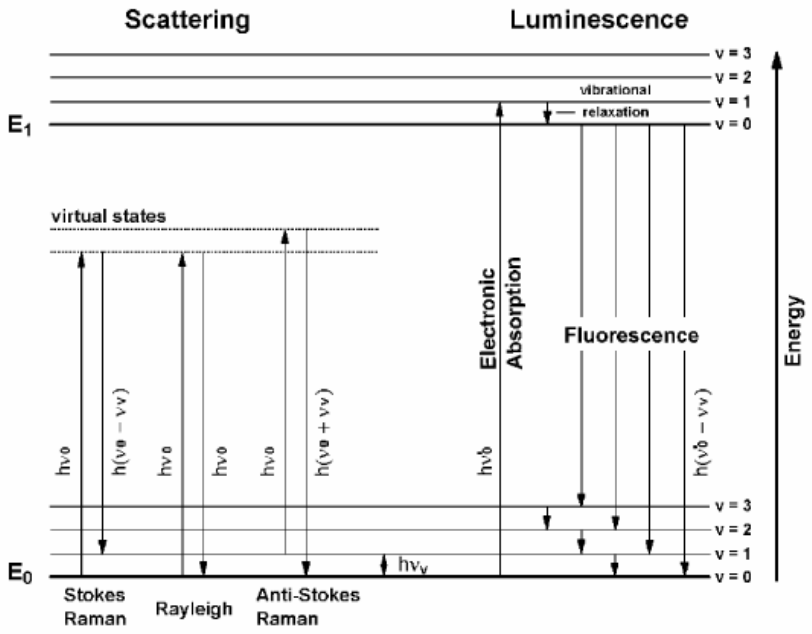


Fig. 1 Example energy level diagram depicting, on the left-hand side, molecular scattering processes (Rayleigh and Raman) and, on the right-hand side, a photoluminescence process (fluorescence) that is particularly problematic for practical applications of Raman spectroscopy. E_0 and E_1 are the ground and first excited electronic energy levels, respectively, for the example molecule. v = 0 and v = 1 are the ground and first excited vibrational levels, respectively, for a given electronic level, and v = 2 and v = 3 are higher vibrational energy levels.

Figure 2.1 – A copy of a diagram produced by Collette and Williams (2002)
The above diagram, copied from Collette and Williams (2002) shows the fundamental difference between the type of light scattering that produces Raman and fluorescence (Collette and Williams, 2002).

The versatility of Raman spectroscopy owing to the relative ease of integrating it into portable sensors and its capacity to identify structural information about those substances has led its adoption in many different fields. Some examples of this are as follows:

1) Astrobiology – Raman spectroscopy had been proposed as a method of identifying "key spectral markers" that can be used to identify both organic and inorganic compounds and could be used to study samples of different sizes (Villar and Edwards, 2006). Although no probes have yet been sent to Mars (or other planetary bodies) equipped with a Raman spectrometer there is a strong possibility of the European Space Agency (ESA) missions to Mars – the ExoMars flights being equipped with Raman laser spectrometer (RLS) (Hutchinson et al., 2014).

2) Drug detection – Customs officials require quick and reliable methods of identifying drugs with minimal interference from either the substance or the packaging (Garrido et al., 2007, Hargreaves et al., 2008). One benefit of Raman spectroscopy that makes it very appealing for portable sensors is that samples can

often be analysed with little to no preparation. Samples can even be analysed when stored inside of transparent material, making it potentially useful for analysing a substance without disturbing the contents of that substance.

3) Diabetes measurements – there are many potential medical applications for Raman spectroscopy but one of the proposed ideas is for the development of glucose sensors – which provides a potential opportunity for measuring blood glucose levels (Shafer-Peltier et al., 2003, Robertson and Harmon, 2006, Shao et al., 2012)

4) Graphene studies – One of the most interesting uses for Raman spectroscopy is to study the structure of graphene – in particular being able to identify a particular sample's purity (Frank et al., 2011, Ferrari and Basko, 2013). Raman spectroscopy can be used to determine properties such as the number of layers of graphene present, the quality and types of edges (Nanda et al., 2016).

One of the most important developments in the evolution of Raman spectroscopic analysis has been the discovery and development of surface enhanced Raman spectroscopy (SERS) which enables the enhancement of Raman signals (Fleischmann et al., 1974). SERS is achieved by having the target substance adsorbed onto a metallic surface (usually though not always copper, silver or gold) where that surface is rough at a nano-scale level (Campion and Kambhampati, 1998). Initial experiments were achieved through the use of roughened silver electrodes but later approaches used colloids and custom designed substrates.

This project seeks to advance the understanding of how to make measurements of DOC using Raman spectroscopy and how to interpret those results. From this a determination could be made on whether Raman spectroscopy is a suitable tool for analysing the composition of DOC. The approach to doing this was to review approaches published in the literature that had been done before and then to perform practical experimentation in the laboratory to verify those results and to, if possible, refine them.

2.5.2 Suitability of Raman Spectroscopy for DOC analysis

The compositional analysis of DOC has been possible for some time through the use of resins (Aiken et al., 1979, Leenheer, 1981, Esteves et al., 2009), fluorescence spectroscopy (Senesi et al., 1991, Matthews et al., 1996, Baker, 2001, Fellman et al., 2010) and nuclear magnetic resonance NMR spectroscopy (Rodríguez et al., 2016). In particular fluorescence spectroscopy has been used since the early 1990s (Coble et al., 1990, Mopper and Schultz, 1993) and is still a widely used analytical technique (Fellman

et al., 2010). Given fluorescence spectroscopy has several decades of research backing it up it would only really make sense to adopt Raman spectroscopy if presented some analytical benefit over these other methods. Here are presented the results from Raman experimentation designed to determine whether Raman spectroscopy is potentially a suitable method for differentiating between different components of DOC.

Resins such as XAD-8 are a useful way for chemically separating out different components of DOC. The use of resins is the technique used to produce humic acid and fulvic acid standards by the International Humic Substances Society (IHSS) (Bolton, 2003). Resins work by adsorbing the less polar fractions of DOC such as humic and fulvic acids onto the resin's surface, where these molecules are subsequently desorbed with NaOH and then the fulvic and humic acids are separated by precipitating the humic acid at a low pH (Bolton, 2003). However, there is some criticism that resins or other techniques based on isolation and fractionation are considered to potentially impose artificial distinctions between humic and fulvic acids (Hautala et al., 2000, Bolton, 2003). Furthermore, it has been noted that there is no strict chemical division between humic and non-humic substances but instead a natural continuous transition (Peuravuori and Pihlaja, 1998).

Fluorescence spectroscopy has been arguably the leading method for analysing the structure of DOC since the late 1980s. It has also been used to analyse diurnal changes to the composition of DOC (Spencer et al., 2007). The main benefits of fluorescence are that it has a high sensitivity when compared with some other methods such as NMR (Pickup et al., 2005) and can be used to analyse at natural concentrations (Patterson et al., 1992, Bolton, 2003). Another benefit of fluorescence spectroscopy is that it is non-invasive, which can speed up throughput and minimise preparation time (Kalbitz and Geyer, 2001, Bolton, 2003, Pickup et al., 2005). The development of excitation-emission matrices also greatly expanded the analytical capabilities of fluorescence spectroscopy for investigating DOC (Coble, 1996, Chen et al., 2003). The main drawback with fluorescence spectroscopy is that it is not as sensitive to molecular structures as NMR (Bolton, 2003) and Raman spectroscopy (Vankeirsbilck et al., 2002a). Therefore, one of the main motivations for investigating the effectiveness of Raman spectroscopy to analyse DOC is whether it can provide more detailed structural information about DOC than fluorescence spectroscopy.

The primary benefit of NMR is that it is highly-specific with regards to the structural information that it can reveal about substances such as humic substances (Simpson et al., 2001, Šmejkalová and Piccolo, 2007, Rodríguez et al., 2016). The drawbacks of the technique are that it requires more sample preparation than techniques such as

fluorescence (Bolton, 2003) and the technique has relatively poor sensitivity and also extremely high costs when compared with other methods (Chatham and Blackband, 2001).

There are indeed some papers that discuss the use of Raman spectroscopy for measuring humic substances (Leyton et al., 2005, Sanchez-Cortes et al., 2006, Corrado et al., 2008, Heighton, 2013). However, as I will discuss in chapter 6 there are questions over whether these experiments were truly identifying humic substances. One of the reasons for investigation the use of Raman spectroscopy for analysing DOC is that it is a technology that can be made portable with relative ease compared to NMR and has been shown for some materials to have greater capacity for identifying structural information about some materials (Kneipp et al., 1999, Vankeirsbilck et al., 2002a). The ease with which Raman spectrometers can be miniaturised into portable sensors and its ability to measure the molecular structural details about a variety of substances makes it a technique worth investigating with an eye for future sensor development. As there are already pre-existing methods for analysing the composition of DOC other than Raman spectroscopy it would only make sense to pursue using Raman spectroscopy to analyse DOC if it could be shown to have some benefit over other pre-existing methods e.g. capability to provide more detailed structural details about the DOC composition.

One of the uses for Raman spectroscopy that occurs in academic disciplines is where the a substance does not naturally fluoresce and therefore fluorescence would not be an appropriate technique (Liu et al., 1992). However, in the case of DOC many of the major components, such and humic and fulvic acids, do fluoresce strongly (Matthews et al., 1996). Therefore, one of the key goals of this project is to determine the effectiveness of analysing DOC and consider how it performs with respect to other techniques such as resins, fluorescence spectroscopy and NMR spectroscopy as reported in published studies.

2.5.3 Possible applications of Raman Spectroscopy for field deployment

As discussed in section 2.3 field deployable sensors have an advantage of being able to collect data at a significantly greater rate than can be achieved by manual sampling, leading to significantly denser datasets for significantly less invested time. Raman as discussed in section 2.5.1 has been developed into portable sensors. Also, as discussed in section 2.5.1, Raman spectroscopy is a useful method for identifying structural information about a particular substance. A further benefit of Raman spectroscopy is that minimum sample preparation is required prior to analysis.

Collectively, these three benefits of portability, ability to identify structural information and minimums sample preparation make Raman Spectroscopy an intriguing prospect for the basis of a field based sensor to measure DOC. If a consistently reliable method can be generated to measure DOC using Raman spectroscopy, then this could enable the development of a sensor to measure compositional changes to [DOC] at high resolutions. From this it could be determined if there are composition variations at short time scales such as the composition before and after events.

2.5.4 Measurement of humic and fulvic acid using Raman Spectroscopy

In determining how effective Raman spectroscopy is for analysing DOC, it is critical to assess its capability as a tool for measuring the most abundant forms of DOC. As discussed (section 2.1.1), humic substances account for about 70-80% of organic carbon found in mineral soils (Piccolo, 2002) and therefore they are a major area of interest when analysing the composition of DOC. Indeed, there have been several published studies analysing humic substances using Raman spectroscopy (Vogel et al., 1999, Leyton et al., 2005, Sanchez-Cortes et al., 2006, Corrado et al., 2008, Roldán et al., 2011, Heighton, 2013). All these studies made use of SERS and the most common method discussed of improving the Raman signals of the humic substances was to increase the pH of the substances being analysed (Vogel et al., 1999, Leyton et al., 2005, Corrado et al., 2008). The hypothesis constructed was that as humic acids become more alkaline, the molecules unravel and form a less globular structure, and thereby are more likely to adsorb onto the SERS substrate. The approach suggested by Heighton does involve increasing the pH, but also performs a reduction reaction on the humic acid, considered that this will enable greater Raman spectra to be obtained of the humic acid (Heighton, 2013). Both these ideas were put to the test in this project. One of the reasons for considering the use of SERS as opposed to standard Raman was that in addition to enhancing the Raman signal SERS substrates can supress fluorescence signals (Dulkeith et al., 2005, Whitney et al., 2005).

One of the areas of concern when reviewing these papers, which show Raman measurements of DOC, was that there was some inconsistency in the location and relative intensity of the Raman bands identified. This inconsistency is in contrast to some other substances analysed that seem to show very consistent spectral information by comparison such as the analysis of graphene (Frank et al., 2011, Ferrari and Basko, 2013). To be confident of assigning Raman bands to components of DOC it is necessary that a consistent measurement of those Raman bands can be achieved to enable predictive modelling.

One of the reasons that most Raman studies focus on Stokes-scattering is that the same spectral bands often appear mirrored in Stokes and anti-Stokes regions but the Stokes effect is stronger than the anti-Stokes effect at room temperature (Fantini et al., 2004). In most cases, it is therefore beneficial to analyse the spectral peaks generated from Stokes scattering. However, for this project anti-Stokes Raman spectroscopy may potentially have one advantage and that is the absence of fluorescence spectra interfering – as fluorescence is a Stokes process (Lakowicz, 2013). As discussed in section 2.2.1, the humic component of DOC, which is a major component of most DOC samples, is highly-fluorescent and it is possible that many or all of the Raman spectra of DOC in the Stokes region of the spectra will be obscured by stronger fluorescent signals (Porterfield and Campion, 1988). Anti-Stokes scattering raises the possibility of analysing DOC without the interfering influence of fluorescence spectra but there do not seem to be any studies that have taken this approach yet.

Studies on components of DOC other than humic substances have been conducted successfully. In particular there are several studies that focus on glucose detection using Raman spectroscopy, with much of the interest levels arising in this area as an accurate method for measuring glucose levels in the blood (Shafer-Peltier et al., 2003, Shao et al., 2012). The studies thus far however have been limited by a very-high limit of detection, with most studies focussing on concentrations measured > 100,000 mg/l. By contrast this project is focussed on water samples with DOC concentrations < 100 mg/l and where an individual substance such as glucose would only account for a small proportion of the overall total organic carbon.

One future solution to the problem of the poor limit of detection of Raman spectroscopy is to integrate SERS with metal organic frameworks (MOFs). MOFs are molecules that are both extremely porous and have extremely large surface areas relative to their size and mass – it is the large surface area that gives them the property of extremely high adsorbance (Furukawa et al., 2010). Research has already shown that this is a workable method for producing SERS spectra from molecules that would not normally adsorb onto the surface of a SERS substrate (Sugikawa et al., 2011, Kreno et al., 2014). In these studies, it was shown that glucose could adsorb onto a SERS substrate that was integrated as part of a MOF. The use of a SERS substrate integrated as part of a MOF enabled identification of glucose using SERS where it was not normally possible in low concentrations. In the future as this type of substrate becomes more widely available it may enable easier measurement of materials such as humic substances.

2.6 Summary

This chapter discusses the background research that has led to this project. The composition of DOC, why it is measured and how it is defined are discussed. The measurement of carbon is central to the research, but the goals of the project deal principally with how to measure DOC. The key summary points from the literature reviewed are:

1) DOC has been identified in the literature as a heterogeneous mixture of different components, which makes analysing those compounds and measuring the overall concentration challenging. Several methods of analysis have been attempted over several decades and recent advances in sensor technology have enabled greater ability to monitor DOC concentrations and composition.
2) Peatlands have been studied as they are a major source of terrestrial carbon and are important through the role that they play in the global carbon cycle. In this project, a catchment draining part of a peatland is used to test equipment and it is discussed that these are the systems that we want to gain better understandings of with a long-term view to potentially gain greater insight into short term catchment functioning.
3) Measurements made quantifying how much DOC is being exported from the terrestrial landscape is a major research goal to better understand the role that DOC is having in catchment functioning, ecosystems, the carbon cycle and climate change.
4) The composition of DOC is variable and complex, including components such as sugars, amino acids, humic acids and fulvic acids. Different methods, such as the use of resins and fluorescence spectroscopy, have been utilised over the years by different researchers for analysing the composition of DOC samples. In this project Raman spectroscopy is tested as a technique for potentially analysing the composition of a DOC sample.
5) Wavelet analysis is a useful tool for analysing multiple periodicities in time series simultaneously, and therefore is being used as a means of exploratory statistical analysis in hydrological sciences as can reveal trends previously identified.
6) Raman spectroscopy is discussed to provide background on what it is and how it has been useful in other scientific research applications. Other analytical techniques used to analyse and measure DOC are discussed including UV-visible spectroscopy (mostly in the form of the Spectro::lyser^{TM}), resins, fluorescence and NMR. It is noted that as Raman can sometimes provide more detailed structural information than fluorescence that it is worth studying to determine if it can be used as a technique for analysing DOC.

3.0 Study site and methods for fieldwork, laboratory, Raman and statistical analysis

To measure the [DOC] in natural water systems and to analyse carbon molecules using Raman spectroscopy it was necessary to perform experimental work in both a field and laboratory environment. Subsequently understanding the data required statistical analysis. This chapter outlines the field setup and laboratory methodology and discusses some of the statistical analyses used to analyse the data. Specifically, the chapter details:

1) the location of the fieldwork, why the field site is of interest and why it is a useful site for measuring [DOC] at high resolutions.
2) a description of each of the sensors used in this project and the protocols for deployment and data generation, including particularly the algorithm created for measurement of [DOC].
3) how the discharge time series was generated and the various relationships used to determine the discharge rate.
4) the laboratory experimental methods – in particular a description of filtration, acidification, storage and how the measurements of [DOC] and [POC] were carried out.
5) the methodology used for Raman measurements

Each of the results chapters (4-7) contains a small methods section with methods that were specific to that chapter or where a concept is best introduced just before it is discussed. The methods here are central to more than one chapter.

3.1 Study site – Drumtee water and Whitelee wind farm

The Spectro::lyser^{TM} sensor used to measure [DOC] needed to be deployed in a field environment as the focus of this project was on validating whether or not the measurements made by the sensor were accurate and if so what information could be learned from the sensor that might be difficult to obtain using conventional manual sampling. Owing to the cost of the sensor a single fieldsite was selected for experimentation, where the effectiveness of the sensor to accurately measure [DOC] could be evaluated. To enable an evaluation of the accuracy of the Spectro::lyser^{TM} comparisons were made with manually collected samples that would be returned to the laboratory and analysed using a high temperature catalytic oxidation approach.

The study site for this project was Drumtee Water (Figure 3.1), (55°41’16 N, 4°23’37 W) a 9.4 km^2 catchment draining part of Europe’s largest wind farm, Whitelee. One of the reasons for selecting the site at Drumtee was that it is part of an ongoing project to study

the impact of carbon resulting from Peatland disturbance as a result of the wind farm construction (Waldron et al., 2009, Murray, 2012). The site has also been used as part of a study to help better understand the controls on CO_2 efflux from fluvial systems (Long et al., 2015).

If the [DOC] sensor could be shown to be an accurate method of measuring [DOC] then a network of sensors could be set up to monitor the DOC being exported in greater detail than would be possible via manual sampling. Another reason for selecting Drumtee water as a field site was because of its relative proximity Glasgow University, allowing for regular field trips and additional trips for maintenance with the minimum of delay but far enough away from any towns or cities to significantly reduce the chance of vandalism or theft of equipment. The fieldsite's proximity to a relatively low density human population means that anthropogenic sources of C are minimal meaning that the majority of C contributions are likely to be from primarily naturogenic sources, from activity related to the wind farm or from activities related to the sheep and cattle farm.

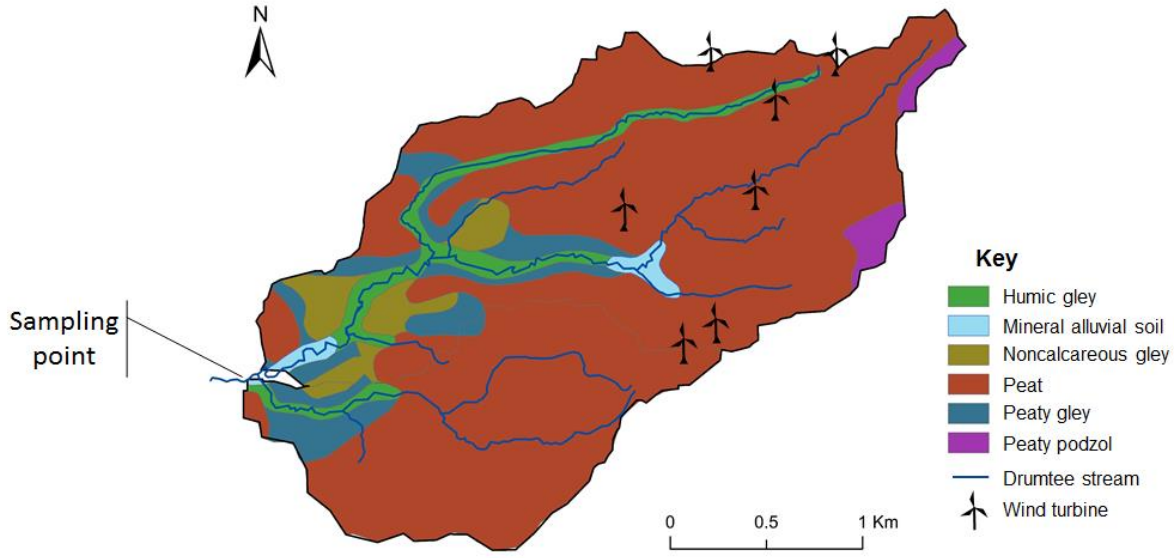


Figure 3.1 - Map of Drumtee catchment
Map of the Drumtee catchment with the colour-coded region representing the catchment and its boundary, including seven wind turbines within the catchment boundary. Soil is predominantly peat, although in the area near to the sampling point there is a greater proportion of gleys that further upstream. The map was produced from the GIS package ArcMap with the data obtained from the James Hutton institute (JamesHuttonInstitute, 2014).

As noted above Drumtee water is a catchment draining part of Whitelee windfarm. As a result of the planning consent required to undertake the wind farm construction the land area has been examined extensively with the Environmental Impact Statement (EIS) prepared by Scottish Power (ScottishPower, 2002) and has been further summarised

(Murray, 2012). Carboniferous porphyritic basalts of the Clyde plateau volcanic group of the calciferous sandstone series dominate the underlying geology. There are many small igneous deposits including small regions of basaltic tuffs and trachytes. Another significant geological feature, this time originating from the tertiary era is the north-west – south-east trending andesite and andesitic quartz-dolerite dykes. The bedrock is predominantly overlain with approximately 3 metres of glacial and recent drift deposits, with some outcrops throughout the fieldsite, with a sandy angular moraine present in the centre of the field site and drumlins present in the south west of the field site. The wind farm is primarily located in areas of peat overlying the bedrock.

The Whitelee area sits on a plateau with a peak height of 376 m, with the wind farm being constructed on the top soil, which consists primarily of peat. At Drumtee water the maximum elevation of the catchment is 260 m (Long et al., 2015). Across the wider Whitelee catchment the peat depth, measured at 161 locations, had a mean depth of 1.9 m (with a standard deviation of ± 1.347 m). The main land use in the Whitelee area has been for forestry plantations (mainly Sitka spruce – Picea sitchensis), with most of the forested areas exhibiting varying degrees of surface drying and damage, although the Drumtee water area has undergone deforestation as shown in Figure 3.3 (Waldron et al., 2009, Murray, 2012). (Long et al., 2015) The structure of the peatlands in the wind farm development areas are blanket bog but with some features associated with intermediate bogs such as sphagnum dominated pools and lawns (Waldron et al., 2009). *Sphagnum*-dominated areas contrast with areas of heather-dominated Calluna vulgaris-Eriophorum vaginatum vegetation (Waldron et al., 2009).

The other major land use in the Whitelee area on the open moorland regions is the rough grazing of sheep and cattle (Murray, 2012), which is relevant to the Drumtee catchment as the sampling point is in a farm used for sheep and cattle grazing. At the Drumtee water catchment the area at the sampling point contains more mineral soil than is indicative of the catchment as a whole (as shown in Figure 3.1). Data from Saughall Meteorological Station courtesy of Michael Chalton has been presented in (Murray, 2012). Measurements taken between 1961 and 1990 showed that over a 29 year period the mean number of hours of actual sunlight per day was 3.61. Also noted was that the mean annual rainfall between 1975 and 2005 was 1342 mm. Between 1988 and 2005 climate data from this station indicated that the mean number of days per annum when there was an air frost was 72.8, with the mean winter soil temperature at 0.3 m was 2.69 °C, with the mean summer temperature at the same depth was 13.5 °C. Finally, the lowest annual mean air temperature recorded between 1998 and 2005 was 4.2 °C, with the high annual mean air temperature recorded during that period being 11.5 °C.

The first 140 turbines at Whitelee became operational in 2009 (ScottishPower, 2009), and a further 75 turbines became operational in 2013 (ScottishPower, 2012, ScottishPower, 2013). The monitoring point at Drumtee is downstream of seven the wind farm's turbines, in a third order stream, where the stream width at the point of monitoring is approximately 3 m. At Drumtee the discharge is very responsive to rainfall events, ranging from 0.001 m^3s^{-1} during base flow to 5.3 m^3s^{-1} measured during event flow. There has been some disruption to the catchment for the construction activities, but in 2008 Scottish Power initiated an ongoing peatland rehabilitation programme within the catchment (ScottishPower, 2008).

Most of the data presented in this study came from the sensors deployed at Drumtee water and the samples collected from there. However, technical issues with the equipment relating to sensors not working during certain periods meant that additional data had to be collected from other sources to compensate for this loss. Temperature data was collected from a weather station at Glasgow airport and discharge data was collected from a SEPA gauging station at Newmilns (grid reference: NS532371, 55°36'22 N, 4°19'52 W). The discharge data obtained from the SEPA gauging station at Newmilns was compared against data measured at Drumtee during the periods when the ISCO flow logger was deployed and the relationship was established between the two measurements as discussed later in section 3.2.4.

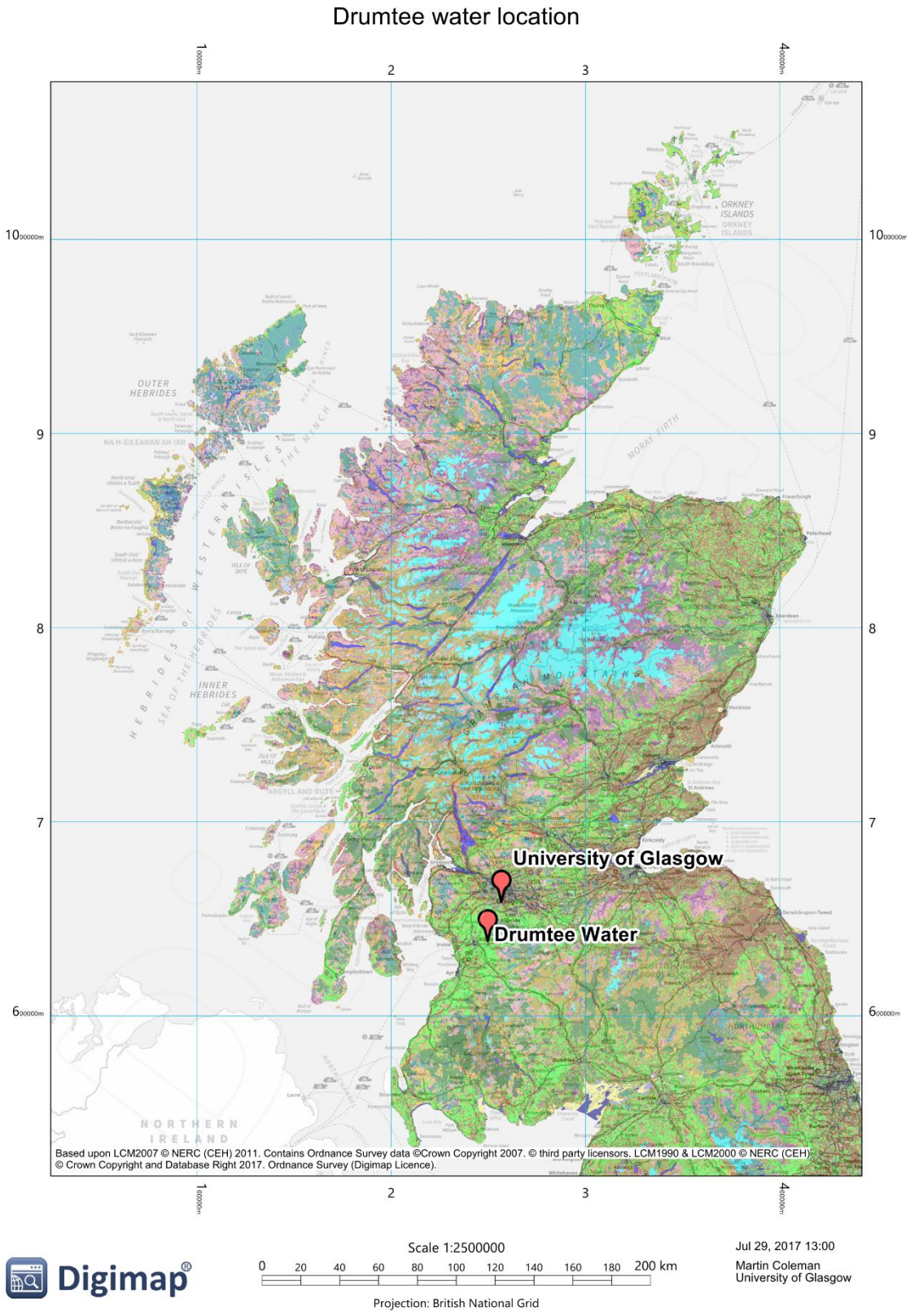


Figure 3.2 – Location of Drumtee water with respect to University of Glasgow
Map showing the location of Drumtee water approximately 30 miles south west of the University of Glasgow. The map was produced using Digimap®.

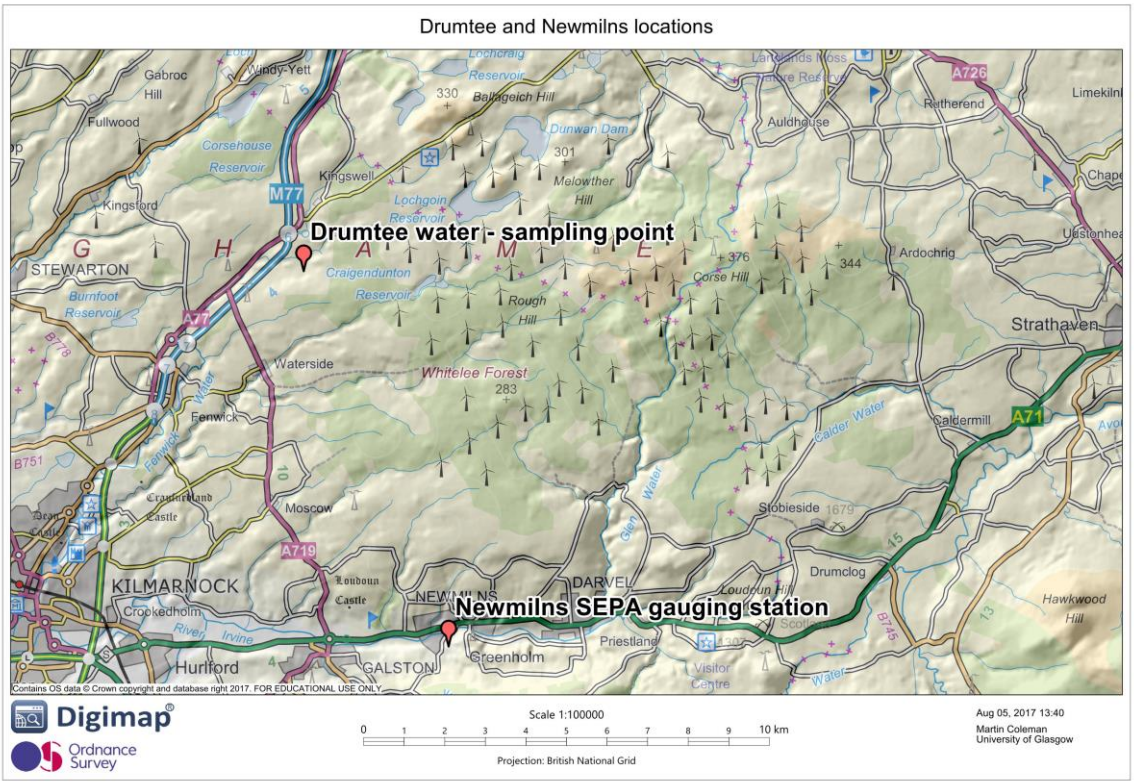


Figure 3.3 – Relative locations of Drumtee water and Newmilns SEPA station
The map of the Whitelee region on Eaglesham Moore SE of Glasgow, Scotland. Annotated on the map are the the relative locations of Drumtee water and the Newmilns SEPA gauging station. The map was produced using Digimap®.

3.2 High resolution sensors

Central to one of the two main goals of the project: evaluating the effectiveness of using a UV-visible spectrometer for measuring [DOC] was tested at Drumtee water. The effectiveness of such a sensor is dependent on the accuracy of the data and how this data can be put into context of the other hydrological variables measured at Drumtee. At the Drumtee field site there have been up to four sensors deployed simultaneously. This project focuses on data that has been collected since May 2012, but data has been collected here since 2007. The earlier data was primarily water quality, through the measurement of [DOC] and water chemistry (pH, specific conductivity) and physical properties of the water (temperature and water pressure). The pressure was used to calculate the depth of the water in the stream and, through a relationship calculated by a deployed flow logger, was used to calculate the discharge of the water and thereby an estimate the total DOC contribution to the annual C budget (Waldron et al, 2008).

The three sensors (described in greater detail in subsequent paragraphs) that were deployed in the Drumtee field location were:

1. a Spectro::lyser™ from S::Can™
2. a 2150 flow logger from Teledyne ISCO
3. a TROLL® from In-Situ Inc.

Additional hydrological data was obtained from the SEPA station at Newmilns, which is the closest gauged catchment to Drumtee. Newmilns is a larger catchment that is on the same river but upstream of the Drumtee catchment.

3.2.1 Spectro::lyser™

The Spectro::lyser™ was deployed in a submerged stainless steel gridded cage with a mesh of approximately 30 mm x 30 mm perpendicular to flow. The Spectro::lyser™ was powered with a 12 V_{DC} deep charge battery that could be used as a car battery. Between measurements the Spectro::lyser™ would go into sleep mode to reduce power consumption and extend the life of the battery. Supplementary power was provided using a solar panel attached to the battery and secured with a breeze block on top of the box in which the battery housing (a large plastic box). Also within the plastic box was the Con::nect™ box that was a control box connecting the battery and the Spectro::lyser™ and also enabling the Spectro::lyser™ to communicate with a computer that contained the software Ana::Pro™. To reduce the chances of any static charges damaging the connect box a piece of thick cardboard was used to separate the battery from the Con::nect™ box (Figure 3.4).

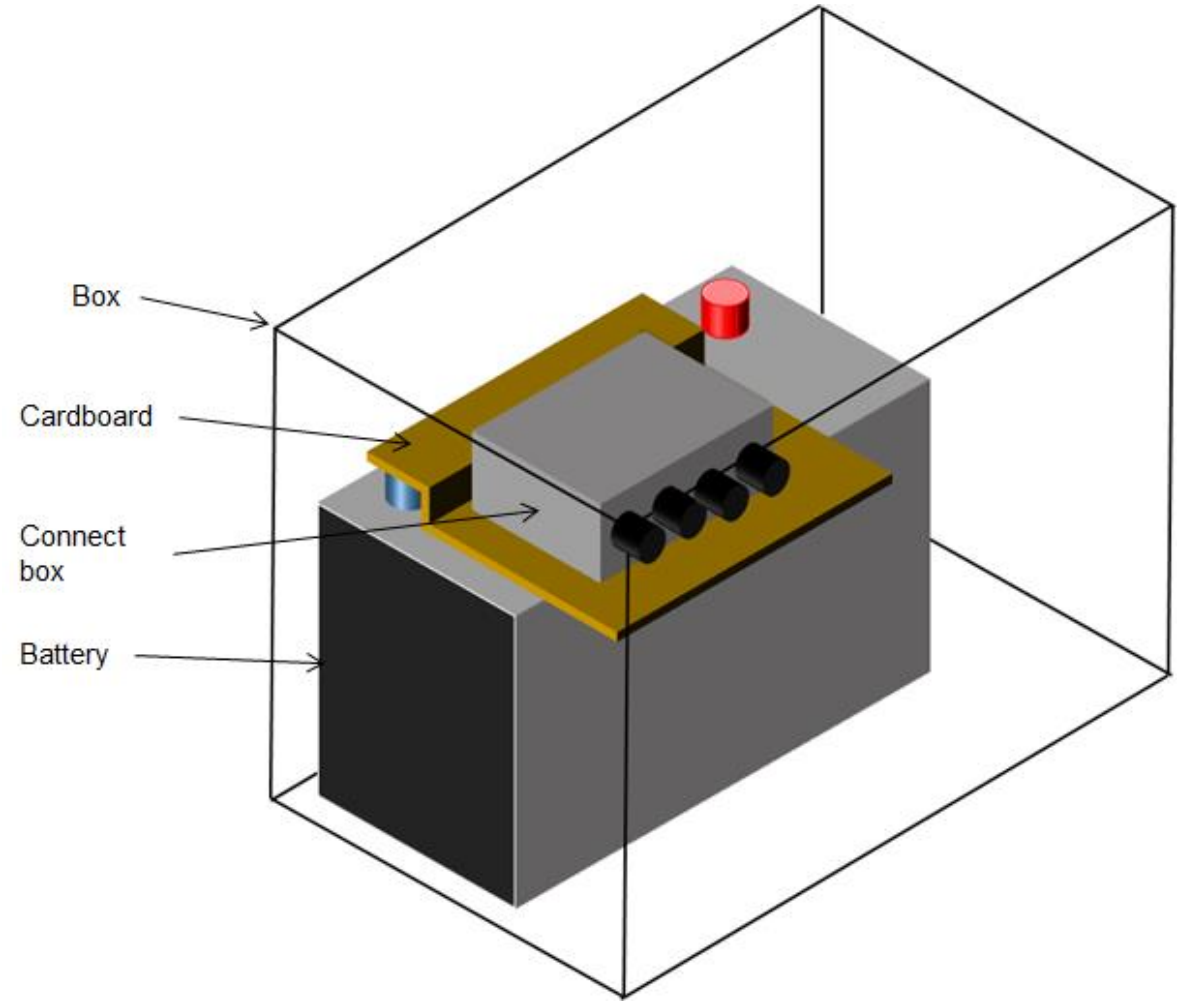


Figure 3.4 – Position of connection box and deep charge battery
Schematic diagram of Con::nect™ **box** on top of battery inside the box with cardboard in between. Dimensions of the box, battery Con::nect™ box and cardboard are all approximate and not necessarily to scale.

The model of the Spectro::lyser™ used here has a 5 mm path length, through which a UV-visible light source is transmitted illuminating the water and dependent on composition, light will be absorbed[1]. The Spectro::lyser™ measures this absorbance from 200 – 735 nm, incorporating both visible and UV light. The Spectro::lyser™ outputs an automated calculated [DOC] value based on internal algorithm that is applied to the measured wavelengths of light. In addition to the calculated [DOC] value the Spectro::lyser™ also outputs the raw data of the absorption values, which could be used to recalculate the

[1] There are different versions of the Spectro::lyser™ with variations on the pathlength. The pathlength options available to me were 1 mm, 5 mm, 35 mm and 100 mm. Where possible the 35 mm lens is recommended because this is the sensor with the greatest accuracy amongst these 3 sensor models. However, as greater amounts of light will be absorbed in longer pathlengths this can result in too much light being absorbed in the water to enable an accurate measurement when the concentration is too high. If all the transmitted light is absorbed in a solution of 40 mg/l then the sensor would no longer be able to differentiate between a 40 mg/l and a sample with a [DOC] concentration greater than this. Following the manufacturer and distributor recommendations that [DOC] at Drumtee had been measured in excess of 70 mg/l it was decided that the 5 mm pathlength represented the best trade off between accuracy and being able to measure at high concentrations.

[DOC]. The Spectro::lyser™ had two internal [DOC] calibrations. The 'global' calibration is required for the sensor to work and is based on a large number of samples from different parts of the world. The 'local' calibration uses samples that have been collected from Drumtee to try and incorporate any locally occurring conditions into the measurement process - the local calibration was performed when the sensor was delivered. The detail of the calibration algorithm is proprietary knowledge I was not allowed access to it by the supplier.

The outputted local [DOC] values were calculated using previously collected field samples, which were used to assess whether the [DOC] was as expected for that time of year (known from this earlier research). When performing the local calibration, the old samples used in the laboratory could not perfectly represent conditions in the field environment as some factors such as turbidity or random material mixed in the water that may influence results would be challenging to simulate in a laboratory-based environment. However, as the system algorithm is not known (being proprietary information) it is not possible to manipulate the equation or to compensate for material that the original algorithm does not, thus the output could not be reinterpreted after a correction for distortion had been applied to the absorption curve. As such I had to generate my own algorithm through the comparison of laboratory measurements of field sample [DOC] and the absorption curve measured by the Spectro::lyser™.

The absorption values were used to recalculate the [DOC] using a modified version of a previously developed equation (Tipping et al., 2009). This original equation used two wavelengths of light at 254 nm and 340 nm. The value closest to 254 nm was selected, which was 255 nm (equation 1). Additionally, the absorption value at 735 nm was incorporated because dirt and biofouling that occurred on the lens of the sensor when deployed in the field, or when the water was turbid under certain flow conditions, caused a distortion to this line and so incorporating this compensated for (equation 2).

$$Eqn.\,1. \qquad [DOC] = x \cdot \lambda_{255} - (1 - x) \cdot \lambda_{340} \qquad where\ 0 \leq x \leq 1$$

$$Eqn.\,2. \qquad [DOC] = x \cdot (\lambda_{255} - \lambda_{735}) - (1 - x) \cdot (\lambda_{340} - \lambda_{735}) \qquad where\ 0 \leq x \leq 1$$

Data was collected from the Spectro::lyser™ every 30 minutes and 1484 data points could be collected before the computer memory reached maximum capacity. The 1484 data points equates to a little over 30 days in terms of data collection before the memory starts to overwrite itself. The time is set from whenever the computer has finished downloading

and the probe is redeployed, which does create a slight offset between data sets. As statistical relationships over very short periods such as an hour are not considered here this is not a major issue.

Furthermore, all of the probes have their own internal clocks and synchronising them precisely is not considered to be a necessary procedure as there is no evidence to suggest that the water chemistry is altering at time scales as short as a couple of hours that would not be obscured by noise.

Data was downloaded approximately every 2 to 3 weeks but was dependent upon being able to get out to the field and access the equipment. Some longer gaps of more than four weeks in length occurred when circumstances prevented access to the field. Such time gaps were never intended because this exceeds the period of deployment before the Spectro::lyser™ starts to overwrite the initial data (Table 3.1). These longer data gaps produced data gaps that required filling in. To reduce the number of data gaps as a result of poor data or no data the sensor was kept as clean as possible.

When the Drumtee river level was sufficiently low during field trips the Spectro::lyser™ was manually cleaned using a brush and 10% HCl acid and water. Between field visits a gas cylinder of compressed nitrogen was used to clean the Spectro::lyser™ by blasting pressurised nitrogen across the path length to remove as much accumulating material as possible from the gap and also the lenses of both the light transmission and receiver. The gas cylinder was set to blast nitrogen at 4-6 bar pressure for 1-3 seconds every 3 and 4 hours. In the winter the water was generally clearer and less material settled so required less cleaning and setting the gas to clean with a single 1 second blast every 4 hours was usually sufficient. Nitrogen gas rather than compressed air was selected because it reacts with very few materials and is not volatile, making it relatively safe to work with and unlikely to bond to the surface of the lenses. If the gas were to react with material in the water or on the lenses it could influence the absorption profile and therefore lead to an incorrect interpretation of the absorbance data. It was identified that nitrogen could distort the UV-visible absorbance signal used to calculate the [DOC] and therefore a 1 minute delay was set into the measurement made directly after the blast of nitrogen gas.

During downloading and maintenance the Spectro::lyser™ had to be shut down so there were data gaps ranging from 30 minutes to 4 hours. As the Spectro::lyser™ could only be accessed during low flow, when significant changes in the [DOC] were unlikely, these short gaps in the time series were filled in by a straight-line projection. For larger time gaps (as described in Table 3.1) straight line projection was also used for calculating budgets and performing wavelet analysis. However, exploratory statistical analysis using

GAM modelling was used to test the usefulness of the GAM model for filling in this data, although this was not used in the statistical analysis of the Spectro::lyser™ data. The reason for not analysing the data produced from the GAM model was because the aim of the analysis was to evaluate how useful UV-visible spectrometry measurements are for measuring [DOC] and splicing modelled data into measured data was deemed to be potentially introducing an additional source of error.

The collection of data from deployed sensors was complimented by the collection of a water sample for laboratory measurement of [DOC]. A one litre sample of water was collected during each fieldtrip and returned to the laboratory where it was filtered, acidified and analysed to create a concurrent alternate [DOC] time series estimated using the more established method of high temperature catalytic oxidation (HTCO). Thermalox® measurements in the laboratory enabled an evaluation of how reliable the Spectro::lyser™ results are in comparison with a tried and tested method.

Inactivity log			
Spectro::lyser™			
Start	End	Duration	Reason
28 July 2013 @ 02:14	30 July 2013 @ 10:59	56 hours	Battery ran out of power
30 December 2013 @ 01:12	8 January 2014 @ 14:53	229 hours	Battery ran out of power
2 March 2014 @ 19:15	3 March 2014 @ 17:44	22 hours	Unknown
17 March 2014 @13:45	25 March 2014 @12:19	190 hours	During this period there was an on-going problem whereby there were unusually large amounts of sediment in the stream that were effectively burying the Spectro::lyser™. When the frequency and intensity of the gas blasts were increased the gas cylinder was depleted preventing its use from further sediment deposits.
8 May 2014 @ 10:16	20 May 2014 @ 02:20	279 hours	
19 June 2014 @ 23:20	25 June 2014 @12:03	132 hours	
11 July 2014 @ 09:15	5 August 2014 @ 10:40	600 hours	Accidently forgot to turn Spectro::lyser™ on
16 December 2014 @10:49	8 April 2015 @10:20	2708 hours	Flooding damaged connect box over Christmas period and after analysis and time spent attempt to repair equipment it was determined that a new connect box was required. At the end of 2014, a large storm (anecdotally described by the farmers working at Drumtee as the worst storm they had ever seen there in 20 years) damaged much of the field equipment including the Spectro::lyser™. It was not until April that this equipment was replaced in the field.

Table 3.1 – Inactivity log for Spectro::lyser™
Detailing lengthy periods when the Spectro::lyser™ was not operational – not including short periods of time when downloading data or cleaning in the field.

3.2.2 Teledyne ISCO 2150 Flow Logger

The Teledyne ISCO 2150 flow logger was not the first sensor deployed to measure the discharge of water passing through the stream but it was the most important sensor for this project because all other measurements regarding stage height or discharge were made with respect to the ISCO flow logger.

The estimations of discharge combine the measurement of the cross-sectional area of the stream with the velocity that the water is moving at, giving the volume of water within a given period. There were three components to this calculation:

1. the mean stream depth
2. the width of the stream at different stage heights
3. the mean velocity of the water passing through the stream

Throughout the project discharge was measured either directly by the ISCO flow logger with its inbuilt pressure sensor, when deployed, or by two pressure-transducers (PT) from In-Situ Inc, one in a hydrochemistry sonde (Section 3.2.3), and later when that was damaged by an independent PT (Section 3.2.4).

The ISCO flow logger was deployed in an area of the stream approximately 15 metres downstream from the Spectro::lyser™. Deploying the flow logger this distance downstream reduced the influence of turbulence arising from the Spectro::lyser™, the cage housing it or the fencing upstream of the Spectro::lyser™. The stream's banks at the point of location for the flow logger were relatively steep, making the cross sectional area easier to measure and calculate. Despite high Reynolds water turbulence being an issue at Drumtee (Long et al., 2015) an area of deployment was selected for the ISCO flow logger that was relatively straight and where visible turbulence was less apparent than in other parts of the stream. The area closer to the Spectro::lyser™ (5-10 metres) was shallower, meaning that during low flow the flow logger could have ceased to be fully submerged in the stream and this area also exhibited a more turbulent flow. By moving the sensor further downstream the turbulence influencing the flow rate being measured and the risk of the sensor not being fully submersed during low flow were reduced.

3.2.3 MD-9000 TROLL®

A hydrochemistry sonde, the In-Situ Inc MD-9000 'TROLL®' was deployed at Drumtee water for this project to measure pH, SC, water temperature and additionally stage heights. Two 'TROLLs®' have been deployed at Drumtee between 2007 and the time of writing. The first was deployed between 2007 until it broke in November 2013 – though this was not picked up fully until early 2014. It was eventually replaced in April 2014, with a slightly different model, the In-Situ Inc TROLL® MD-9000XP, which contains four rather than two batteries. Having four rather than two batteries made the replacement TROLL® slightly bulkier but allowed for a longer battery life. The replacement TROLL® was deployed as near as possible to the position and height above the river bed as the original, although this second TROLL® did not have a pressure sensor transducer so height reproducibility was less important. The implications of this are discussed in section 3.2.4.

The TROLL® was calibrated approximately every two months to ensure that any drifts in the data due to sensor inaccuracy were kept to a minimum and that the measurements were accurate. The TROLL® was removed from the stream for calibration, thus inclement weather could delay this calibration if the TROLL® was submerged due to heavy rain.

Calibration was conducted on the pH sensor using a pH 4 buffer solution and a pH 7 buffer solution. Prior to calibration the sensor was cleaned as much as possible by rinsing in the stream and if necessary pouring some deionised water over the sensor and wiping with a paper towel. Once completed the pH of the pH 4 and 7 buffers were measured and the values noted. A minimum of three values were noted – if the first three values were stable (± 0.2) then this was considered stable and acceptable. If not, the sensor was cleaned again until a stable value was recorded. Once noted, the calibration procedure was completed by letting the pH sensor calibrate itself using first the pH 4 buffer solution and then the pH 7 buffer solution. Once completed the sensor was used again to measure the pH 4 and pH 7 buffer solutions. On second the second set of measurements it was expected that the solutions would be measured as being at 4 and 7 or at least very close to these values (± 0.2). Calibration on the conductivity sensor was achieved by calibrating in a 147 $\mu S.m^{-1}$ solution. As with the pH sensor the SC values before and after calibration were noted and these were measured by testing against the 147 $\mu S.m^{-1}$ solution. Usually the calibrations were quickly successful.

Water temperature was measured using an inbuilt sensor on the In-Situ Inc TROLL®. Alternate temperature data was also obtained at times from a weather station at Glasgow

Airport when there was an issue with the TROLL®. A relationship was established between the temperature recorded at this weather station with the temperature sensor at Drumtee and when temperature data could not be obtained for extended periods this relationship could be used to gap fill these periods (Figure 3.5).

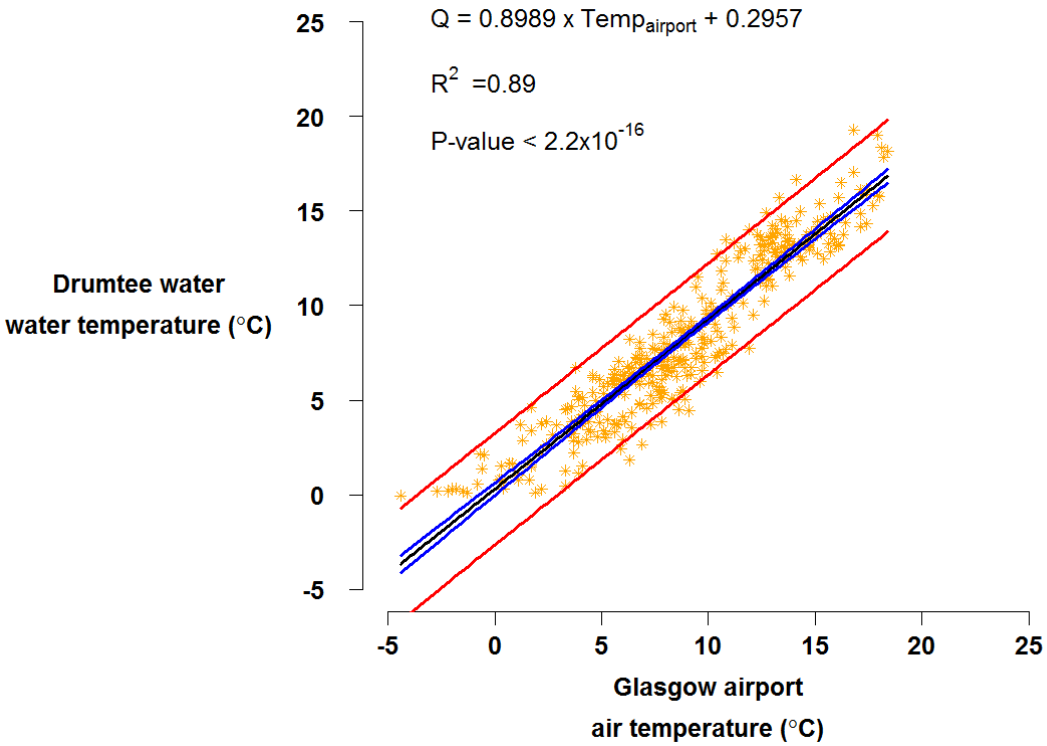


Figure 3.5 – Temperature relationship between Drumtee and Glasgow Airport Mean daily air temperature at Glasgow airport compared with mean daily temperature at Drumtee water. This relationship was used when there were extended periods without any water temperature data due to instances such as the TROLL® data not being collected when it broke. The blue lines show the 95% confidence intervals with the red lines indicating the 95% prediction intervals. The standard error on Temp$_{airport}$ was 0.01706 and on the intercept (of 0.2957) it was 0.17. The residual standard error was 1.5 on 364 degrees of freedom.

3.2.4 Discharge time series

Establishing the discharge time series was crucial for two reasons.

1) to identify when events were occurring to consider [DOC] response and compare to other hydrochemical variables (water temperature, pH, SC and [DOC]).
2) for the C export budget calculations.

Between 2012 and 2014 there were three sensors deployed at Drumtee water capable of measuring water pressure and/or stream velocity. In addition to the ISCO flow logger, a dedicated In-situ Inc pressure transducer and an In-situ Inc TROLL® that incorporated a pressure transducer were deployed. Additional discharge data was also available from the SEPA gauging station at Newmilns, which as discussed in section 3.1 is the nearest

SPEA gauging station to Drumtee. The SEPA data is measured at 15 minute intervals so only every second data point was used to match the frequency of measurements made by the sensors at Drumtee. The time period for which data was available from is shown in Figure 3.6.

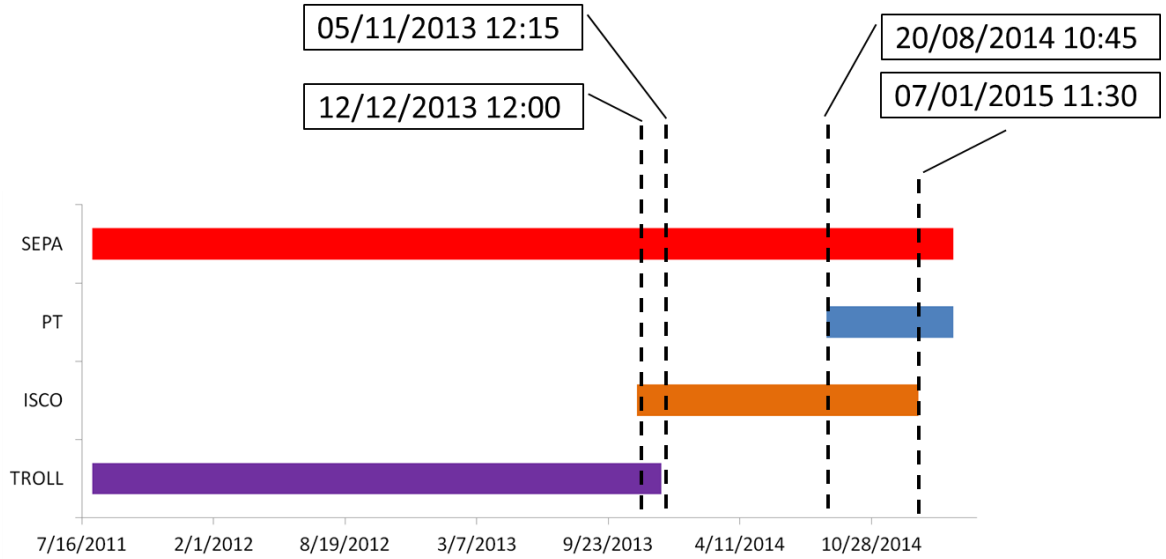


Figure 3.6 – Timing of water flow time series
Data was available for recreating a discharge time series as shown above. SEPA refers to the SEPA data that was obtained from the Newmilns catchment; PT refers to the dedicated pressure transducer from In-situ Inc; ISCO refers to the flow logger from ISCO and the TROLL® refers to the In-situ Inc TROLL® that housed the hydrological sensors pH, EC, water temperature and the pressure sensor.

The pressure sensor in the TROLL® stopped operating on 12 December 2013. Consequently, the overlap of data with the ISCO flow logger was between 5 November 2013 and 12 December 2013. However, during this period the pressure transducer on the TROLL® exhibited some unusual variation in terms of "spikey" data that differed from the usual event characterisation of fast rising limb and slower falling limb. This spikey data was not observed on either the ISCO flow logger or the data from Newmilns. Nevertheless, a relationship between the ISCO flow logger and In-Situ Inc TROLL® stage height data was established (Figure 3.57). A straight-line relationship is not appropriate because stage heights measured below about 0.35 m would end up calculating the discharge as a negative value. Therefore, a power curve relationship was established instead.

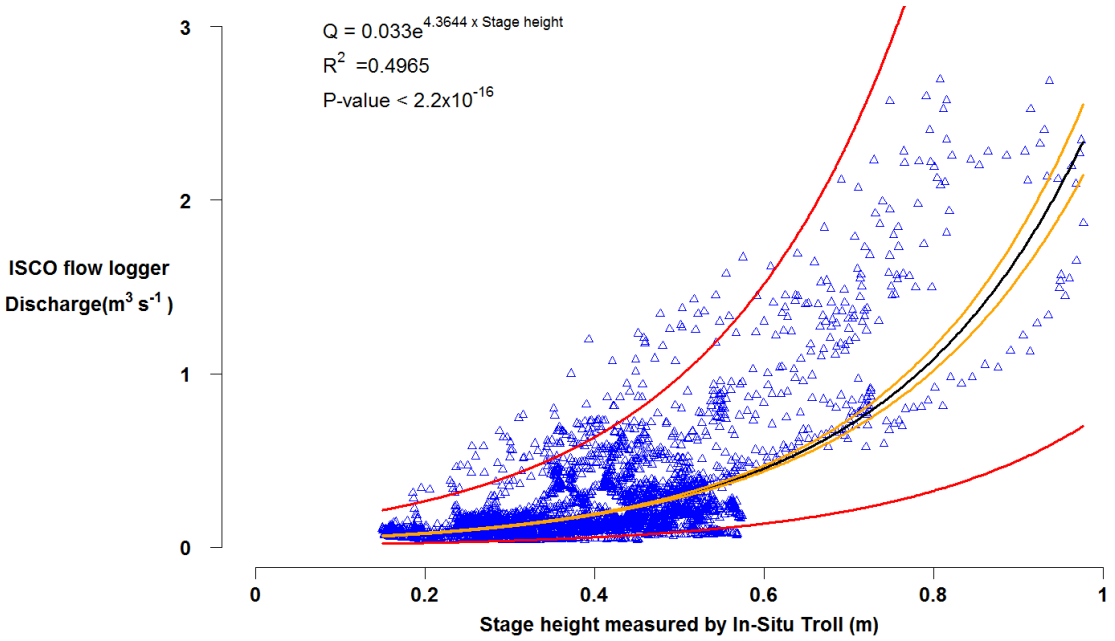


Figure 3.7 – Stage height and discharge
Relationship between stage height measured by the TROLL® at Drumtee with the discharge measured by the ISCO flow logger at Drumtee. Using a power curve relationship, the R^2 is 0.50. The orange lines show the 95% confidence intervals with the red lines indicating the 95% prediction intervals. The standard error on stage height was 0.0007 and on the intercept it was 0.03. The residual standard error was 0.6 on 3550 degrees of freedom.

However, as this relationship was not as strong as would be ideal, an alternative relationship was established between the discharge measured by the ISCO flow logger at Drumtee and the SEPA data from Newmilns (Figure 3.8) and used to calculate discharge for the period prior to 5 November 2013 and then briefly from 25 June 2014 to 20 August 2014.

Data was obtained from the SEPA station at Newmilns from 1 August 2011 at 00:00 to 7 January 2015 at 12:30. Newmilns is a 72.8 km^2 catchment near to Drumtee catchment and a relationship was established between discharge at Drumtee and discharge at Newmilns to determine if this could be used as a suitable proxy for instances when there was no discharge data collected at Drumtee itself.

The Newmilns-SEPA discharge data was compared against the Drumtee-ISCO flow logger data for the period 5 November 2013 12:15 to 7 January 2015 11:30 (Figure 3.6) and used to generate a linear model. During some large events at Drumtee the stream could overflow, with water coming over the banks, which could cause inaccuracies in calculated discharge. In such events the discharges recorded by the ISCO flow logger were less representative, remaining approximately constant at Drumtee, although discharge may be increasing at Newmilns as the river channel is larger. Thus, when the SEPA station at Newmilns recorded discharges above 40 m^3s^{-1} this may not be reflected

in increasing discharge at Drumtee and so the Drumtee discharge was underestimated. In this case it is difficult to estimate discharge, but fortunately values greater than 40 m^3s^{-1} recorded at Newmilns only account for 0.3% of recorded values. If SEPA discharge values greater than 40 m^3s^{-1} are filtered out then the relationship becomes Q_{ISCO} = $0.0885Q_N$ – 0.0071. Therefore, removing the high discharge data from Newmilns still produces a similar relationship to that shown in Figure 3.6 with only a small difference in the slope and intercept of the line. This does indicate that direct measurements by the ISCO flow logger may involve a greater error at this site for large values but this would account for a very small percentage of values.

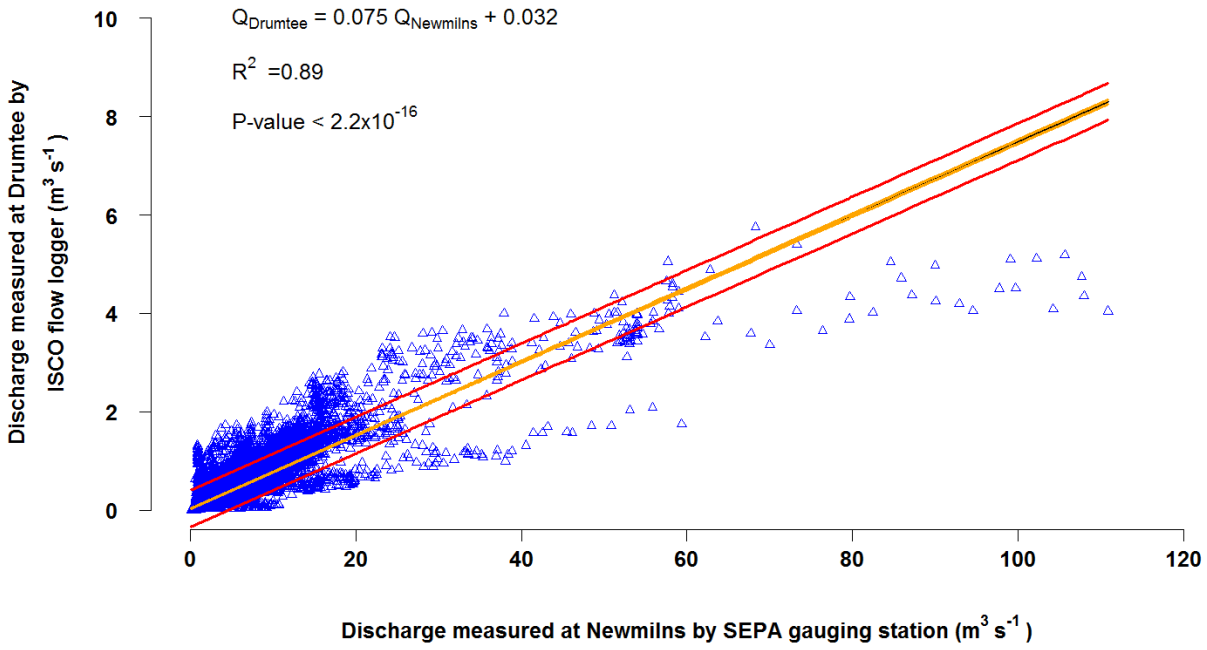


Figure 3.8 – Relationship between Drumtee and Newmilns
Relationship between discharge measured by SEPA at Newmilns with data recorded at Drumtee water using the ISCO flow logger. The orange lines show the 95% confidence intervals with the red lines indicating the 95% prediction intervals. The standard error on $Q_{Newmilns}$ was 0.0002 and on the intercept (of 0.032) it was 0.001. The residual standard error was 0.2 on 35405 degrees of freedom.

A third sensor used to measure discharge at Drumtee was an In-Situ pressure transducer that was deployed on 20 August 2014, by attaching it to the front of the steel mesh cage that housed the Spectro::lyser™. It was attached with a steel clamp to the front of the cage that was downstream from the Spectro::lyser™ so that is did not interfere with stream flow through the Spectro::lyser™. A polynomial relationship was selected for modelling the data as logarithmic, exponential and power curve models produced poor predictive results at the larger stage heights and discharge values. The relationship between the pressure transducer and ISCO flow logger is good with an adjusted R^2 of 0.92 and a p-value of less than 2.2×10^{-16}. Therefore when the ISCO flow logger data

could not be used directly the In-Situ pressure transducer was used when available as it provided a stronger relationship than the SEPA or Troll® data.

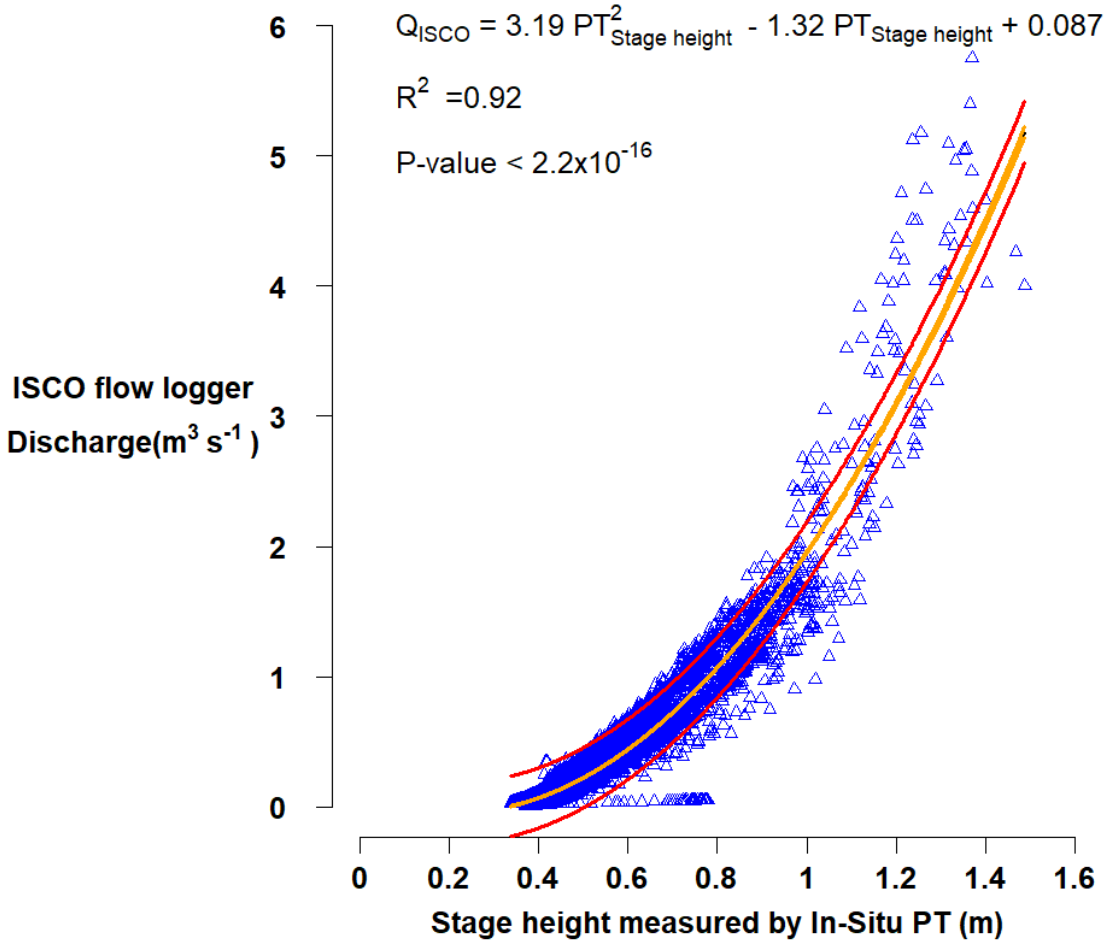


Figure 3.9 – Relationship between Pressure Transducer and ISCO flow logger Relationship between discharge measured by Pressure Transducer at Drumtee water with data recorded at Drumtee water using the ISCO flow logger. The orange lines show the 95% confidence intervals with the red lines indicating the 95% prediction intervals. The standard error on PT2 was 0.01, on PT it was 0.04 and and on the intercept it was 0.01. The residual standard error was 0.118 on 13063 degrees of freedom.

The adjusted R^2 value between the discharge recorded at the SEPA station at Newmilns and the data recorded by the ISCO flow longer is considerably greater than the adjusted R^2 value recorded between the discharge recorded at the TROLL® positioned at Drumtee and the ISCO flow logger. Therefore, the relationship shown in Figure 3.7 was in the analysis involving the extended GAM model (section 2.6) because here the model needed data from 2008, but the Newmilns data I had access to only started from 2011. As will be shown in chapter 4 the bivariate model of discharge and [DOC] was extremely poor so the influence on the GAM was minimal. The discharge time series was produced, where possible, from the ISCO flow logger. When this data was not available the second choice was the pressure transducer and finally when neither of these was an option then the discharge calculated from the SEPA data was used as illustrated in Figure 3.10.

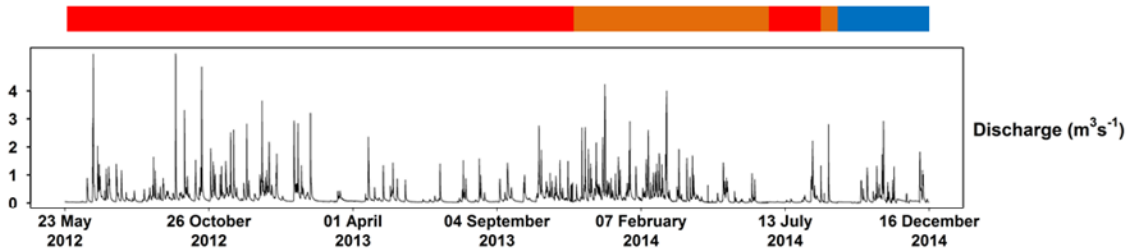


Figure 3.10 – Spliced discharge profile used in project
The colour bar shows the period in red where the discharge was calculated using data from the SEPA data at Newmilns (red bar), data from the ISCO flow logger (orange bar) and the In-Situ Pressure Transducer was used. The SEPA data was used from 23 May 2012 till 25 November 2013 and then from 25 June 2014 till 20 August 2014. The ISCO data was used from 25 November 2013 till 5 June 2014 and then from 20 August 2014 till 8 Septmeber 2014. The In-Situ pressure transducer data was used for the remainder of the time series.

3.3 Laboratory methodology for measuring [DOC] and [POC]

Laboratory work for this project consisted of the following processes:

1. processing of samples upon return from the field
2. measurement of those samples for [DOC] and [POC]

3.3.1 Initial processing

One litre field water samples were vacuum-filtered through a 0.7 µm glass fibre filter, which had been ashed at 500 ºC for 8 hours to remove organic C. When filter papers became blocked, a fresh GF filter paper was used. The volume of water filtered was recorded in order to calculate [POC] (mg/l). There was a gap of one to six hours between collecting the sample and processing it in the laboratory.

After filtration filter papers were oven-dried (60˚C) and retained in plastic petri dishes for further analysis. Two 50 ml samples of filtrate were retained in plastic centrifuge tubes (A and B); the second was a back-up in case the first bottle was either spilled or there was suspected contamination. Inorganic carbon was removed from the A sample through acidification of the solution to pH 4 using 0.05 mol hydrochloric acid. The acidification converted the inorganic carbon into CO_2, which was then degassed by placing the samples into an ultrasonic bath for 12 minutes. After the acidification and sonification the water samples were stored in a fridge at 5°C until such time as [DOC] measurements were conducted. The above procedures were usually conducted within 2 months of sampling.

3.3.2 [POC] measurements

The [POC] was measured through the process of loss on ignition. Here the filter papers were dried to remove water (at 105ºC for a minimum of four hours), weighed on a balance accurate to 0.001g and then furnaced at 395°C overnight, for a minimum duration of 16 hours although usually it was closer to 20 hours. The filter papers were weighed again and from this the loss of material can be calculated; this mass is subsequently corrected (van Bemmelen factor) for the fact that only 58% of the lost mass lost is C (Equation 3) (Ball, 1964). The loss on ignition method for measuring [POC] is well-established method for measuring [POC], with early analysis comparing combustion at 800°C for a duration of less than an hour to combustion at 375 °C for a duration of 16 hours (Ball, 1964) and finding that the lower temperature reduced error by eliminating additional loss on ignition that might be attributed to POC.

Equation 3

$$[POC] = \frac{(Mass\ of\ filter\ paper\ at\ 395℃ - Mass\ of\ filter\ paper\ at\ 105\ ℃)}{Mass\ of\ water\ filtered\ (g)} \times 1000 \times 58\%$$

Brian Schumacher of the United States Environmental Protection Agency produced a report (Schumacher, 2002) identifying that although 58% is used as an assumption on the basis that 58% of organic soil organic matter is comprised of organic C, based on a conversion factor of 1g of organic C per 1.724g of organic matter. However, the report cites literature that suggests conversion factors ranging from 1.724 to 2.5 (Burt, 1992, Nelson and Sommers, 1996), with reference being made to a paper, which recommends conversion factors of 1.9 for surface soils and 2.5 for subsurface soils (Broadbent, 1953). Therefore, the 58% for calculating [POC] based on the loss on ignition is an assumption. However, the predominant focus of this project is on [DOC] and as the dissolved fraction of organic carbon dominates at Drumtee errors in the POC exports will be minimal. This approach has also been used in other research studies (Hope et al., 1997, Dawson et al., 2002, Dinsmore et al., 2013). As research into the composition of carbon exports expands then a possible future area of research might be to better establish how consistent this figure is and if so what is the best % figure to use.

3.3.3 [DOC] measurements

[DOC] was calculated using high temperature catalytic oxidation (HTCO). HTCO is a method of measuring [DOC] that has been used widely in academic studies since the late 1980s (Sugimura and Suzuki, 1988) though the concept had been experimented with in the 1960s (Van Hall et al., 1963, Menzel and Vaccaro, 1964)and 1970s (Jonathan H, 1973). The original method was developed for sea water but it is applicable to fresh water

samples as well. The concept behind this technology is to oxidise the DOC component of the carbon pool in the presence of oxygen and a catalyst at a high temperature after it has been isolated from other sources of carbon such as particulate carbon or inorganic carbon. The oxidisation converts the carbon in the sample to CO_2 and this can be measured using a CO_2 detector. A helpful paper by Spyres provides a concise, well referenced description of the key procedures of HTCO and the development of this method (Spyres et al., 2000).

Following filtration, acidification and sonification it was possible to measure the [DOC] content of the water solution using HTCO. A Thermalox® from Analytical Sciences Ltd was used as the HTCO system and was equipped with a computer controlled arm that could inject samples using a pump and a syringe into a furnace designed to measure total organic carbon concentrations [TOC]. The Thermalox® was delivered with software written on LabVIEW, which provided a GUI to control its operations. The furnace was temperature controlled to 680°C and contained a platinum catalyst and was supplied with a flow of O_2. When a sample was injected into the furnace the carbon in the solution was converted in the hot environment under the presence of the catalyst to CO_2. The CO_2 passed out of the furnace along a tube to an infrared spectrometer and the data from the infrared spectrometer was read by the computer, which converted the measured absorption of IR light into a measurement of the quantity of CO_2. By comparing the quantity of CO_2 measured to the injection volume of solution the computer converted this value into a [DOC] value.

The [DOC] measured in the laboratory was compared with the [DOC] estimated in the field by the Spectro::lyser™. The time of a sample was noted when it was collected in the field. On occasions where the time was not noted then a time of 12:00 noon was used as this was the time that most samples were collected at. The fieldsite was located about 40 minutes drive away from Glasgow. A typical field trip would start by leaving the university about 10:00 in the morning and fieldwork would start at about 11:00. A water sample was collected after the Spectro::lyser™ was redeployed so 12:00 was considered representative. The Spectro::lyser™ value recorded at the corresponding time could then be compared with this value.

If there was a significant difference from the measurements made by the Spectro::lyser™ or an unexpectedly low or high sample then that sample would also be retested. Ideally samples were analysed within a few months of being returned from the field, which is fine as immediate filtration and refrigeration thereafter allows storage for up to three months without compromising sample composition (Gulliver et al, 2010). However, the analyser was malfunctioning for a lengthy period during 2014 preventing analysis being conducted

until a maximum of about nine months after collection, so it is possible the HTCO [DOC] measurements generated from the 2014 samples may be less accurate than normal.

To ensure the system did not accumulate an analytical blank throughout the run and to account for drift in the measurements the [DOC] samples were interspersed with three types of solution (Figure 3.8):

1. UV-filtered, de-ionized water that should give a carbon value very close to 0 mg/l.
2. standard solutions of known concentration covering the full range of expected [DOC] values. Therefore, a wider range of standards was used in the summer time unless the Spectro::lyser™ indicated greater than normal concentrations for a particular time of year. In event that a sample from Drumtee was outwith the range of standards that sample would be retested using an increased range of standards. The standard solution was a diluted solution of hydrogen potassium phthalate, which was stored in stock at 1000 mg/l.
3. an internal standard, termed '20-check' was to ensure that the machine was consistent from one run to the next. The 20-check was made from 1000 mg/l stock solution of hydrogen potassium phthalate. The same 20-check solution was used for multiple runs to ensure that the results of one run were comparable to another.

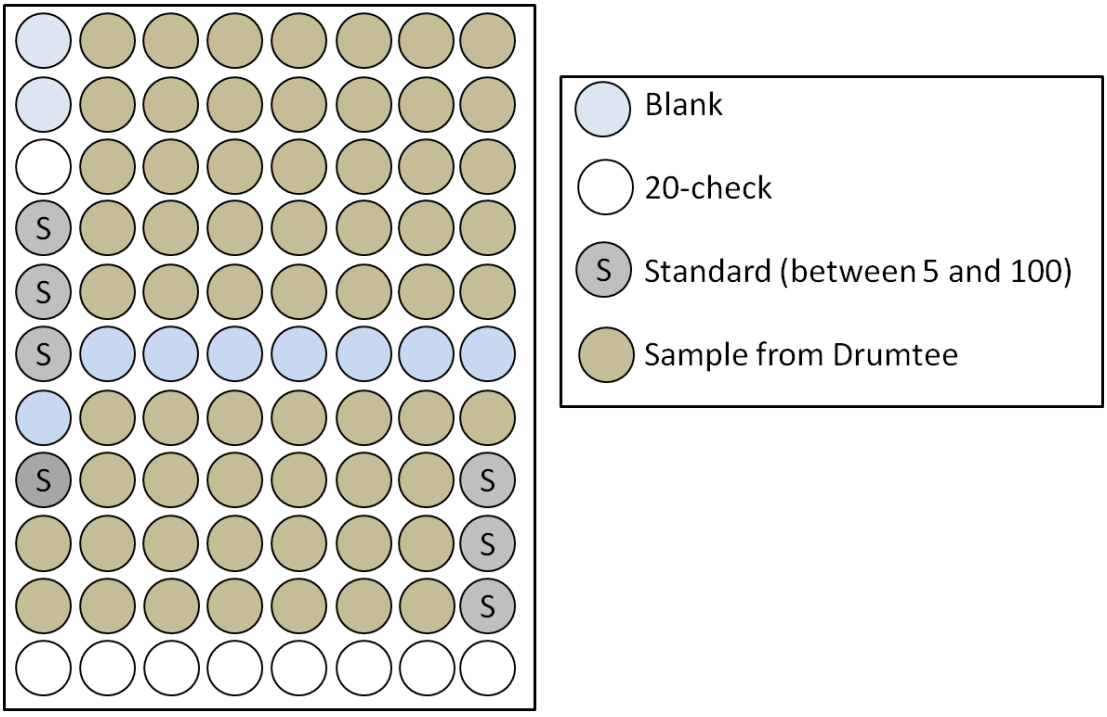


Figure 3.11 – A typical layout of a set of samples on the Thermalox®
A mixture of samples from Drumtee were interspersed with blank water and [DOC] samples of known concentration prepared using hydrogen potassium phthalate.

The standards of known concentration are prepared by the Analytical Sciences Ltd Thermalox® from dilution of a 1000 mg/l of hydrogen potassium phthalate stock solution. The Thermalox® was programmed to prepare standard solutions from most concentrated to least concentrated.

Collectively the standards enabled the correction of the [DOC] values returned by the Thermalox® to give more accurate [DOC] values that would be more representative of the actual [DOC]. A linear regression line was used in the comparison between the output measurements of the standard solutions made and the known values of these solutions (an example of which is shown in Figure 3.12).

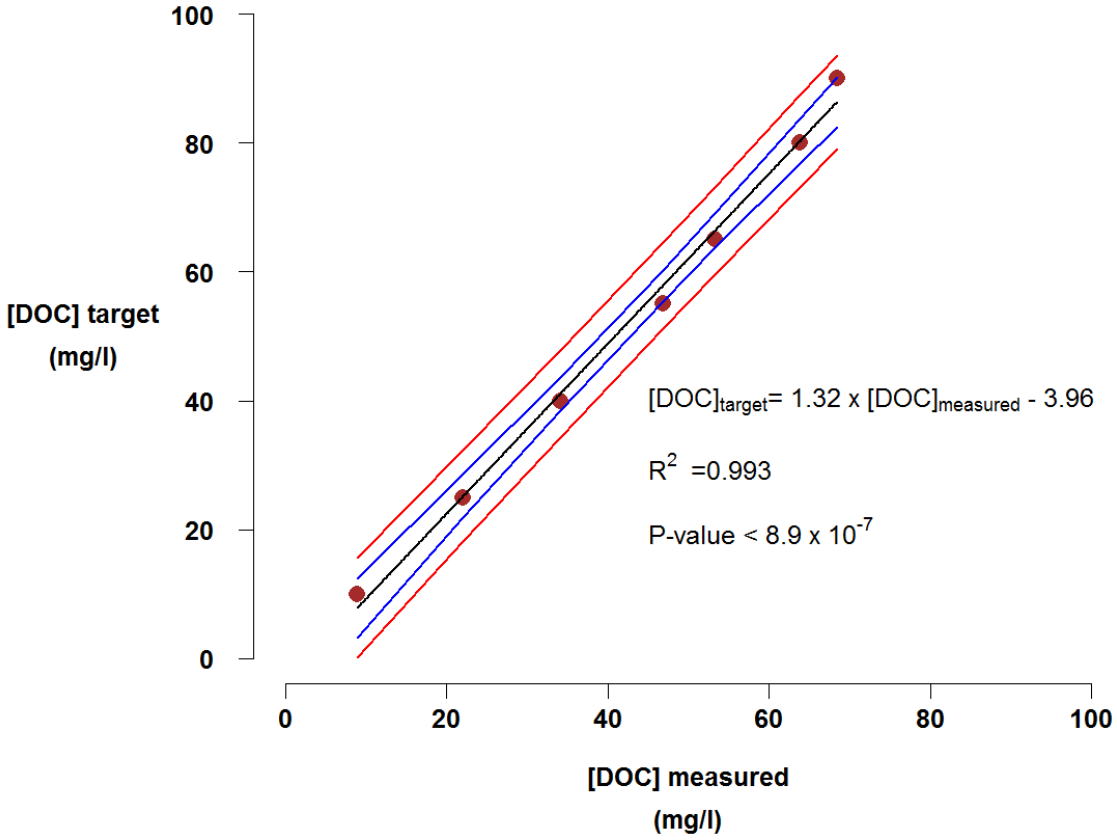


Figure 3.12 – Analysis of standards
Comparison and modelling of measured concentration with the known concentration of those standard solutions. Data was taken from a set of experiments run on 19 May 2015. The model line is shown in black, with the 95% confidence intervals shown in blue, with the 95% confidence intervals shown in red. The standard residual error of [DOC] measured was 0.045 and the standard error on the intercept was 2.1. The residual standard error was 2.4 on 5 degrees of freedom.

Two injections of 30 µl of water were injected into the furnace for each sample. If there was a significant difference, indicated by a coefficient of variation > 5% between the measured concentration of the two samples then the computer controlled equipment

injected a third sample. (The coefficient of variation is the standard deviation of the samples analysed divided by the mean of the samples analysed). In cases where there was still a high coefficient of variation of the [DOC] measured then the samples would be retested at a later date, which reduced the likelihood of significant error in the laboratory measurements. The inability to retest samples in the field using the Spectro::lyser™ does mean that there is a greater error associated with the field based measurements compared to the laboratory based measurements.

Finally, the dilution resulting from the acid added during titration was corrected for (equation 4) to produce the [DOC] used in results and analysis. Typically, the resulting correction was very small, with the mean volume of acid added to reduce the pH to a value of 3.9 being 0.5 ml with a standard deviation of ± 0.5 ml. Alkalinity at Drumtee was low with a mean value of 35.4 mg/$CaCO_3$ and a standard deviation of ± 43.3 mg/$CaCO_3$ (Table 3.2).

	Acid added (ml)	Alkalinity (mg/$CaCO_3$)
Minimum	0.1	3.5
Maximum	2.8	299.4
Mean	0.5	35.4
Median	0.3	21.8
Standard deviation	0.5	43.3

Table 3.2 - Statistical summary of volume of acid added and the alkalinity of samples
Summary table of the acid added to a sample to acidify the sample to a Ph value of 3.9 and the alkalinity value of that water sample.

Equation 4

$$[DOC] = [DOC_{thermalox}] \times \frac{Volume_{acid} + Volume_{solution}}{Volume_{solution}}$$

The [DOC] measurements made with the Spectro::lyser™ were compared with the [DOC] measurements made in the laboratory by the Thermalox® (Figure 3.13). The model generated produced a reasonably tight relationship between the two. If the data point at 81.1 mg/l is removed then the R^2 improves to 0.78. The p-value of the relationship was 7.1×10^{-15} with this data point left in but only changed to a value of 4.3×10^{-18} when the data point was removed. The extraction of data points from the Spectro::lyserTM time series to compare with the laboratory measurements was achieved by identifying the data point closest in time on the Spectro::lyser™ time series to when the field sample had been taken.

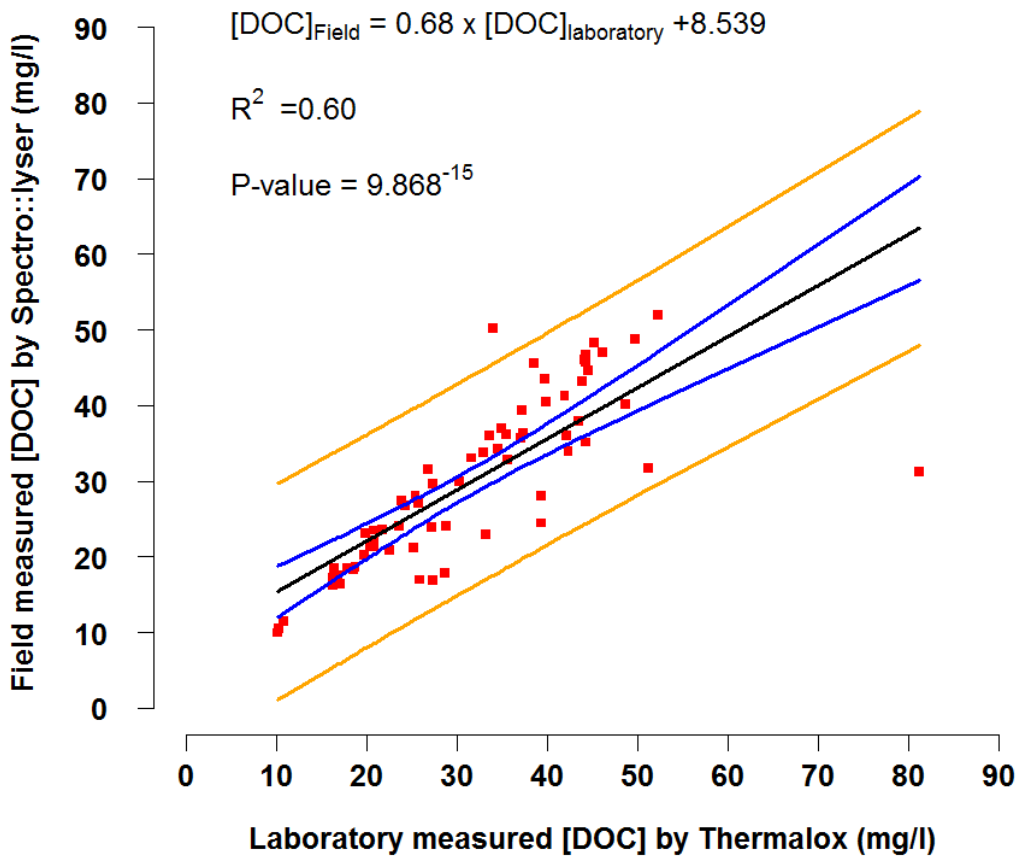


Figure 3.13 – comparison of laboratory and field measured [DOC]
Relationship for measurements made in the field and the laboratory. The black line represents the best fit straight line, with the blue line showing the 95% confidence intervals, with the yellow lines showing the 95% prediction intervals. The measurement error for the laboratory measured data is small because of the sample is measured under laboratory conditions and can be measured multiple times. The error on the Spectro::lyser™ is harder to determine but the standard error is presented. The p-value is low. The standard error on $[DOC]_{laboratory}$ was 0.068 and on the intercept (of 8.539) it was 2.3. The residual standard error was 6.9 on 65 degrees of freedom.

To create a separate continuous time series the data between the different laboratory points was projected as a straight line (Figure 3.14). This straight-line projection approach to gap filling was also used to create a continuous [POC] time series between data points based on a straight line between measured values.

The true error present in the measurements made is likely to be greater than the standard error as it is challenging to identify all sources of error. It is possible that some of this error is as a result of the measurements made by the Thermalox® but as these measurements are made under laboratory conditions and can be retested it is likely that most of the error is from the Spectro::lyser™. As previously discussed the errors could be arising from issues such as imperfections with the algorithm for converting the UV-visible light profile to a [DOC] value or other materials in the stream interfering with the measurements. Greater errors tended to be recorded when the Spectro::lyser™ was not kept clean, as was the case before the installation of the nitrogen gas cylinder.

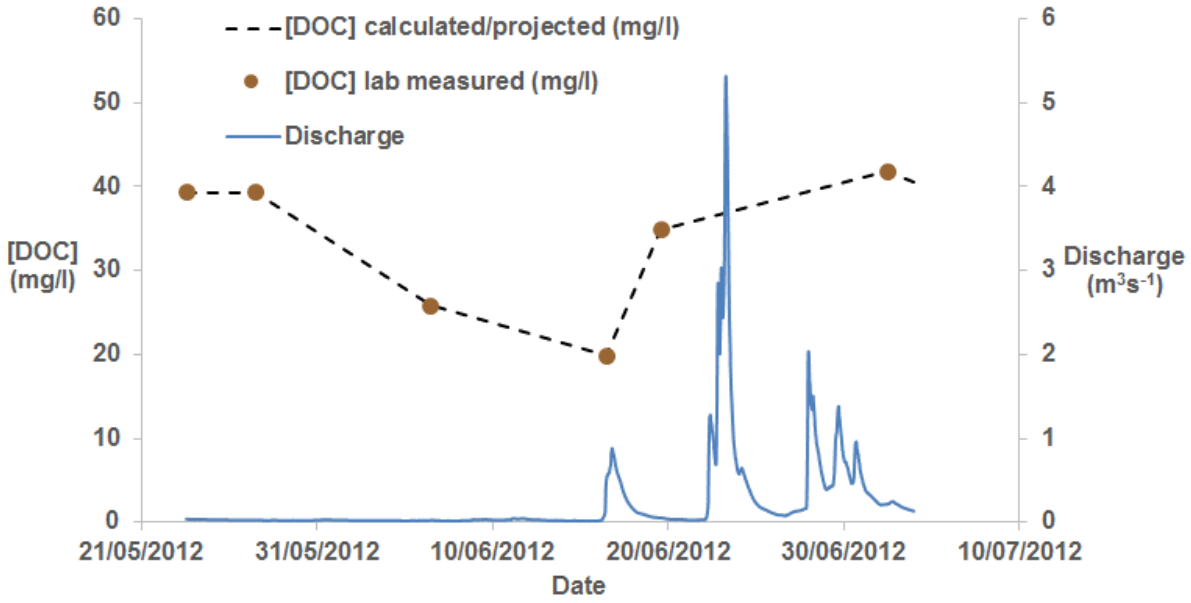


Figure 3.14 – Calculated [DOC] between manually collected samples
For manual samples a straight line was assumed to be representative of [DOC] between successive [DOC] and water samples and the calculated value using this straight line was extracted for each 30 minute period and used in equation 5.

As a result, there were three organic carbon time series:

1) [DOC] generated using the Spectro::lyser™
2) [DOC] generated using manual sampling but projected to 30 minute intervals using the method shown in Figure 3.14
3) [POC] collected using manual sampling but projected to 30 minute intervals using the method described above.

3.3.4 Uncertainty in budget calculations

As discussed in section 3.3.3 it is difficult to fully determine the errors associated with the [DOC] measurements and in particular those made by the Spectro::lyser™. These errors will in turn introduce uncertainty into the budget calculations. The sources of error can be varied: errors can result from the [DOC] measurements; the [DOC] algorithm when using the Spectro::lyser™; from the [POC] measurements and also potentially from errors in the discharge measured.

The discharge time series used was the same for all budget calculations, whether the DOC components were based on laboratory data or the field measured data. The time series of the [POC] measured using the loss-on-ignition method was used for all budgets regardless of whether the [DOC] time series being used was generated from field or laboratory data. As the POC accounts for a smaller percentage of the organic carbon at Drumtee than does

the DOC this potential disparity in the frequency of measurements is not as significant as it might have been.

One source of error when using the Spectro::lyser™ can lie with the measurements themselves. For example, is the water passing through the path length representative of the stream? This issue of the water not being representative is not likely to be a major source of error as the [DOC] that is measured is usually stable over several successive measurements unless there is a rainfall event. If this were an issue then we would expect to measure greater fluctuations than were observed at each successive half-hour measurement. Errors with the algorithm represent a more likely source of error. The composition of [DOC] is complex and therefore the spectroscopic properties of some components might not be being identified as some less common components may not have been identified via the local calibration. Less common components of DOC not identified during the calibration might account for why the laboratory value in November 2014 shows an unusually large value of 81.1 mg/l but was only measured as 31.2 mg/l by the Spectro::lyser™. This sample was tested twice in the laboratory due to the unusually large value – the first analysis on 19 January 2015 produced a result of 76.5 mg/l and the second result produced the value of 81.1 mg/l.

However, another consideration is that although the error associated with each [DOC] measurement taken in the laboratory is smaller than the [DOC] measurement taken in the field the field measurements of [DOC] are made more regularly, generating a denser dataset. Therefore, the laboratory data is not able to pick up on short term variations in [DOC], particularly during events when the largest export of organic C will occur (Raymond and Saiers, 2010). These long gaps in the data likely represent at source of error for budget calculations based on the laboratory data as there is the potential for significant fluctuations in the [DOC] in the periods between sampling. In contrast, the largest source of error associated with the field measured [DOC] is regarding the accuracy of the measurements.

There is also an uncertainty associated with the POC components used in the budget calculations. Unlike the [DOC] there is no known method for calculating [POC] using the UV-visible spectroscopy data and consequently there may be a significant amount of detail missed and therefore potentially a large export of POC may have occurred during some events but not measured.

Errors associated with the discharge will also influence the budgets calculated. The ISCO flow logger is set up to measure the discharge at Drumtee as directly as possible. There are errors with the measurement of the equipment, with turbulence in the water and any errors when calculating the cross-sectional area of the stream. When discharge data from

Newmilns has to be used to gap fill the discharge profile this will introduce additional errors into the budget calculations. Firstly, there may be localised variations in the weather, which may introduce a source of error. Secondly, as shown in Figures 3.7, 3.8 and 3.9 the relationship is not 1:1 and therefore there is an error with the model converting the other measurements of stage height or discharge to a discharge value for Drumtee water.

3.4 Statistical approaches to analysing data

Large volumes of data can necessitate changes to the way that analysis is conducted. Specifically, large volumes of data can make manually sifting and analysing that data impractical, especially if that data is still being collected during the time when analysis is being conducted. Therefore, it can be advantageous to automate analysis so that the data can be processed more quickly and allow more time for interpreting the outcomes of the analysis. In this project, the open source software R was used extensively within the integrated development environment (IDE) R Studio.

One of the purposes that R was used for was to automate the process of identifying and analysing events. This automation using R enabled the effect(s) of events on the catchment to be identified and determined. Identifying events by eye when looking at the graph of the discharge is relatively straight forward as the events are identifiable as rapidly increasing discharge values that maximise and then decrease (usually at a slightly slower rate than they increased at).

There are approaches for identifying leads and lags within data such as hysteresis, which has been used in hydrology research and published papers before (Basile et al., 2003, Beven, 2006, Fovet et al., 2015). However, the decision was made to not use the hysteresis approach and to write new code for analysing the data because when it was written it was not known exactly what was going to be looked for. Comparison of the time series during different segments of the events was done using custom written code to give greater flexibility so that different comparisons could be extracted as ideas arose.

3.4.1 Identifying events and their influence on catchment chemistry

To meet the challenge of effectively utilising the data collected via high-resolution sensors in a system as hydrologically active as Drumtee there is a significant benefit to automating the analysis using computer code. An important question to consider in particular was what influence events were having on the [DOC], pH and SC. By writing code to identify the timing of significant points on the hydrological time series such as [DOC] minimum and discharge maximum it would be possible to explore detail questions such as: i) does

the rising limb of the discharge profile affect the stream in a different way from the falling limb? ii) Do larger events affect the catchment differently than small events?

Creating a statistical analysis of the [DOC] response to discharge is of particular interest because, on visually inspecting the time series, the response appeared to vary from one event to the next with no obviously clear pattern. It was necessary to identify rules to enable a consistent analysis of the hydrological data. The simplest method of event analysis would have been to set a single threshold value for identifying event flows. When discharge was above this value then the stream would be considered to be in event flow and when the stream was below this value the stream would be identified as being at baseflow. The problem with this approach is that it is not especially informative and would fail to provide detail.

To automate the process of identifying events it was necessary to establish common data features that could be used to establish simple rules that could be identified by the computer. A characteristic of peatlands is that they have flashy high floods with short lag times (Holden and Burt, 2002), a pattern observed in the discharge profile generated from Drumtee. In practice, the events are characterised on the discharge time series by a rapid increase in value, with the subsequent decrease in discharge occurring at a slower rate. Four points were selected in the discharge time series for characterising events as being points in events (Figure 3.15):

1) the start of the event
2) the point in the event where the increase in discharge was at its most rapid. This increase would be identified by subtracting each data point from the one that preceded it and the largest positive value would be the point of most rapid increase
3) the point of maximum discharge during an event
4) the end of an event

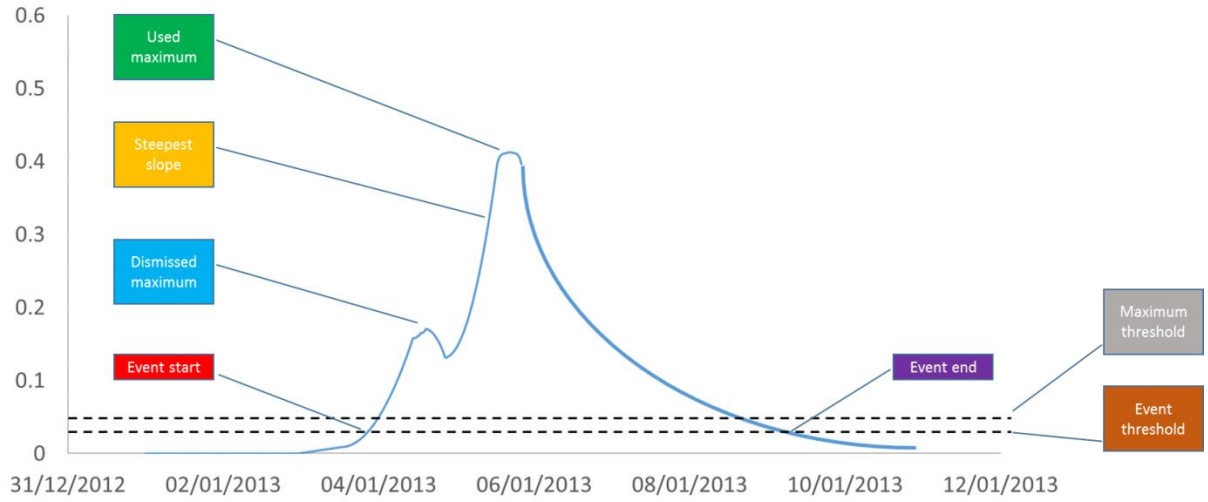


Figure 3.15 - A hypothetical example of an event
There are two dotted threshold lines (event threshold at 0.03 and maximum threshold at 0.05). If the discharge is recorded as greater than maximum threshold the code will search for the first point after maximum discharge that is lower than the event threshold. Next, the code would search the time series for the first value prior to maximum discharge which is greater than the event threshold. The code will then search for the period between the maximum discharge and the start of the event for the steepest slope on the rising limb.

The data points were not identified in sequential order. The first step in the analysis was to normalise the stream baseline. To compensate for change in the measured base flow values during the period of study a baseline correction was applied to the data to set using the "R" package "baseline", which set all the baseline values to zero or values very close to zero. Consequently, if low discharge was 0.01 m^3s^{-1} in the summer but 0.1 m^3s^{-1} in the winter then both of these values would be reduced to zero. Completing the baseline correction enabled a consistent baseflow to be identified within the data. This baseline correction would alter the magnitude of the discharge values so these were not used for any budget calculations. This analysis was generated to determine if the events were triggering any changes in the hydrology of [DOC], pH or EC.

Having established a baseline, the next step in the analysis was to extract the events. The discharge maximum was identified first and all the other parts of the event were identified with respect to this. A discharge value was identified as having a magnitude value of 0.05 m^3s^{-1} or greater. This value was chosen based to be large enough to reduce the likelihood noise being interpreted as an event. Anything smaller than this was discarded as not being a true event. Therefore, if a value greater than 0.03 m^3s^{-1} was identified as being above the event threshold but no event maximum was recorded above 0.05 m^3s^{-1} before the discharge decreased below 0.03 m^3s^{-1} then this data point would be discarded as noise. A second condition for an event maximum being identified was that it was the largest discharge value recorded during the 12 hours duration before or after this

point. If two discharge peaks were recorded within this time frame then the larger of the two was identified as the event maximum and the two were treated as a single event (as is the case in the hypothetical event presented in Figure 3.15).

The second data point identified was the end of the event. The first thing that the code did was to determine the minimum value between two maxima. If the minimum discharge value was equal to or greater than 0.03 m^3s^{-1}, then the time of this minimum value was identified as the end of the event. However, where this value was less than 0.03 m^3s^{-1} then the code would search for the first data point after the maximum that was less than 0.03 m^3s^{-1} and this would be taken as the time of the end of the event. The third value identified was the start of the event. The start of an event would either be the first value that was equal to or greater than 0.03 m^3s^{-1} prior to the maximum discharge value being recorded or the minimum discharge value between two events when baseflow was not reach between two events. The last of the discharge profile data points extracted from the discharge time series is where the increase in discharge during an event was at its most rapid. The point of most rapid increase in the discharge was identified by subtracting each data point in the time series from the one before it (n – (n-1)). Collectively this data was put into a matrix in R such as the following example (Table 3.3). It should be noted that even although the magnitude of these values was also recorded the primary function of the R-code was to identify the timings of the event start, the most rapid increase, the maximum discharge and the end of the event.

Event	Start of event	Largest increase in discharge	Maximum discharge	End of event
1	1206	1255	1289	1406
2	1488	1501	1525	1593
3	1593	1617	1633	1701
…				
n				

Table 3.3 – Example of data points of discharge time series identified and stored
Hypothetical examples of discharge event data from the start to the end of an event. Event 1 starts at data point 1206 and finishes on data point 1406. In real terms 1206 translated to 17 June 2012 at 15:30 and 1406 translated to 21 June 2012 at 19:30.

3.4.2 Wavelets – MODWTs and wavelet coherence analysis

Wavelet analysis was used in this project as a form of exploratory analysis. Wavelet analysis enables the various frequency components of a signal to be identified individually (Graps, 1995). They are used because they provide exact scale based decomposition results (Serroukh et al., 2000). In this project three types of wavelet analysis were used:

maximal overlap discrete wavelet transforms, continuous wavelet analysis and wavelet coherence transforms.

The maximal overlap discrete wavelet transforms (MODWTs) were carried out using the package wmtsa. The MODWT provided a decomposition of a time series to enable different frequency (or periodicity) components to be identified. The decomposition consists of "details" (which are the low frequency components) and a single "smooth" (which is the low frequency component) (Percival and Mofjeld, 1997). The MODWT approach has been used in hydrology papers before (Kallache et al., 2005, Bogner and Kalas, 2008) so there is precedence for their use in this type of research. The MODWTs were applied to the following individual time series: water temperature, pH, SC, discharge and [DOC].

The continuous wavelet transforms (CWTs) and wavelet coherence transforms (WTCs) were calculated using the biwavelet package. Within this, the Morlet wavelet package was selected as this has been used by other researchers in an environmental context and is suggested to be a useful tool for feature extraction (Grinsted et al., 2004). The coherence of a relationship is indicated by the colour coding within the plot at various regions, with the areas in warm colours such as red indicating periods of strong coherence, while the regions highlighted in cooler colours such as blue have weaker coherence (Schaefli et al., 2007). The output of the wavelet coherence plot includes a cone of influence (COI) which exists because the time series is finite in length. The data beneath the COI has a larger error associated with it (Grinsted et al., 2004). CWTs have been applied to the analysis of different types of time series including hydrology datasets (Smith et al., 2004, Lane, 2007). In this project CWTs were applied to the same time series as the MODWTs.

WTCs have also been applied to hydrology datasets (RIVERA et al.). Wavelet coherence analysis is dependent on synchronised data. For example, two time series are recorded with one hour between each measurement – in the first hypothetical signal the data might be recorded at 0 minutes passed each hour (e.g. 01:00, 02:00, 03:00) but in the second hypothetical time series the data was recorded at 30 minutes passed each hour (e.g. 01:30, 02:30, 03:30). To enable synchronisation of the data obtained from the Spectro::lyser™ and MD 9000 TROLL® were rounded to the nearest half hour to generate contemporaneous and equally spaced measurements. WTCs were used to compare the [DOC] with the other variables measured: water temperature, pH, SC and discharge.

Further details of how the MODWTs, CWTs and WTCs were applied to the data sets are discussed in sections 5.2 and 5.3.

3.4.3 Back-casting - Use of bivariate and multivariate analysis to model [DOC]

Back-casting (the opposite of forecasting) uses a model to estimate the value of a variable that was not previously measured by using a model generated from other data that was recorded at that earlier time period. Hydrology sensors have been positioned at Drumtee since 2006 (Waldron et al., 2009), but the Spectro::lyser™ was not deployed until 2010 and so there are several years during which little or no high resolution [DOC] measurements exist at this site, but for which there is abundant half hourly discharge, pH, SC and water temperature data. Bivariate and multivariate statistical models (both of which are explained in the following paragraphs) were generated to determine a suitable model for [DOC]. For both the bivariate and the multivariate models the data used was from 23 May 2012 to 12 December 2013 because there were no significant data gaps in this period (see table 3.1).

The simplest statistical model type is a bivariate model, which relies on just a single variable to predict another variable. A bivariate model was generated between the [DOC] and each of the other variables measured at Drumtee and from this a list was generated of the adjusted R^2 value for each bivariate model.

To allow more flexible regression models to be built, I used “mcgv” to fit a generalised additive model (GAM). The variables used to model against [DOC] were the: pH, SC, month, discharge, the current water temperature, the mean temperature for that day, the mean temperature during the previous 200 hours, the time elapsed since the end of the last event and finally the duration of the current event. The model from this data was generated and the adjusted R^2 was calculated and compared with the bivariate results to determine if the GAM gave a better result.

The generation of the GAM model (in particular the choice of the smoothing) is done automatically in mcgv package (as is the case with most GAM packages). The benefit of automated smoothing is that it enables a relatively complex model to be generated, which was especially useful for the modelling of [DOC] (because no consistent relationship was found between [DOC] and any other individual variable) by means of linear regression by comparing two variables such as [DOC] and discharge. The disadvantage of using GAM models is that they do not have the concise easy to understand summary that would be obtained in the form of a linear equation of the form $y = ax + b$ if standard linear modelling were used instead.

3.4.4 Comparing with the laboratory data

The model that generated the tightest adjusted R^2 value (which as it turned out was the GAM) was then compared with the laboratory data. The model had generated half hourly data from 23 May 2012 to 12 December 2013 and from this the individual points closest in time to the laboratory values were extracted and compared with the laboratory data.

For example, if there was a water sample collected from Drumtee at 28 May 2012 at 12:00 then the data point in the generated model corresponding to 28 May 2012 at 12:00 would be extracted and be put into a new data matrix. Between 23 May 2012 and 12 December 2013 there were 45 water samples collected and therefore 45 comparisons to be generated in the scatter plot.

In some cases, the time was not noted for the collection of the sample at Drumtee. In such cases a time of 12:00 noon was used when comparing the laboratory value against the output of the model. Sometimes the time that a sample was collected at was lost but for older projects it was not necessary to record the time as all sampling was done manually and used to compare a combination of months across different fieldsites around the Whitelee wind farm and therefore knowing the precise time was not necessarily a priority.

Once this was completed the model was applied to all the continuous hydrology data that could be found from Drumtee water – this extended back to 20 November 2008. Therefore, the GAM model was applied to all the hydrology data between 20 November 2008 and 16 December 2014. Once again, the data points that corresponded to the date and time at which a manual sample had been collected from Drumtee water were extracted from this model time series. In this case there were 88 samples in total (including the original 45 data points). Laboratory data collected at earlier dates before this project allowed for a comparison with data that was not used to generate the model. The process of generating a model and testing it against laboratory collected data is summarised below (Figure 3.16)

The benefit of the back-casting is shown as the data from the Spectro::lyser^{TM} can allow us to see high resolution [DOC] data in response to different events over a longer period of time and may be able to provide inside as to whether there has been any change in the [DOC] over a longer period of time without relying intermittent field samples. The data collected directly from the Spectro::lyser^{TM} can be analysed and will have a less

uncertainty that [DOC] time series generated by the GAM model for an earlier time period, but there are two further potential applications that make it valuable to create this model:

1. if GAM or similar models can be established between [DOC] measured by the Spectro::lyser™ and hydrochemistry variables then the Spectro::lyser™ (or other [DOC] sensor) could be moved from one location to another to enable the creation of high resolution spatial time series, when budgetary constraints would otherwise preclude in-situ sensor deployment in each system. For example, if a three-year project involves the monitoring of six fluvial systems within a catchment and there are two [DOC] sensors then these sensors could be positioned within a fluvial catchment for one year and then moved to another site so that each site gets the benefit of the high resolution [DOC] sensor for one year to generate sufficient data to generate a GAM relevant to that particular fieldsite. It is likely that at least one year of data will be required at least in temperate zones to account for the season variation in temperature and biological production of catchment areas.
2. if similar studies are conducted at other sites then it could be determined whether models could be used for catchments of different sizes and/or predominant soil types within different catchments.

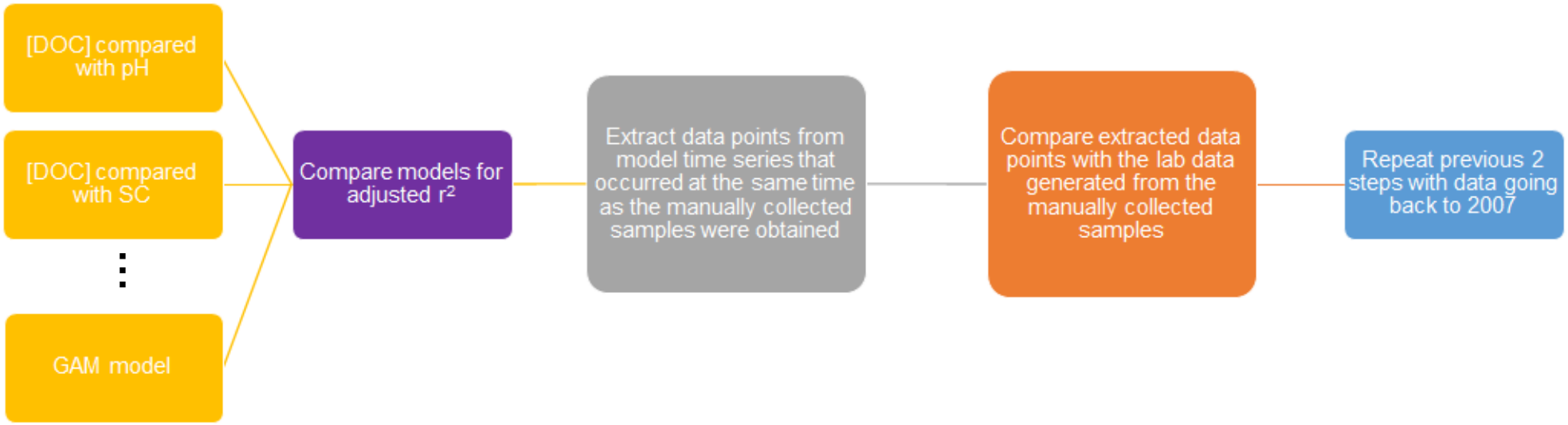


Figure 3.16 – Method for verification of [DOC] models

Comparing model with laboratory data. Flow chart showing process of determining the effectiveness of different models from left to right. First different models are compared against sample data and then the model with the best adjusted R^2 is compared against laboratory data that has been measured independently from the data used to generate the model.

3.5 Raman spectroscopy methods

The Raman chapters (6 and 7) are focussed on methodological development and so most of the method detail is retained in these chapters as needed to understand the research context. For example, the substrates on which experimentation was carried out and the focussing of the lasers is discussed in greater detail within those chapters. Here only a few brief points are discussed regarding the underlying methodology.

3.5.1 Equipment

Raman spectroscopy was conducted under laboratory conditions using both 514 nm and 785 nm lasers on an In-Via Raman spectrometer from Renishaw. The Raman spectrometer is situated in a secure, air-conditioned laboratory that is temperature controlled to 22°C. Although variations of a couple degrees are to be expected this enables confidence that the laboratory conditions during summer and winter months are not significantly different and therefore unlikely to affect the Raman results. Initial experimentation revealed that the 785 nm was the most effective laser for most of the samples being analysed in this research due to producing lower fluorescence (Vítek et al., 2012).

Samples were prepared and analysed using both solid samples and solutions. Operation of the Raman spectrometer was simple. There were five lenses available:

1. 5 x objective lens
2. 20 x objective lens
3. 50 x objective lens
4. 50 long x objective lens
5. 100 x objective lens

For my experiments, the 50 long x objective lens was used as this enabled a strong optical zoom for targeting the solid particles, but also could be positioned a couple mm above where the sample was positioned. The 20x, 50x and 100x objective lenses had to be positioned so close to the target sample that on occasion they would come into contact with the solution being analysed which could potentially deposit water, dirt or other substances onto the lens. Using the 5x or 50 long x objective lenses enabled the focussing of the laser without risking the objective lens. The spectrometer is equipped with both an eye piece and a camera. During analysis, the camera was always used for focussing the lens. The samples sit on top of a plate that can be controlled from the computer either up/down, forward/back or left/right.

As discussed in the literature review experimentation was conducted using both standard Raman and surface enhanced Raman spectroscopy (SERS). The former was used to determine if we could obtain signals from the target carbon materials without enhancement and the SERS was used to determine if it was possible to obtain enhanced signals and thereby detect carbon at lower concentrations than would otherwise be possible. There are three main settings that could be varied on the Raman spectrometer:

1. laser power. The laser power could be set as high as 300 mW. Controlled using the computer, this could be reduced using a pre-set list of percentages. Liquids and non-volatile solids (e.g. sugars and amino acids) were analysed using standard Raman where the full 300 mW was used. However, when conducting SERS analysis, the power was reduced to 5% (15 mW) because settings more powerful than this would burn the substrate and this would result in erroneous Raman spectra measuring burned substances such as charcoal. The powdered humic substances were volatile and therefore power settings of 1% (3 mW) were used to avoid the incineration likely with higher power readings. Fortunately, in solution even the humic acids and fulvic acids could be analysed with the 300 mW power settings and this meant that clearer spectra with superior signal to noise ratio (SNR) could be obtained.
2. length of exposure time. The length of time that the laser was focussed continuously on the target. Longer exposure times can enable a clearer spectra to be obtained (Siesler, 2009). Using a longer exposure time is similar to a camera that uses a longer exposure to collect more photons. Unfortunately, this can sometimes lead to the overheating and damage of a sample.
3. number of accumulations. Accumulations refers to the number of times the laser is shone on a target for the exposure time mentioned above. As with increasing exposure time increasing the number of accumulations can also improve SNR but is less likely to lead to overheating.

The relevant power, exposure times and the number of accumulation settings for different experiments are discussed in chapters 6 and 7. In this project I measured signals using both Stokes Raman scattering and anti-Stokes Raman scattering. The physical distinction between these two types of measurement is discussed in the literature review but in practical terms the Stokes spectra is measuring positive valued wavenumbers and the anti-Stokes spectra are measuring negative valued wavenumbers.

3.5.2 Sample preparation

Raman analysis was conducted on samples obtained from the following sources: Drumtee water; humic and fulvic acids from the Suwannee river, obtained from the International

Humic Substance Society (IHSS); three types of sugar (glucose, fructose and sucrose) from Sigma Aldrich and four types of amino acid (glycine, tyrosine, tryptophan and phenylalanine) also from Sigma Aldrich. The individual compounds were chosen because they were considered likely to represent the individual components of DOM (Leenheer and Croué, 2003). During analysis both solid phase and solutions were tested. Glucose and sucrose produced clear Raman spectra when in high concentration solutions and so the limit of detection for these samples was measured.

Samples of stream water were analysed in concentrations found in the stream (5 – 60 mg/l) and as a direct comparison with the humic and fulvic acids purchased from the IHSS a powder was extracted from the Drumtee water catchment by taking a sample obtained on 14 November 2013 and this was concentrated using rotary evaporation and then freeze dried to remove the water but not the DOC itself, leaving a powder.

The experiments conducted were to identify the anti-Stokes bands generated using the powders of various substances (humic acid, fulvic acid, glucose, fructose, sucrose, phenylalanine, glycine, tryptophan and tyrosine). Subsequently limit of detection tests were conducted on those substances which produced visible anti-Stokes Raman spectral bands.

For solutions of humic and fulvic acids the starting concentration was 10000 mg/l. More concentrated solutions would be ideal but since clear Raman spectra were obtained at this concentration and only a small quantity of humic and fulvic acids were available for testing the quantity of material used in any experiment was minimised. Concentrations were diluted until the signals could no longer realistically be extracted from the background noise and signals.

3.5.3 Measuring humic and fulvic acids

When the signals corresponding to the humic and fulvic acids were established then an experiment was set up to determine if it could be determined using the spectra whether a solution contained proportionally more humic acid or fulvic acid. Solutions with varying proportions of humic acid and fulvic acid were set up in order to establish whether or not the relative intensity of these peaks differs between humic and fulvic acids. Solutions were prepared at a target concentration of 10,000 mg/l using different proportions of humic and fulvic acids in 10% increments. For example, solution 1 would contain 100 % humic acid and 0 % fulvic acid, whereas solution 11 would contain 0 % humic acid and 100 % fulvic acid and solution 6 would be 50% humic acid and 50% fulvic acid.

4.0 – High resolution [DOC] and hydrochemistry time series

4.1 Abstract of chapter

Here I present the time series and the first-stage analysis of data collected from high-resolution hydrochemistry sensors deployed at Drumtee water between 23 May 2012 and 16 December 2014. The hydrological variables measured were [DOC], specific conductivity (SC), pH, water temperature and water discharge. As the implementation of in-situ sensors for measuring [DOC] using absorbance spectroscopy is still a relatively new technique, [DOC] was also measured in the laboratory to confirm the accuracy of the Spectro::lyser™ data. The understanding that these sensors offer into catchment hydrology is discussed. During events pH and SC both decreased, while the response of [DOC] was more complex and inconsistent. These variables exhibited a high degree of co-dependence with regards to stimulus from events. In tandem with the presentation of results is a discussion about how the data is analysed.

Using the data from the field-deployed sensors the distribution properties of the different variables are presented. The discharge was most regularly at values less than 1 m^3s^{-1} and the [DOC] was most regularly measured between 15 mg/l and 25 mg/l. Separate budgets for DOC export were calculated using the Spectro::lyser™ time series and the laboratory 'snapshot' data, using matrices of the data and a commonly-used analytical method. No reliable method was identified for producing a [POC] time series and therefore only laboratory measured values were used for budgets, with either [DOC] data from the laboratory or the Spectro::lyser™.

A statistical analysis of the quantitative data produced was necessary to develop an understanding and interpretation of the fluvial dynamics of the hydrology in the Drumtee catchment. The stream generally exhibited complex hydrological relationships that were difficult to characterise by using simple techniques such as linear regression. In particular the large datasets produced by the field deployed sensors (such as the Spectro::lyser™ and TROLL®) revealed trends occurring at short time scales that could not be explained through linear regression after plotting data points on a scatter plot. Consequently, a GAM model was found to be a better representation of the data.

4.2 Introduction

The export of [DOC] from peatland catchments represents a significant and important source of C to the global carbon cycle (Limpens et al., 2008). Concerns have been raised over the destabilisation of peatlands and how this destabilisation may lead to these stores of C being depleted and exported when peatlands are disturbed (Freeman et al., 2004, Armstrong et al., 2010, Moore et al., 2013). Increased development of wind farms in recent years in Scotland has led to increased monitoring of C exports to determine the impact on these sites (Grieve and Gilvear, 2008, Waldron et al., 2009).

The development of sensors that can measure [DOC] in the field represents an opportunity to create a more detailed picture of carbon concentrations and exports than has been possible using traditional sampling methods. The volume of data generated from high-resolution sensors provides an opportunity to develop a better understanding of catchments and event dynamics within those catchments. Field-deployed, high-resolution sensors for monitoring other variables in hydrological systems have been shown to be advantageous in terms of generating a better understanding of the systems being studied. (Kirchner et al., 2004, Neal et al., 2012). However, this volume of data also provides an analytical challenge in that the quantity of data is more difficult to manage and to making full use of this data requires the use of more sophisticated statistical tools than may be required when analysing more limited data sets.

The over-arching aim of this part of the research was to deploy an available [DOC] sensor, the Spectro::lyser^{TM}, in a field site where it is important to gain a better understanding of C exports – a site that was previously subject to disturbance and fluvial concentrations of [DOC] were known to be high (e.g. Waldron et al, 2009; Murray, 2012), but where fine detail of C export was missing. Considered likely to be particularly revealing would be hydrochemical response to event flow as this could reveal key responses through considering: how did the [DOC] vary during events; how much organic carbon was being exported during events, relative to the intensity of events and how did the other hydrochemical time series respond?

Further, as the use of [DOC] sensors is not commonplace, a secondary aims were to: to compare a budget using data collected from traditional manual sampling with a budget constructed using the high-resolution data from the Spectro::lyser^{TM}; to explore the temporal resolution needed to generate a representative budget and to explore how the [DOC] time series covaried with other components of catchment hydrochemistry that are measured more routinely and could act as a proxy for DOC export.

The rationale for the last aim is that at Drumtee the discharge, pH, SC and water temperature data have been collected for longer than the Spectro::lyser™ has been collecting [DOC] data. The [DOC] data was collected using the Spectro::lyser™ intermittently between October 2010 and April 2012, but became fully operational for long periods of time from May 2012 onwards. If a model can be created of the [DOC] data based on the other hydrological variables, then the [DOC] time series can be extended backwards in time (a technique called backcasting) to 2007. A further motivation for this is that real-time [DOC] sensors are expensive and there may be more field sites in a particular research group than there are sensors available. Therefore, the [DOC] sensor could be used for a limited time until a strong relationship was established and a model could be generated based on hydrological variables that are cheaper to measure long term, enabling the high-resolution [DOC] sensor to be moved to a different site.

However, although shelf-bought sensors exist, they are still not commonplace and thus in generating the time series, prior to advanced analytics of the data, first the sensor accuracy had to be considered through comparison with [DOC] collected via manual sampling. If there were differences, it had to be established if this was result of the sampling rate, or a technological problem with the sensor.

4.3 Methods

The field description and related information is given in chapter 3. Here I focus on the production of the data presented in this chapter, including validation of the high-resolution time series and the following tasks:

1. Aligning all the data recorded by sensors so that it could be synchronised for analytical purposes
2. Gap-filling data in an appropriate manner where possible, or alternately stating where gaps in the data cannot be filled.
3. Calculating carbon budgets for Drumtee.
4. Exploring the use of the hydrochemistry data as a proxy for [DOC] reconstruction through statistical modelling using other measured variables at Drumtee water.

4.3.1 Alignment of data

The data collected by the Spectro::lyser™, TROLL®, ISCO flow logger and the data from Newmilns SEPA station is measured at 15 or 30 minute intervals. The Spectro::lyser™ collects data at 30 minute intervals to preserve battery life and consideration of the time series shows the variation at this time interval is usually very small and therefore

significant details in variation are unlikely to be missed. The ISCO data and the data from Newmilns are both recording at 15 minute intervals. Visual inspection of the time series showed little change at the 30 minute interval. Therefore, to align the data sets, the data at 0 minutes and 30 minutes past the hour was selected and the data at 15 minutes and 45 minutes past the hour was filtered out. Filtering the data in this way meant that all the recorded time series were now showing data with 30 minute intervals, making comparative analysis between the data sets easier.

Sometimes the loggers were not started at 0,15, 30 or 45 minutes past the hour. Thus, the second stage of the alignment was achieved by rounding of each data point to the nearest half hour, for example 09:17 was moved to represent the 09:30. Aligning the data in this way introduces a minor, but unlikely to be significant, error into the data as there is little variation at the 15 minute interval (differences occur at the hourly scale) and data is only offset by up to 15 minutes. Shifting the data allowed the different time series to be more easily analysed (and this was also necessary for wavelet analysis, chapter 5).

4.3.2 Calculation of C export

The [DOC], [POC] and discharge data collected were used to generate carbon budgets for the Drumtee catchment for the 2012/13 and 2013/14 hydrological years using two different methods. The first made use of all the high-resolution [DOC] data and the second used the less frequent laboratory-measured samples. There was no high-resolution data for [POC], and in 2 out of 27 instances in 2012/13 no [POC] data were generated as the field samples used were collected by a fellow student from Edinburgh University (Phin, 2015) and arrived pre-filtered. Two different methods were used to calculate carbon budgets as described below.

Equation 5

The first method (Equation 5) calculated the sum product of the total organic carbon and the discharge across a hydrological year (1 October at 00:00 to 30 September at 23:30) and this was calculated for the hydrological years 2012/13 and 2013/14. This time period equates to 17520 data points spaced at 30 minute intervals. The [DOC] and [POC] time series values recorded at 30 minute intervals were added together and multiplied by the discharge. Discharge was multiplied by 1800 prior to use in this equation because discharge in m^3s^{-1} need to be scaled up to the half hour period (1800 seconds). Similarly, [DOC] and [POC] are converted from mg/l to $g.m^{-3}$, but as this is done by multiplying by 1,000 to convert from litres to m^3, and by dividing by 1,000 to convert from mg to kg the numerical values of the [DOC] and [POC] remain the same. Finally, this quantity is divided by the area of the catchment to give the total kg of carbon exported per m^2 of the catchment.

$$Organic\ C\ budget = \sum\left(([DOC] + [POC]) \times Discharge\right)$$ Equation 5

Equation 6

The second method (Equation 6) calculated the budget using the mean discharge in conjunction with the individual concentration and discharge measurements relative to the total time period measured (Walling and Webb, 1985), an approach common in other studies (Clark et al., 2007a, Pawson et al., 2008, Stutter et al., 2008):

$$F = K \cdot Q_r \left(\frac{\sum_{i=1}^{n} Q_i C_i}{\sum_{I=1}^{n} Q_i}\right)$$ Equation 6

Here K refers to the number of data points in the discharge time series (in this case each data point represents 30 minutes so equates to 8760 hours); Q_r refers to the mean discharge measured over the period from 25 September 2012 to 4 October 2013 in m^3s^{-1}; C_i represents the TOC measured from samples returned to the laboratory and Q_i represents the discharge at the same as the water sample was taken. Unlike equation 5, an adjustment has to be made for time series being longer than 17520 data points because sampling did not start and end the exact beginning and end of the hydrological year. For 2012/13 the value produced by equation 6 is multiplied by 17520 and divided by 17943. For 2013/14 the value produced by equation 6 is multiplied by 17520 and divided by 17856.

To assess what would be the minimum number of [DOC] and [POC] data that would be needed to construct a representative budget the outputs of both methods were calculated using sequentially fewer data points to determine how the estimates of export would alter.

4.3.3 Backcasting to reconstruct a longer [DOC] time series

Two methods of modelling were attempted: bivariate and multivariate analysis. Bivariate analysis was conducted by creating a scatter plot between [DOC] and each of the other variables measured at Drumtee. Bivariate analysis is a simple method of modelling and would enable the easiest method of calculating [DOC]. The adjusted R^2 value produced between [DOC] and each variable was noted.

The tool selected for multivariate analysis was a generalised additive model (GAM), with the R package “mcgv” selected as the tool to analyse and generate GAMs of the data. The GAM model was generated using the following variables: water temperature; mean daily water temperature; water temperature over the previous 200 hours; pH; SC; month of the year; discharge; the time since the previous event and the time since the start of the

current event. These values were chosen to capture catchment productivity as season will be broadly captured by temperature, and then catchment hydrological response.

Using all the hydrochemistry data obtained between 2012 and 2014 the [DOC] was modelled against all available data for the different variables mentioned. The data calculated using this model was then compared against [DOC] measurements made in the laboratory that had been made using the HTCO technology, giving an independent method of testing the accuracy of the model (as the laboratory data was not a component used in building the model). To test whether the model could be extended back to the period between 2007 and April 2012 the outputs of the model were compared against laboratory data measured for earlier studies at this site (Waldron et al., 2009, Murray, 2012). Using these older results increased the number of samples tested from 58 to 88. In some cases, there is no specific time noted for the collection of the sample at Drumtee. In such cases a time of 12:00 noon was used when comparing the laboratory value against the output of the model as this was typical of the time field work took place after travelling from Glasgow.

4.4 Results

The time series generated between 23 May 2012 and 14 December 2014 by the field-deployed sensors and laboratory measurements are show in Figure 4.1. Where there were data gaps of more than a few hours these are shown as grey areas but some of these have been filled in where there were other sources of data available:

a) some of the temperature data has been gapped filled using the relationship described in chapter 3 – methods with data obtained from Glasgow airport.
b) Missing SC data was modelled based on a relationship with pH.
c) [DOC] has been gap filled by fitting a straight line between the last recorded data point before the missing data starts and the first data point after the missing data finishes

Each time series is now discussed in detail and key characteristics are summarised (Table 4.1) to support this. These were generated using simple summary statistic analytical tools and these allow the data to be compared with other studies and to establish baseline values for analytical purposes.

	Min.	Max.	Mean	SD	median	n
Discharge (m^3s^{-1})	0.001	5.3	0.23	0.36	0.11	46031
Temperature ˚C	-0.2	25.8	9.5	5.2	9.8	39544
pH	4.5	8.8	6.9	0.7	7.0	39543
SC ($\mu S.m^{-1}$)	14.9	246.5	68.0	40.7	55.0	37624
[DOC] (mg/l) (Spectro::lyser™)	8.0	55.7	28.2	10.5	26.9	38973
[DOC] (mg/l) (Laboratory measurements)	10.2	81.1	30.9	11.9	28.9	70
[POC] (mg/l)	0.5	26.2	1.6	3.6	1.0	61

Table 4.1 – Summary of results
Summary values for the relevant variables discussed for data sets collected during the period from 23 May 2012 to 16 December 2014. The [DOC] generated by the Spectro::lyser™ and manual sampling are given separately. There are 70 [DOC] samples measured and 61 [POC] samples because the data was collected for more than one project; therefore [POC] data is collected for most but not all samples. The discharge data has more data points because this was not affected by the extended periods where the TROLL® did not collect data.

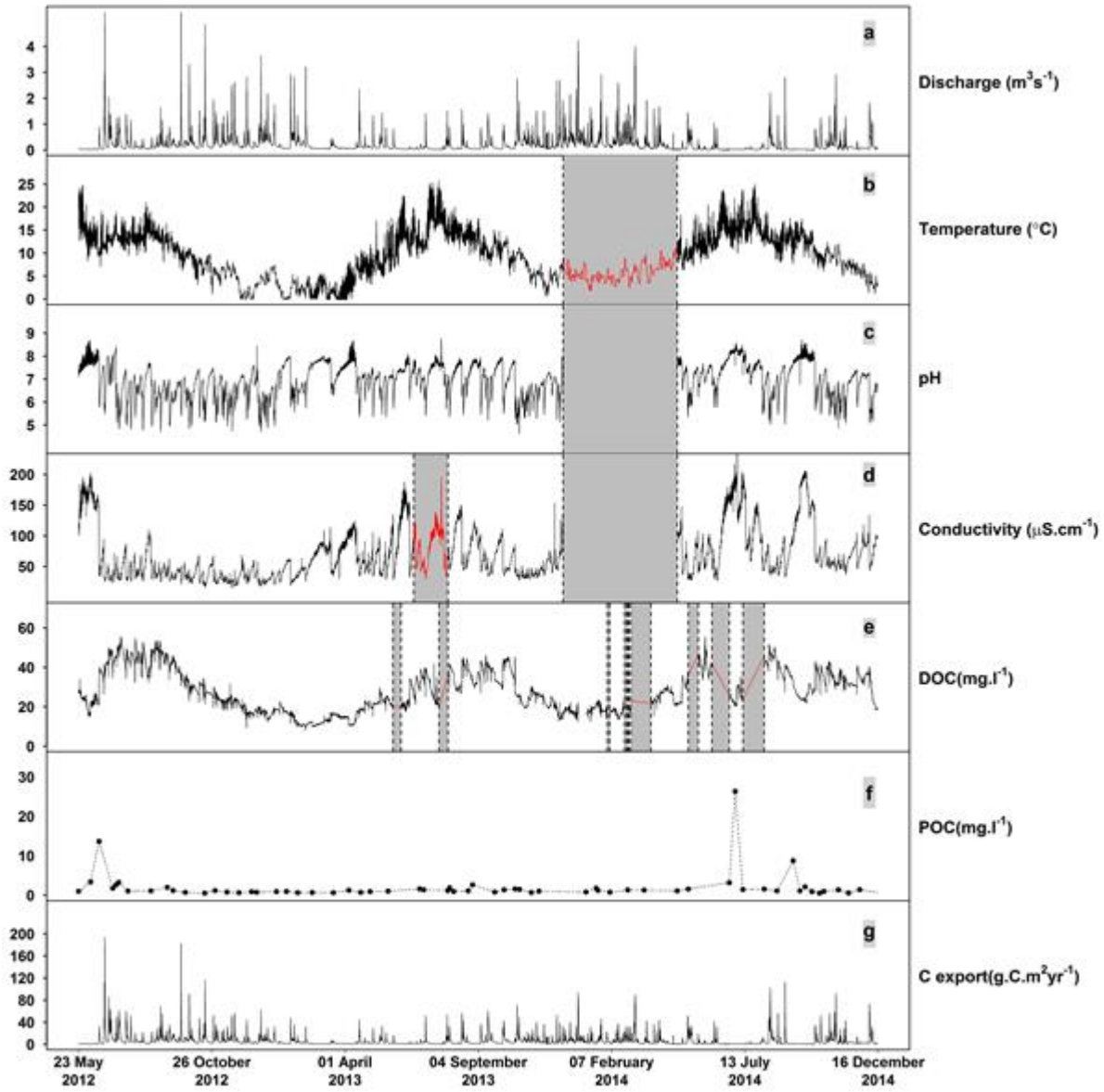


Figure 4.1 – Compiled time series
The time series for (a) discharge (as shown in Figure 3.10); (b) water temperature; (c) pH; (d) specific conductivity (with data gap filled using relationship with pH data); (e) [DOC] generated by the Spectro::lyser™; (f) [POC] and (g) C export for the period 23 May 2012 till 21 December 2014. Data gaps are highlighted with grey boxes. The [POC] time series consists of discrete points from manual sampling. The large data gap from 12 December 2013 to 25 April 2014 occurred due the water leakage sustained by the TROLL® originally positioned here. The C export value is calculated using the discharge value; the Spectro::lyser™ [DOC] data and the [POC] data. Gap-filled data is shown in red, within the grey boxes for (b), (c), (d) and (e).

4.4.1 Discharge

Discharge recorded at Drumtee ranged between 0.001 m^3s^{-1} and 5.3 m^3s^{-1} with a mean value of 0.23 m^3s^{-1} and a standard deviation of 0.36 m^3s^{-1}. The median value was 0.11 m^3s^{-1}. The mean value of discharge is 2.5 times greater than the median value and this indicates that high discharge values occurred less frequently but also had a proportionally greater influence on the discharge mean than did the low discharge values. The flow duration curve (Figure 4.2) illustrates that the largest 10% of flows account for discharge values between 0.5 m^3s^{-1} and 5.3 m^3s^{-1}. It is important to consider the range of discharge values as the largest 10% of flows also accounted for 44.2% of DOC exports, which fits in with other studies that show the largest flows correspond to the largest exports (Raymond and Saiers, 2010).

As seen in Figure 4.1 the summer of 2012, at the end of the 2011/12 hydrological year (HY2011/12), defined as 00:00 on 1 October 2011 to 23:30 on 30 September 23:30, was notable for a relatively large number of events. The only period in the HY2011/12 year considered in this project is 23 May 2012 to 30 September 2012 as this coincides with the collection of the [DOC] data. Of the 2262 hours where discharge was ≥ 0.547 m^3s^{-1} 9.4% of these instances were in the period from 23 May 2012 to 30 September 2012. In contrast during the equivalent time period (23 May to 30 September) in the HY2012/13 and HY2013/14 the percentage of discharge values falling into the top 10% of discharge values (≥ 0.547 m^3s^{-1}) was only 3.4% and 3.2% respectively. Furthermore, of the top 1% of discharge values, 22.3% of these occurred during the period between 23 May 2012 and 30 September 2012. During the equivalent periods in HY2012/13 and HY2013/14 the percentage of discharge values in the top 1% of discharge values (≥ 2.57 m^3s^{-1}) were 0% and 2.2% respectively.

Generally, the winter months (October through March) coincided with greater discharge values than the summer months (April through September). In total, 14.0% of the largest 10% of discharge values occurred during the winter period (October through March) during the HY2012/13. Similarly, 18.5% of the largest 10% of discharge values occurred during the winter period during the HY2013/14. Additionally, 22.3% and 40.8% of the largest 1% of discharge values recorded during the period of study occurred during the HY2012/13 and HY2013/14 winter months respectively.

Altogether the wet summer of the HY2011/12 provides an interesting contrast with the relatively drier summers of HY2012/13 and HY2013/14. The summer of HY2011/12 is more comparable in terms of discharge values to the winter months of HY2012/13 and HY2013/14 than it is to the summer months of HY2012/13 and HY2013/14. There is more

missing data in the summer months of HY2013/14 than the two earlier years, making it more challenging to offer a more definitive comparison.

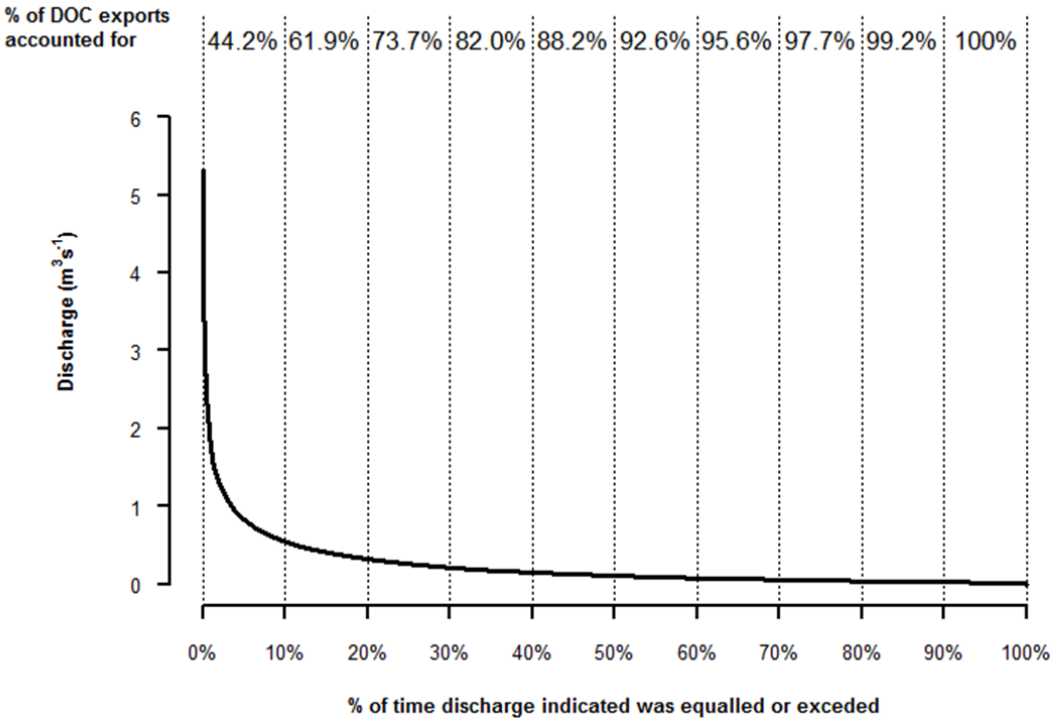


Figure 4.2 – Flow duration curve
Flow profile indicating the % of time that a particular discharge rate was equalled or exceeded, with flow ranging between 0.001 m^3s^{-1} and 5.3 m^3s^{-1}. Vertical lines segregate the graph into 10% segments and the % of DOC export in each segment is given by the % value at the top of the graph. For example, the top 10% of flows (values equal to or greater than 0.514 m^3s^{-1}) account for 44.2 % of the TOC exports.

4.4.2 Water temperature

Fluvial water temperature ranged from -0.2 °C to 25.8 °C between 23 May 2012 and 16 December 2014. The water temperature has a strong seasonal component to it, which is to be expected given the field site location in a temporal region in Scotland. The highest monthly average temperature recorded between May 2012 and December 2014 (Figure 4.3, Table 4.1) was in May 2012, although only data between 23 May 2012 and 31 May 2012 was taken into account for this period (to coincide with the Spectro::lyser™ data) and therefore the warm period of weather may have resulted in the calculation of a mean temperature that is not entirely representative of the full month of May 2012. In both 2013 and 2014 the largest mean temperature was recorded in July. In 2012 July was measured to have the third largest mean temperature.

In addition to the mean daily temperature the mean diurnal variation during each month has been calculated. The diurnal temperature of the water was more prominent during

warmer months than cooler months (Figure 4.4). The largest mean diurnal variations measured in 2012, 2013 and 2014 were in May, July and July respectively, which are also the three months during which the largest mean temperatures were recorded (Figure 4.3). As shown in Figure 4.5 is the comparison between these two figures, which shows the relationship between the diurnal variation and the average temperature. As shown the general relationship is that the greater the mean monthly temperature then the greater the diurnal variation.

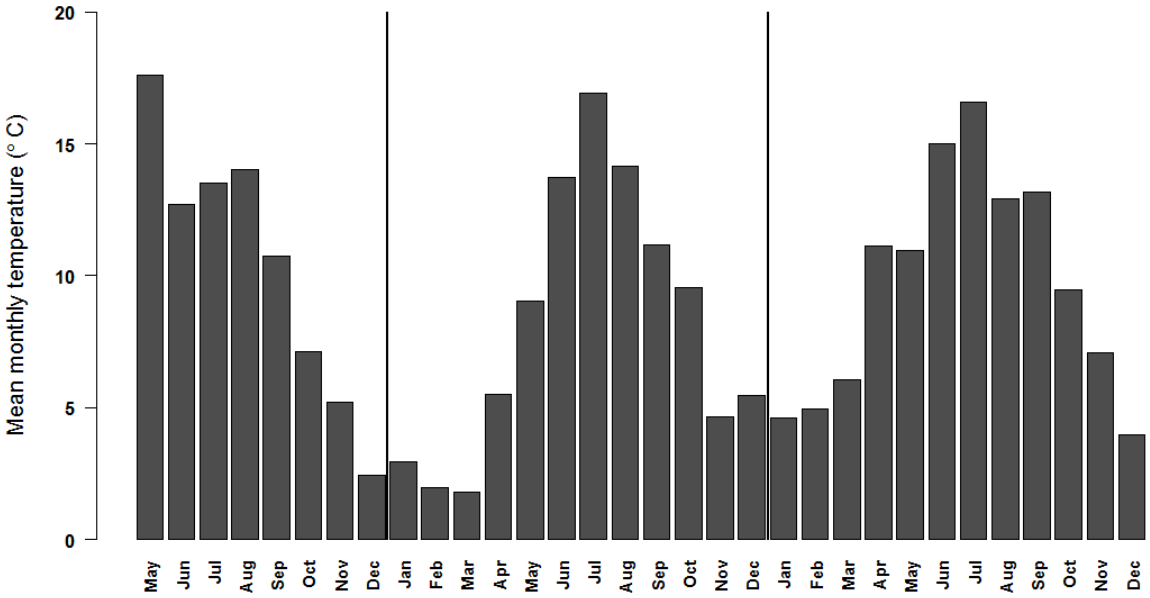


Figure 4.3 The mean monthly water temperature at Drumtee
The mean monthly water temperature (ranged from 1.8 °C to 17.6 °C) from the data recorded at Drumtee for years 2012, 2013 and 2014. January, February and March in 2014 year were reconstructed from Glasgow airport temperature data.

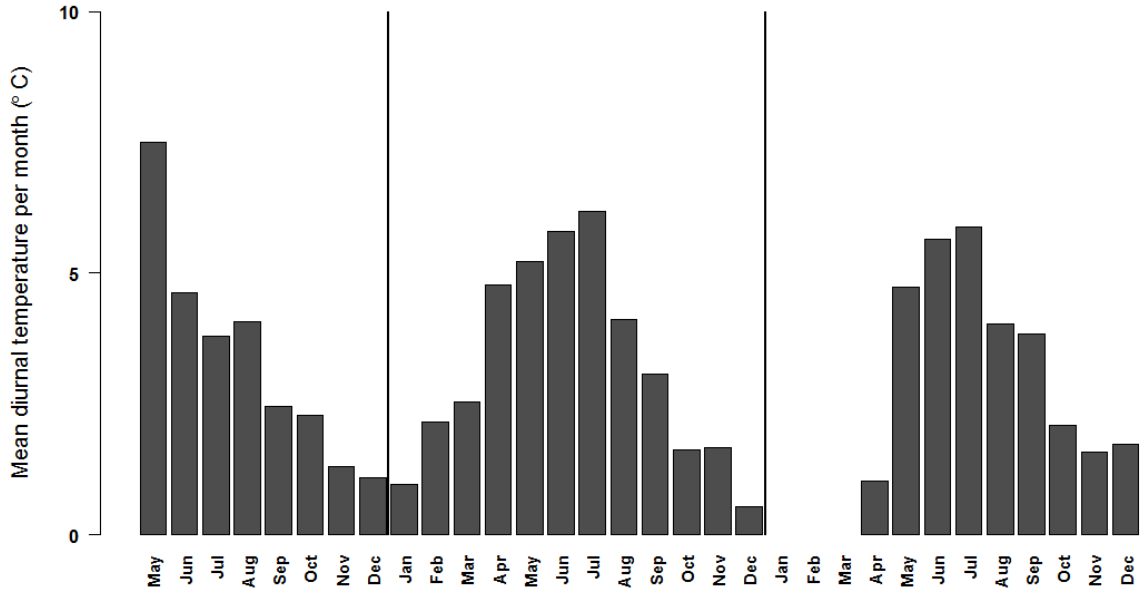


Figure 4.4 - The mean diurnal variation in temperature
Mean diurnal water temperature (ranged from 0.5 °C to 7.5 °C) from the data recorded at Drumtee for the years 2012, 2013 and 2014. Unlike Figure 4.2 the mean diurnal variation in temperature could not be calculated because the model was between the mean values – relationships were tested between minimum stream temperature at Drumtee and minimum air temperature at Glasgow airport but the relationship was poorer.

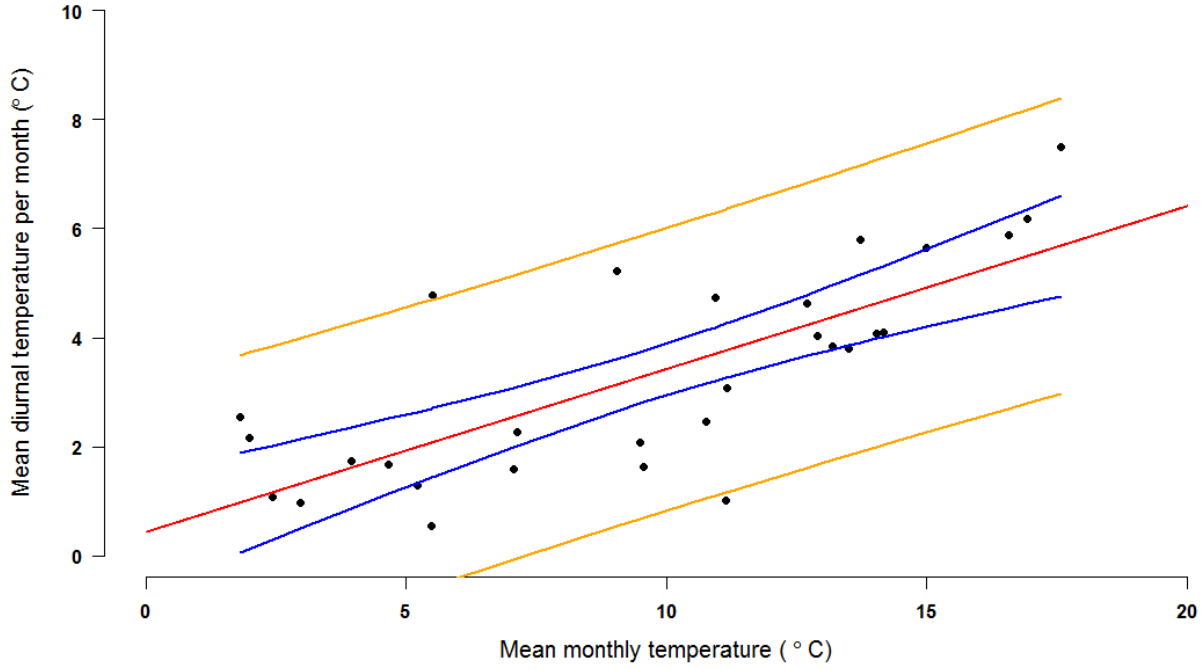


Figure 4.5 – Mean monthly temperature
The data from Figure 4.3 and 4.4 is compared to indicate that warmer months also have greater associated diurnal temperature variation. The model line is shown in red, with the 95% prediction intervals shown in blue, with the 95% confidence intervals shown in orange.

4.4.3 pH

The pH ranged from 4.5 to 8.8 between 23 May 2012 and 16 December 2014. The largest pH values typically occurred during periods of lower discharge but would decrease rapidly during an event. During an event, the pH decreases to a minimum value and subsequently increases. In most instances, the pH continues to increase until the next event (Figure 4.1) when the pH value recorded decreases again. After extended periods without an event the pH, when above a value close to or greater than 7 would tend to exhibit a diurnal variation, consistent with photosynthetic/respiratory variation identified in other studies (Simonsen and Harremoës, 1978, De Montety et al., 2011).

There is no obvious seasonal variation in the pH data. However, if during a particular month or season there are fewer events then this will influence the pH profile, with fewer decreases observed in the pH profile and a greater proportion of the time when pH is closer to 7 than to the pH 5 to pH 6 that is regularly measured after an event. Therefore, it may be anticipated that during the summer months the pH would be recorded at values that were close to neutral pH and that during the winter there would be a greater number of instances where the pH would be more strongly acidic. This pattern of higher pH in the summer than in the winter was the case for the 2013/2014 year but this was not the case starting from May 2012 through to the end of the 2012/2013 hydrological year, where the winter was relatively dry and during the summer there was reasonably significant rainfall as indicated by Figure 4.1 of the stream hydrograph.

4.4.4 SC

In the period between 23 May 2012 and 16 December 2014 SC ranged between 14.9 $\mu S.m^{-1}$ (19/11/2012) and 246.5 $\mu S.m^{-1}$ (22/07/2013). The SC data recorded increased between events and decreased rapidly during an event (Figure 4.1). The minimum SC recorded during an event was a median of 31.8 $\mu S.m^{-1}$ with a standard deviation of ± 14.2 $\mu S.m^{-1}$. During the summer months, which coincided with higher water temperatures, the rate of SC increase seemed to be greater than for an equivalent period during the winter. The response in SC to an event is similar to the response in pH: during an event, the SC decreased is a short period of time before slowly increasing. The data indicates that the rate of increase of SC between events will slow as the time period between events increases. However, whereas pH tends to settle between a pH value of 7 to 8 post-event, the SC data recorded shows a continued increase between events (although it does so at a reduced rate). However, during most events the SC would decrease to a value of about 50 $\mu S.m^{-1}$.

4.4.5 [DOC]

[DOC] measured by the Spectro::lyser™ between 23 May 2012 and 16 December 2014 ranged between 8.0 mg/l and 55.7 mg/l with a mean value of 28.0 mg/l and a median value of 26.5 mg/l (n=44973). In the same period, the [DOC] measured in the laboratory ranged between 10.2 mg/l and 81.1.1 mg/l with a mean value of 30.9 mg/l and a median value of
28.9 mg/l and a standard deviation of ± 11.9 mg/l (n=70). (Both sets of values are summarised in table 4.1). Both of the time series produced by these measurements indicate seasonal variation in the [DOC]. The data collected using the Spectro::lyser™ indicated that [DOC] was regularly observed to decrease at, or just after, the start of an event, with a subsequent increase to a maximum value. However, this decrease and subsequent increase did not occur during every event and therefore it is explored in section 4.7.

The summary data (Table 4.1 and Figure 3.10) shows that [DOC] measured in the laboratory is similar to the [DOC] measured by the Spectro::lyser™. However, the Spectro::lyser™ does not seem to have picked up on the largest [DOC] value as measured by the HTCO method in the laboratory. The means and medians of both time series may be higher than if there were two full years of data as the time series used 7523 hours during this time period between April and September and only 6108.5 hours during this time period between October and March. Therefore, more of the time series data was taken during the summer than the winter, which generally has lower [DOC].

In the [DOC] time series measured by the Spectro::lyser™ there was an unusual decrease on 5 April 2013 (Figure 4.6). This decrease coincides with the Spectro::lyser™ being cleaned and after cleaning the data is more reliable. However, the data prior to this cannot simply be corrected by introducing an offset. In the section of data prior to this, the field-collected and laboratory analysed water sample (17.4 mg/l), compares well with the Spectro::lyser™ estimate of 16.3 mg/l. Therefore, as it is unknown when the data has begun to drift no correction has been applied to the data because this may introduce an artefact into the data that is no more accurate than the original data.

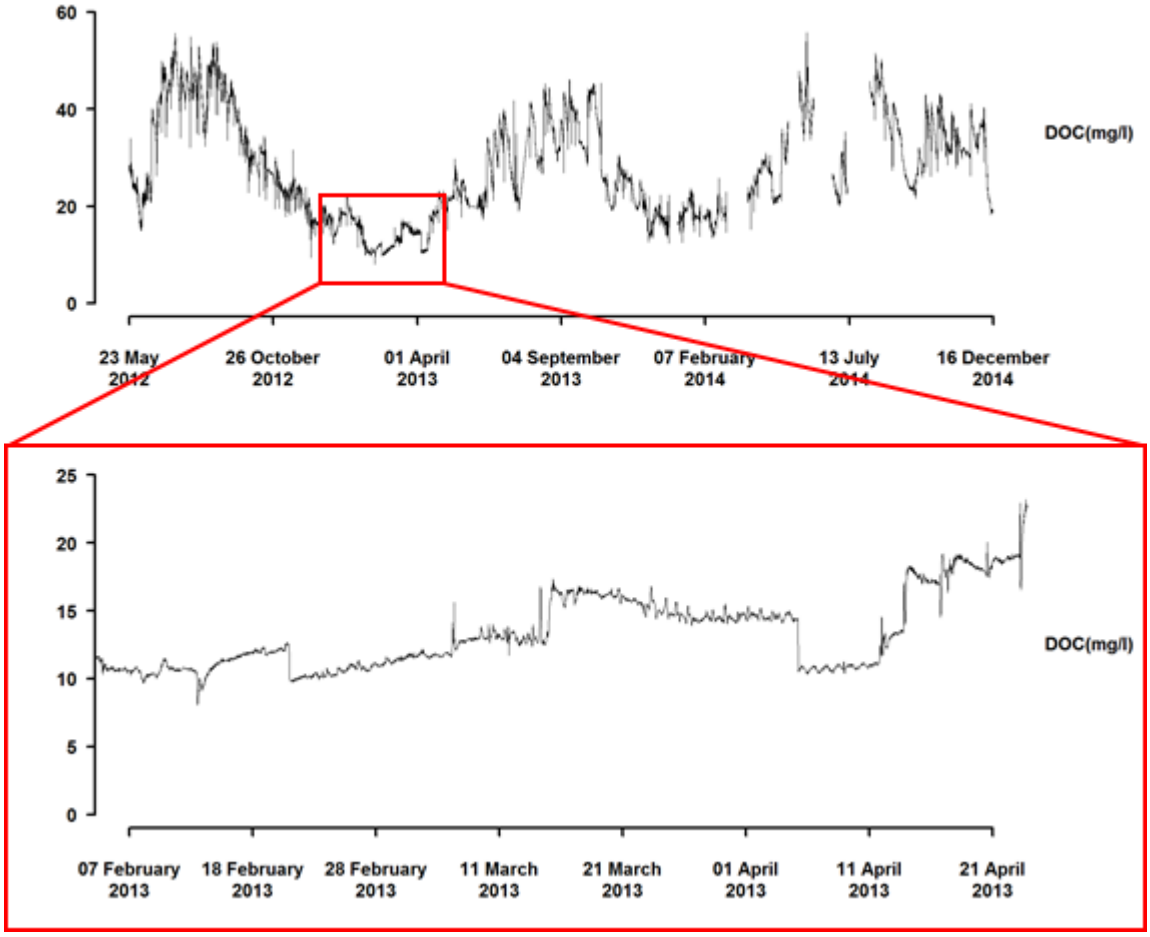


Figure 4.6 – [DOC] profile including unusual drop in [DOC]
The upper graph shows the full [DOC] profile between 23 May 2012 and 16 December 2014. The bottom graph is a zoomed in region of this graph that shows an unusual series of increases and decreases in [DOC].

The [DOC] was plotted against the discharge using a log-log scale (Figure 4.7). The plot observed is not typical of a relationship between discharge and [DOC], with the relationship producing an adjusted R^2 value of 0.0001 and a p-value of 0.032. In this instance, the low p-value is effectively driven by the high volume of data as the low R^2 value indicates that this model has little to no predictive power. It was expected that the [DOC] would generally increase in line with discharge. However, when the 18 month time series used to produce Figure 4.7 was divided into 33 sections of 250 hours each the median p-value for the 250 hour segments was 4.1×10^{-16} with a standard deviation of ± 0.08. Splitting the time series into 250 hour segments removes the seasonal component of the [DOC] and therefore the [DOC] does increase as the discharge increases.

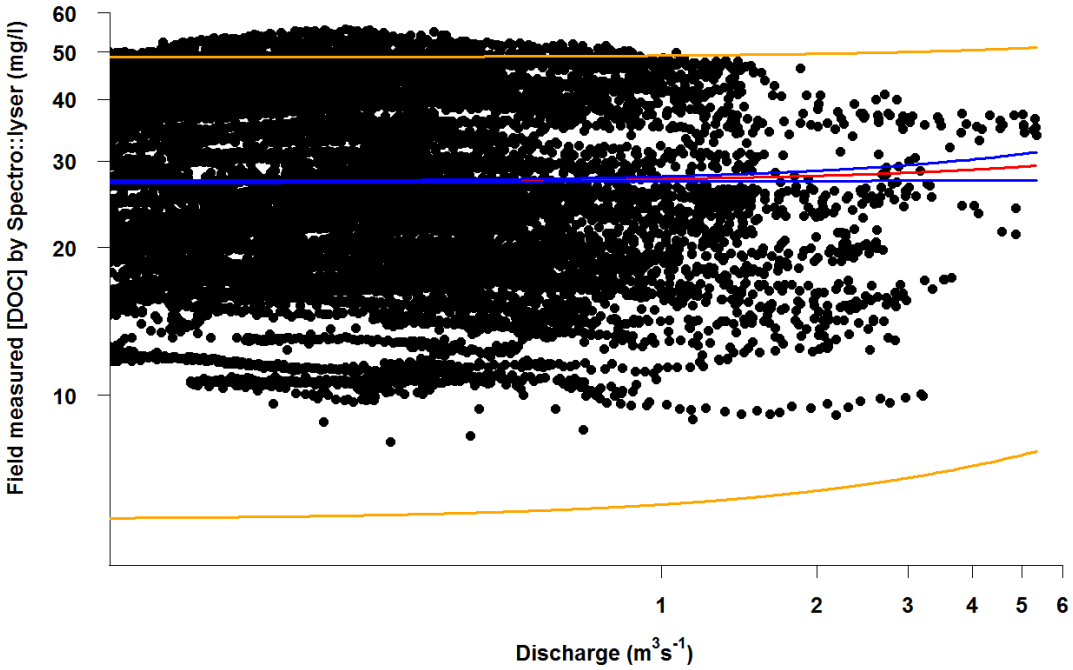


Figure 4.7 – [DOC] and discharge
Log-log relationship between discharge and [DOC]. The adjusted R^2 is low at 0.0001 and a relatively high p-value of 0.03. The model line is shown in red, with the 95% confidence prediction intervals shown in blue, with the 95% prediction intervals shown in orange. The 95% prediction intervals are extremely wide indicating little to no predictive power in this relationship.

4.4.6 [POC]

[POC] ranged from 0.9 mg/l to 45.4 mg/l from with a mean value of 3.5 mg/l, median value of 1.9 mg/l and a standard deviation of ±6.4 mg/l (n= 61). The median value is much closer to the minimum [POC] value than it is to the largest [POC] recorded indicating that most of the time the POC concentration is low and occasionally there is an outlying value. There are 3 main data points that deviate significantly from the norm. A value of 23.5 mg/l measured on 16 June 2012, 45.4 mg/l on 2 July 2014 and 15.1 mg/l on 8 August 2014 were significantly above the normal mean and median values. If these three data points are excluded from the mean calculation then the mean calculated is reduced to 2.2 mg/l, producing a value more similar to the median value calculated.

As a method for calculating [POC] using the UV-vis spectra data has not been established the only source of [POC] data is from the manually collected field samples. Consequently, there is a disparity in the amount of data available from the [DOC] time series and the [POC] time series for constructing the time series and performing budget calculations. In the case of [DOC] budgets calculated using two methods: the high resolution time series generated via the Spectro::lyser™ and the time series generated using the laboratory data. In the case of the [POC] data only the laboratory data is available. Therefore, there

are uncertainties is the data with respect to what is happening at short time scales, especially as there are a few large [POC] values.

As both high resolution field measured data and laboratory measured data can be utilised for [DOC] measurements this could result in a more accurate DOC component of the overall organic C budget. In contrast, the POC component of the overall organic C budget may have had a greater error associated with it as there may have been other large [POC] values not captured by the manual sampling (similar to the three large [POC] values measured), which would potentially result in the understating of the POC exports.

However, it should be noted that most studies until recently would have relied on manual sampling for both DOC and POC components of the budget calculations. Therefore, the POC is being reported with the same frequency as previous studies would have done but there have been advances in the DOC sensors, allowing for improved DOC sensors. Had this not been available then the [DOC] would have been presented at the same frequency as the [POC].

It is important to incorporate [POC] into the results as not doing so would not give a full picture of the best calculated value of exported organic carbon export budget. The results indicate that the largest proportion of organic C exports are in the dissolved form and although there is some uncertainty about this due to potentially missing periods of large POC exports this still gives the reader a more complete picture. Without this information it would not be clear whether the POC is accounting for 10% of exported carbon or 90% of exported carbon.

4.4.7 C Export

The C export time series presented in Figure 4.1 was calculated by multiplying the discharge by the concentration of carbon, producing the C export time series (Figure 4.1g). The C export profile is most similar to the discharge time series, with the rising and falling limbs observed in the graphs occurring simultaneously, which is particularly clear during the largest events. However, the magnitude of the discharge and C export maximums for different events is not completely synchronous as it is also influenced by [DOC] and [POC]. The largest C exports occurred at times not only when events happen, but at times when [DOC] is relatively high (such as during summer months). The full C budgets are discussed in section 4.6.

It should be noted that C budgets have a relatively large associated error because it is not measured directly but is calculated using a combination of the discharge, the [DOC] time

series and the [POC] time series. The time series shown in Figure 4.1 (g) used the Spectro::lyser™ [DOC] time series values. (For reference see chapter three and in particular Figure 3.7, Figure 3.8 and Figure 3.9 and Figure 3.12 for the errors associated with discharge and [DOC] error).

The discharge measured ranged from a minimum of 0.001 m^3s^{-1} during low flow to 4.7 m^3s^{-1} during high flow, indicating a strong influence on the amount of C exported as a consequence of events. The mean value 0.28 m^3s^{-1} is greater than the median value of 0.11 m^3s^{-1} because more of the discharge values are closer to the minimum than the maximum but when maximum values occur they deviate more significantly from the norm. The summary data indicates that the water temperature mean (9.5 °C) and median (9.8 °C) values are very similar to one another as there are not a significant number of outlying data points that are skewing the analysis or indeed having a disproportionate influence on catchment dynamics. The pH ranges from 4.5 to 8.8, with very similar mean and median values of 6.9 and 7.0 respectively, which ties into observations made from Figure 4.1 that during events pH decreases as the stream becomes acidic but that between events pH increases and the stream becomes more neutral and eventually slightly alkaline. SC ranges from 14.9 $\mu S.m^{-1}$ to 246.5 $\mu S.m^{-1}$ which ties in with the observations made from figure 4.1 that SC decreased during events but increased between events.

The [DOC] ranged from 8.0 mg/l to 55.7 mg/l as measured by the Spectro::lyser™ and between 10.2 mg/l and 64.1 mg/l as measured by laboratory analysis. Using the Spectro::lyser™ data the mean was calculated as 28.2 mg/l and the median was calculated as 26.9 mg/l. Similarly, the mean calculated using the laboratory data was 30.9 mg/l and the median value was 28.9 mg/l. The laboratory data produces slightly larger values, which potentially indicates that the Spectro:;lyser™ has a tendency to underestimate [DOC]. The [POC] ranged from 0.9 mg/l to 45.2 mg/l and the mean of 2.7 mg/l was greater than the mean of 1.8 mg/l, owing to most of the [POC] being measured closer to the minimum than the maximum value but where there are also a few really large [POC] values that disproportionately influence the mean value.

When comparing maximum [DOC] at successive events the [DOC] can be identified as increasing or decreasing. There is a mean increase of 2.5 mg/l of [DOC] where there is an increase in [DOC] (n=83) between events. Similarly, a mean decrease of 2.1 mg/l was measured where there is a decrease in [DOC] (n=85) when comparing successive events.

4.5 Assessing controls on [DOC], DOC export and generating models

Here the major controls on both [DOC] and DOC exports are identified and the distribution properties of the different variables ([DOC], DOC export, pH, SC and water temperature) are considered.

4.5.1 Distribution Properties

The data collected for discharge, water temperature, pH, SC and [DOC] are shown in histograms (Figure 4.8) to help visualise whether a particular composition tends to be closer to the minimum or the maximum value or somewhere in the middle. For example, the most common band for water temperature is between 12 and 13 °C, with temperatures above this increasingly uncommon, tending towards the high twenties. The distribution for pH, is almost a normal distribution with the largest proportion of data points occurring between pH 7 and 7.5, which fits in with the mean and median values (Table 4.1). The histogram for specific conductivity indicates that SC below 20 $\mu S.m^{-1}$ is very uncommon, with values greater than 60 being less common. Larger SC values become less common as these values will decrease as soon as an event occurs. The histogram for DOC indicates that most data falls within the 15-25 mg/l range, but that concentrations from 10 – 40 mg/l are common and that [DOC] outside this range are less common.

The cumulative density functions (CDFs) of the same five variables are shown along with the probability density functions (PDFs) (Figure 4.9). The PDFs provide a similar output to the histograms with the skew value for each distribution shown in Table 4.2. The CDF for discharge shows that > 90 % of the discharge values are occurring at values < 1 m^3s^{-1}. The distribution of the other variables is more spread over the range of variation. The CDFs for temperature and pH, which have more normal distributions than the other variables show a more consistent slope over most of the data range. The slopes of the SC and [DOC] also exhibit steep slopes in the CDF plots in the first half of the data. The CDF for SC indicates that approximately 80% of the SC is of a value of 100 $\mu S.m^{-1}$ or less. The skewness towards lower SC values is indicative of the high frequency of events in the catchment as SC will decrease during events but increase after and between events. The maximum of the PDF and the steep slope of the CDF show that the [DOC] is most commonly about 20 mg/l but there is still a lot of variation between 10 mg/l and 50 mg/l. There is not a small but not major skew away from a normal distribution on the [DOC] data.

The normal-probability plots (Figure 4.10) can give a further indication of whether the data for each of the five variables is normally distributed. These indicate that the discharge and SC are right skewed. In both cases this skewness is rather extreme and is easy to

see by eye by viewing the histograms (Figure 4.8). The normal-probability plots for temperature, pH and [DOC] indicate that there is less variance than would be expected from a normally distributed series and that there is a more even spread of data for the water temperature, pH and [DOC].

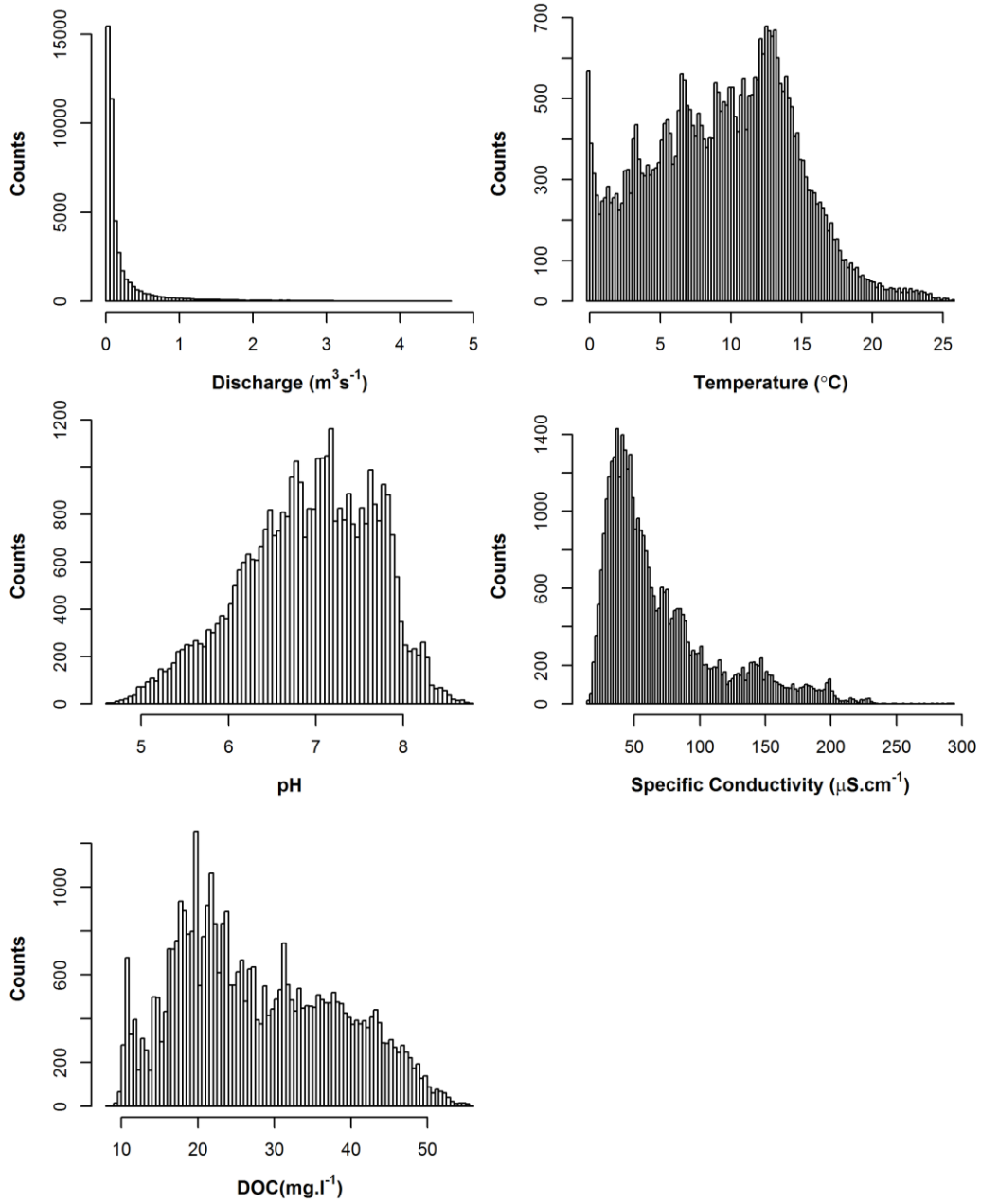


Figure 4.8 – Histograms distribution properties.
Histograms of (a) discharge (b) temperature (c) pH (d) SC and (e) [DOC] for the period between 23 May 2012 and 16 December 2014. The minimum, maximum, mean, median, standard deviation and skewness values are given in Table 4.2.

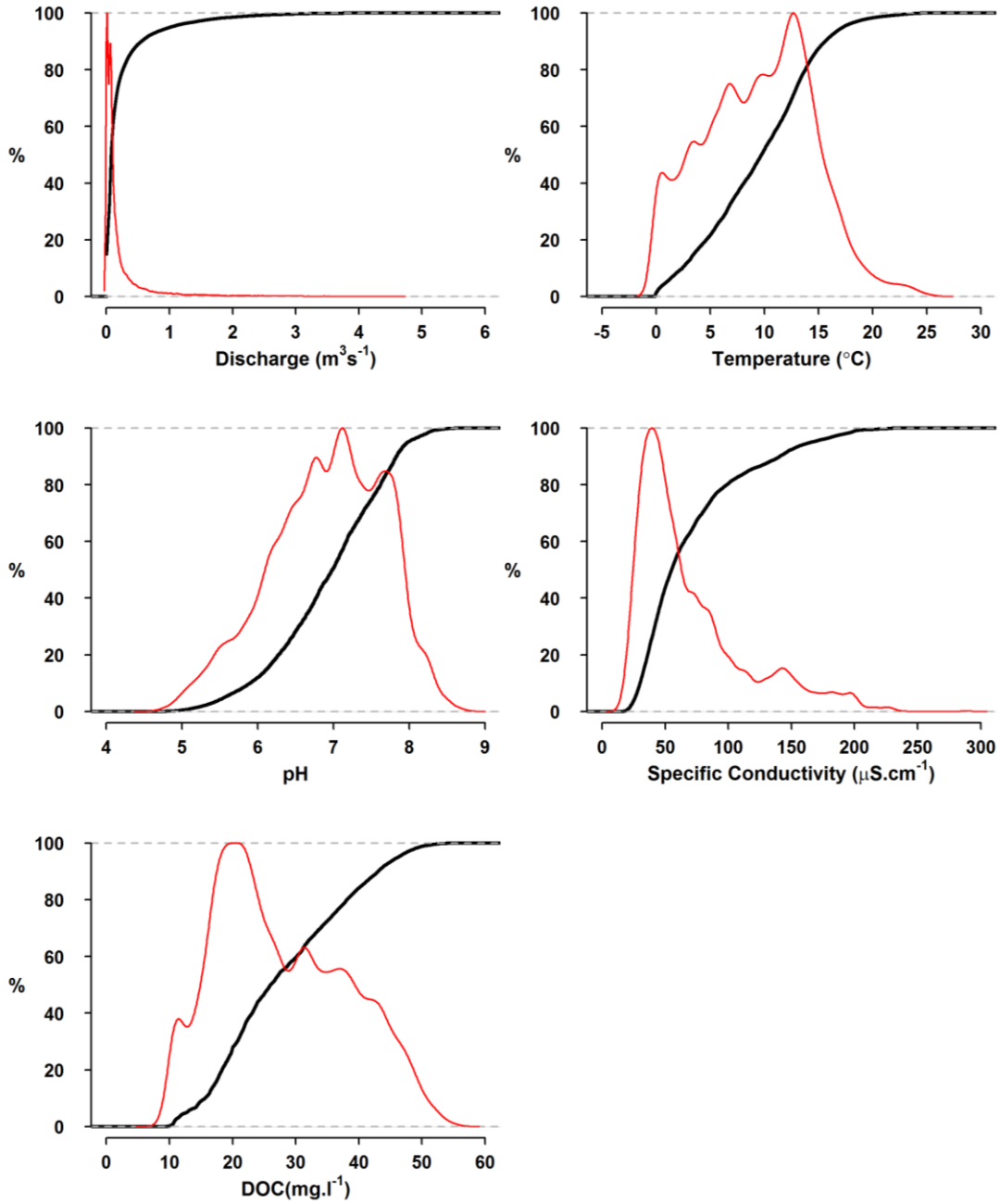


Figure 4.9 – Cumulative density functions and probability density functions
The cumulative density functions (black) and probability density functions (red) of (a) discharge (b) temperature (c) pH (d) SC and (e) [DOC] for the period between 23 May 2012 and 16 December 2014. The minimum, maximum, mean, median, standard deviation and skewness values are given in Table 4.2.

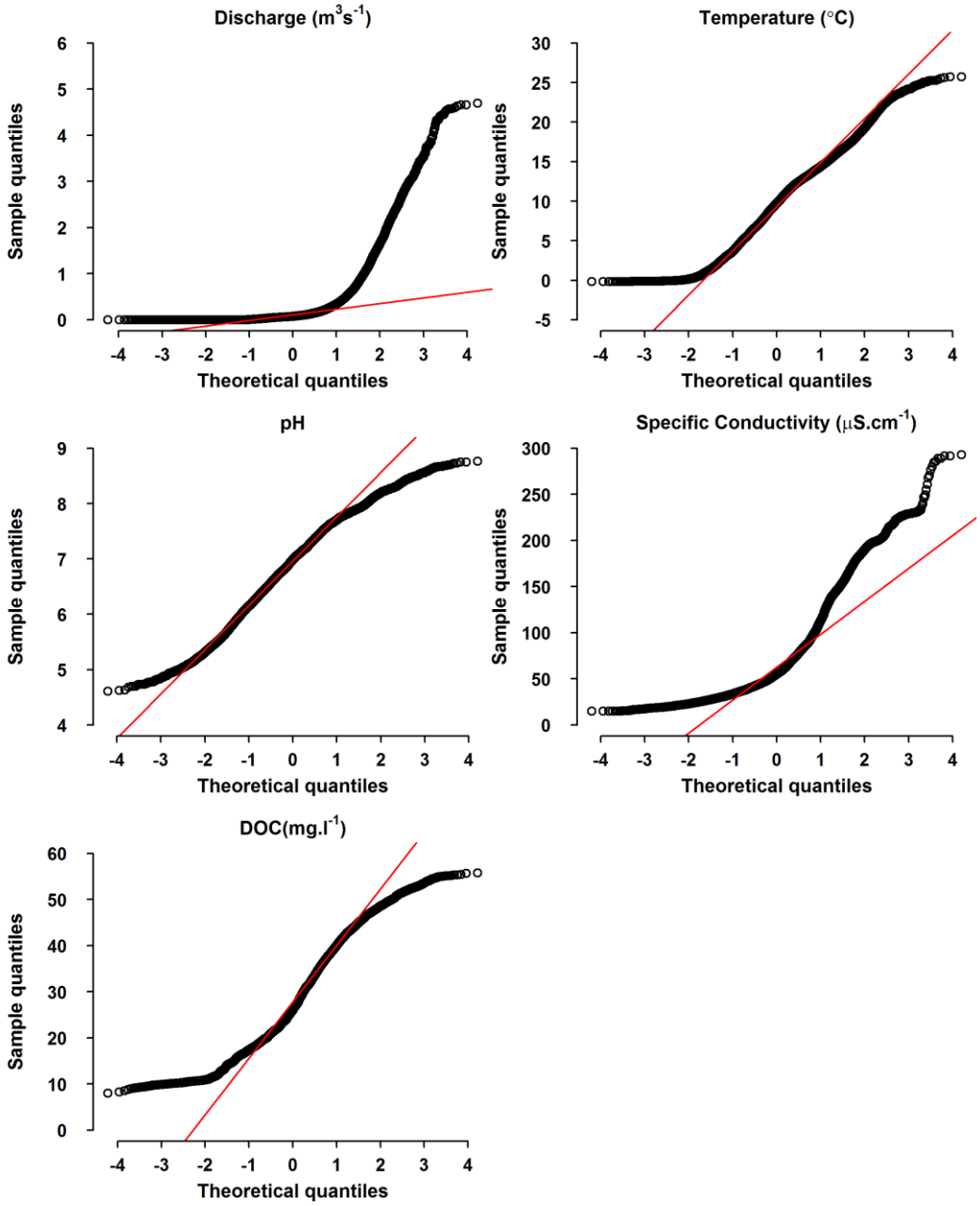


Figure 4.10 – Normal probability plots - distribution properties.
Population composition of (a) discharge (b) temperature (c) pH (d) SC and (e) [DOC] for the period between 23 May 2012 and 16 December 2014. The skewness values are given in table 4.2.

4.5.2 Variation within events and the major control on export

The high-resolution data collected by the Spectro::lyser™ and the In-Situ Inc. TROLL® are summarised below (Table 4.2) to identify how much particular variables varied during events (of which there were 169 as a sample size). Due to the large gap in the TROLL® data, events after this were not analysed in this way. The largest change to [DOC] during an event was 22.7 mg/l, which is equal to 47.6% of the [DOC] range for the full time period analysed between 23 May 2012 and 12 December 2013 i.e. the largest variation within one event was measured to be equivalent to half the variation caused by seasonal variation. For all five variables analysed the mean values were greater than the median values. The minimum change in SC recorded during an event was 0.53 $\mu S.m^{-1}$. Small decreases in SC were more common than large ones, primarily due to how frequent events are at Drumtee, which will prevent SC from increasing to particularly large values.

	Mean	Median	Min	Max	Standard Deviation	Skew
Discharge (m^3s^{-1})	0.99	0.58	0.04	4.4	0.99	4.2
Temperature (°C)	2.5	2.3	0.05	9.3	1.5	0.05
pH	0.91	0.81	0.06	2.5	0.62	-0.4
Specific conductivity ($\mu S\ cm^{-1}$)	22.0	12.8	0.53	118.5	22.8	1.4
[DOC] (mg/l)	5.7	4.9	0.28	22.7	4.5	0.4

Table 4.2 – Summary of variable range during recorded events
Summary of the ranges over which variables change during events. That is calculated by subtracting the minimum value the event from the maximum value so is independent of the direction of change.

The C export calculated for each event was plotted against both the maximum discharge recorded during that event and maximum [DOC] recorded during the same event to identify whether the discharge or [DOC] had a greater influence on C export at Drumtee water (Figure 4.11). The relationship is stronger between discharge and C export than it is between [DOC] and C export, indicating the in this catchment C export is more strongly controlled by discharge than by the concentration itself.

Finally, a flow duration profile was used to understand how discharge relates to C export (Figure 4.12 and Figure 4.13). The flow duration curves shown in both the linear and logarithmic scale both show that the largest 10% of flows accounted for a disproportionately large percentage of the TOC exports. The graphs show that 41.3% of the exports occurred during the largest 10% of flows and 61.9% of the exports occurred during the largest 20% of flows. The flow duration curve provides a contrast with the PDFs (Figure 4.9) showing that most of the discharge is occurring at small values but it is the less frequent larger discharge values that are important when considering exports.

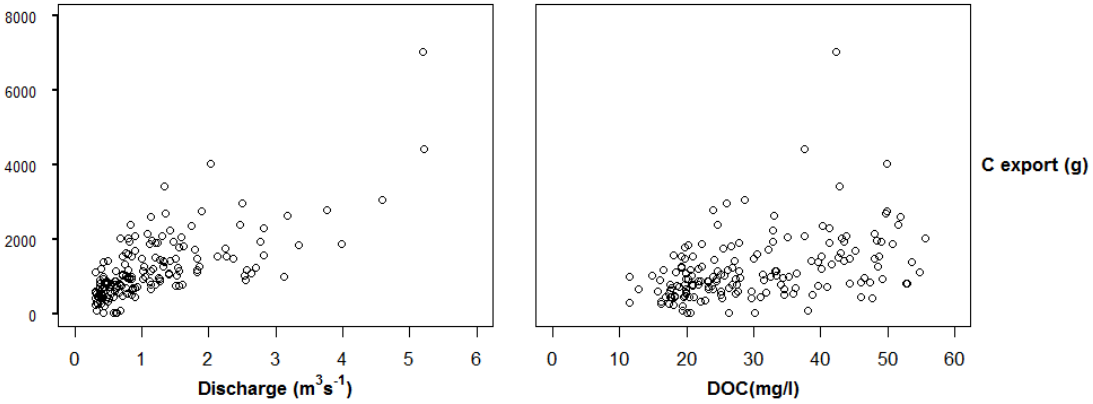


Figure 4.11 – Control on C export
Relationships between C export and (a) maximum Q and (b) maximum [DOC] recorded during events. The adjusted R^2 calculated when comparing the influence of maximum discharge and maximum [DOC] recorded during an event on C export were 0.5 and 0.17 respectively, indicating that discharge has a greater influence on C exports than does [DOC].

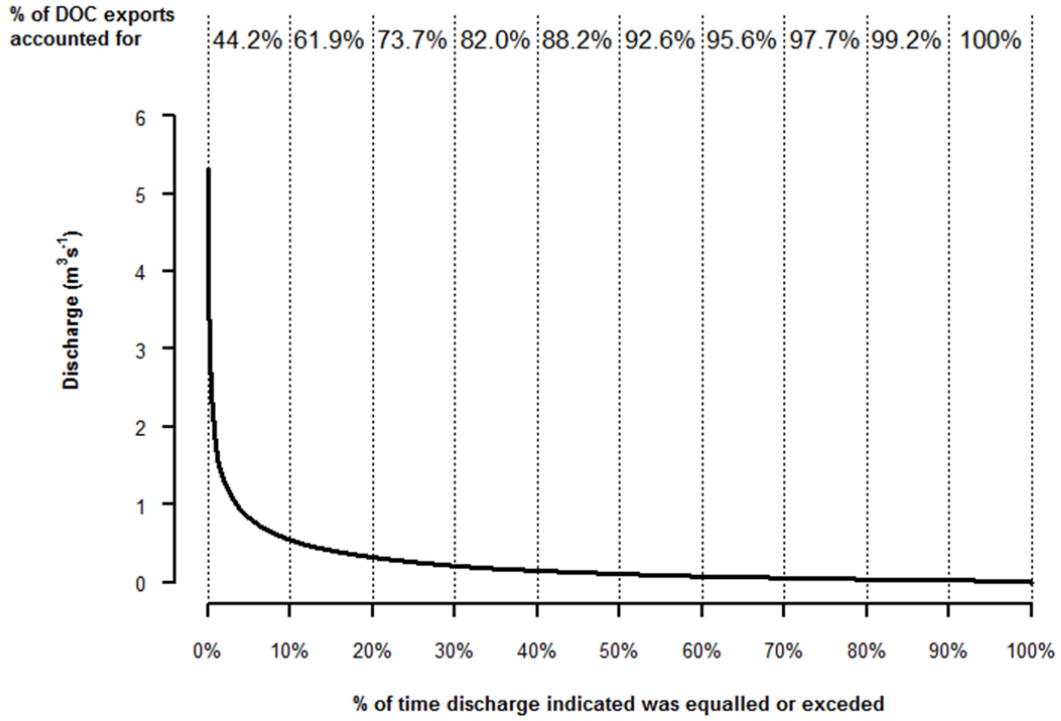


Figure 4.12 – Flow duration curve
Flow duration curve indicating the % of time that a particular discharge rate was equalled or exceeded, with flow ranging between 0.001 m^3s^{-1} and 4.485 m^3s^{-1}. Vertical lines segregate the graph into 10% segments and the % of DOC export in each segment is given by the % value at the top of the graph. For example, the top 10% of flows (values equal to or greater than 0.514 m^3s^{-1}) account for 44.2 % of the TOC exports.

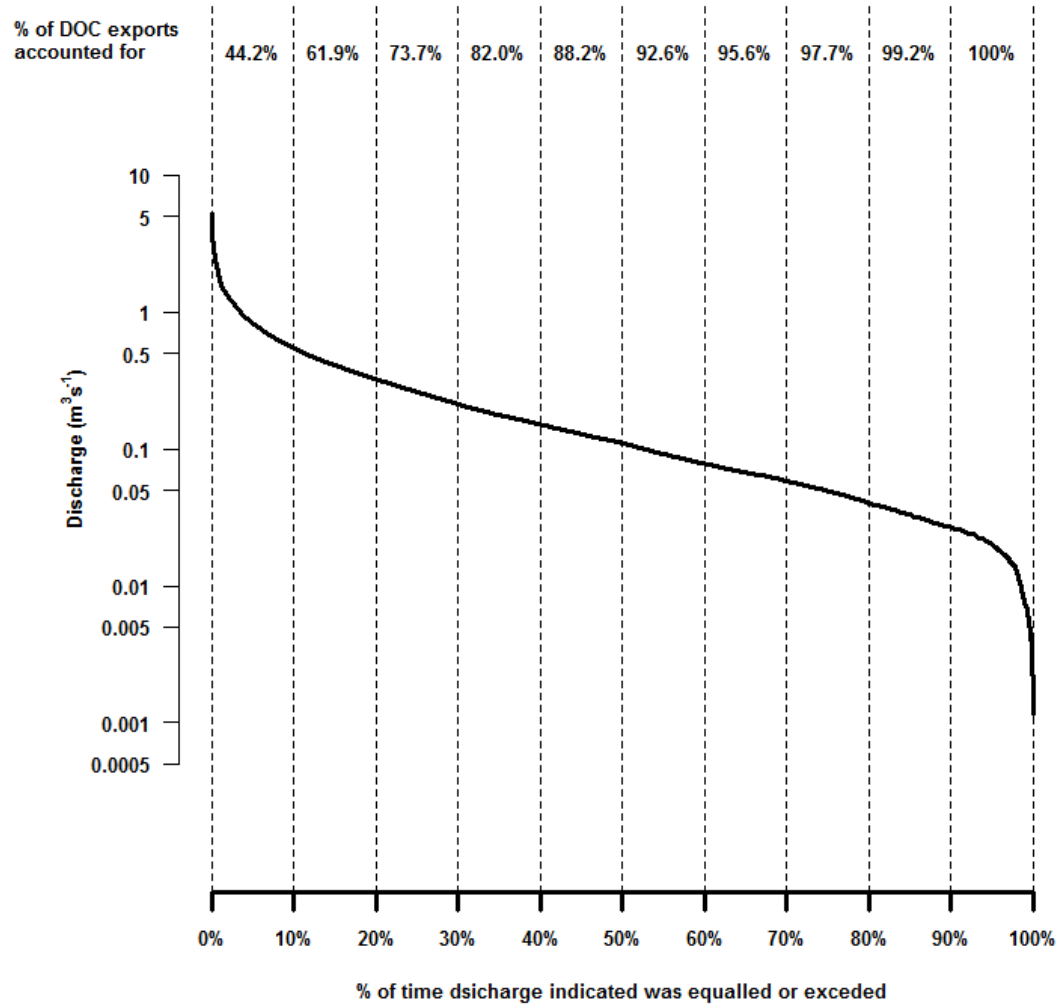


Figure 4.13 – Flow duration curve (log (Discharge))
Flow duration curve indicating the % of time that a particular discharge rate was equalled or exceeded, with flow ranging between 0.001 m^3s^{-1} and 4.485 m^3s^{-1}. Vertical lines segregate the graph into 10% segments and the % of DOC export in each segment is given by the % value at the top of the graph. For example, the top 10% of flows (values equal to or greater than 0.514 m^3s^{-1}) account for 44.2 % of the TOC exports.

The flow duration curve is the inverse of the probability density function for discharge, where the largest percentage of flows are the smallest ones, whereas the flow duration curve shows that when the largest flows occur these are the most significant in terms of C exports. There is also a moderately significant linear relationship between the total C export per month and the total water discharged per month (Figure 4.14). This relationship between monthly discharge and total month exports of organic C hides a clear seasonal pattern in monthly C export (Figure 4.15). However, despite the clear seasonal pattern, there are inter-annual differences. For instance, September 2012 produced the largest C export for of the study period but September 2014 had the lowest C export.

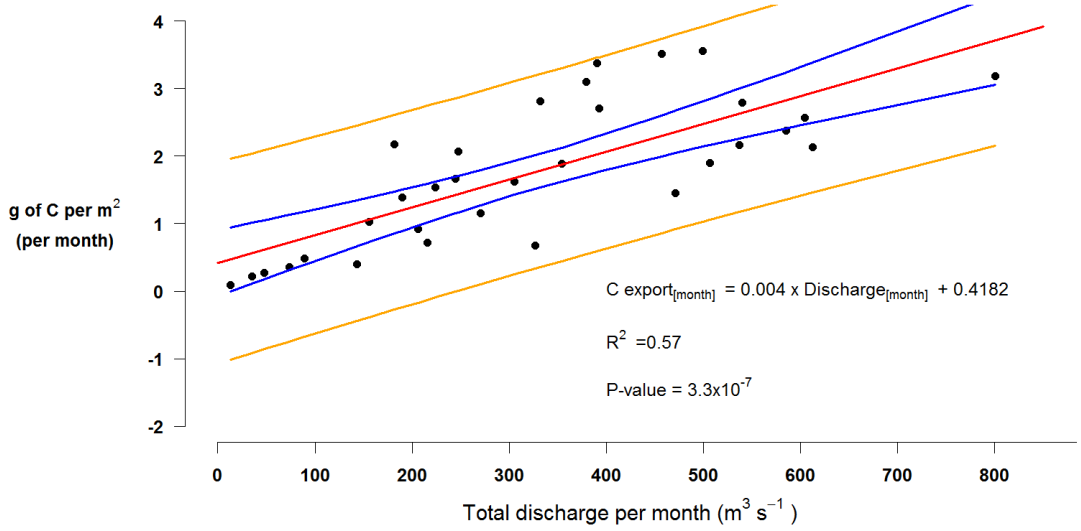


Figure 4.14 – Monthly C export relative to discharge
Total discharge and total C exported for each month analysed in this study. The monthly C export values are the same as presented in Figure 4.12. The 95% confidence interval is shown in blue, with the 95% prediction interval shown between the orange lines. The standard error of the discharge values is ± 0.00063 and the standard error of the C export values are 0.24.

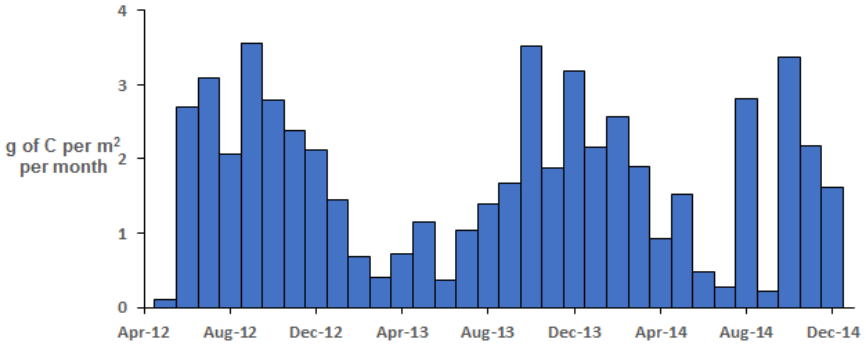


Figure 4.15 – Monthly C export on bar chart
Histogram showing monthly exports of C based on Spectro::lyser™ measured [DOC] and laboratory measured [POC]. This analysis indicates that there is a seasonal pattern.

4.5.3 Bivariate analysis for modelling [DOC]

The consideration so far shows that organic C exports are strongly influenced by discharge, but this represents only the control on delivery from the catchment to the river system and there may be other factors that are important in controlling [DOC] and by extension these other variables should also be considered when building a statistical model. For example, river water temperature is likely broadly representative of air temperature (see Figure 3.5) and so catchment vegetation productivity and surface soil temperature. To take this consideration further, the average temperature over the previous 200 hours was incorporated into the GAM model.

The reason for using three different temperature measurements is as follows:

1) The half hourly recorded temperature is the most accurate reflection of conditions on the catchment and is therefore included.
2) The mean daily temperature removes the diurnal component of the temperature, giving a more accurate reflection of the catchment in day to day terms. For example, the minimum temperature during the summer might be close in temperature to the warmest part of the day on a mild day in the winter.
3) The temperature over the last 200 hours represents a time period which is representative of recent climactic conditions that might influence DOC production and released. For example, if there is a week in spring of unseasonably warm weather this may result in increased DOC production. A single warm day is unlikely to generate this type of DOC production.
4) The month is used as a proxy for seasonality. The month is used rather than temperature because although similar temperatures might occur during the spring and summer seasons the biological processes ongoing during each season are different: plant growth during the spring and plant degradation during the autumn.

The temperature measured over the previous 200 hours gives an indication of the recent climactic conditions of the catchment, which will reflect more strongly on current biological productivity and [DOC] within the catchment. For example, warm conditions over a 1-week period could support greater bacterial production of DOC and increase [DOC], regardless of whether a sample being made is collected during a warm day or a cold day.

There is a lot of variation in the predictive capacity of each different variable to calculate [DOC] (Figure 4.16). In terms of the adjusted R^2 values the model between the mean temperature recorded during the preceding 200 hours produced the strongest relationship with [DOC] with an adjusted R^2 value of 0.53, whereas the model between and [DOC]

produced the smallest R^2 value at 0.0009. When the data is binned into groups of 100 (Figure 4.17) the random variation is reduced and relationships are easier to distinguish. With the exception of the [DOC] discharge relationship the p-values for all the binned data is small. The months data is not binned (as there isn't really a feasible method of doing this that is in line with the other variables) but the other eight graphs have been binned and are shown in Figure 4.17. The strongest relationships with respect to the [DOC] are the temperature measurements (half hourly, mean daily temperature and the mean temperature over the previous 200 hours). Some of the other variables (such as the length of the current event) do have low p-values but as shown by very small R^2 values the predictive power of most of these variables is very limited. It is not clear whether the relationships that do exist between [DOC] and other variables are linear or non-linear. However, at first glance the relationships do appear to be non-linear suggesting that the flexibility granted by non-linear modelling may be appropriate in this instance.

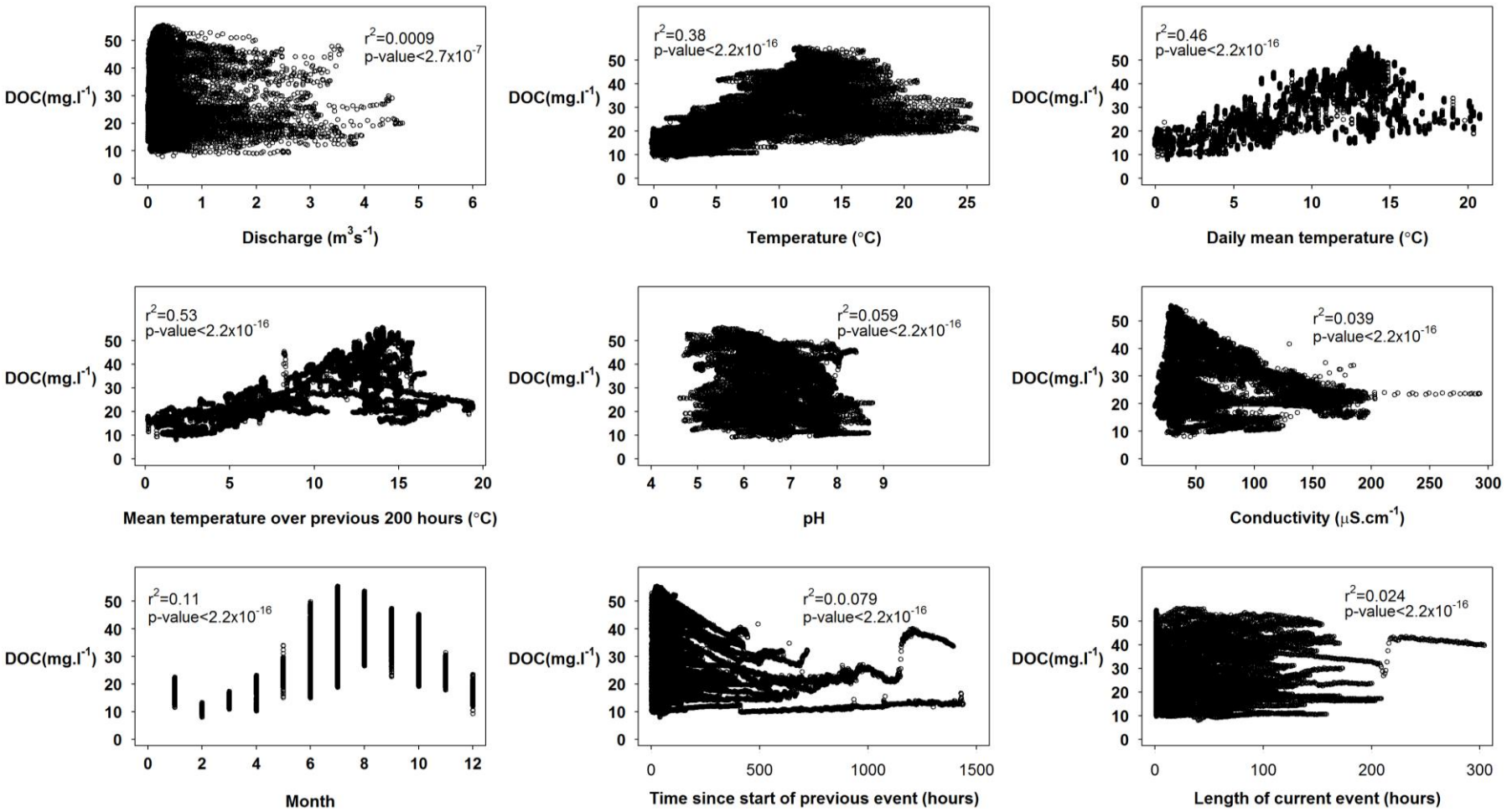


Figure 4.16 – Bivariate relationships between [DOC] and other 9 variables
The adjusted R^2 value is given in each plot with the average temperature recorded during the previous 200 hours resulting in the greatest adjusted R^2 of 0.53 with the discharge resulting in the lowest adjusted R^2 of 0.0009.

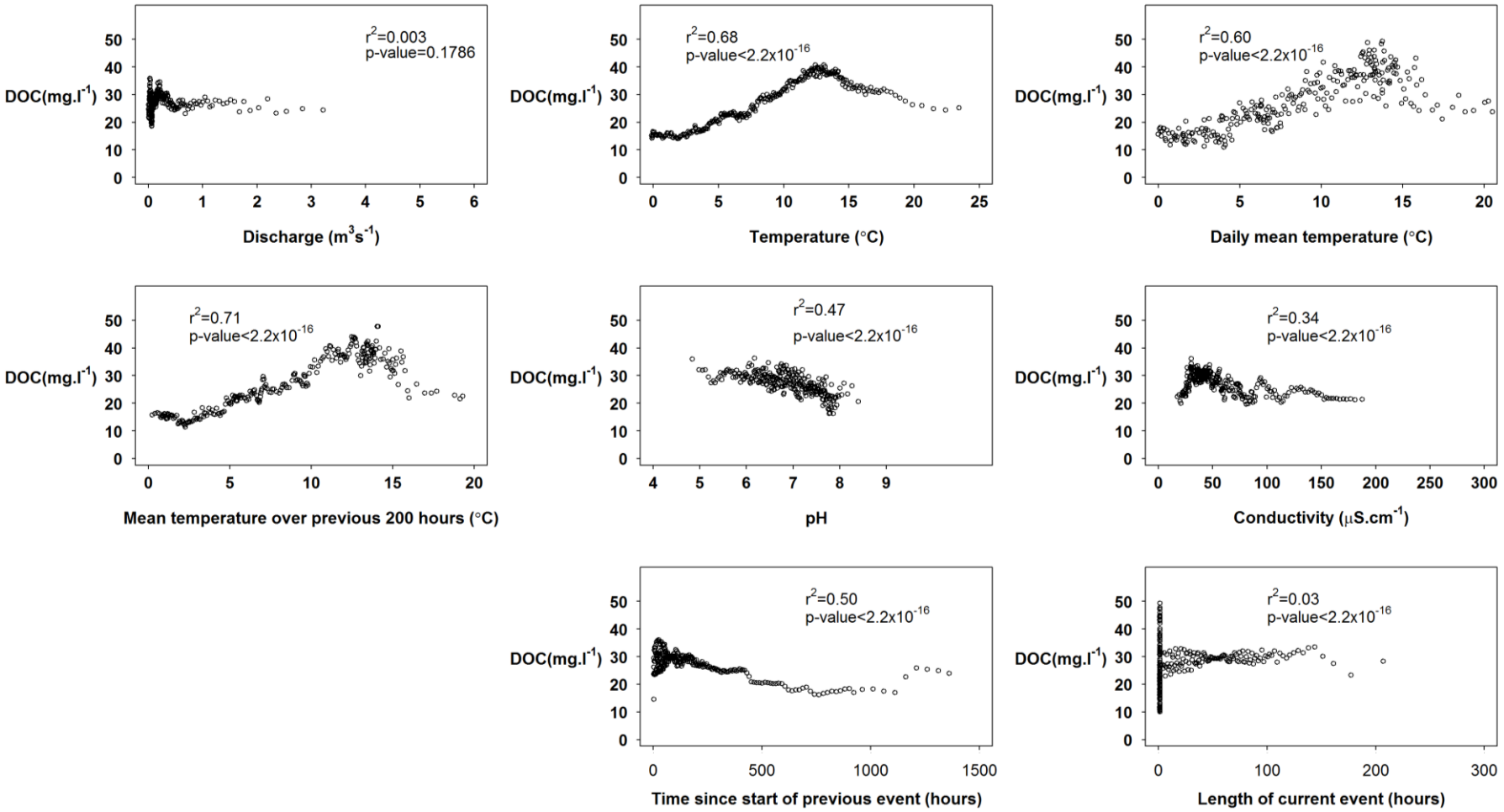


Figure 4.17 – Bivariate relationships between [DOC] and other 8 variables
The adjusted R^2 value is given in each plot with the average temperature recorded during the previous 200 hours resulting in the greatest adjusted R^2 of 0.71 with the discharge resulting in the lowest adjusted R^2 of 0.03.

4.5.4 GAM model - backcasting

As the bivariate analysis did not show promise to model [DOC] multi-variable analysis was explored to determine if a better relationship can be established between [DOC] and other time series of water hydrochemistry. This could have been done through a process such as multi-variable linear regression but as some of the relationships between [DOC] and other variables appears to be potentially non-linear a generalised additive model (GAM) has been selected as the appropriate multi-variable modelling tool.

The GAM model generated using the Spectro::lyser™ data in the mgcv package between [DOC], discharge, pH, SC, water temperature, time between events, length of events, month of the year, average daily water temperature and the average temperature over the previous 200 hours produced an adjusted R^2 value of 0.927 (Figure 4.18). As the Spectro::lyser™ time series will be subject to some uncertainty, the laboratory data was compared with the model data. This relationship had an adjusted R^2 value of 0.72, which is less than the GAM model, but still large and suggests this approach is capturing descriptors of the catchment controls on [DOC].

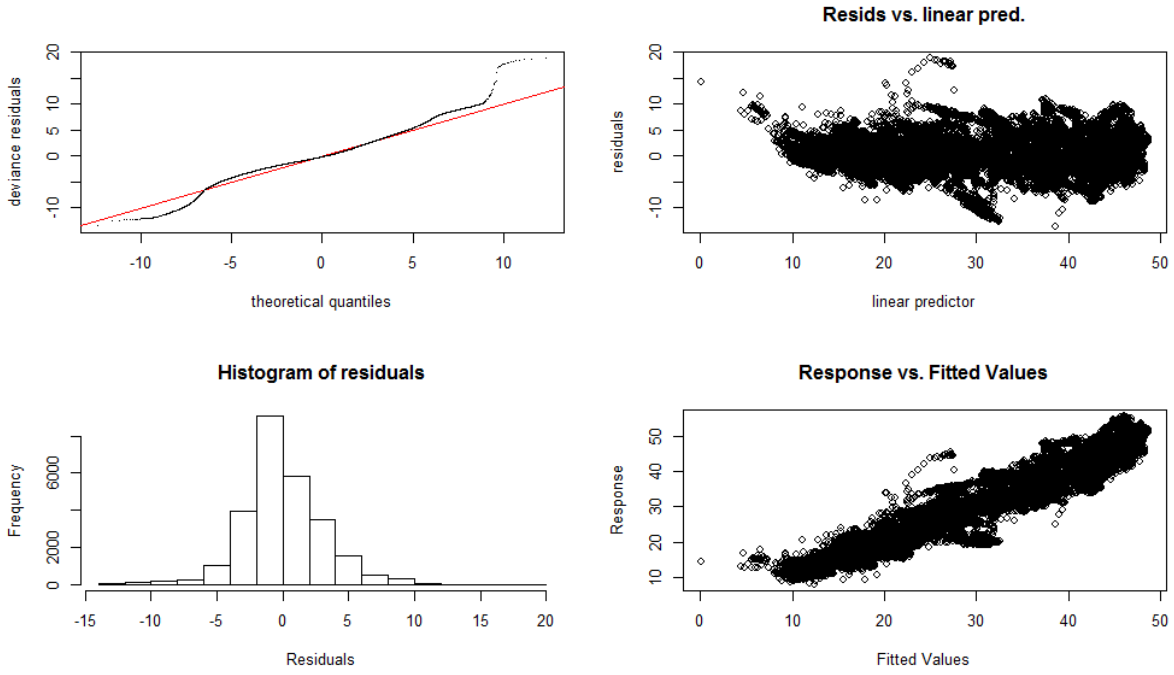


Figure 4.18 – Outputs of GAM model
The outputs of the GAM model show that the residuals do not have any distinct pattern and that there is a normal distribution of residuals when comparing the output of the Spectro::lyser™ to the output of the model and also that (bottom right graph) there is a reasonably tight straight line relationship between the Spectro::lyser™ [DOC] and the [DOC] calculated by the GAM model.

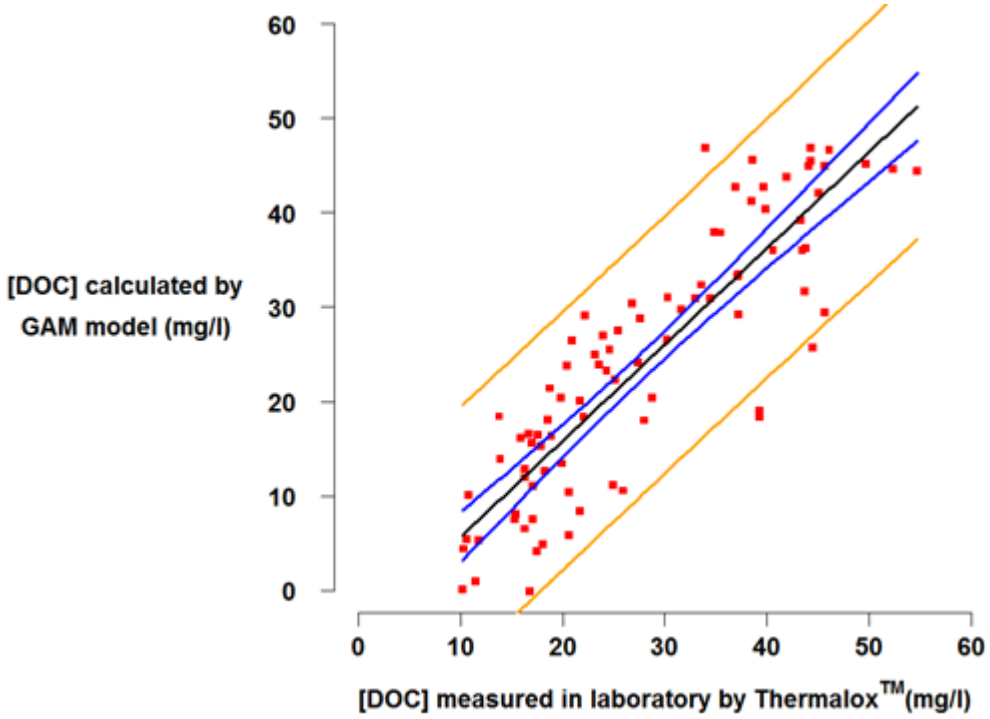


Figure 4.19 – Laboratory [DOC] compared to GAM model
Comparison of the GAM model outputs generated from hydrochemistry time series with laboratory data measurement of [DOC] that extend back to 2007.

The model appears to underestimate the largest values (when measured laboratory [DOC] is greater than 50 mg/l). Consequently, this could mean that there is an underestimation of summer [DOC] values calculated by the GAM model. This underestimation likely occurs because the amount of time that Spectro::lyser™ [DOC] is greater than 50 mg/l is small and that similar conditions at different times produced [DOC] of lower concentration. Potentially data collected over many years would be able to identify values that are outliers as the volume of data collected would increase and the model would be able to use this additional data.

The second stage of the analysis was to test the GAM model using data not used to generate the model, namely using the laboratory [DOC] values collected between 20 November 2008 and August 2014 inclusive. For this second test, there were a total of 88 laboratory samples (including the original 32 samples used in the initial GAM model and 56 other samples). The results produced by the GAM model resulted in an adjusted R^2 value 0.75 and a p-value of $< 2.2x10^{-16}$. The relatively high adjusted R^2 value and low p-value was important because it indicates that at Drumtee water this model can be applied to longer time periods than the 2012-2014 data period including time periods when the Spectro::lyser™ was not deployed at Drumtee water. That the incorporation of the older laboratory data produced a =adjusted R^2 value was unexpected.

There is a drift in the calculated data with the model producing large numbers of negative values during the winter. The [DOC] calculated by the GAM model produces a linear model with the laboratory measured concentrations:

[DOC] calculated = 1.1517 * [DOC] measured – 7.3024

Using this best fit straight line linear equation the following correction was applied to the data.

$$[DOC]corrected = \left(\frac{[DOC]calculated + 7.3024}{1.1517}\right)$$

Applying this correction did not alter the adjusted R^2 value but it did produce larger [DOC] values in the summer and winter, more in line with what was observed in the original [DOC] time series. This correction is especially important during the winter months because it removes all but a few negative [DOC] values.

4.6 DOC and TOC export budgets estimated using several methods

Using the approaches described in section 4.3.2, g of C exported per m^2yr^{-1} were calculated for different hydrological years (table 4.3). Equation 5 makes use of every measurement and equation 6 uses mean C and discharge concentrations. Approaches A-G reduce the number of data points considered to explore how decreasing measurements resolution changes understanding.

As discussed in chapter 3 the budgets are subject to several sources of error, both in terms of measurement and also in terms of the calculations. It was also noted that as high resolution data was not available for measuring [POC] it is difficult to determine how stable this value is during an event with the data available. This can generate a larger than preferred error associated with the POC component of the budget.

2012/13								
Export 2012/13	**Number of data points used**		**Equation 5**			**Equation 6**		
			DOC $\left(\frac{gC.m^2}{yr}\right)$	**POC** $\left(\frac{gC.m^2}{yr}\right)$	**TOC** $\left(\frac{gC.m^2}{yr}\right)$	**DOC** $\left(\frac{gC.m^2}{yr}\right)$	**POC** $\left(\frac{gC.m^2}{yr}\right)$	**TOC** $\left(\frac{gC.m^2}{yr}\right)$
A	17520	25	16.6	1.1	**17.7**	18.1	1.1	**19.3**
B	27	25	16.2	1.1	**17.3**	21.2	0.8	**22.1**
C	25	25	16.2	1.1	**17.3**	21.1	0.9	**22.1**
D	14	14	16.1	1.1	**17.3**	22.5	1.2	**23.7**
E	8	8	16.5	1.0	**17.5**	24.3	1.2	**25.5**
F	6	6	15.9	1.0	**16.8**	26.4	1.2	**27.6**
G	2	2	27.9	1.2	**29.1**	30.6	1.5	**32.1**
2013/14								
Export 2013/14	**Number of data points used**		**Equation 5**			**Equation 6**		
			DOC $\left(\frac{gC.m^2}{yr}\right)$	**POC** $\left(\frac{gC.m^2}{yr}\right)$	**TOC** $\left(\frac{gC.m^2}{yr}\right)$	**DOC** $\left(\frac{gC.m^2}{yr}\right)$	**POC** $\left(\frac{gC.m^2}{yr}\right)$	**TOC** $\left(\frac{gC.m^2}{yr}\right)$
A	17520	23	19.8	1.9	**21.7**	19.5	1.8	**21.3**
B	23	23	21.5	1.9	**23.4**	24.5	2.0	**26.5**
C	14	14	20.1	1.7	**21.8**	24.2	1.8	**25.9**
D	8	8	21.4	2.7	**24.2**	22.3	2.4	**24.6**
E	6	6	19.3	5.7	**24.9**	16.8	2.7	**19.4**
F	4	4	20.0	2.5	**22.5**	21.0	2.3	**23.3**
G	2	2	26.4	1.1	**27.5**	29.5	1.1	**30.6**

Table 4.3 – C budgets calculated using different equations
DOC, POC and TOC export generated using equations 5 and 6 for the 2012/13 and 2013/14 hydrological year. The DOC portion of budget A was generated using the Spectro::lyser™ data with the remaining budgets being generated using manually collected field samples. The manual field samples were collected between 25/09/2012 and 04/10/2013, so the budget is apportioned to the equivalent of 17520 data points rather than the 17943 covered between these two sampling times.

4.6.1 Comparing 2012/13 budgets with 2013/14 budgets

For each hydrological year (2012/13 top of Table 4.3; 2013/14 bottom of Table 4.3) seven values of unit area C export are calculated. Whilst there is variation across the budgets, the estimated C exported is similar for most scenarios, although more variable for equation 6. However, in 13 of the 14 carbon budgets equation 6 produces a larger carbon budget than does equation 5. The exception to this is for budget E in the 2013/14 year, where equation 5 is strongly influenced by a sample collected on 2 July 2014, which produced the largest [POC] value recorded during the study period of 45.2 mg/l.

The results indicate that >90% of the budgets are made up of dissolved rather than particulate material. As the number of samples used to construct the DOC budgets decreased the more erratic the carbon budgets calculated were with the tendency being that the calculated C export tended to be calculated as larger when there were fewer C samples used in the budget calculations . The calculated increase in exported C when fewer samples are used is particularly true for budget G in both 2012/13 and 2013/14 regardless of whether equation 5 or equation 6 was used. The issue with budget G about because this budget uses two [DOC] data points at the start and the end of the hydrological year when the [DOC] is just coming off the high summer values and therefore they are higher than the [DOC] time series mean of 28.2 mg/l when all obtained samples are considered.

From approach A (this calculation uses the most data) and using equation 5, organic C budgets of 17.7 gC m^{-2} yr^{-1} were calculated for the 2012/13 hydrological year and 21.7 gC.m^{-2} yr^{-1} for the 2013/14 hydrological year. However, all approaches using both equation 1 and equation 6 indicate that more C was exported per unit area in 2013/14 than was exported in 2012/13, with this being between 4.0 gC m^2yr^{-1} (equation 5) to 2.0 gCm^2yr^{-1} (equation 6) greater.

The POC exports for HY2013/14 show more variation for the different scenarios A-G than for HY2012/13. The reason for the more varied outcome of the different calculations may be in part due to the drier 2013/14 summer as C would be able to accumulate between events in greater quantities rather than being regularly flushed out. In comparison, the [DOC] is steadier between the different scenarios. In HY2012/13 the regular rainfall during the summer resulted in the regular export of C, which resulted in continually high [DOC]. Consequently, this meant that [DOC] would generally increase and decrease in line with the seasonal variations of the water temperature for the HY2012/13 year but in HY2013/14 this was not the case as a more complex relationship formed, where [DOC] decreased between events to relatively low values during warm but dry weather. The more complex relationship

between seasonal variations of [DOC] and water temperature for HY2013/14 is evidenced by instances such as the [DOC] measurement for a sample collected on 2 July 2014 being measured at 22.6 mg/l, at a time of year when it would be expected that a [DOC] sample would be > 40 mg/l at Drumtee. In the run up to this period there had been no significant events for over a month. The method for gap filling data used to produce the data to be inputted into equation 5 was gap filled using a straight line between successively measured [DOC] points on the time series. If the number of data points used to produce the annual [DOC] time series is insufficient to capture important details at short time scales not captured as seasonal variation then this will result in errors in the calculation.

The final source of variation across the 2013/14 budget calculations arises from POC budgets that more exhibited greater variation during the HY2013/14, which influences the budget calculations, especially when using equation 5. The 2 July 2014 sample that produced a [DOC] of 22.6 mg/l generated [POC] of 45.2 mg/l, exceptionally large for Drumtee water. The budgets in which this value was incorporated (B, D and E) were strongly influenced by this value and these showed some of the larger POC export values. In particular, budget E for 2013/14 using equation 5 produced a POC budget of 5.7 g $Cm^2.yr^{-1}$ as there were only 5 other [POC] samples used to produce the budget.

Equation 5 calculated that there were exports of 17.7 and 21.7 gC $m^{-2}yr^{-1}$ for 2012/2013 and 2013/2014 respectively. The discharge time series generated and discussed in section 3.2.4 was used for this analysis. If discharge estimated using SEPA data from Newmilns and an earlier discharge relationship established at this site (Murray, 2012) between Newmilns and Drumtee is used, then Q_r increases to 27.7 $gCm^{-2}yr^{-1}$ for 2012/2013 (an increase estimate of 31%) and 32.4 $gCm^{-2}yr^{-1}$ for 2013/2014 (an increased estimate of 24%). Clearly if understanding exactly how much C is lost from a landscape is important, then accuracy in discharge profiling is paramount. The Walling and Webb method (Walling and Webb, 1985) (equation 6) may produce a greater estimate of C export than using the high-resolution data as sampling in the wetter summer in 2012 may have over-represented high flow, higher concentration conditions.

4.7 Profiling [DOC] during events and the influence on export

When visually inspecting the [DOC] time series one feature that was observed was that there were sudden rapid decreases in value and that these seemed to occur during events and then followed by a rapid recovery. These rapid decreases are the sort of detail that could only be identified through high resolution monitoring and these sudden decreases and subsequent increases indicated a key advantage of high resolution sensors for monitoring [DOC]. These decreases and increases led to the following questions:

1) Did these decreases correspond to events?
2) Did these decreases occur more commonly during large events or small events?
3) Did these [DOC] minimums correspond to any particular point on the discharge time series?

For comparison purposes the timing of the maximum [DOC] was also analysed in terms of when it was happening with respect to discharge.

In the methods section 3.4.1 method used to identify events was described, which enabled the identification of the start of the event; the largest step increase on the rising limb; the point of maximum discharge and the end of the event. For clarification, the largest step increase on the rising limb is calculated by taking each value on the rising limb and subtracting the preceding value and the largest difference is the point at which the discharge is increasing most rapidly. This point of most rapid increase in discharge is referred to here as the largest step increase on the rising limb. Using the code described in 3.4.1 each event was split into three sections:

1) The time between the start of an event up to and including the point of the largest step increase in discharge
2) The time after the largest step increase in discharge up to and including the point of maximum discharge
3) The time after maximum discharge until the end of the event

The total time the between 23 May 2012 and 12 December 2013 when the stream was in event flow was 6009.5 out of a total of 13631.5 hours. Of those 6009.5 hours section 1 (between start and the most rapidly increasing point on the rising limb of the discharge profile) accounts for 1445.5 hours (24.1%) of the time spent in event flow; section 2 accounts for 720.5 hours (12.0%) of the time spent in event flow and section 3 accounts for 3843.5 hours (64.0%) of the time spent in event flow. If the distributions of [DOC]

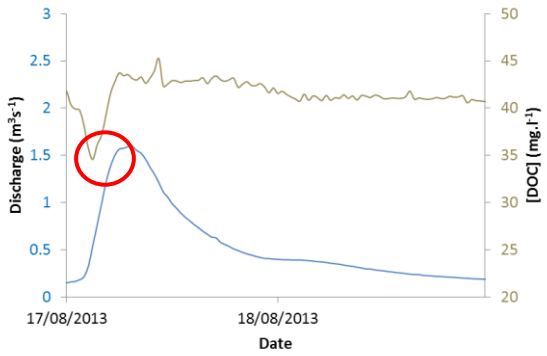
minimums and [DOC] maximums have similar distributions (e.g. if 24.1% of [DOC] minimums occurred during section 1 of events then this is what we would expect by chance and indicate that there isn't a mechanistic link between [DOC] minimums and any particular part of the discharge time series.

To determine this the time at which the lowest [DOC] value recorded during an event was recorded and it was identified whether the [DOC] minima were occurring during one of the three sections listed above. This same process was also done with the [DOC] maximum values and this created six different scenarios from which data was tallied. The six scenarios were:

1) When the [DOC] minimum was measured between the start of an event up to and including the point of the largest step increase in discharge was measured
2) When the [DOC] minimum occurred between the point after the largest step increase in discharge up to and including the point of maximum discharge
3) When the [DOC] minimum recorded during the event occurred at any point after the maximum discharge had been recorded
4) When the [DOC] maximum was measured between the start of an event up to and including the point of the largest step increase in discharge was measured
5) When the [DOC] maximum occurred between the point after the largest step increase in discharge up to and including the point of maximum discharge
6) When the [DOC] maximum recorded during the event occurred at any point after the maximum discharge had been recorded

The above six scenarios (numbered in the bullet points above) are the basis for tables 4.4 and 4.5 on the following pages. On table 4.4 and table 4.5 an example of each scenario is provided. These scenarios are provided in sequential order. For example, the first scenario is on table 4.4 and scenario 6 is on table 4.5.

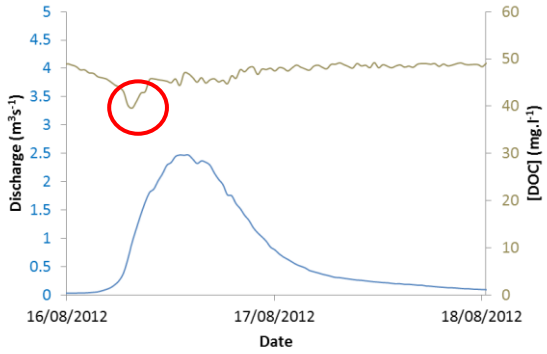
Scenario 1



	Discharge (m^3s^{-1})	DOC exported during event (gC)
Min	0.08	28.3
Max	4.3	5110.0
Mean	1.1	1250.2
Median	0.9	931.8
Standard Deviation	0.9	1081.1

Number of scenario 1 events	Total	Summer	Winter
	89 (52.7%)	45	44

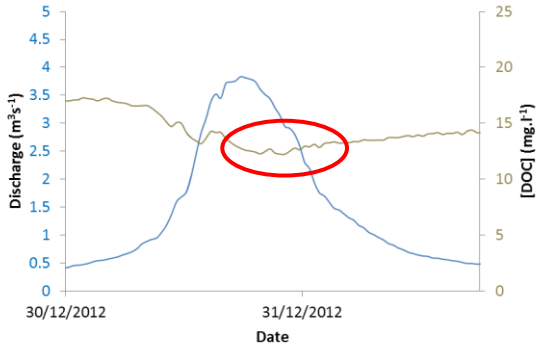
Scenario 2



	Discharge (m^3s^{-1})	DOC exported during event (gC)
Min	0.05	103.3
Max	4.4	5812.5
Mean	1.0	1164.8
Median	0.3	652.1
Standard Deviation	1.2	1288.6

Number of scenario 2 events	Total	Summer	Winter
	45 (26.7%)	29	16

Scenario 3



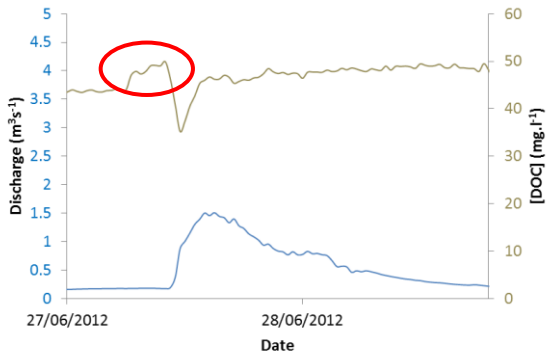
	Discharge (m^3s^{-1})	DOC exported during event (gC)
Min	0.05	41.8
Max	3.4	3304.2
Mean	0.7	800.9
Median	0.4	528.8
Standard Deviation	0.8	720.6

Number of scenario 3 events	Total	Summer	Winter
	35 (20.7%)	10	25

Table 4.4 – Timing of [DOC] minimums during events

Examples of event summaries under scenario's 1-3 (as presented in 4.7) are presented with an illustrated example of each and the total number of events falling into each category is shown as both a total number and as a percentage. For example, there are 89 events (45 during the summer and 44 during the winter) where the smallest [DOC] value recorded, which accounted for 52.7% of [DOC] minimums during events. A summary of the discharge (m^3s^{-1}) and carbon export (gC) measured corresponding to each scenario is presented.

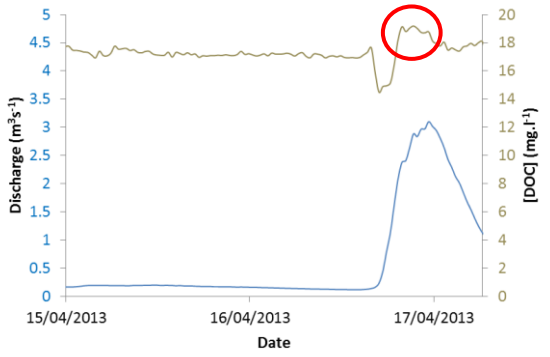
Scenario 4



	Discharge (m^3s^{-1})	DOC exported during event (gC)
Min	0.05	41.8
Max	4.4	5812.5
Mean	1.5	1623.5
Median	1.4	1125.6
Standard Deviation	1.3	1403.8

Number of scenario 4 events	Total	Summer	Winter
	44 (26.0%)	16	28

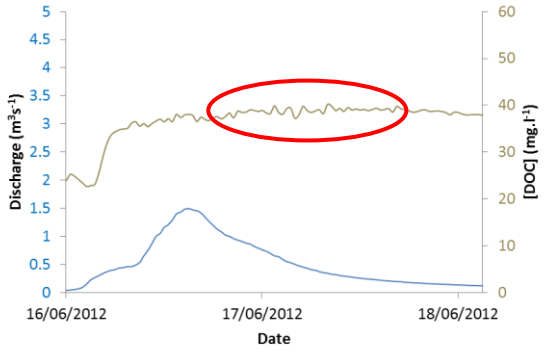
Scenario 5



	Discharge (m^3s^{-1})	DOC exported during event (gC)
Min	0.05	78.2
Max	3.0	2301.1
Mean	0.6	658.0
Median	0.2	294.2
Standard Deviation	0.9	773.6

Number of scenario 5 events	Total	Summer	Winter
	13 (7.7%)	9	4

Scenario 6



	Discharge (m^3s^{-1})	DOC exported during event (gC)
Min	0.05	28.3
Max	2.8	4162.5
Mean	0.9	997.6
Median	0.6	763.7
Standard Deviation	0.77	906.4

Number of scenario 6 events	Total	Summer	Winter
	112 (66.3%)	59	53

Table 4.5 – Timing of [DOC] maximums during events

Examples of event summaries under scenario's 1-3 (as presented in 4.7) are presented with an illustrated example of each and the total number of events falling into each category is shown as both a total number and as a percentage. For example, there are 112 events (59 during the summer and 53 during the winter) where the largest [DOC] value recorded, which accounted for 66.3% of [DOC] minimums during events. A summary of the discharge (m^3s^{-1}) and carbon export (gC) measured corresponding to each scenario is presented.

4.7.1 Event characteristics

Of the 169 events identified during the period between 23 May 2012 and 12 December 2013 there were 84 events identified in the summer and 85 events identified during the winter months meaning that there is unlikely to be a bias in the analysis brought about by seasonality.

The quantity of data presented between tables 4.4 and 4.5 is quite large but there are some interesting points that can be drawn out from them. The first task was to determine whether the [DOC] minimums and [DOC] maximums were occurring any more or less regularly than would be expected by chance (Table 4.6). It can be seen from this table that the frequency with which [DOC] maximums occur more or less coincides with the duration of time that each phase of the discharge time series accounts for. For example, during events 24.1% of that time is accounted for in the time period between the start of the event and the largest step increase in discharge. During this time 26.0% of the largest [DOC] maximums were recorded. By comparison the [DOC] minimums occurred most frequently during this in initial phase of the events, indicating that the distribution of [DOC] minimums is not random. Therefore, it was necessary to explore the link between events and [DOC] minimums further.

	% of discharge event flow occurring during this part of an event	% of minimum [DOC] recorded during this part of an event	Maximum [DOC] recorded during this part of an event
The time between the start of an event up to and including the point of the largest step increase in discharge	24.1%	52.7%	26.0%
The time after the largest step increase in discharge up to and including the point of maximum discharge	12.0%	26.7%	7.7%
The time after maximum discharge until the end of the event	64.0%	20.7%	66.3%

Table 4.6 – Summary of distribution properties

The above information is available disparately in the beginning of section 4.7 and then in tables 4.4 and 4.5 but here the percentage of instances for different scenarios is more succinctly summarised for ease of reading. The amount of time that the discharge profile between different events is given and can be compared with how regularly minimum [DOC] during an event and maximum [DOC] during an event is recorded during this time period.

It is also notable from table 4.4 that instances where the[DOC] minima thatwere recorded after maximum discharge were more likely to occur during the winter (25) than in the summer (10), whereas from table 4.5 it is notable that maximum [DOC] was more likely to be recorded during between the start of the event and the largest step increase during the winter (28) than during the summer (16). Therefore, the data indicates that in the winter months there is a depletion in the DOC stock available for export, which is being removed via events. However, as the number of data points related to these scenarios is limited it reduces the confidence of the conclusion that DOC is being depleted in the catchment during winter months.

If the median DOC export values calculated using the analysis presented in tables 4.4 are compared, then the largest DOC exports are occurring in scenario one and the smallest DOC exports are occurring in scenario three. It is also notable from table 4.4 that the median discharge value is 0.9 m^3s^{-1} for scenario one but only 0.3 m^3s^{-1} for scenario two and 0.4 m^3s^{-1} for scenario three. It may be that these rapid decreases in [DOC] are being brought on by large events where the volume of water is sufficient to overcompensate for the increased C coming into the stream and cause a dilution, whereas in smaller events this might be less likely to occur.

A chi-squared (χ^2) test was conducted on the data in Table 4.6. The χ^2 value for the minimum [DOC] was 19.2 with 2 degrees of freedom, while the χ^2 value for the maximum [DOC] was 0.5 with 2 degrees of freedom. By comparing with a χ^2 distribution table the value of significance when there are 2 degrees of freedom is typically taken as 5.991. As 19.2 is greater than 5.991, this indicates that the phases of an event and minimum discharge are not independent. This analysis provides additional supporting evidence that minimum [DOC] is more likely to occur at the start of events and that the processes of events and [DOC] minima are linked. By contrast the χ^2 value of 0.5 is lower than 5.991 and this indicates that maximum [DOC] and the phases of an event are independent from one another. The implication is that the maximum [DOC] value during an event is not linked to the identified phases of an event: period between the start and the largest step increase in discharge; the time after the largest step increase in discharge and maximum discharge and finally the period after the maximum discharge has been recorded.

4.7.2 Histograms

Histograms are a simple but visually useful approach to profile when event hydrochemistry minimum and maximum values are occurring (Figure 4.20 and Figure 4.21). The [DOC] minimum is typically within two hours of the largest 30 minute step increase in discharge, with a median time difference of 0.5 hours. In contrast [DOC] maxima do not show a consistent relationship with the discharge profile. Instead the [DOC] maximums appear to be more evenly spread, with the largest percentage (~12%) being observed within 2 hours of the discharge maximum. This analysis of the timing of the [DOC] minima and maxima with respect to the discharge solidifies the conclusion drawn from table 4.6 and the χ^2 test that there is a mechanistic control on [DOC] minimums but when [DOC] maximums are occurring is more random.

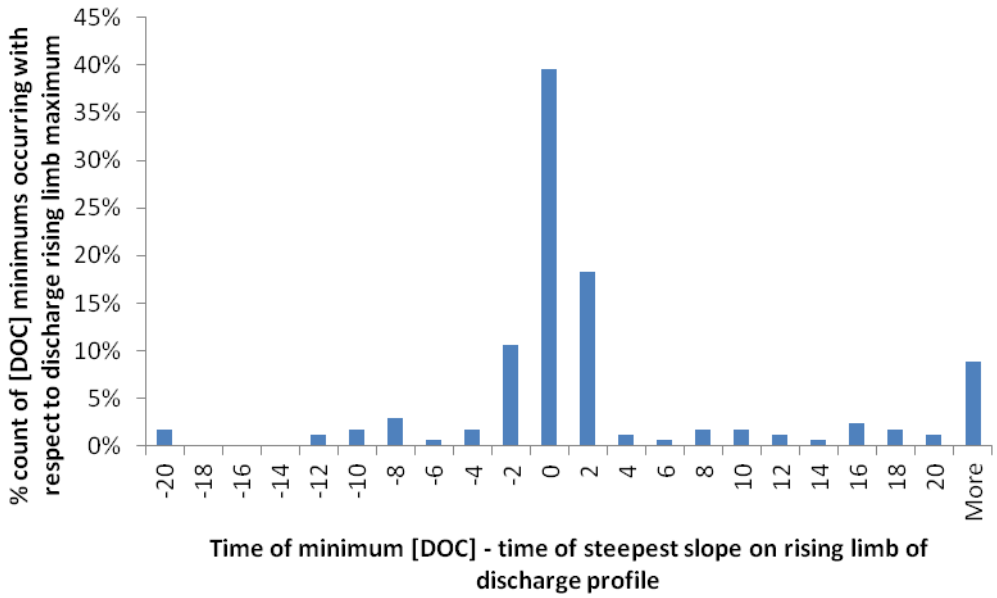


Figure 4.20 - Histogram of minimum [DOC] with respect to discharge
The minimum [DOC] recorded during an event with respect to steepest slope on rising limb of the discharge profile. The largest percentage of [DOC] minimums are occurring at about the same time as the largest step increase on the rising limb of the discharge profile over a 30 minute time period.

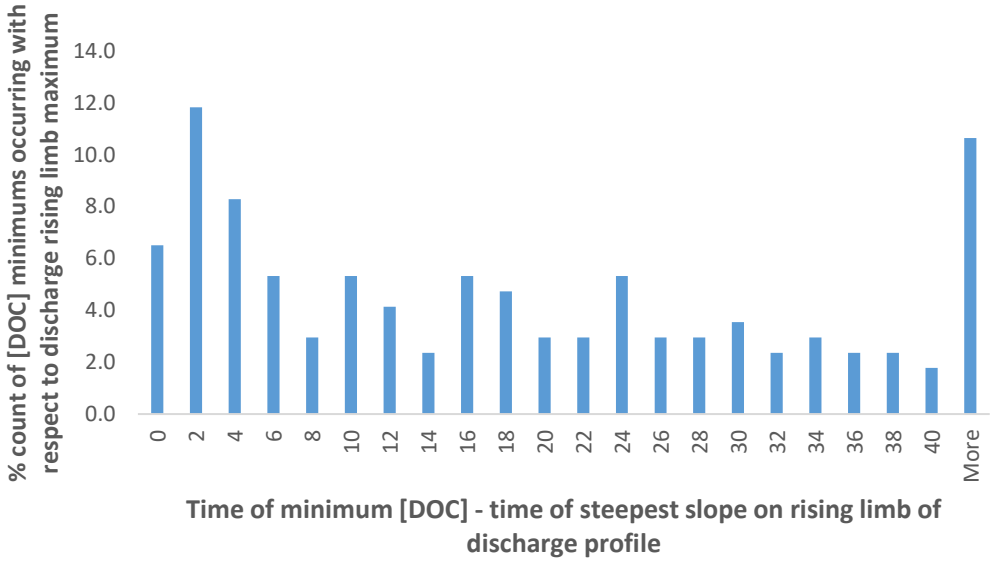


Figure 4.21 - Histogram of maximum [DOC] with respect to discharge
The maximum [DOC] recorded during an event with respect to steepest slope on rising limb of the discharge profile. Where the difference is not more than 40 hours. The general trend is that maximum [DOC] is slightly more likely to occur just after the time of steepest slope on the rising limb of the discharge profile.

4.7.3 Influence of time between events on [DOC] locations

Using the same code as used to determine the time of the events a final test was to establish if the time between successive events had a bearing on whether [DOC] minimums fell into scenarios 1-3 or if successive events had any bearing on whether [DOC] maximums fell into scenarios 4-6.

Time elapsed between events each of the six scenarios was characterised (table 4.7), and the median values show that the time between events was longest when [DOC] minimum was observed as the event started (scenario 1, median 57.5 hours ± 122.7 SD). However, the standard deviations are so large that it is difficult to say conclusively if the time between events has any major bearing on whether [DOC] minimum will be recorded during scenario 1, 2 or 3. Similarly when the median time between the start of successive events was compared then scenario 6 was the most likely to occur. However, again the standard deviation is too large to have any major confidence in this regard.

	Time between start of successive events (hours)	
Scenario 1	57.5 ± 122.7	Minimum [DOC]
Scenario 2	52.0 ± 73.5	
Scenario 3	37.5 ± 39.1	
Scenario 4	43.0 ± 53.8	Maximum [DOC]
Scenario 5	41.5 ± 21.2	
Scenario 6	54.8 ± 115.8	

Table 4.7 – Influence of time between events on timing of minimum and maximum [DOC]
The median number of hours and the standard deviation that elapse between the starting time of one event and the start of the next (i.e. not the end of one event and the start of the next).

4.8 Relationships to previous studies and other field sites

One of the major goals of this project was to monitor catchment dynamics at short time intervals over a long period of time in a way that is not practical unless sensor technology is employed. The major control on [DOC] over long time scales was seasonal variation, manifest here as water temperature, but including also air and soil temperature and the amount of sunlight available. However, at shorter timescales the dominant variable that influenced everything else was discharge and particularly when events occurred. A characteristic of peatlands is that they have flashy high floods with short lag times (Holden and Burt, 2002).

One of the most common tools now for measuring [DOC] is the S::can™ Spectro::lyser™. There is precedent in environmental research for the use of the S::can™ Spectro::lyser™ technology (Waterloo et al., 2006, Bass et al., 2011, Grayson and Holden, 2012, Bass et al., 2013, Strohmeier et al., 2013). Research was also conducted by (Wilson et al., 2013), where a Turner Designs Cyclops-7 coloured dissolved organic matter (CDOM) fluorometer was used as the main tool for measuring and calculating [DOC]. They also generated a model that identified the largest DOC exports as being associated with large events during the "summer and fall". That study corresponds to the findings of this Drumtee water study where the [DOC] is greater during the summer than the winter and that events are strongly influenced by discharge. This project was made possible because of such advances in sensor technology over the last few decades.

The [DOC] measured at Drumtee would indicate that [DOC] at this field site is unusually high when compared to many other bodies of water found in the UK, for example the 22 sites measured by the UK Acid Monitoring Network all had lower maximum [DOC] (Kernan et al., 2010) than Drumtee. The 2012-2013 annual report by the Acid Monitoring Network, measuring 12 streams states that for the 2012/2013 year the mean [DOC] recorded ranged from 1.46 mg/l ± 1.12 mg/l (Narrator Brook) to 11. 48 ± 1.46 mg/l ± 6.72 mg/l (Danby Beck) and with only the Old Lodge site registering a [DOC] value greater than 50 mg/l (Shilland et al., 2014). By comparison the mean value at Drumtee of 27.3 mg/l is much larger. The Acid Monitoring Networks 20 year report published in 2010 indicates that [DOC] has increased at all of their monitored sites over this period (Kernan et al., 2010). As only two annual budget calculations have been calculated here an overall trend cannot be given for long term increases (based on years).

However, a truer picture of the Drumtee [DOC] is to compare with bodies of water where the surrounding soil type is of similar composition (peatlands). Peatlands and soils with high organic content generally results in greater [DOC] than those with lower organic carbon content (van den Berg et al., 2012). Here, Van den Berg compared four different habitats and five different soil types across 29 sites and identified that moorlands contained a greater average [DOC] value than found at woodlands, heathlands or grasslands. The [DOC] measured at Drumtee is greater than at most streams found in the UK but as part of the Eaglesham Moorland the concentrations are within the range of 25.53 mg/l and 74.86 mg/l identified by van den Berg, although the mean [DOC] mean of 28.8 mg/l is at the lower end of this range.

The data presented in this study indicates a smaller C budget than those reported earlier from this field site (Waldron et al., 2009). However, this is partly down to using new data to recalculate the discharge values measured at Drumtee. For example, the total exported DOC for 2012/2013 would have been calculated at 27.7 $gC.m^{-2}yr^{-1}$ using the previous stage height/discharge relationship (Murray, 2012) instead of the 17.1 $gC.m^{-2}yr^{-1}$ and 19.3 $gC.m^{-2}yr^{-1}$ calculated using equation 5 and equation 6 respectively. Other differences may be the result of changes to land management practice brought on by the wind farm construction at Whitelee wind farm (ScottishPower, 2009, ScottishPower, 2012, ScottishPower, 2013) which could have involved construction work on the peatland relating to the wind farm construction as well as restoration work. The summer of 2012 was also extremely wet whereas the winter of 2012/13 was relatively dry as evidenced by few events and increased pH and SC for various points during this period.

However, despite differences the [DOC] recorded at this site during this study are comparable with the two earlier studies (Waldron et al., 2009, Murray, 2012) in terms of

[DOC] and export indicating that different analytical approaches have not had a significant influence on the outcome of the analysis. Additional assurance of the validity of the [DOC] time series generated by the Spectro::lyser™ is that the data measured in the laboratory from field-collected water samples had a magnitude median difference of 1.1 mg/l with a standard deviation of ± 3.2 mg/l for the measured [DOC] range of 8.0 to 55.6 mg/l (chapter 4.4).

Achieved here and in other projects making using of high-resolution sensors (e.g. hydrochemistry (Kirchner et al., 2004, Waldron et al., 2009, Neal et al., 2012) or [DOC] measurements (Bass et al., 2011, Jones et al., 2014)) is the ability to make measurements at very short time scales (< 1 hour) over an extended period (> 1 year). From these high resolution sensors, it has been possible to determine when variations in the different hydrochemistry variables have been occurring with respect to one another. Further, the histogram revealed (Figure 4.20) that the minimum [DOC] was occurring in approximately 40% of the cases at the same time that the discharge was recorded as having its most rapid increase, indicating that when an event was at its most intense the rate of C delivery to the stream did not scale with the increase in run-off and so there was DOC dilution.

For most events the maximum [DOC] occurred after the point of maximum discharge (Figure 4.21). However, for the very largest events this changed and the maximum [DOC] occurred before the point of maximum discharge. One possible mechanistic explanation for this is that when large events occurred they initially produce a large amount of surface runoff (one of the components of the [DOC] measured) but that there may be a saturation point where the quantity of DOC available from surface runoff is reduced as so much of it has already entered the stream in the early part of the event.

At Drumtee the proportionate difference between the largest and smallest discharge is a factor of 4700 and the proportionate difference between the largest and smallest [DOC] is 6.3 (Table 4.1). This disparity in the differences between the minimum and maximum values of discharge and [DOC] respectively indicates that large discharges will have a more significant influence on C exports than will [DOC] values at this particular field site. For example, the maximum [DOC] multiplied by minimum recorded discharge yield an export of 0.23 gCm^2hr^{-1} but the minimum [DOC] multiplied by the maximum discharge yields an export of 172.6 gCm^2hr^{-1}.

Using the high-resolution data produces a C budget of 17.7 gCm^2yr^{-1}, which is smaller than an earlier calculation of 23.1 $gC.m^2$ for the 2012/13 hydrological year (Waldron et al., 2009). However, the 2013/14 hydrological year export budget of 21.7 gCm^2yr^{-1} is more similar. However, using the Walling and Webb method means that a similar budget would

be calculated for this period in 2012/13 of 19.3 $gC.m^{-2}$ and a value of 21.3 $gC.m^{-2}$ in 2013/14. However, I do not think this is a more accurate approach because the latter method makes use of mean concentration and discharge. The former approach takes full advantage of the high resolution data and therefore likely represents a more accurate budget in this case for this catchment.

The budgets of DOC calculated using both the high-resolution data and the data collected and analysed in the field yielded similar results. However, this has not been observed in other studies. In the Lehstenbach catchment in southeast Germany the budget calculated using the Spectro::lyser™ data, generated larger DOC exports than manual field sampling (Strohmeier et al., 2013). The major difference between Lehstenbach and Drumtee is that the [DOC] profile at Lehstenbach increases during an event synchronously with discharge, whereas at Drumtee the [DOC] typically decreases and then increases resulting in a smaller net change in [DOC] during events. Additionally, there is a more pronounced seasonal variation with greater [DOC] in the summer than in the winter at Drumtee, which is not apparent in the Lehstenbach data. This seasonal variation is the dominant trend in [DOC] at Drumtee (especially in 2012/13) whereas this is not a noticeable trend from the data presented in the Strohmeier paper (Strohmeier et al., 2013). Thus, it may be the case that where the variation caused by events influences [DOC] more than seasonal variation then it is more important to use high-resolution sensors to measure [DOC].

SC at Drumtee (mean of 68.0 $\mu S.m^{-1}$) is typical of moorland regions. For example a study of 33 sites in the River Severn catchment, where the soil has been mapped as a mosaic of stagno-podsol, peat, brown earth and stagno-gley (Neal et al., 1997) – so a not too dissimilar soil map from Drumtee) and at these sites the electrical conductivity means ranged from 30 $\mu S.m^{-1}$ to 79 $\mu S.m^{-1}$ (Hill and Neal, 1997). The response of SC in events in this study is also typical of moorland sites discussed in other studies. A 2007 (Tetzlaff et al., 2007) study described two peaty sub-catchments in Scotland (Water of Dye and Brocky Burn) where the low conductivity was measured between 50 – 60 $\mu S.m^{-1}$ at Brocky Burn but with measured values increasing to between 60 – 100 $\mu S.cm^{-1}$ during events. In comparison 60 – 80 $\mu S.m^{-1}$ was regularly measured during base flow at the Charr catchment would decrease during events to values close to 40 $\mu S.m^{-1}$ (Tetzlaff et al., 2007).

The Tetzlaff study showed that conductivity changes during events frequently as different sources of water mix (Tetzlaff et al., 2007). At Drumtee the trend is for a rapid decrease in SC during events down towards the lowest SC. The only exceptions to this were when the events in question were extremely small and the preceding SC values were towards the upper end of the measured range. The predominant soil type in the Drumtee catchment is peat (see Figure 3.1 in methods), which means that the SC values are lower than that

measured in catchments with more mineral soils (Martini et al., 2007). In comparison a study from 2000 focussing on rivers draining into the North Sea (Neal and Robson, 2000) which would give annual median electrical conductivity values ranging from 212 $\mu S.m^{-1}$ to 1172 $\mu S.m^{-1}$across 18 sites with non-specific soil types.

Another observation is that at Drumtee water SC is often higher during the summer months as these are drier periods. However, by comparing the temperature (Fig 2b), discharge (Fig 2a) and SC (Fig 2d) it is notable that drier periods in higher temperatures tend to result in larger increases in SC than do dry periods of similar duration in colder temperatures. More rapid increases in SC occur during warm weather because the flow becomes progressively less-enriched in lower conductivity soil water with time and so groundwater, which has more dissolved ions (West, 2010) becomes proportionally more important. Again not all sites show that response, for example in Canada (with latitude between 54° and 55° compared to Glasgow's 55.9°) seasonal variations on SC were not apparent and landscape controls such as soil type had a much greater influence (Vitt et al., 1995). In India (a very different climate) conductance was shown to vary with temperature (Ahipathy and Puttaiah, 2006).

The trend for decreasing pH during events has been observed in other studies (Tetzlaff et al., 2007) Here, the loading of DOC may play a role, as from peatlands is rich in humic and fulvic acids and is generally acidic and therefore as it enters the stream in greater quantities during events the pH reduces (Wellington and Driscoll, 2004). However, the median and mean pH values observed at this site were unusually high when compared to any of the stream values noted by the UK Acid Water Monitoring Network in a 15 year review published in 2001 (Evans et al., 2001). An additional factors here is that soil CO_2 will also be exported during event flow (Long et al., 2015) and this may reduce the pH of the stream during events, as will other acidic compounds e.g, sulphate, so the exact pH will reflect the interaction of these processes. Again however, the sensor technology provides the best opportunity for understanding these systems and dynamics by providing the most detailed time series currently available.

4.9 Conclusions/ Summary

This chapter has demonstrated the usefulness of high resolution sensors and datasets for measuring [DOC] and their potential for expanding knowledge of DOC in fluvial systems. One of the things identified in this project is that although [DOC], pH and SC all are highly responsive to events, they respond on similar but crucially not identical time scales. If the timescale of consideration is compressed then it appears that all of these variables respond instantaneously to event flow. However, as the sensors provide half hourly timescale detail the time series generated enabled identification that the responses were

not occurring concurrently and that the [DOC], pH and SC minimums are not changing fully synchronously.

Surprisingly, the high-resolution [DOC] data has had a negligible impact upon the budget calculations for the year. By being able to record the [DOC] at more regular intervals it was anticipated that additional, large [DOC] maximums would be recorded during events but as the dilution reduced the [DOC] at the start of the event the maximums had a lesser influence on the budget than anticipated. Additionally, the analysis showed that discharge was the greater predictor of C exports than were the [DOC] values themselves at Drumtee, which did not necessarily reflect the results from other locations around the world. The [DOC] was more strongly influenced by seasonal variation than it was by events, with the largest variation between summer and winter measured to be 47 mg/l but the largest variation in [DOC] recorded during an event was 22 mg/l.

The analysis of the event data reveals a dynamic and complicated system in terms of [DOC]. In the 199 events measured [DOC] reaches its minimum and maximum values in a more inconsistent manner than pH and SC, but during larger events this inconsistency in behaviour is less apparent. For the majority of events the minimum [DOC] was recorded when the discharge was at the point of most rapid increase. An event may lead to an increase in [DOC] but this study suggests that (at least in certain instances) [DOC] will decrease when an event is at its most intense (as indicated by the steepest increase in discharge during an event). The results also indicated that maximum [DOC] can occur before the maximum discharge is reached, especially during larger events. However, during smaller events it was more common for maximum [DOC] to be recorded after maximum discharge had occurred.

Modelling [DOC] is complicated and it was noted that multi-variable models may enable greater predictive power than bivariate models for calculating [DOC]. A GAM model successfully implemented and shown to be a useful method for reconstructing [DOC] using other hydrological variables (discharge, water temperature, pH, SC, time between events, time between the start of events, the mean daily temperature and the mean temperature during the previous 200 hours). The GAM has produced a working model that can be used to calculate at high-resolutions across longer periods of time. Although this method is less preferable to the use of a [DOC] sensor such as the Spectro::lyser™ it provides a useful tool for time periods where [DOC] data could be collected and enables the calculation of [DOC] at a time when the In-Situ Inc TROLL® was deployed but the Spectro::lyser™ was not. Thus, the [DOC] has been modelled quite successfully although the R^2 is currently at 0.751 and there is an issue with small values in particular being calculated lower than the laboratory measurements indicate. The collection of additional

data and the continued analysis may enable improved understanding and reduce the error margins of the model and I suggest here measurements of water table depth on the peat soils to better characterise events and measurements of dissolved oxygen may aid in the improvement of predictive statistic models as some fluvial systems do show a relationship with [DOC] (Daniel et al., 2002).

5.0 Wavelet analysis of time series data

5.1 Abstract

Wavelet analysis was used as an exploratory tool to enable improved understanding of the stream hydrology at Drumtee. Three methods of wavelet analysis (maximums overlap discrete wavelet transforms; continuous wavelet transforms and wavelet coherence transforms) were applied to the time series data to offer a different perspective on this data and indicate how best to process the data contained within these complex hydrological time series.

Maximum overlap discrete wavelet transforms (MODWTs) were used to breakdown the time series datasets into discrete decompositions of the original signal at different periodicities. The MODWT analysis was used to identify short- and medium-term signals within the time series (1 hour up to 512 hours with respect to the oscillating periodicities). Using MODWTs, long period oscillations in the pH, SC and [DOC], at periodicities of about 256 hours to 512 hours were identified, which corresponded to extended periods between events when particular long-term responses were observed. Additionally, MODWTs were used in conjunction with continuous wavelet transforms (CWTs) to identify diurnal variations in the water temperature and pH data sets. In this project the MODWTs and CWTs were used to analyse the same datasets as a form of complementary analysis.

Finally, wavelet coherence transforms (WTCs) were used to compare the [DOC] time series with four other hydrological time series (water temperature, pH, SC and discharge). The main feature extracted using the WTCs was that [DOC] had a stronger relationship over a short period (spring to early summer) with all of water temperature, pH and SC than during the remainder of the year, although a satisfactory mechanistic reason for this was not identified.

5.2 Introduction

Having generated the high-resolution time series data (as presented in chapter 4) consisting of thousands of data points the next question was how can this data be effectively analysed? Comparing two variables on a scatter plot is the simplest method for identifying differences between different time series. However, this approach is only useful for two variables that are strongly linked to one another over the full length of the time series. A different approach is required that is capable of identifying potentially transient relationships in time series or instances where signals are composed of multiple signals and where the total signal might not reflect a particular component's contribution.

To identify specific components of a signal it is often beneficial to analyse datasets in both the time domain and the frequency domain. In the previous chapter I focussed on time domain analysis and here I present a frequency domain analysis. Comparisons of time and frequency domain data analysis are probably most common in fields such as electrical engineering (Mallat and Hwang, 1992, Gu and Bollen, 2000) as signals passed through electrical networks and components often have steady-state signals where it is useful to analyse in the frequency domain but there have also been examples in diverse research fields including wind research and simulation (Yamada and Ohkitani, 1991, Kitagawa and Nomura, 2003, Alam et al., 2014) and soil sciences (Wolf, 1987, Darbre and Wolf, 1988, Bernal and Youssef, 1998). Frequency domain analysis is useful in some situations for probing regularly occurring frequencies within a signal.

Two of the most common methods of statistical analysis in the frequency domain on signals are Fourier analysis and wavelet analysis. Fourier analysis describes the decomposition of a time series into a sum of sinusoidal components (Bloomfield, 2004). Fourier analysis has already been used with success to analyse CDOM using the Spectro::lyser™ deployed in lake Tämnaren (a eutrophic, humic lake in Sweden) (Müller et al., 2014), where the Fourier analysis was used to show that rapid variations in the weather had little impact on the in-lake DOC production. However, for this project a decision was made to analyse the [DOC] signals using the less explored wavelet analysis.

An interesting and informative paper by Sifuzzaman discusses the advantages of wavelet analysis for many systems compared to Fourier analysis and lists a multitude of different applications that Wavelet analysis can be applied to – including data compression, cell membrane recognition and in the financial sector (Sifuzzaman et al., 2009). Wavelet analysis has the benefit, when compared to Fourier analysis, of being able to analyse non-stationary time series (Huang et al., 1998, Sifuzzaman et al., 2009). In colloquial terminology stationary time series refers to time series where properties such as mean variance and autocorrelation remain constant over time (Nason, 2006b, Jebb et al., 2015). In a time-series analysis stationarity can be beneficial because it makes predictions easier and more robust (Jebb et al., 2015). One feature of stationary time series is that the distribution is not dependent upon time (Stock and Watson, 1988). It should be noted that non-stationary is not the same as random – for example white noise is a random process but its mean, autocorrelation and variance remain largely unchanged over extended periods (Jebb et al., 2015). Most time series generated by measuring natural environmental variables will therefore fall into the non-stationary category as they are subject to change. For example, large variations in discharge are measured at different timescales depending on local weather and climactic conditions. As a consequence of the variables measured here ([DOC], discharge, water temperature, pH and SC) all falling into

the non-stationary time series category wavelet analysis was chosen over Fourier analysis. Further with wavelet analysis it is possible to analyse a signal over a localised section of a larger time series (Grinsted et al., 2004, Alam et al., 2014), but not so with Fourier analysis.

Wavelet analysis enables the various frequency components of a signal to be identified individually (Graps, 1995) and is therefore a powerful tool for analysing complex data sets in the frequency domain, identifying recurring patterns in the data and categorising these by periodicity. Periodicity describes the time period over which a pattern at a particular frequency or wavelength manifests itself such as an hour, day, month, year or any other time period (Schleicher, 2002). For example, temperature will usually have a diurnal cycle where temperatures will increase during the day and decrease at night. This diurnal cycle would produce a signal with a periodicity of 24 hours. However, in temperate regions there is also usually a seasonal variation in temperature where the summer months are warmer and the winter months are colder. Therefore, seasonal variation will produce a signal with time periodicity of 1 year. However, there may be other patterns within the signal that are occurring over shorter periods of time and that do not exist throughout the entirety of the signal. Generally, wavelet analysis is conducted by a computer because the large number of calculations required for a useful output make calculating the answers manually too time consuming.

Therefore, although it may be possible to filter out the seasonal variation and other signals of different periodicities before analysing the higher frequency components of the time series as an alternative approach it is not done here. The filtering of the signal might be achieved through the use of harmonic regression or splines or some other technique. However, wavelet analysis is automatically filtering different periodicities and this also saves the need to model different periodicities. Filtering data prior to performing a wavelet analysis also overlooks the main benefit of wavelet analysis which is the ability to analyse multiple periodicities simultaneously. Additionally, although some components that we may wish to filter out such as seasonality would be relatively easy to identify and model some other components may be harder to filter out, especially if they are transient relationships.

As the wavelets can analyse across a range of periodicities by identifying frequency variations embedded within a signal there is no need to de-trend the signal for the wavelet analysis itself. Furthermore, any errors in modelling a pattern intended to detrend the signal risk introducing artefacts into the time series, which could result in erroneous coherent regions being identified.

Wavelet transforms were developed in the 1980s as a method of analysing seismic waves (Morlet et al., 1982, Grossmann and Morlet, 1984, Kumar and Foufoula-Georgiou, 1997), which proved hugely beneficial as seismic data contains data with many different frequencies, some of which overlap one another for transient periods. CWTs have been used for many different types of studies including for the analysis of the measurements of rainfall and runoff over an extended period of decades (Nakken, 1999), flood plain analysis of the river Nile (Fraedrich et al., 1997) and the analysis of stream discharge (Bradshaw and McIntosh, 1994).

One type of wavelet analysis used here is wavelet coherence transforms (WTCs) which allows for the analysis of the dependencies of two signals (Cazelles et al., 2008). Wavelet coherence phase analysis is similar to Fourier analysis but can analyse coherence as a function of time (Lachaux et al., 2002). WTCs are used for the comparison of two variables and identifying the strength of relationship that exists at a variety of temporal time scales between these two time series and this approach has been used in the context of a number of geographical studies (Torrence and Compo, 1998, Grinsted et al., 2004). WTCs are used here to help identify the strength of relationships between different variables such as [DOC] and temperature at a range of temporal time scales.

Some specific examples of WTCs being used are: for ecological studies (Cazelles et al., 2008, Polansky et al., 2010); analysis of the influence of weather and vegetation on soil CO_2 production (Vargas et al., 2010) and in oceanographic research (Jevrejeva et al., 2003, Doglioli et al., 2007).

There are two types of wavelet analysis: continuous and discrete (Heil and Walnut, 1989). Continuous wavelet analysis is better suited to signal analysis and therefore might seem to be the most logical approach used for this chapter, whereas discrete wavelet analysis is often used for applications such as signal encoding and image compression (Rioul and Duhamel, 1992). (Cazelles et al., 2008). However, discrete wavelet analysis in the form of MODWTs was found to be a useful form of analysis as it provided a useful visual tool for identifying spectral features at approximate periodicities.

5.2.1 – Purpose of using wavelets

Here wavelet analysis is used as a form of exploratory analysis to identify transient relationships, periodicities and modes of variation. whereby after reviewing the data initially wavelet analysis is used to analyse the data in the frequency domain to identify transient relationships (such as those resulting from variables such as seasonality) that were not identified in the initial analysis, which looked at the entire time series at once. Once these

patterns had been identified the original data was reanalysed to gain a better understanding of what was occurring at these shorter timescales.

5.2.2 – Interpretation of MODWTs

MODWTs provide an interesting analysis of time series data in that they enable a visual breakdown of different periodicities into the individual components. MODWTs are arguably the easiest form of wavelet analysis to read. In the example provided (Figure 5.1) there are twelve rows shown. The top row "Sum" is the original signal being analysed. The following 10 rows are listed D1 to D10. Each of these relates to variation at a particular periodicity, with each number relating to a decomposition level of the form $D_n = 2^n/2$. (The value is divided by two because data points are spaced half an hour apart so there are two data points per hour). For example, D1 is the decomposition at the periodicity 2^1, which is equal to 1 hour. D2 is the decomposition at periodicity of 2^2 (2-hour periodicity) and D3 is the decomposition at periodicity of 2^3 (4-hour periodicity) and so on.

It should be noted that a signal does not have to fall exactly on the target periodicity to be identified. In Figure 5.2 a synthesised signal at a single periodicity is most strongly identified by decomposition level D5. However, a weaker signal is picked up in components D4 and D6. In effect, the discrete decomposition levels will pull in periodicities that are close to that value.

Finally, the Sn number (S10 in Figure 5.1 and S6 in Figure 5.2) show the remainder of the signal that has not yet been broken down into a decomposition periodicity. In this chapter, the signal decomposition is taken up to D10 for the hydrological variables [DOC], pH, SC, discharge and water temperature, which is equivalent to 512 hours. Not showing decomposition levels greater than D10 prevented the visualisation of the MODWTs from becoming too compressed with additional D levels to show. Furthermore, although MODWTs could be used to show annual variation it was deemed unnecessary here as the hydrological signals that contain seasonal variation ([DOC] and water temperature) are clearly observable within the time series analysis without using wavelet analysis. Both have greater values in the summer months than winter months.

MODWTs are a very useful method for visualising data because the time series is decomposed into easy to read segments. Rather than using colour coding (as for CWTs and WTCs) the various periodicities are shown individually as time series one above and below each other. MODWTs enable the dissection of a time series to determine frequency components where there are patterns. The drawback with this approach (especially as

we look at longer components of the signal) is that it will not reveal specifically what the frequency of oscillation is. It can offer an approximation but although a signal is identified by D10 it does not mean a signal is oscillating at exactly 512 hours – it may actually be 442 hours but it is being picked up by the D10 decomposition level. This potential drawback in terms of accurately determining the periodicity using discrete wavelet analysis has been discussed in other environmental studies (Lark, 2006).

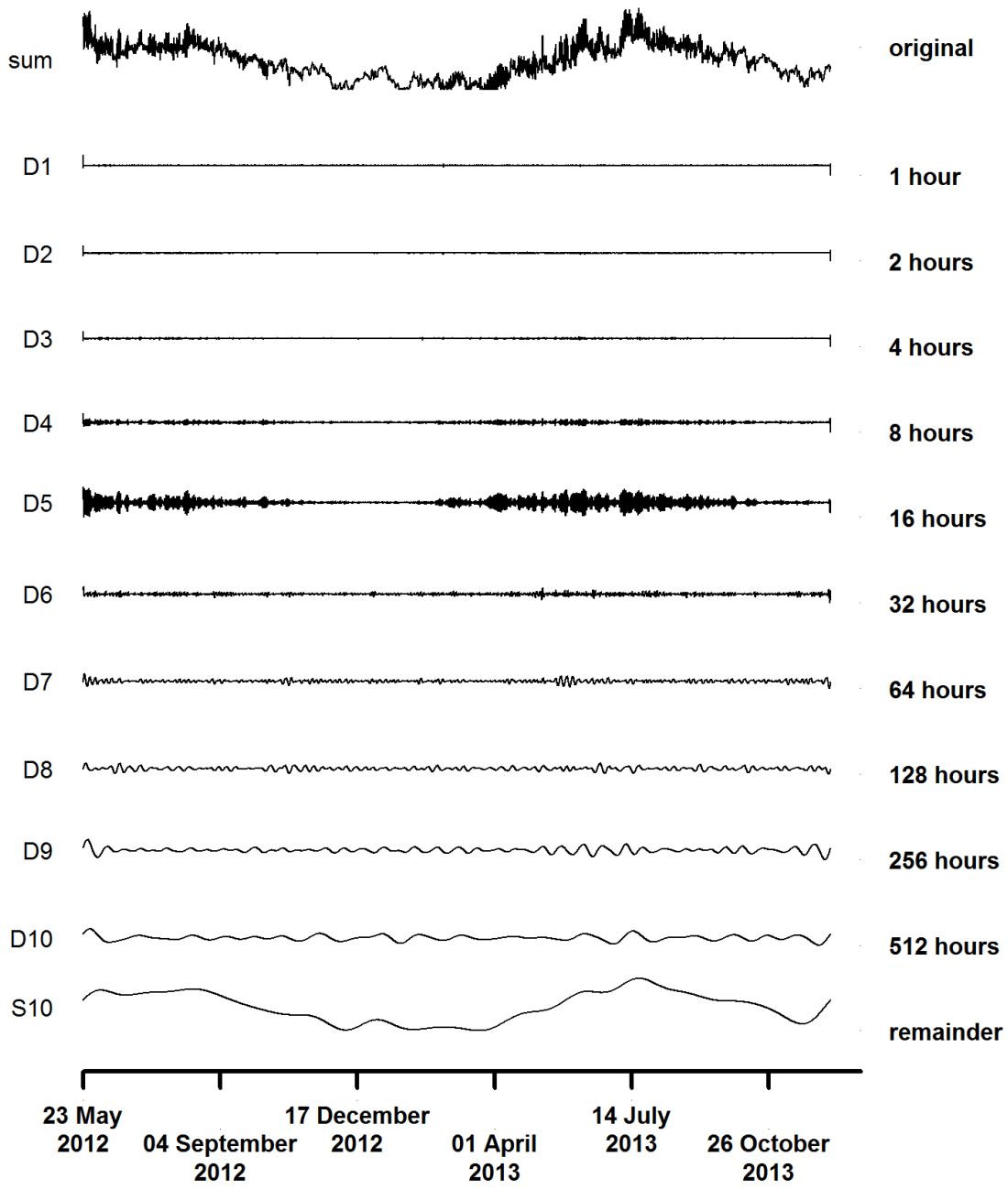


Figure 5.1 – Example wavelet analysis of water temperature
The signal shown at the top is broken down into various levels of decomposition showing different periodicities representing different lengths of time. The signal at the bottom "Remainder" is the left-over signal that has not been broken down yet.

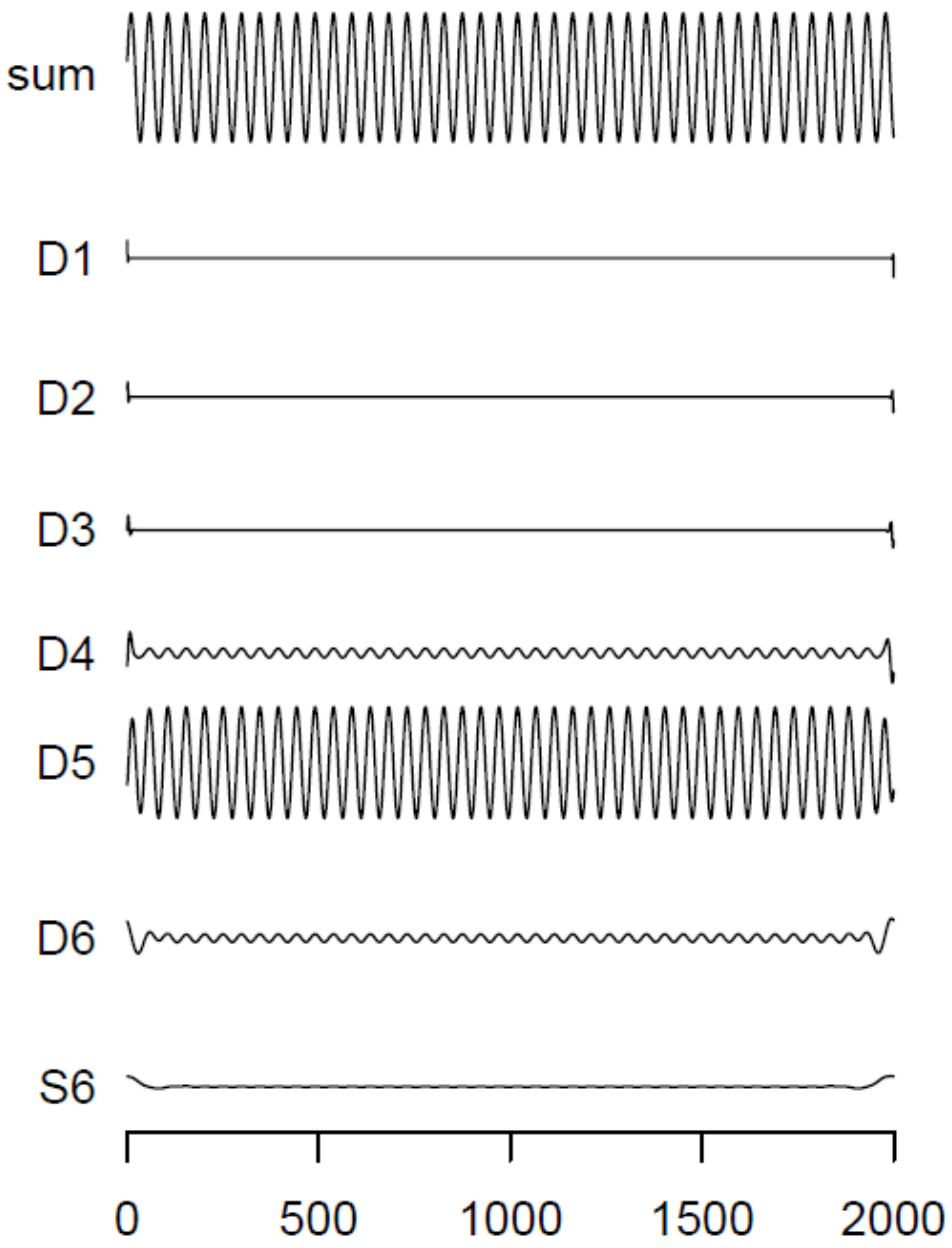


Figure 5.2 – Synthesised 24 hour signal
Showing the decomposition of a synthesised 24 hour sinusoidal signal into the different periodicity components. As can be seen the signal is picked up most strongly by the D5 component (16 hour oscillation) but it is also identified more weakly by the D4 (8 hour oscillations) and D6 components (32 hour oscillations).

5.2.3 – Interpretation of CWTs and WTCs

CWTs (continuous wavelet analysis) and WTCs (wavelet coherence analysis) produce very similar looking graphs, although the interpretation of these graphs is slightly different. CWTs are the product of a univariate analysis, while WTCs are comparing the coherence between two different time series. Both WTCs and CWTs use colour coding to indicate the strength of relationships and are also similar as they both have the original time series scale along the horizontal x-axis and the periodicity on the y-axis. The difference is that two variables are being compared on this occasion and it is their similarity at a particular periodicity that is being compared. The colour coding is defined as colder blue colours indicate a poor relationship (low coherence) and warmer red colours indicate a stronger relationship (high coherence).

To better constrain the coherent regions, Monte-Carlo simulations are used to highlight regions of high coherence by surrounding them in black contours. Monte Carlo simulations are important in this context because they allow the assessment of the statistical significance against red noise background (Torrence and Compo, 1998, Grinsted et al., 2004). To achieve these contours outlining the regions of significance in the wavelet output is compared with random spectra (RIVERA et al., Grinsted et al., 2004). This comparison enables the areas outlined in red with a strong relationship in the output wavelet to be assessed to determine whether they are in fact statistically significant. As the comparison is with random spectra there is a possibility of errors when identifying the regions of significance. To reduce the likelihood of errors associated with the contours the Monte-Carlo simulations are repeated more times. The more times the simulations are repeated the greater confidence can be ascribed to the contours of significance.

CWTs can be used to extract similar features to MODWTs in that they will identify if a signal has a component at a particular periodicity. However, they have a greater scalability in that they can identify variation at almost any periodicity. Here, the coherence of a relationship is indicated by the colour coding of the plot, with the areas in warm colours such as red indicating periods of strong coherence, while the regions highlighted in cooler colours, such as the blues, have weaker coherence. For CWTs colour coding is used rather than the time series decomposition diagram approach of showing the strength of relationship at each individual periodicity as used for MODWTs as there is no practical way to show hundreds or even thousands of periodicities on a single page using that same layout. A CWT or WTC graph can be thought of as 3D plot, but where the colour is replacing a visual z-axis. As indicated previously hot colours such as red and orange indicate high values (i.e. strong coherence) and cold colours such as blue and green

indicate low values (i.e. weak coherence). There are examples in papers of CWTs presented in 3d plots (Kingsbury, 1998, Gdeisat et al., 2006, Rucka and Wilde, 2006).

One feature of CWTs and WTCs is the cone of influence (COI). Data within the cone of influence (COI) is shown in the relevant figures and interpreted but data outwith the COI interpreted more cautious approach has to be taken when interpreting the data below the cone of influence (Grinsted et al., 2004). The relationships outwith the COI are not necessarily wrong – they simply have a greater potential error associated with them.

The usefulness of CWTs is illustrated in Figure 5.3. Here a signal combined from two components – a randomly-generated signal and a sinusoidal signal- are spliced together. The CWT is able to identify the periodicity of the sinusoidal signal and also identify the time period of the graph where it is present. In the example presented in Figure 5.3 a WTC analysis was not conducted because there wasn't a second time series to compare it with. However, an example using a WTC is shown in Figure 5.4

The benefit of the WTC over CWTs is illustrated by figure 5.4. Two signals with an identical middle phase but with no signal at a fixed periodicity are compared with one another. The CWT produces a less defined outline of significance than when the sinusoidal signals was added to the original signal because there is no oscillation at a particular periodicity. In contrast the CWT was able to identify a signal in Figure 5.3 because there was an oscillation at a particular periodicity.

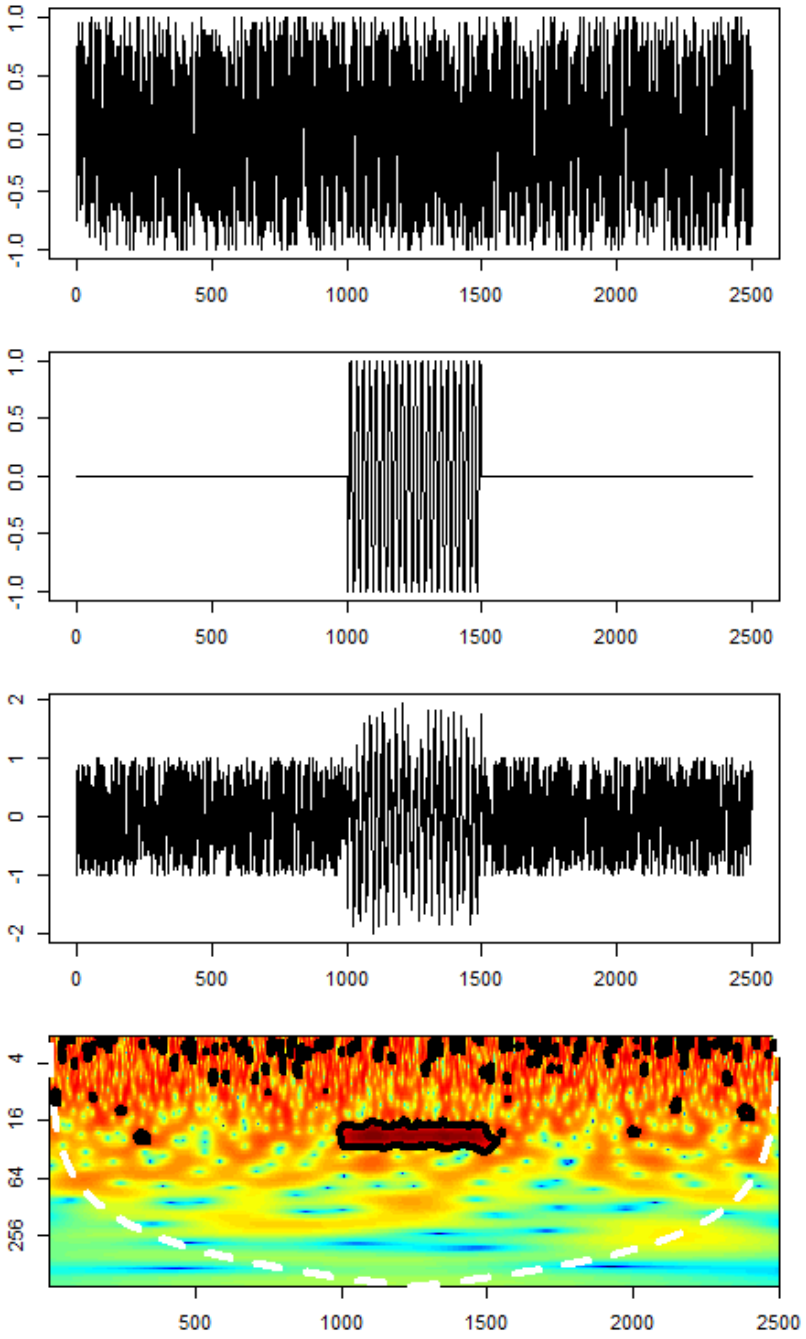


Figure 5.3 - CWT of synthesised signal
Four plots are shown. The top signal is randomly generated in Excel, where the second plot is a sinusoidal signal with a time period of 24 with respect to the x-axis. The third plot is the two signals added together. The fourth plot is the CWT of the third graph, which highlights the area between 1000 and 1500 on the x-axis where the sinusoid is present within third signal.

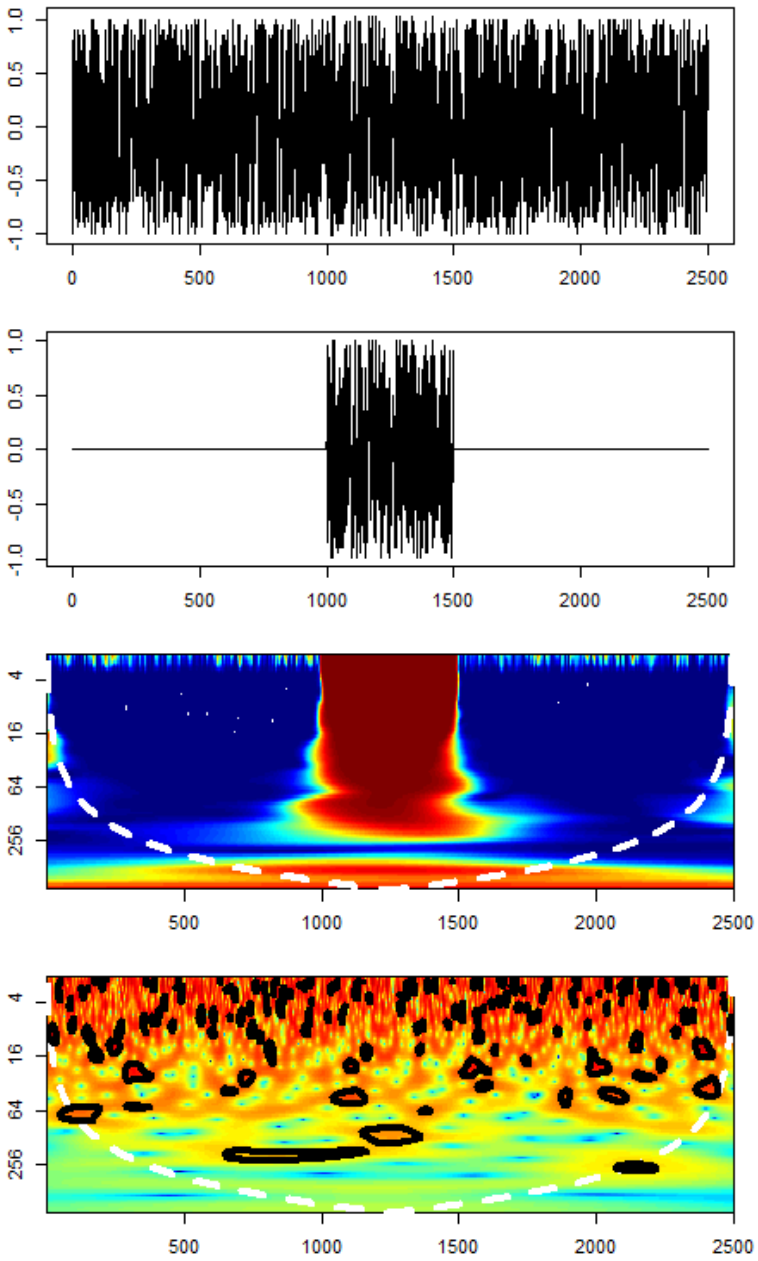


Figure 5.4 - CWT and WTC of synthesised signal
Four plots are shown. The top two signals show randomly generated data in Excel, where the data between 1000 and 1500 on the x-axis is identical. Using WTC in the third graph the region of coherence highlighted in red (strong correlation) can be clearly observed, which identifies the time period over which the two signals are identical. By contrast, the bottom image, which is a CWT of the top signal shows several highlighted regions where a signal of a particular periodicity is identified. As the signal being analysed is made up of random values the identified regions only correspond to small areas of the wavelet plot and no specific information is periodicities are identified.

5.3 Methods

5.3.1 Selecting the data for analysis

The data used to for the wavelet analyses are taken from the same data sets as those presented in chapter 4. Data collection from the sensors at Drumtee is not contemporaneous and is therefore rounded to the nearest half hour as noted in chapter 3 to allow for the time series produced to be synchronised for the wavelet analysis. The rounding of data to the nearest half hour was described in chapter 3 but it is repeated because it was crucial for wavelet analysis because the analytical packages such as biwavelet cannot perform the calculations if the timings in the data series are not synchronised. There is no reason to suspect that the error introduced into the time series by adding or subtracting 15 minutes will significantly influence the analysis.

Having established how to synchronise the time series the next step was to make a decision as to how much of the time series was to be analysed. There is data available from 23 May 2012 up to 16 December 2014 but as discussed in the previous chapter there are data gaps. Prior to 12 December 2013 there was almost 20 months of data where the largest data gap was only 56 hours in the Spectro::lyser™ data, from 28th to 30th of July 2013. This 20 month period contains over 100 events and means that almost all the data has been recorded directly in the field. Although this amount of data is not likely to clearly illustrate annual variability it is the short term changes that are of primary interest in this instance. Gap filling was required for the SC for the period from 20 June 2013 to 30 July 2013 when the SC sensor was not turned on. Using the relationship between pH described in chapter 3 this section of SC data was gap filled. Modelling the SC on pH data could potentially introduce an erroneous artefact into the SC data but this trade-off for a longer time series was considered worth it and interpretation of relationships between [DOC] and SC for this section of the data should be treated more cautiously.

The data is presented for the three types of wavelet analysis with the x-axis presented in a linear scale and the y-axis presented in a 2^n axis. For the x-axis, representing the time series it is important to equally space the time series without compressing the latter half of the graph to preserve the same detail during the initial period of the time series and the latter part of the time series. By contrast it is not as important to gauge the precise periodicity on the y-axis for longer time periods. For example, strong coherence at a periodicity of 1 hour compared to 4 hours is important to know as this is likely representing a different process in the catchment. However, strong coherence at 101 hours compared to 104 hours is a less significant difference.

Two different packages were used for wavelet analysis within this project. MODWTs were produced by using the wmtsa package. CWTs and WTCs were produced using the biwavelet package. For WTCs The Morlet wavelet package was selected as this has been used by other researchers and suggested as a useful tool for feature extraction (Grinsted et al., 2004).

5.3.2 Discrete wavelet analysis

The MODWT analysis was applied to each of the five variables measured at Drumtee: [DOC], SC, water temperature, pH and discharge. The annual seasonal variations are already known from [DOC] and water temperature so features with time frames this long are not analysed because it makes the graph more difficult to read. (Every MODWT component adds a line to the graph, which compresses the vertical detail of the different components). As the longer term trends are well-established as they can be monitored using manual sampling. The greater interest in the wavelet analysis is the short term trends, so not going above a decomposition level of D10 has been deemed unnecessary. A decomposition level of D10 is equivalent to 512 hours or about 3 weeks. Any signals of particular interest are highlighted in the graphs and these results are discussed.

5.3.3 Continuous wavelet analysis

CWT was also applied to the same five time series as the MODWT analysis: [DOC], SC, water temperature, pH and Discharge. The CWT was applied to determine whether there were any repeat oscillations at particular periodicities. Biwavelet automatically colour codes the graph with coherence (repeat patterns at particular periodicities) being assigned warm red colours and outlined in black contours generated using Monte-Carlo simulations. Areas where these patterns are not identified are represented by cooler colours and they do not have any contours outlining these regions.

The data is collected at 30 minute intervals and therefore coherence measurements below about 4 hours are hard to identify visually on the wavelet coherence plots and do not necessarily provide powerful analysis because the amount of data used to calculate each area of coherence is too short.

5.3.4 Wavelet coherence analysis

A WTC was applied to the [DOC] time series in conjunction with the four other variables (SC, water temperature pH and discharge). To more clearly identify the strongest relationships 100 Monte-Carlo simulations were carried out for each of the WTCs produced. Monte-Carlo simulations are represented using black lines as contours around the regions in red in the WTCs. The greater the number of Monte-Carlo simulations the

more these lines will be constrained and the greater confidence can be placed in the analytical results.

The strength of the relationship in a WTC plot (which is outputted in colour) is stored as a numerical value in the root mean squared component of the wavelet coherence analysis as a matrix and is accessible using the command: wtc.1$rsq[n1,], where wtc.1 is the name of the wavelet analysis between two variables and n1 is the row number being measured. The extracted rsq values were scaled between 0 and 1. The 0 value represented the smallest coherence measured between the two variables and 1 represented the strongest coherence between those two same variables. Data was extracted from the time series by identifying the data in the time series that corresponded to a period with an rsq value above a certain threshold by identifying a period where it was consistently above a certain value (see Figure 5.5). From this a scatter plot was produced, using the data from these more coherent time periods, and this could be compared to the scatter plot produced using all the data to identify if there was a stronger relationship between the segment of the time series that corresponded to the high coherence. To further aid comparison, a segment of the time series that corresponded to a period of very low coherence in the WTC was extracted and compared with the scatter plot produced from all the data. This comparison is shown in the results from Figure 5.21 to Figure 5.24

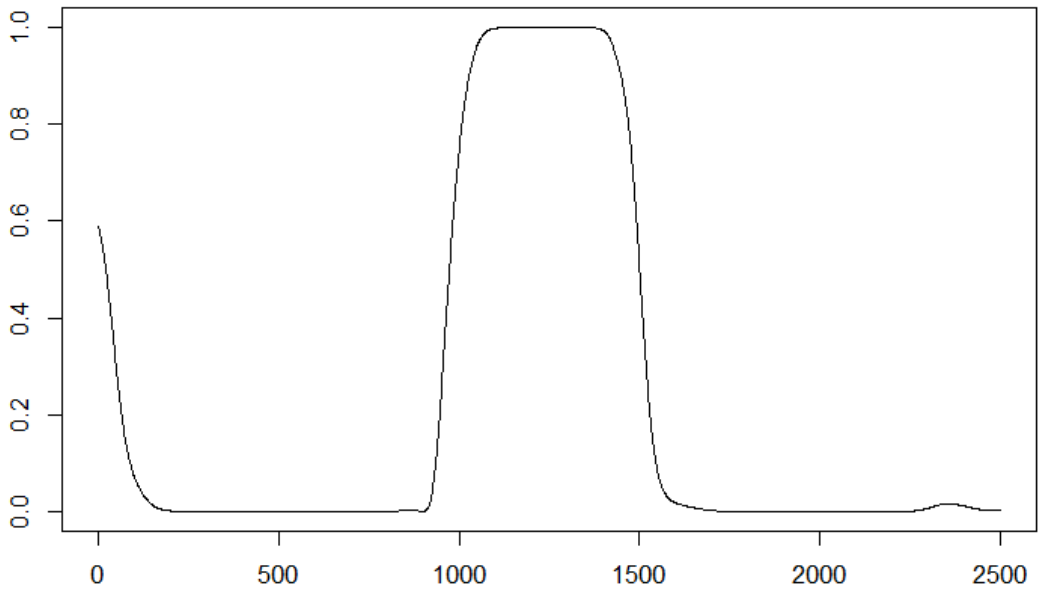


Figure 5.5 - RSQ extract from a WTC
Cross sectional area of high coherence from WTC shown in Figure 5.4 with a periodicity of 64. The region in dark red that was in Figure 5.4 is now represented in this figure as the high values at or near to one. The dark blue area in Figure 5.4 is represented here as values on the y axis as being at or close to zero.

5.4 Results

As a reminder, MODWTs were applied to single variables to explore variation at different frequencies (or periodicities), CWTs to explore single variables but with less restrictions on the range of periodicities analysed although this can make them harder to interpret and WTCs to explore relationships between two variables at different periodicities.

5.4.1 Maximal overlap discrete wavelet transform

MODWTs were produced for each of the five time series (water temperature (Figure 5.6); pH (Figure 5.7); SC (Figure 5.8); discharge (Figure 5.9) and [DOC] (Figure 5.10). Regions of interest are highlight with coloured circles. Different periodicities are highlighted using different colours but a consistent colour grading between the different figures was not possible because the regions of high and low coherence are relative to one another.

The MODWT for water temperature (Figure 5.6) does not identify any strong signals at decomposition levels with periodicities of 8 hours or shorter. However, it does identify an intermittent strong signal at a decomposition level of 16 hours (as noted before, the variation may not be exactly at this frequency). There are signals identified at the oscillating periodicities of 256 and 512 hours but most of the signal is still contained within the remainder of the original signal that has not already been decomposed, which includes the annual variation of the water temperature.

The MODWT for pH (Figure 5.7) identified many different component signals of interest. A signal at the D5 component (16 hours) is once again observed and but this only exists for short time periods so may be coincident with a diurnal control by photosynthesis consuming DIC influencing the pH signal. It is noticeable that this signal is stronger during the summer. Periodicity components are identified at almost all the different decomposition levels with some of the longer periodicity signals highlighted in the plot seeming to coincide with lengthy gaps between events when pH remained at an elevated value. The remaining signal (S10) is almost unrecognisable from the original signal, indicating that most of the features have been extracted by decomposition.

The MODWT for SC (Figure 5.8) is largely featureless at small periodicities but more features are observable up until the largest periodicity of 512 hours. The regions highlighted in red in the graph show a strong periodicity signal at 512 hours that coincides with much of the 2013 summer, which was relatively dry and the green circle indicates an area of this time series where there were more frequent events and the maximum SC values recorded were lower.

As the discharge is respondent primarily to rainfall we might not expect to see many periodic frequency functions so the wavelet analyses for discharge are can be quite hard to interpret. However, the MODWT plot for discharge (Figure 5.9) showed signals at the decomposition levels between 16 hours (D5) and 64 hours (D7). When the distribution of event lengths is identified (Figure 5.11) this time period, 16 – 64 hours, dominates event length. The remainder of the signal that was not decomposed (S10) shows a series of "humps" that occur during the times when there were persistent events causing the baseflow of the stream to increase (see Figure 4.1 in chapter 4).

The MODWT for [DOC] (Figure 5.10) does not identify strong frequency components at the shorter decomposition levels (10 days or shorter time periods) indicating that there are few, if any, repeat oscillating signals at high frequencies within the [DOC]. The strongest signals identified in the decomposed signal are the oscillations occurring a periodicity of about three weeks (D10) which appear to correspond to the decrease in [DOC] that is measured when there are long gaps between events during drier periods. A large proportion of the signal is still present in S10, although this is now a much smoother signal than the original with all the short frequency components stripped out.

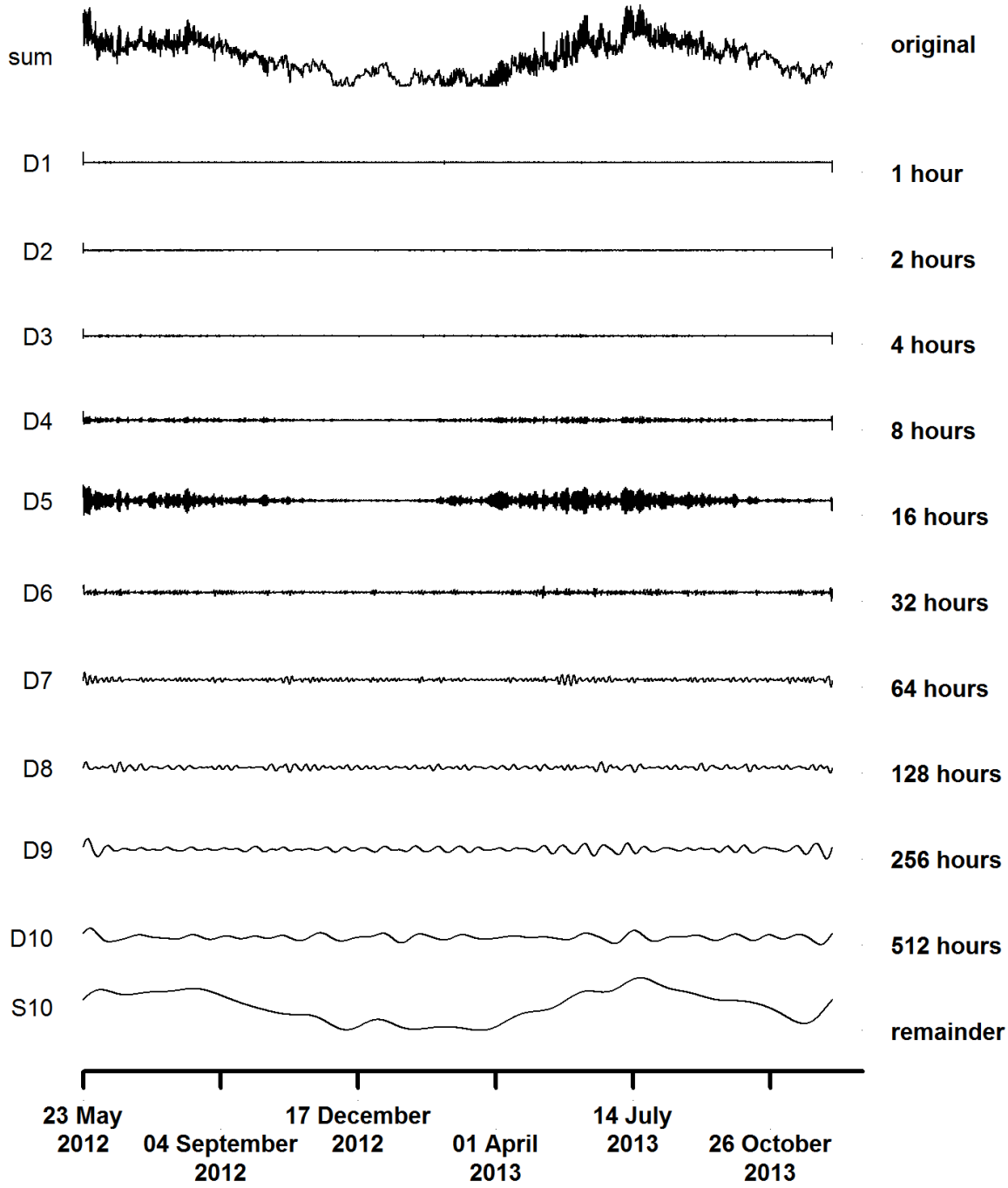


Figure 5.6 – MODWT of water temperature at Drumtee
The D5 component extracted from the signal is identifying the diurnal variation of the cycle. The S10 component at the bottom of the graph is the remainder of the signal that has not been decomposed into the various D components, including seasonal variation (at what would be D15 if it were included). S10 is essentially the same signal as the original (denoted by "Sum") as the signal at the top of the graph but with the high frequency data removed into the different D components.

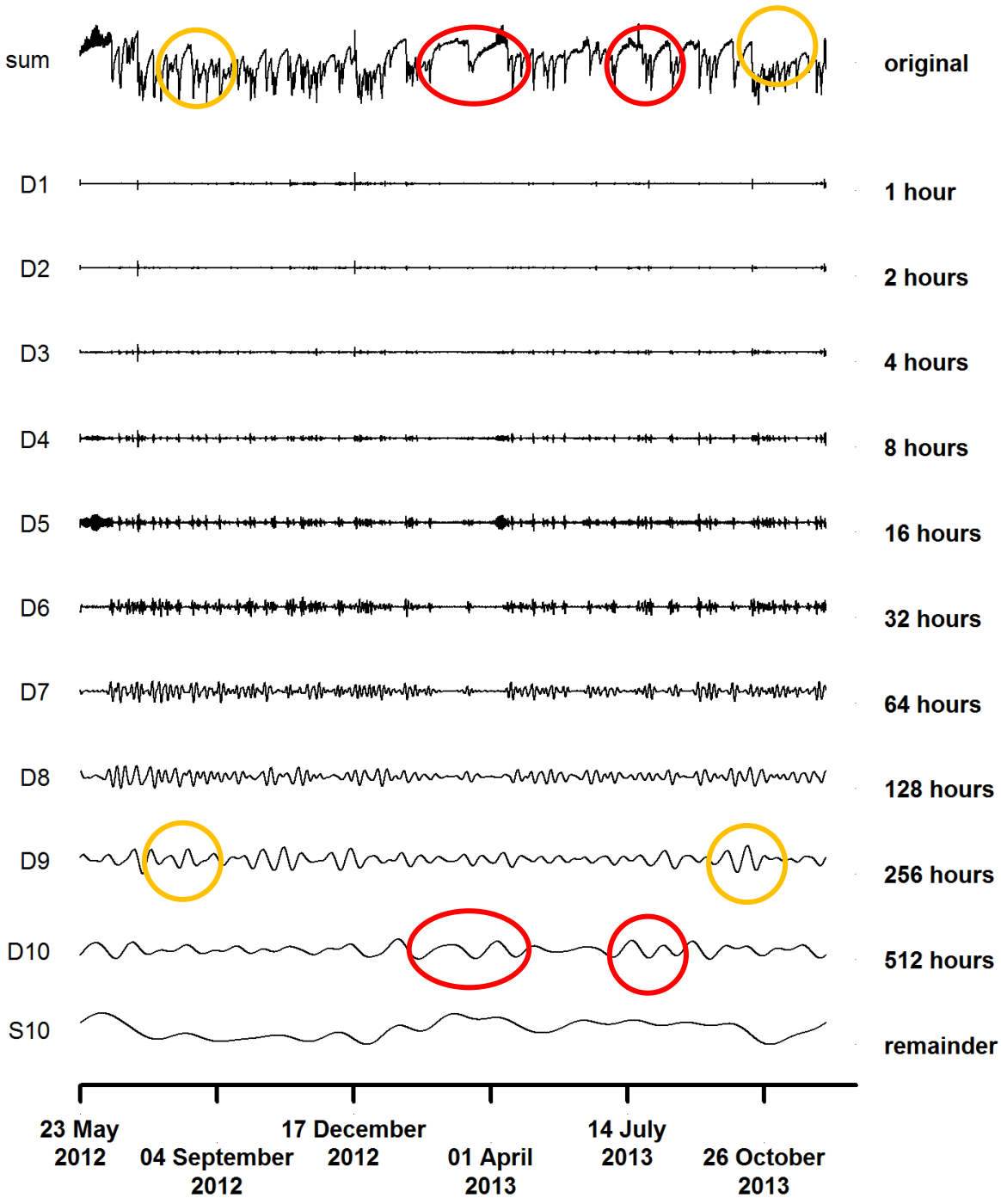


Figure 5.7 – MODWT of pH at Drumtee
The D5 component extracted from the signal is identifying the diurnal variation of the cycle. Signals identified between D8 and D9 (128 and 256 hours) may be linked to individual time gaps between events with a few examples at D9 and D10 identified. The red circles at D10 and the original signal (top) are the longer wave signals that occur when high pH's are measured after periods without events. The orange circles linked to D9 identify signals in the pH time series that are shorter than those in D10 but that are still long enough to be visibly observable on the MODWT and time series.

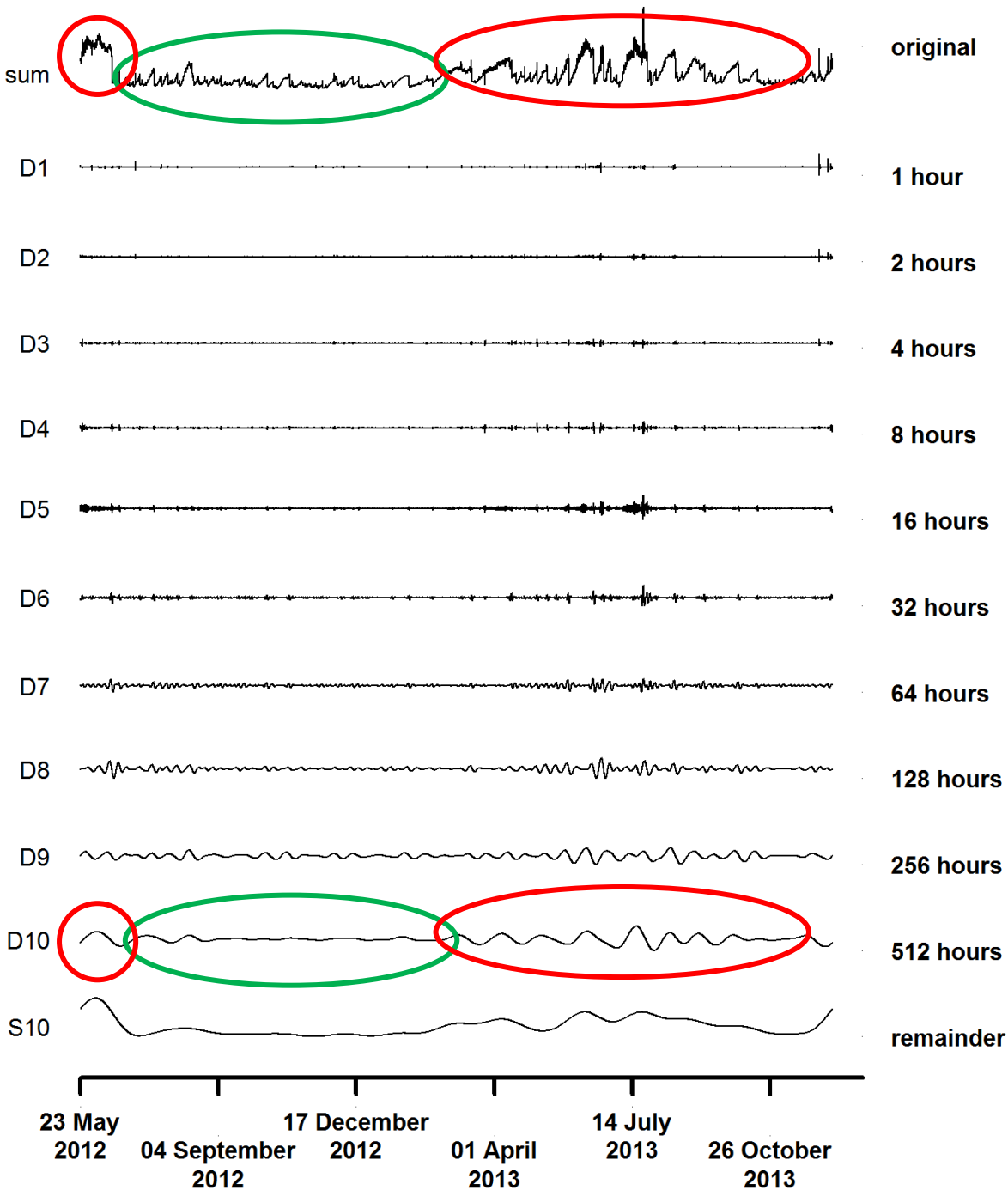


Figure 5.8 – MODWT of SC at Drumtee
The main decomposition level of interest here is D10. During dry periods with no events the continued upwards trend in recorded SC leads to low frequency signals that are identified in D10 and are circled in red, with large amplitude signals. By contrast the area circled in green is characterised with almost no large SC values as frequent events kept SC low and consequently D10 is much flatter at this section.

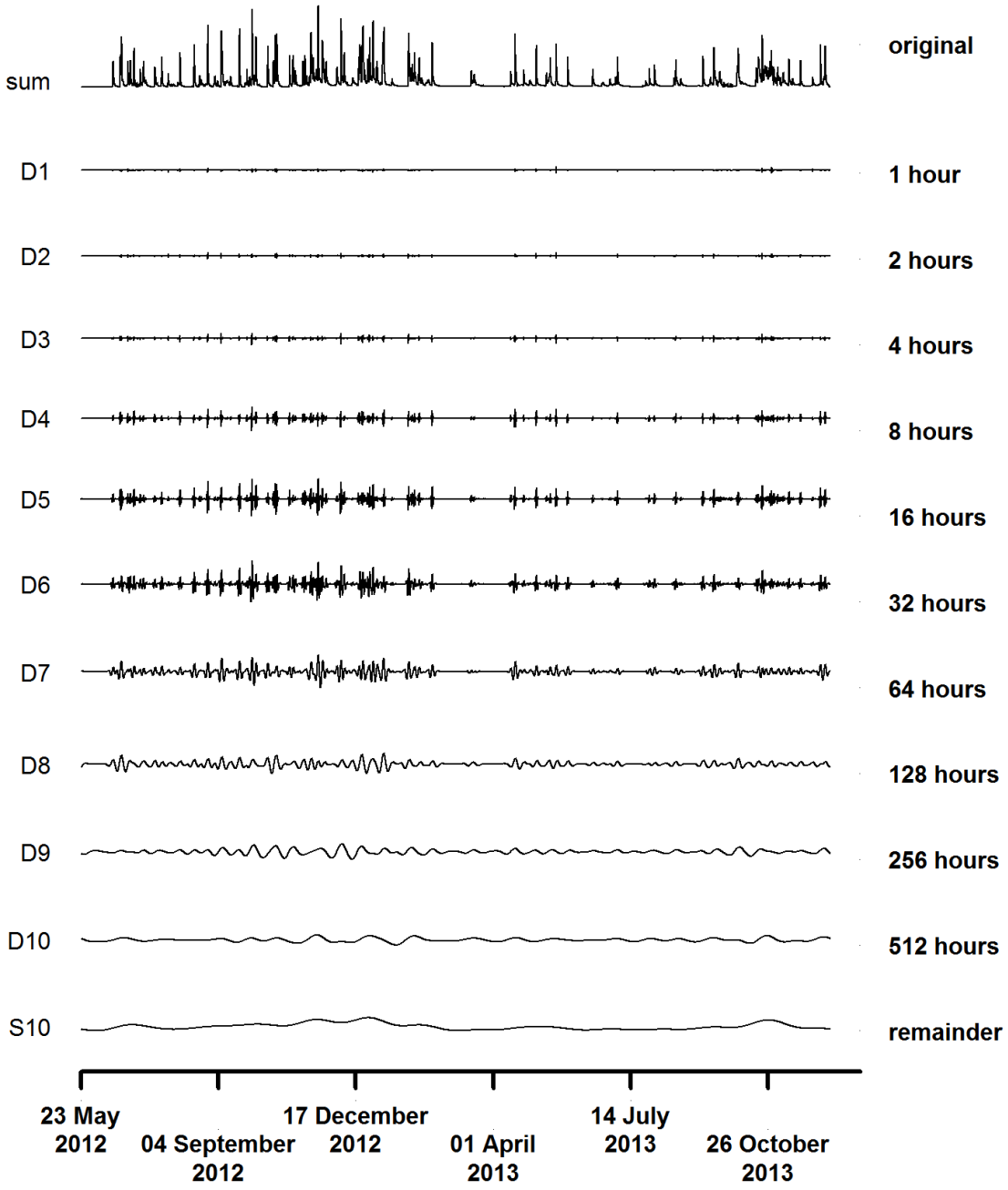


Figure 5.9 – MODWT of the discharge recorded at Drumtee
The MODWT of the discharge recorded at Drumtee is the least interesting of the MODWTs. D4 – D7 (8 to 64 hours) shows signals coinciding with periods at which events can be observed on the time series presented at the top. However, as these are occurring at random intervals there is little of this information that would be of use for a model either of predictions or understanding catchment dynamics.

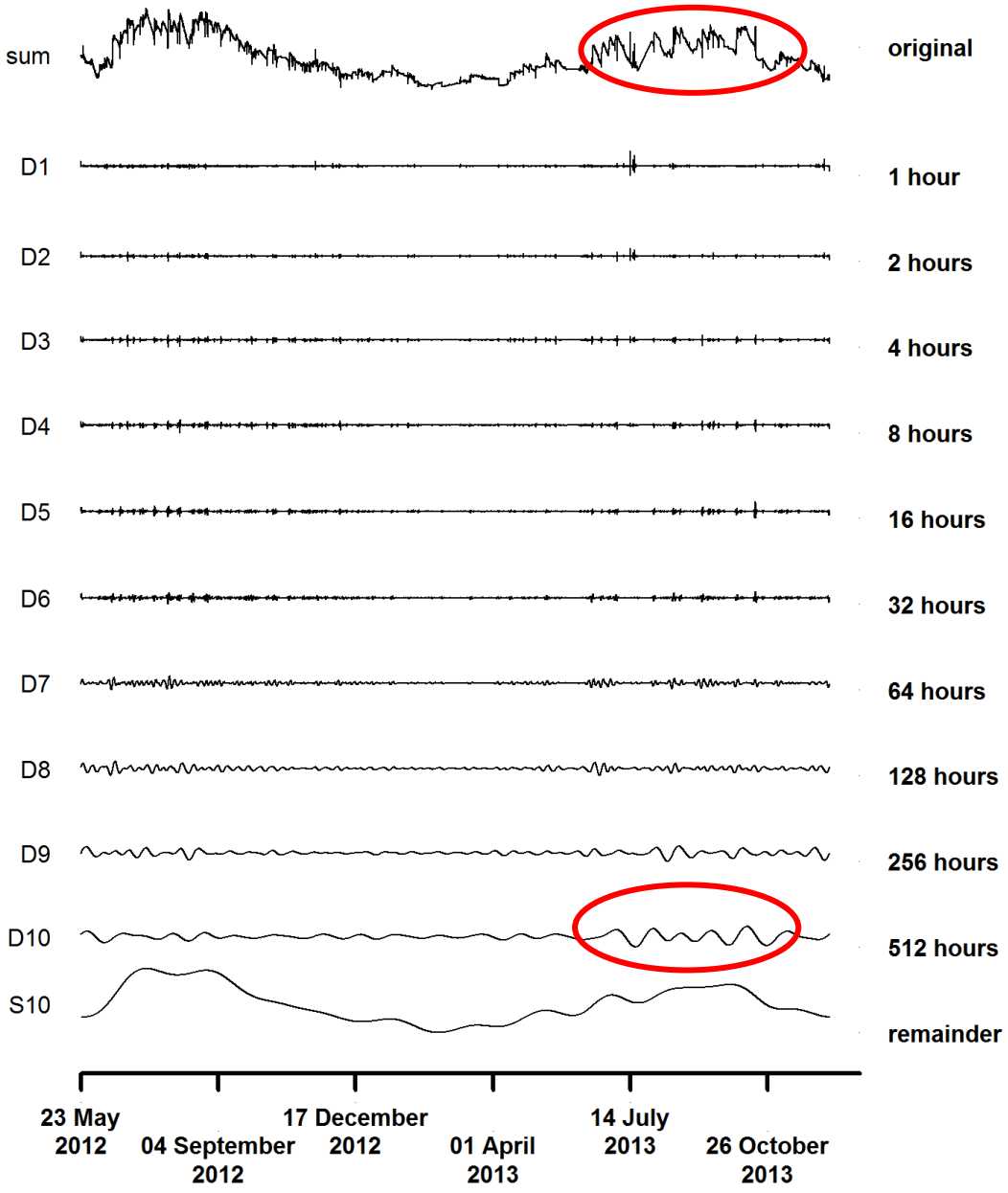


Figure 5.10 – MODWT of [DOC] at Drumtee
The D10 decomposition has been linked for the summer of 2013 with the original time series. During the summer long periods between events coincided with high [DOC] values that gradually decreased during the period without an event until the next event and this produces low frequency signals observed in the D10 decomposition.

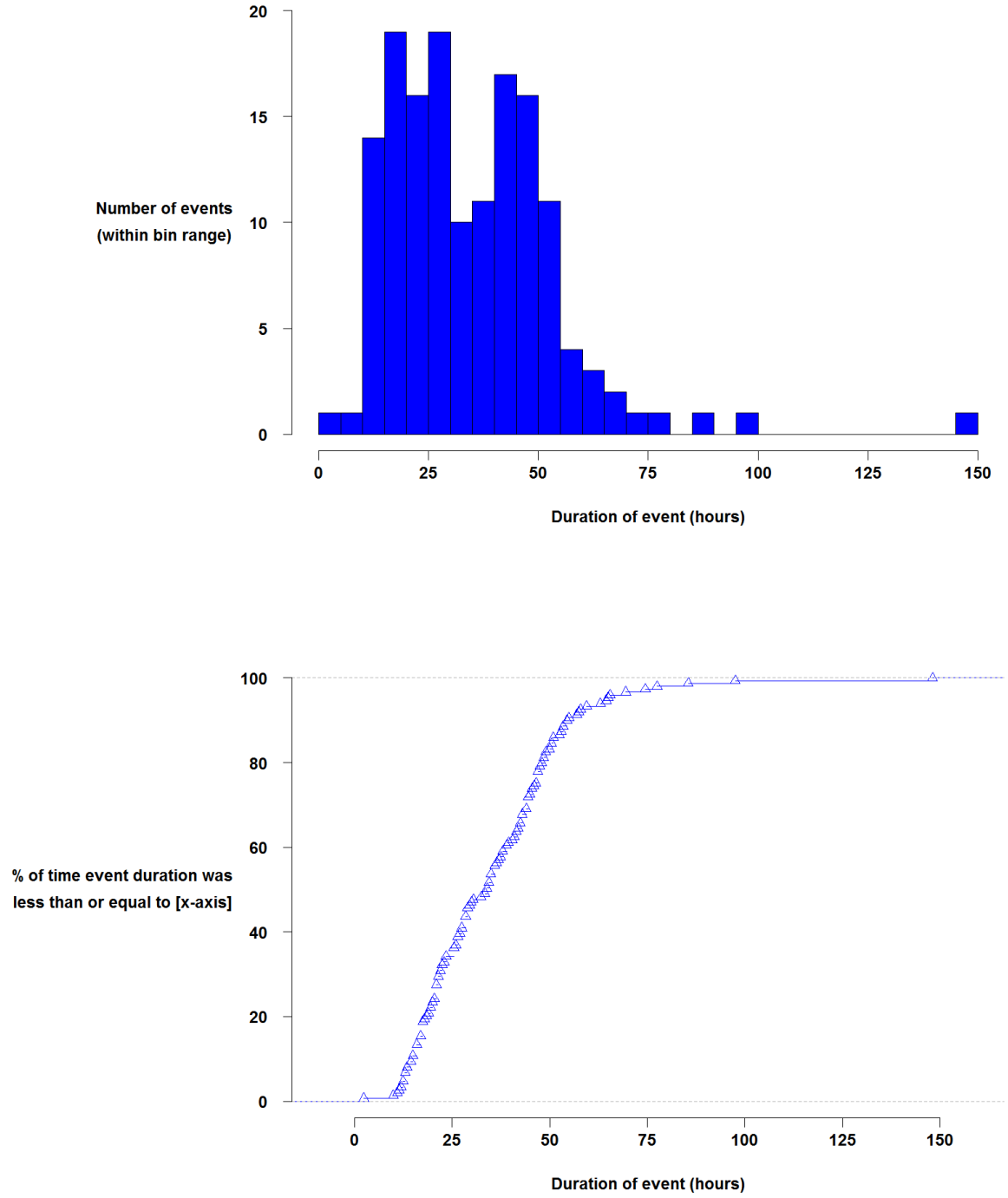


Figure 5.11 – Histogram of discharge durations and CDF
The x-axis is in hours and the number of instances is the number of occasions that an event lasted a particular time. Bin size in this case is 20 hours. The most common events are therefore last for between 40 and 60 hours, which would average out at just over 2 days.

5.4.2 Continuous wavelet analysis

CWTs were obtained for the water temperature, pH, SC, discharge, and [DOC] time series data and the results of these are presented on the following pages.
The CWT related to the water temperature identifies a diurnal variation of the water temperature (Figure 5.12). However, this signal is only identified during the summer months of the hydrological year (April to September), whereas the signal is weaker or not identified in the winter period (October to March). There are a number of other regions identified by the CWT, highlighted in red and with the black outline of the Monte-Carlo simulations, such as a couple of points in the 2013 spring at periodicities between 5 days and 3 weeks but these are small regions.

The CWT produced for the pH data showed the most regions of red, outlined by regions of significance identified by Monte-Carlo simulations (Figure 5.13). The outlined contours of significance appear to be identifiable from vertical streaks that indicate many signal components with periodicities ranging from 32 hours to 3 weeks in duration. These periods of significance also seem to be blending into one another, indicating that there are many periodicity components overlapping one another in the pH time series. A couple of transient diurnal cycles appear to be observable in the pH CWT. These regions seem to coincide with the D5 components observed in the MODWT (Figure 5.7). Indefinable signals with periodicities greater than 3 weeks are rare, with the exceptions being from November 2012 through to March 2013 with a periodicity of slightly less than 6 weeks. This identified signal coincided with a very dry period with only a small number of events and extended periods of increased pH.

By comparison, the SC analysis produced surprisingly few areas that were identified via the contours of significance (Figure 5.14). There is a strong set of signals identified from May 2013 through to September 2013, with a periodicity stretching from 1 week to 6 weeks. Regions of significance were identified as occurring during the summer months of 2012 and 2013. The former does fall outside the COI for most of these regions but as the region of significance is observed at a similar time period on both years it provides reassurance that this is a recurring relationship.

The CWT of the discharge identified contours of significance stretching vertically from periodicities between 8 hours to 3 weeks (Figure 5.15). These signals relating to discharge occurred during periods when there were events. During the summer of 2013 there were relatively few areas identified by the Monte-Carlo simulations and therefore the fewest signals of a particular periodicity. It is noteworthy that the Monte-Carlo simulations in the CWTs identified regions of a strong signal in discharge (Figure 5.13) not SC (Figure

5.14), indicating that there is a stronger link at various time scales between pH and [DOC] than there is between SC and [DOC].

The [DOC] produced the most featureless of the five CWTs (Figure 5.15). Only a very small number of bands were identified via the Monte-Carlo simulations. However, the colouring of the graph indicates that relationships are stronger the longer the periodicities of consideration. For example, the shorter periodicities are represented primarily by cool blue colours, whereas the longer periodicities are represented by warmer red colours.

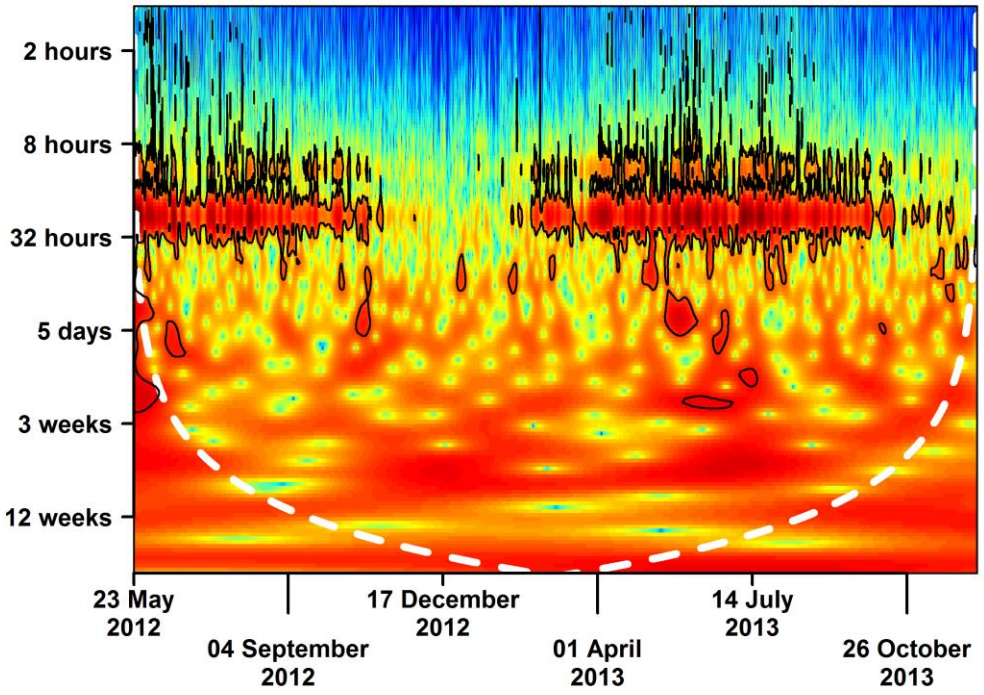


Figure 5.12 – CWT of water temperature
The regions outlined in black mark areas where a signal of particular periodicity has been identified most strongly relative to the rest of the signal.

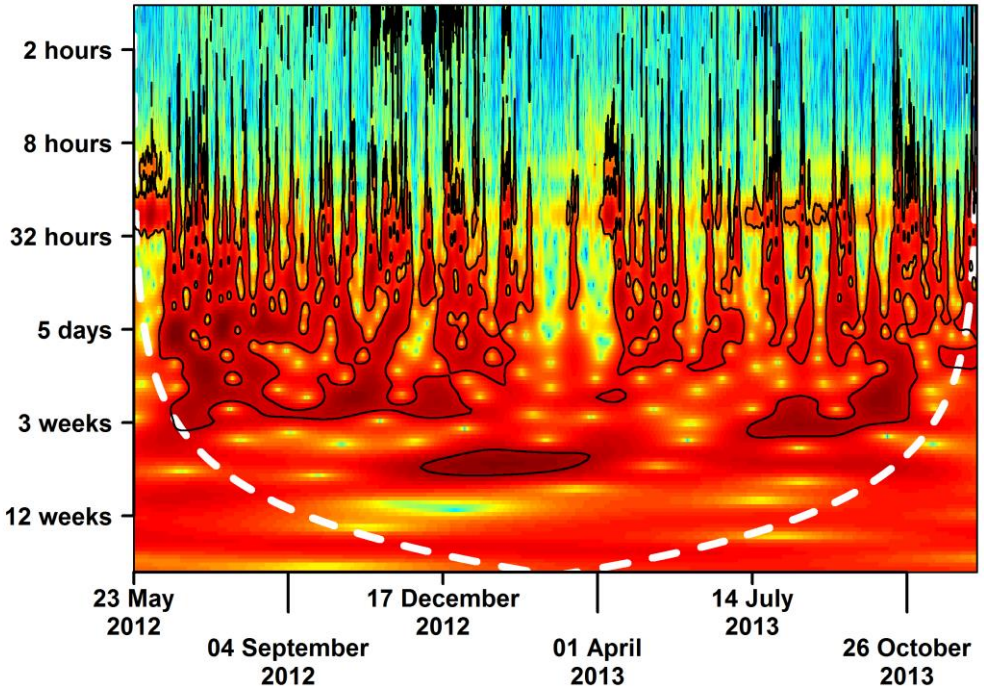


Figure 5.13 – CWT of pH
The regions outlined in black mark areas where a signal of particular periodicity has been identified most strongly relative to the rest of the signal.

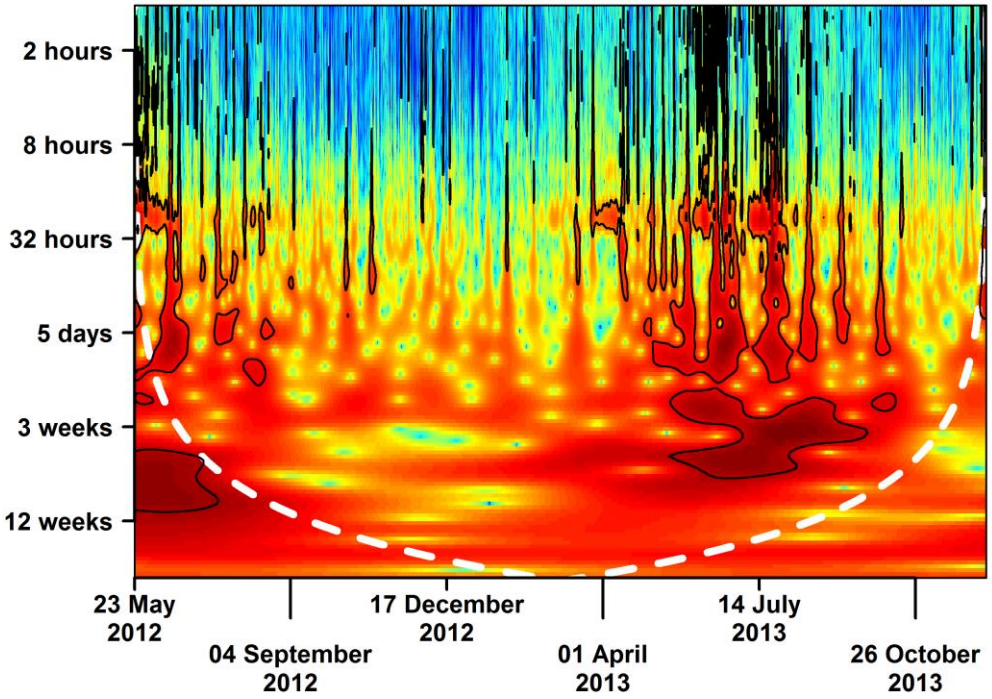


Figure 5.14 – CWT of SC
The regions outlined in black mark areas where a signal of particular periodicity has been identified most strongly relative to the rest of the signal.

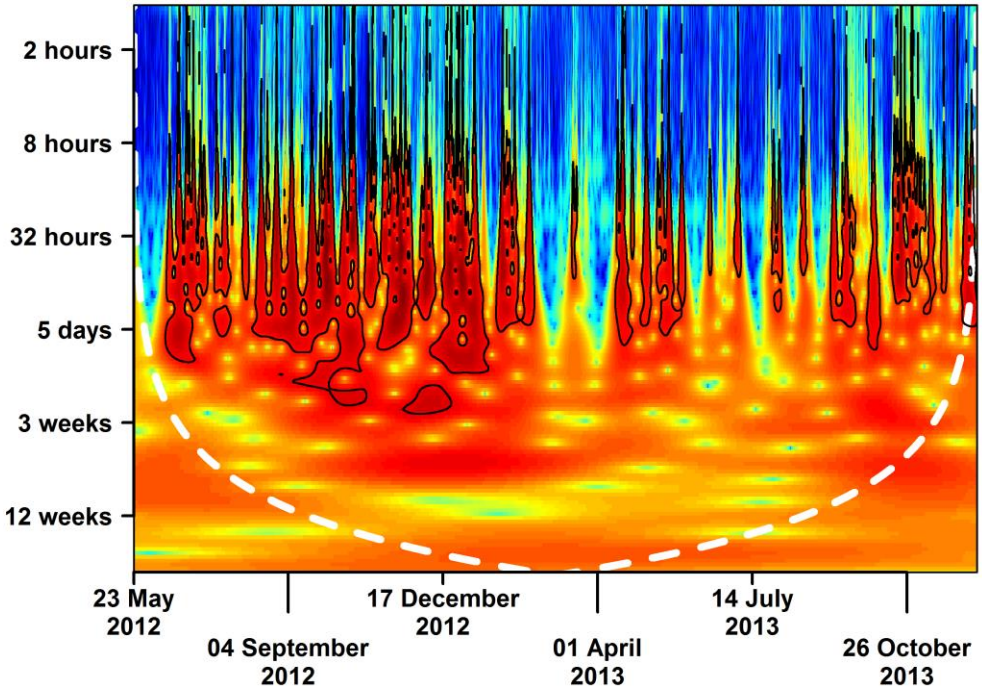


Figure 5.15 – CWT of discharge
The regions outlined in black mark areas where a signal of particular periodicity has been identified most strongly relative to the rest of the signal.

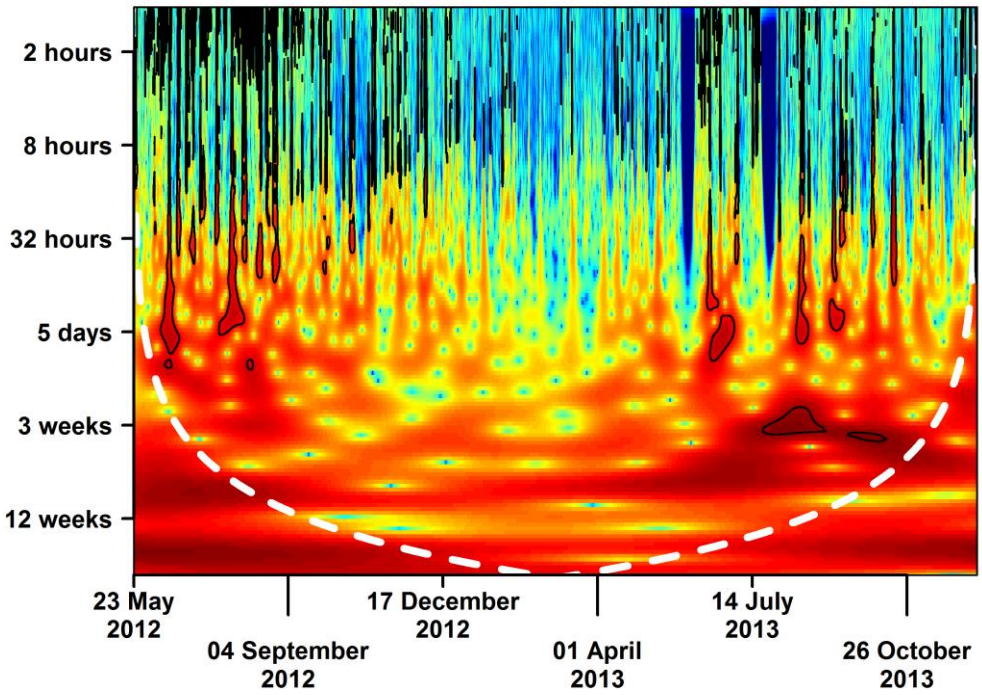


Figure 5.16 – CWT of [DOC]
The regions outlined in black mark areas where a signal of particular periodicity has been identified most strongly relative to the rest of the signal.

5.4.3 Wavelet coherence phase analysis

This section describes the results from wavelet coherence phase analysis that was used to compare the [DOC] data with the high resolution discharge, pH, SC and water temperature data. The two time series being compared are presented one above the other, with the resulting WTC presented beneath both.

The WTC generated between the [DOC] and water temperature identifies that there is a region of strong coherence between 19 June 2013 and 30 June 2013 with a periodicity of about 6 weeks (Figure 5.17). The coherence values were extracted by viewing line 86 of the RSQ matrix. The implication from analysing this section of the WTC was that the relationship between these two variables was stronger during this time period than at most other times in the time series. The Monte-Carlo simulated contours identify a couple other regions but nothing particularly significant or that appear to be relatable to other variables such as event occurrences.

The WTC generated between the [DOC] and pH also identifies that there is a region of strong coherence, this time between 26 June 2013 and 17 July 2013 (Figure 5.18). There are numerous other regions that show red regions indicating high coherence with black contours indicating where the Monte-Carlo simulations are identifying as significant. There are several regions of significance that stretch from about 8 hours down to about 5 days but don't extend across extended time periods. Relating to the discharge time series these seem to coincide with when events are being measured. Finally, there is a region of strong coherence identified with a periodicity of about 12 weeks that seems exist from the start of the time series in 23 May 2012 right to the end of 2012, but this falls below the COI so it is difficult to determine how significant this is.

Similarly, the WTC generated between the [DOC] and pH identified regions of strong coherence in the spring and summer period. In this instance, the region of significance stretched over a longer period and starts at a longer periodicity of about 3 to 6 weeks and then shifts to a shorter periodicity of about 1 to 3 weeks (Figure 5.19). As with the pH WTC there are tall vertical sections of the WTC stretching from 8 hours to 5 days that seem to coincide with events.

The WTC between [DOC] and discharge (Figure 5.20) again contained a short period of high coherence April and August at a periodicity of about 3 weeks. However, it should be noted that as the minimum and maximum coherence is scaled between 0 and 1 the coherence recorded between [DOC] and discharge is lower than for the other variables.

In chapter 4 poor R^2 values were obtained between [DOC] and other variables when performing bivariate analysis. However, if the data for only the periods of recorded high coherence (Figure 5.21) is extracted then the R^2 values for these relationships improves. This next section explores this.

The inverse relationship, between [DOC] and pH, identified after extracting the period from 25 June 2013 to 17 June 2013 from the wavelet coherence plot between pH and [DOC] is a stronger and more consistent relationship than is observed over most other parts of the time series (Figure 5.21 and Figure 5.22). The scatter plot produced using the data between 25 June 2013 and 17 July 2013 yielded an adjusted R^2 value of 0.9142, whereas a time period identified as having poor coherence was extracted and the adjusted R^2 value was recorded as 0.07803. Finally, a scatter plot was produced using all the SC and [DOC] data for the time periods where recorded stream temperature was within the range measured between 25 June 2013 and 17 June 2013 and this produced an adjusted R^2 value of 0.5097.

The inverse relationship, between [DOC] and SC, identified after extracting the period from 7 August 2013 to 7 October 2013 from the wavelet coherence plot between SC and [DOC] is a stronger and more consistent relationship than is observed over most other parts of the time series (Figure 5.21 and Figure 5.23). The scatter plot produced using the data between 7 August 2013 and 7 October 2013 yielded an adjusted R^2 value of 0.8547, whereas a time period identified as having poor coherence was extracted and the adjusted R^2 value was recorded as 0.1212. Finally, a scatter plot was produced using all the SC and [DOC] data for the time periods where recorded stream temperature was within the range measured between 7 August 2013 and 7 October 2013 and this produced an adjusted R^2 value of 0.3461.

There is an inverse relationship between temperature and [DOC] between 19 June 2013 and 30 June 2013 (Figure 5.21 and Figure 5.24). The relationship extracted from the wavelet coherence plot is counter-intuitive because it implies that for a brief period as the temperature was decreasing the [DOC] was increasing. During this period there were three minor events. It is likely that these events were significant enough to mobilise DOC from the soil to the stream but because the discharge and flow rate did not significantly increase then the removal of DOC was slower that some other events. However, [DOC] does have a tendency to decrease between events so increasing temperature may coincide with accumulation of organic material within the catchment. The adjusted R^2 for the region of high coherence was 0.47. For comparative purposes Figure 5.24 also shows the scatter plot using data extracted from the main time series that was identified as having a poor coherence and surprisingly this produced an adjusted R^2 value of 0.44. Finally, a third scatter plot was produced using all the data that was recorded when the temperature in the

stream was between the range recorded between 19 June 2013 and 30 June 2013, which produced the poorest R^2 of all three scatter plots of 0.01762.

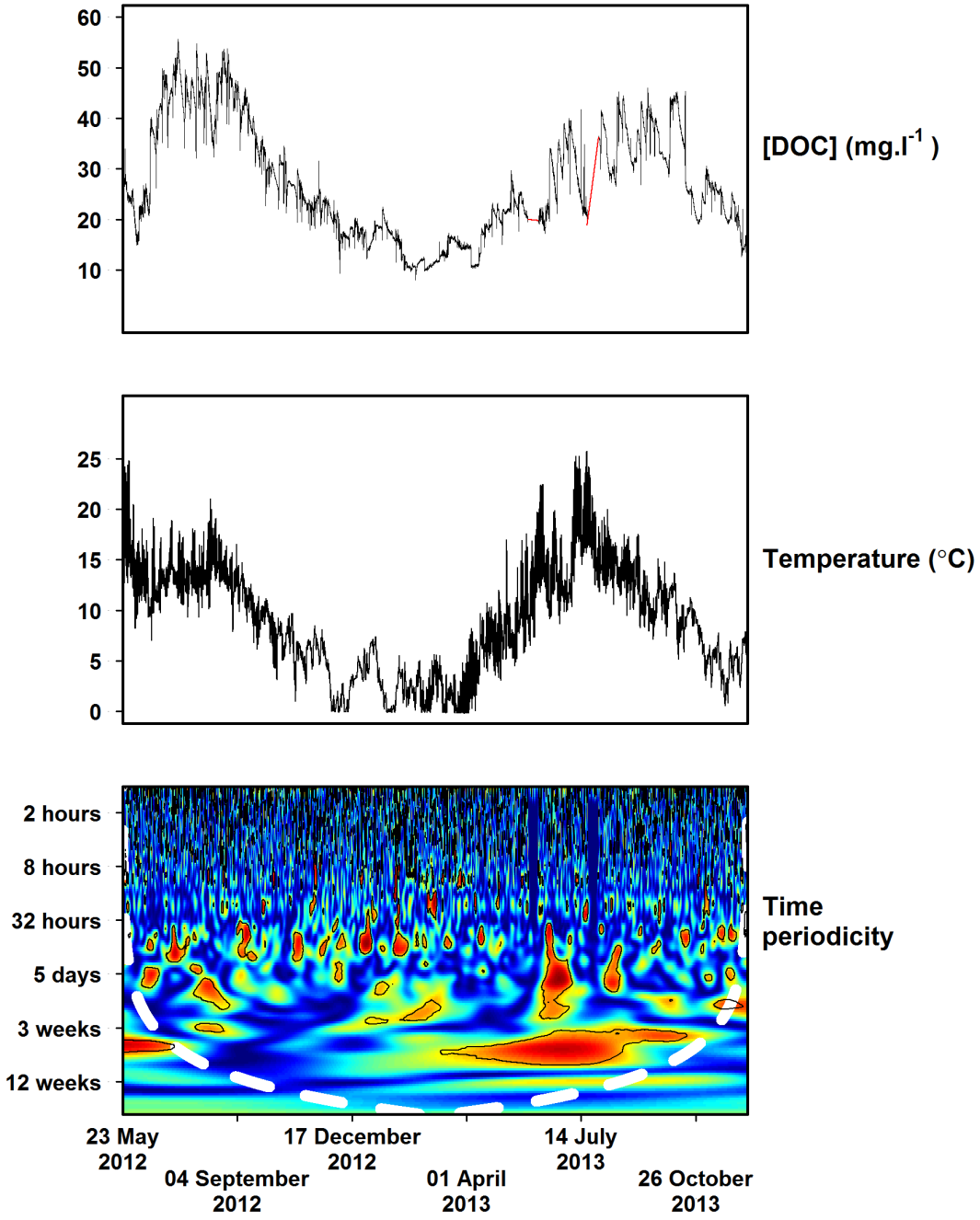


Figure 5.17 – WTC of [DOC] and water temperature
WTC of [DOC] and water temperature with 100 Monte-Carlo simulations, which highlight regions of coherence outlined in black. The coherence period does overlap with the gap filling in the [DOC] shown as a red line in the [DOC] which may influence the large region of coherence between April and October 2013.

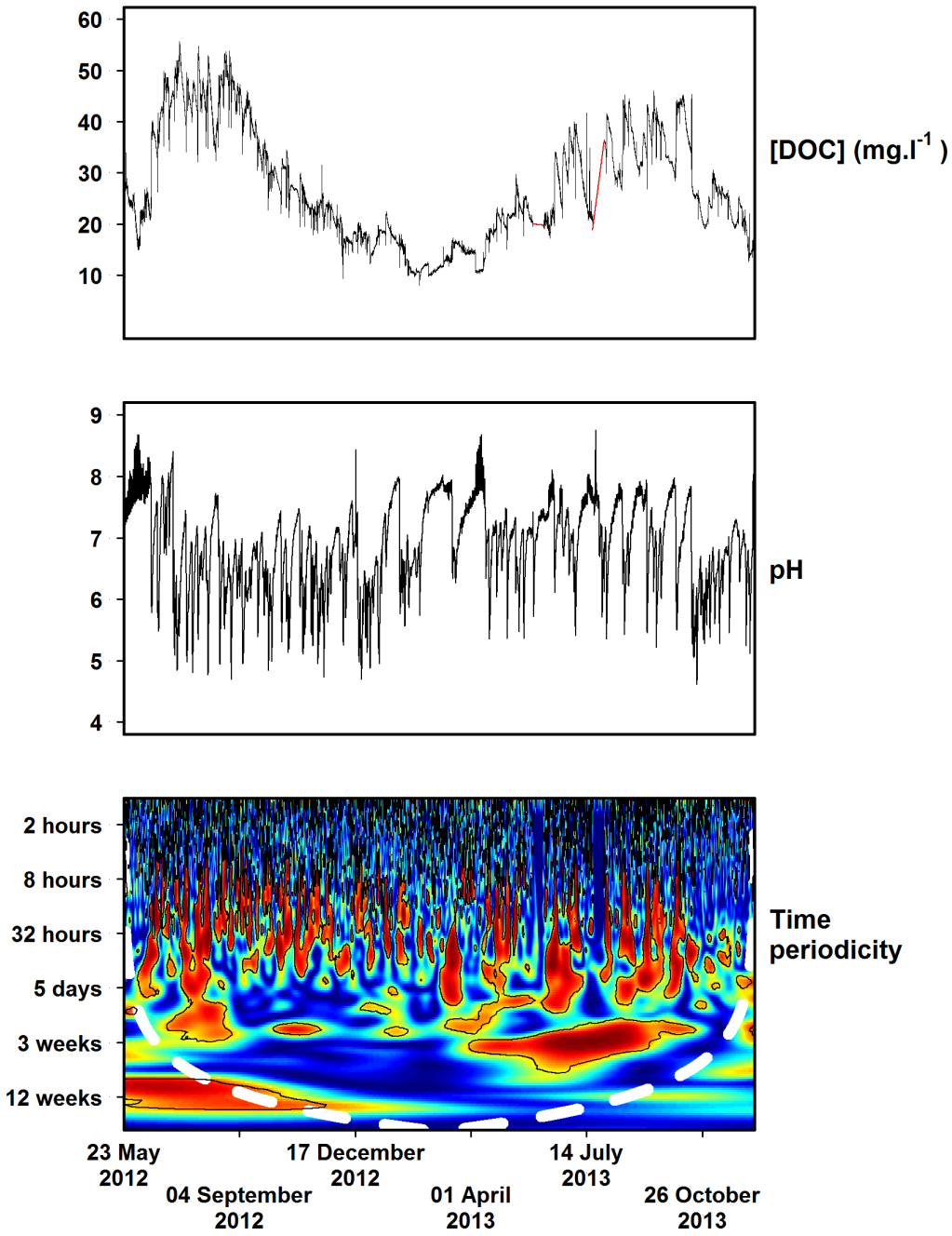


Figure 5.18 – WTC of [DOC] and pH
WTC of [DOC] and pH with 100 Monte-Carlo simulations, which highlight regions of coherence outlined in black. Red areas indicate greater coherence than cooler, blue colours.

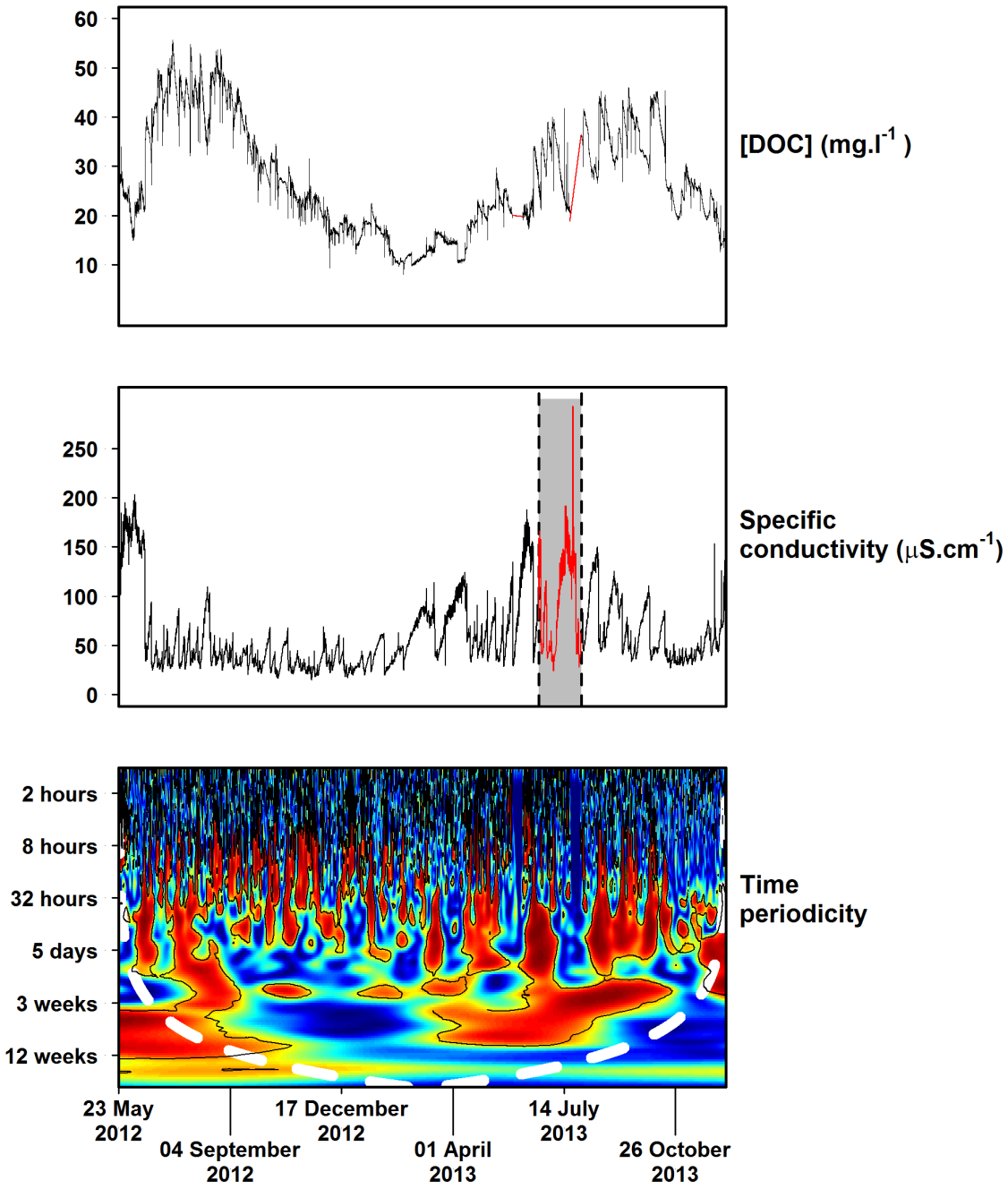


Figure 5.19 – WTC of [DOC] and SC
WTC of [DOC] and SC with 100 Monte-Carlo simulations, which highlight regions of coherence outlined in black. Red areas indicate greater coherence than cooler, blue colours. The region in red on the [DOC] and SC plots indicates data that is in the time series via gap filling.

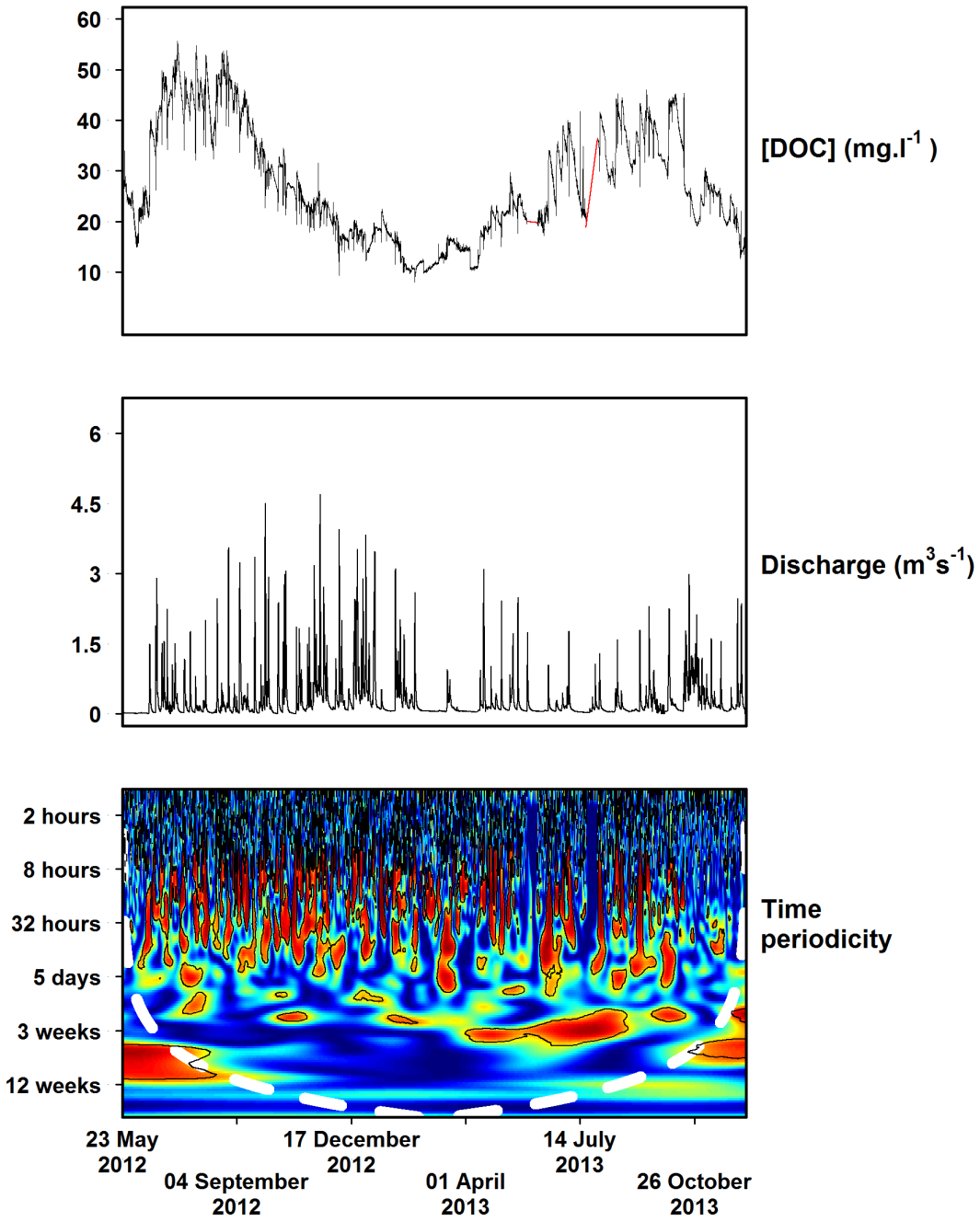


Figure 5.20 – WTC of [DOC] and discharge
WTC of [DOC] and discharge with 100 Monte-Carlo simulations, which highlight regions of coherence outlined in black. Red areas indicate greater coherence than cooler, blue colours.

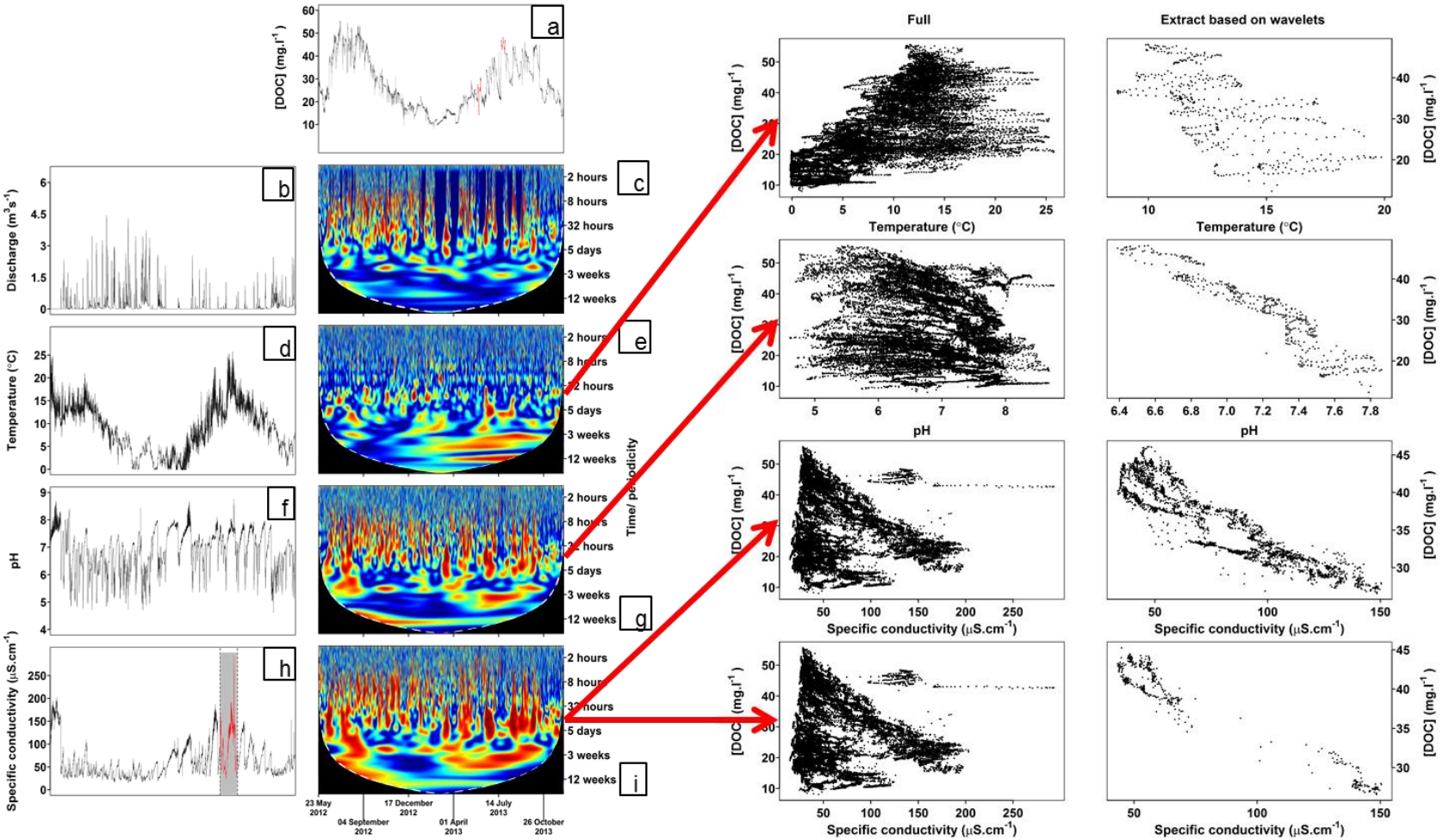


Figure 5.21 – WTC analysis used to extract linear relationships

The scatter plots on the left are the scatter plots for the complete time series with the right-hand scatter plots being the values extracted from the periods of high coherence. No relationship was identified in this manner between the [DOC] and discharge but relationships are identified between the [DOC] and the temperature, pH and SC.

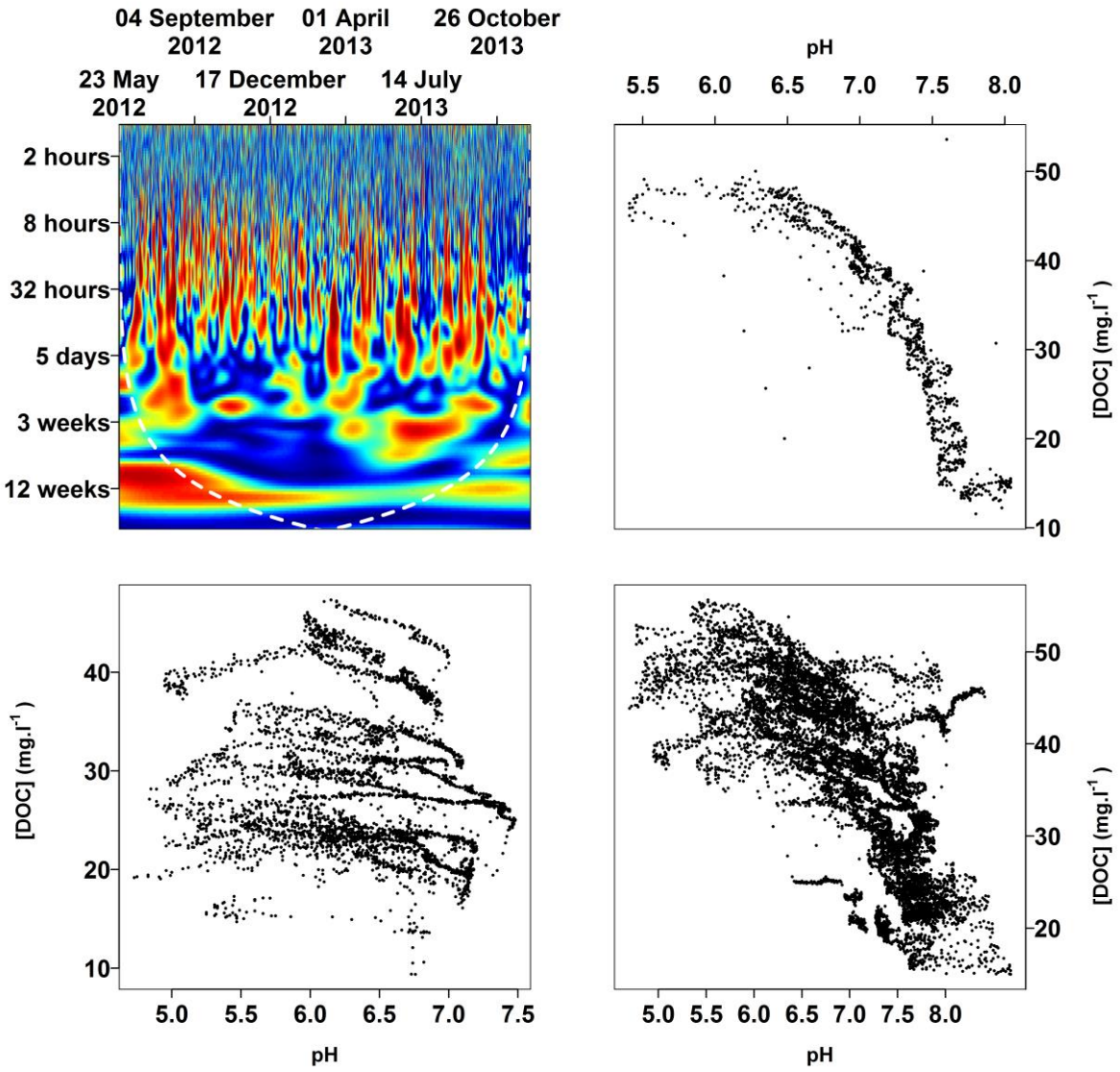


Figure 5.22 - pH and [DOC] WTC.
The top left graph shows the WTC for [DOC] and pH; the top right shows the relationship between [DOC] and pH for a period of identified high coherence, which produced an adjusted R^2 value of 0.91; the graph on the bottom left shows the relationship between [DOC] and pH for a period of identified low coherence, with an adjusted R^2 value of 0.08 and the bottom right shows the relationship between [DOC] and pH for all the data points where water temperature was between 9.85 °C and 25.27 °C, which produced an adjusted R^2 value of 0.51. The p-value for all three relationships are $< 2.2 \times 10^{-16}$.

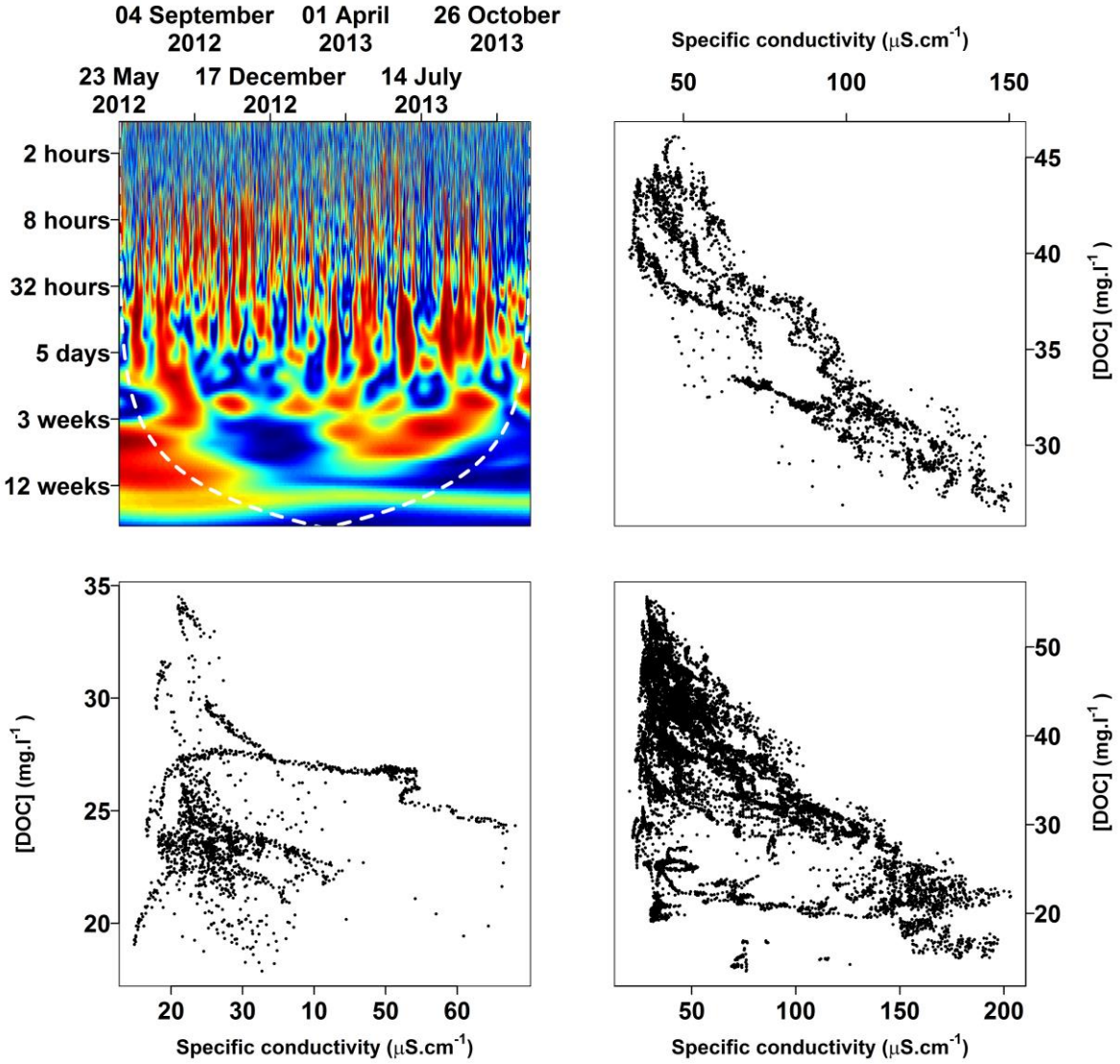


Figure 5.23 - SC and [DOC] WTC.
The top left graph shows the WTC for [DOC] and water SC; the top right shows the relationship between [DOC] and SC for a period of identified high coherence, which produced an adjusted R^2 value of 0.85 and a p-value of $< 2.2 \times 10^{-16}$; the graph on the bottom left shows the relationship between [DOC] and SC for a period of identified low coherence, which produced an adjusted R^2 value of 0.12 and a p-value of 0.0008 and the bottom right shows the relationship between [DOC] and SC for all the data points where water temperature was between 7.11 °C and 19.98 °C, which produced an adjusted R^2 value of 0.35 and a p-value of $< 2.2 \times 10^{-16}$.

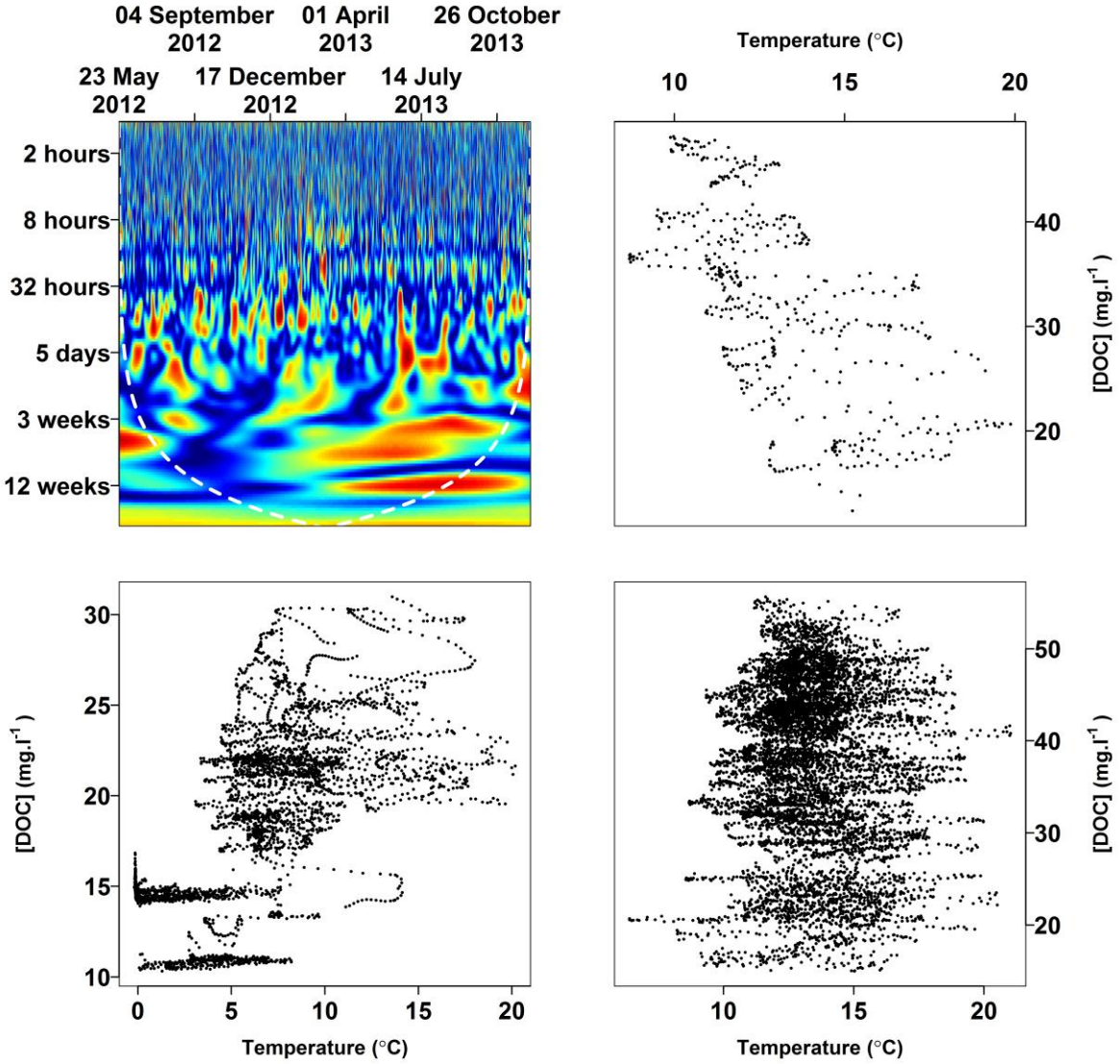


Figure 5.24 - water temperature and [DOC] WTC
The top left graph shows the WTC for [DOC] and water temperature; the top right shows the relationship between [DOC] and water temperature for a period of identified high coherence, which produced an adjusted R^2 value of 0.45 and a p-value of $< 2.2 \times 10^{-16}$; the graph on the bottom left shows the relationship between [DOC] and water temperature for a period of identified low coherence, with an adjusted R^2 value of 0.44 and a p-value of $< 2.2 \times 10^{-16}$ and the bottom right shows the relationship between [DOC] and water temperature for all the data points where water temperature was between 8.69 °C and 19.87 °C, with an adjusted R^2 value of 0.003 and a p-value of 1.6×10^{-7}.

5.5 Discussion

Wavelets were successfully used as an exploratory tool for gaining an insight into the data collected by the high resolution sensors at Drumtee water, revealing some complex and intermittent relationships in the time series that were not apparent from comparing two variables on a scatter plot or from a visual inspection of the time series. For example, wavelet analysis revealed more clearly and in a strong visual format that revealed the time of year and periodicity that transient relationships between variables such as [DOC] and pH or [DOC] and SC were occurring at. This understanding confirms that single variant analysis in the previous chapter is of limited value, but the GAM model presented is an appropriate vehicle to capture the more complex interactions.

5.5.1 CWTs and MODWTs – a comparison

The MODWTs and CWTs presented here appeared to show quite similar trends in the data. However, the results from these analyses were not identical but are nevertheless useful for complementing one another. The MODWT is probably easier to read and interpret. For example, the D5 component of the water temperature identified a signal of about 16-hour periodicity. However, as each decomposition level will pick up signals that are not exactly equal to the time periodicity represented by the decomposition level it is important to not assume that the signal being identified is of the precise periodicity of the decomposition level that identifies it. In this instance it is likely picking up the 24 hour signal from the diurnal variation in temperature.

This is not the first research to discuss the advantages and disadvantages of continuous or discrete wavelet transforms, where MODWTs are a form of the latter type. In a study of soil data it was observed that CWTs have greater scalability, which in simple terms means that the gaps between the different frequencies identified is smaller and therefore periodicities can be more precisely identified (Lark, 2006, Rucka and Wilde, 2006). In other words, CWTs actually provide more information at a larger number of periodicities. However, although CWTs offer the ability to analyse a greater number of periodicities the author (Lark, 2006) emphasises that the interpretation of a CWT is essentially one of image analysis because the CWT does not offer a decomposition of the time series (Percival and Walden, 1993). In this project the MODWTs produced analysis that was more simplistic than the CWTs as it was easier to interpret.

5.5.2 MODWTs as an analytical tool and possible method of data correction

Following on from that discussion it can be stated that the MODWTs provide a convenient breakdown of the different signal components. However, as was stated in the introduction, the decomposition levels are not exact and there is usually some “bleed”

between the different decomposition levels. This bleed is not necessarily disadvantageous as decomposition levels that only identified signals at very specific frequencies would leave large gaps – indeed it is likely that more frequencies would be ignored than identified. For example, if we have two decomposition levels at 8 hours and 16 hours then a perfect decomposition would mean no decomposition level would identify any periodicities between these two numbers. When a signal is decomposed, it can identify frequencies close to it so the same signal component may be picked up by more than one of the decomposition levels.

This "bleeding" of signal components into different decomposition levels meant that when a signal was identified as being in D1, D2…D10 etc it did not mean the signal had to be oscillating at precisely that periodicity. For example, there was a strong signal picked up by the D5 component for water temperature and pH, which corresponded to oscillations of 16 hours. However, when these signals were looked at in more detail these components were actually the diurnal cycles in the water temperature and pH and this driver varies in length as daylight varies. The diurnal cycle in the pH was only measured at neutral or alkaline pH values, so not during events when the stream was acidic. It was already known that water temperature had a diurnal cycle in its signal and this diurnal pattern can be observed in the CWT of water temperature.

One of the potential future uses of MODWTs for fluvial systems is that they could potentially be used for removing signals that might appear to be erroneous within the signal. For example, the diurnal signal that appears in several pH measurements (when pH is greater than 7 for extended periods) is identified in component D5 of the MODWT for pH. By removing the D5 (16 hour) components from the signal this diurnal variation has been removed (Figure 5.25). It should be noted that this process was not applied to the pH data before any analysis because there isn't evidence to indicate that this data is erroneous. However, this is being explored here in the discussion as an example of how this process could be applied to a hydrological time series into which there was a known artefact.

Another hypothetical example of where this could be applied would be to correcting baseline drift in the discharge data. It was observed that when multiple large events occurred within a short time period the baseflow of the stream could slightly increase. It was identified in Figure 5.10 that the S10 component (which is the remainder of the discharge signal that had not been decomposed) showed variation at periods where the baseline had increased through repeated high event flow. In chapter 3 the R package "baseline" was used to correct this but there may be a possibility of using specific decomposition levels of the MODWT to do this without risking distorting other components

of the line shape because it would only eliminate the long term shifts in baseline. However, only the pH is used to demonstrate this principle of stripping a frequency component from a time series because the change to the discharge time series would be extremely subtle, whereas the change to the pH should be more clearly observable. The obvious danger with this approach is that in using the wavelets to eliminate suspect data it may also remove components at similar frequencies that are actually important results.

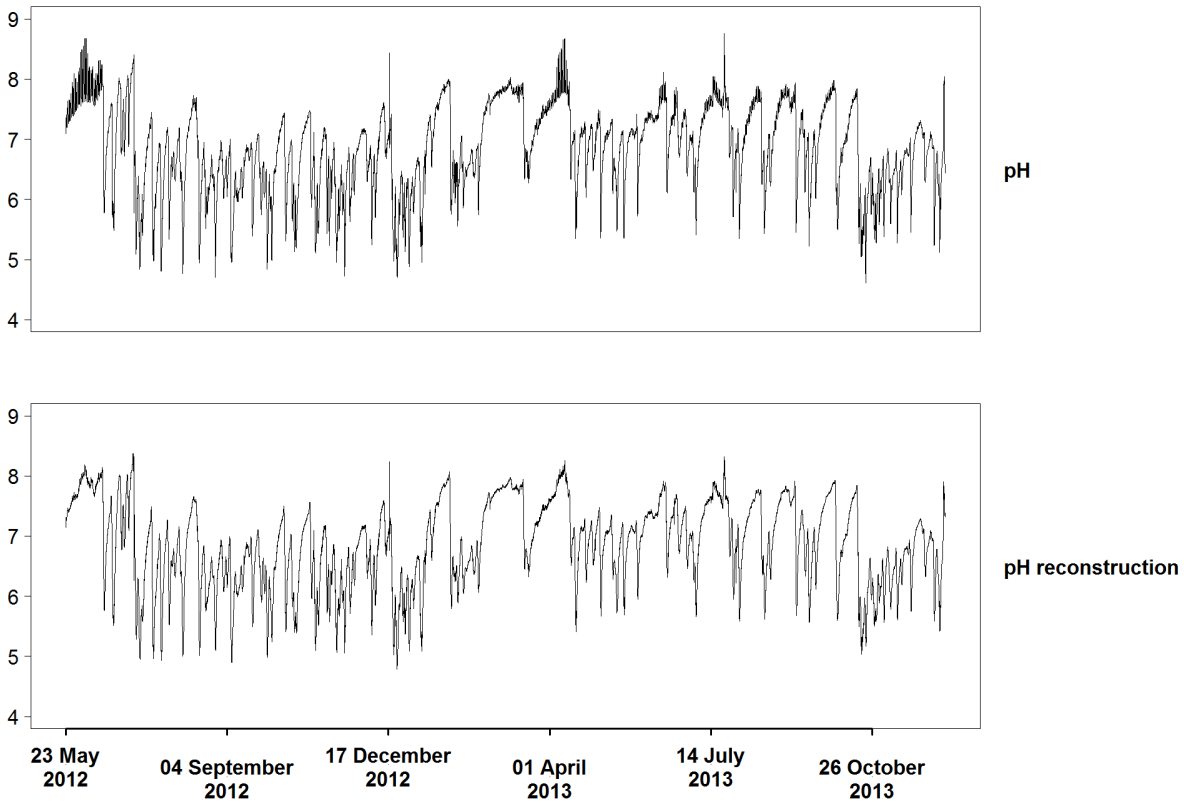


Figure 5.25 – Reconstruction of pH data.
The original signal is shown at the top but the signal components at 8 hours and 16 hours have been filtered out and this has removed the diurnal cycle measured and that was strongly picked up by the D5 component of the decomposition.

5.5.3 Links between [DOC] with SC and pH in the spring and early summer

The use of wavelet analysis has already been discussed as a form of exploratory analysis but one question that might be asked is to what extent wavelet analyses be used to inform our understanding of different hydrological behaviours, particularly with regard to responses that cannot be detected using more conventional analyses. It is somewhat redundant to use wavelet analysis to discuss topics such as seasonal variation in [DOC] or water temperature as this sort of information can be gleamed from simpler and less time-consuming analyses, especially as wavelet analysis is a largely exploratory tool that does not produce a concise data summary. In this regard, the most interesting knowledge

gained from the wavelet analyses were the short-term relationships existing in the time series between [DOC] and three hydrological variables with a periodicity of about 3 weeks as shown in the WTCs comparing the [DOC] with pH, SC and water temperature. The WTCs indicate that relationships exist during periods in the spring and early summer. These short-term relationships were not identified using the other types of analysis in this project as presented in chapter 4. The wavelet analyses do not identify the mechanistic causes of the hydrological relationships but were very useful for identifying the transient relationships and the periodicities could at least yield clues as to the source of these relationships – e.g. events. It is not the case that there are no events during the periods of high coherence in the spring/summer – therefore the periods of high coherence are not simply resulting from an absence of events. It may be the case that the strength of disturbance caused by events during the period of high coherence is insufficient to tip the system of coherence during this period when other factors may be of greater influence – for example during this time of year the organic C available is increasing due to increased biological activity.

Many components of DOC are acidic e.g. humic and fulvic acids. It therefore may seem logical that when [DOC] is greater that the relationship between it and pH will be stronger but this relationship doesn't hold up during the summer when [DOC] continues to be high, so this cannot be the sole reason why the pH and [DOC] are linked at the spring and early summer time period. An almost equally strong relationship, during a similar time frame was identified between [DOC] and SC. The relationship was identified during a period when [DOC] was increasing within the catchment, suggesting that the DOC concentration of the soil water is not constant and so it is not a simple mixing model with fixed DOC proportions. Rather, the stronger coherence between [DOC] and SC is likely the result of the fact that the [DOC] is soil (surface) water, which is varying throughout the year, increasing in summer to autumn reflecting increased soil productivity, and this variation in that end member composition is difficult to capture in this form of data analysis.

5.6 Conclusions

What this chapter primarily demonstrates is that wavelet analysis is a powerful exploratory tool that can be used to identify relationships over time and at different frequencies that are hidden through visual or simple linear exploration of data interaction. This makes wavelet analysis a powerful tool in the analysis of fluvial systems, particularly as sensor technology becomes more common and sophisticated, generating bigger data sets that can quickly become unwieldy.

It is noticeable that the information obtained from MODWT's and CWTs is very similar but the two outputs were complementary. MODWTs provided a more accessible visual insight of whether signal periodicity existed. The CWTs complemented this through revealing at greater graphical resolution of the specific periodicities that the signals (picked up by the MODWT) were oscillating at.

The results presented from the WTCs indicate a dynamic fluvial environment with complex relationships between different hydrological variables. The full implications of some of these relationships represent intriguing research opportunities moving forward including linking periodicity components to other variables or factors within the catchment. The results presented indicated that relationships do exist between [DOC] and three other variables (water temperature, pH and SC) that would otherwise be difficult to identify due to only existing for transient time periods and not necessarily across the entire time series.

Although the wavelet analysis has been presented after the time series analysis of chapter 4 the information presented here was used in an iterative feedback process to inform the time series analysis conducted and presented in chapter 4. So, in practice the work on the time series and on the wavelet analyses occurred concurrently. The WTCs provided the most consistently useful information from the three types of wavelet analysis, particularly with regards to [DOC] as they indicated not only regions of strong coherence at particular periodicities but because they are comparing the [DOC] with other variables. This comparison meant that not only was a region of coherence identified but that it indicated that either the [DOC] was being influenced by that other variable or that they were varying at a similar timescale. The WTCs can't be used to determine what is causing the variation in [DOC] but they can useful for identifying a correlation between two variables. For example, the transient relationship identified between pH and [DOC] (Figure 5.22) does not necessarily confirm that the increase in pH is the result of a decrease in [DOC] but there is a strong transient relationship between pH and [DOC] between 25 June 2013 and 17 July 2013. This type of information can be very useful for constructing statistical models as the model could be programmed to take greater account of particular variables depending on the time of year. Similarly, a transient relationship was identified between SC and [DOC], while the relationships identified by the WTCs between water temperature and [DOC] were weaker. The seasonal pattern between [DOC] and water temperature was out of range of the WTC so only the transient relationships were identified. The seasonal variations being out of range of the WTC is a likely reason why this analysis did not identify a strong correlation but the temperature was an extremely useful component of the GAM model presented in chapter 4.

Although the reason for relationships identified through wavelet analyses can be inferred by drawing on understanding of catchment functioning (e.g. how photosynthesis may modulate pH, or that reduced surface inflow in dry period can reduce DOC, but increase SC as groundwater becomes more dominant), there is scope to understand better the detail, particularly when coherent relationships do not break-down although there is evidence of a change in catchment functioning e.g. a small event. Further these are headwater systems, which may be highly responsive compared to larger catchments, so understanding what some signals occur at long time scales is intriguing and may merit further investigating.

The use of MODWTs has been shown in this project to be a useful method for visually inspecting the data and for filtering out erroneous signals if this was required. The potential for MODWTs is to not only be used as a simple method of filtering out components of complex signals but also inspect individual elements of a signal. Continuous wavelet analysis was a useful tool for identifying any regions with repeating signals within a time series. Consequently, CWTs can be deemed useful for identifying why the wavelet coherence analysis was identifying regions of strong coherence at time periodicities that were not immediately obvious.

6.0 Raman analysis of dissolved organic carbon

6.1 Abstract

This chapter explores the replication of published approaches to using Raman Spectroscopy to identify humic substances and the difficulties encountered in doing so. Such analysis was undertaken as preparatory activity to consider the viability of Raman spectroscopy in a field sensor.

The first step was the processing of isolates: sugars (glucose and sucrose), and humic substances (humic acid and fulvic acid). The latter are particularly important as humic substances represent one of the most prevalent carbon components of many natural fluvial systems and are therefore important to consider when analysing the carbon content of water. Humic substances are however complex mixtures and therefore producing reliable experimental procedures for either measurement or analysis can be difficult. The initial results were generated using powdered forms of the sugars and humic substances. After the powdered humic acids were analysed the same substances, this time dissolved in water, were analysed using Raman spectroscopy to assess whether these substances were identifiable in solution.

`Sugars produce very clear and distinct Raman spectra when in a powder or crystalline form but in solution the Raman spectra are only visible in very high concentrations solutions (thousands of mg/l). Experiments conducted using surface enhanced Raman Spectroscopy (SERS) were conducted on solutions of sugars but the sugars did not produce enhanced signals. The difficulty in producing SERS spectra for sugars is consistent with other research that indicates it is challenging to produce Raman spectra of sugars. The same three experiments were also performed on humic and fulvic acids (powders, standard Raman and SERS). However, measurements of the Raman spectra of humic and fulvic acids produced inconsistent results.

As peatland catchments are often dominated by humic substances this was the substance where most effort was focussed in developing methods to try and produce Raman signals from humic substances. Two methods were identified in the academic literature as potentially offering solutions for amplifying the Raman spectra of humic substances to enable clear visualisation of SERS signals. The first method was to increase the pH of the humic and fulvic acid solutions and the second method was to perform a reduction reaction on those solutions. Despite following these guidelines as near as possible with the equipment available, I was unable to reproduce the spectra the publications had

reported: very few Raman spectra of high intensity were observed and those that were observed were not consistently reproducible.

Reproducibility becomes an important consideration as contamination present could account for intense Raman spectra in a single sample. The failure to reproduce the results presented in the papers discussed does not discredit their experiments, as it may be that it is inherently difficult to reproduce results consistently, and the measurement of humic substances using Raman spectroscopy may require refinement before it can be widely adopted for DOC composition insight. Furthermore, as experimental conditions could not be replicated precisely (use of a 785 nm laser instead of a 633 nm laser) this may be a potential reason for not replicating the Raman spectra reported. From the results presented here it would be difficult to recommend Raman spectroscopy as the basis for a sensor for measuring DOC components.

6.2 Introduction

One important aspect of DOC is that it consists of many different molecules. Some of the important components include sugars, amino acids and humic substances (Sondergaard and Middelboe, 1995, Thomas, 1997, Haei et al., 2012). Some DOC molecules are very labile and can be acted upon by microbes in the water such as sugars and amino acids, whereas many other components are non-labile and can't be broken down as easily (Moran and Hodson, 1990). Some DOC components may even be chemically refectory but biologically labile, indicating that some microbes can break down organic carbon usually resistant to breakdown (Ogawa and Ogura, 1992). The composition of DOC has been studied far less than the concentration in part because it is much more difficult to analyse the composition of a particular DOC sample. At present most studies or the composition of DOC are conducted by extracting components of DOC using resins (Kitis et al., 2002, Peuravuori et al., 2002, Chow, 2006) or the use of fluorescence spectroscopy (Coble, 1996, Baker and Spencer, 2004)

In the measurements and analysis that I have presented in the previous two chapters I have not considered this issue because the techniques I have used for those measurements could only determine the concentration of DOC and not its composition. Determining the composition of DOC is more challenging than measuring its concentration of DOC. However, as the lability of exported C can influence how and when this C is taken up by the carbon cycle means that it is important to develop some understanding of its composition and how it changes over time.

Sugars form an important component of organic matter and there have been several studies making use of Raman spectroscopy to measure sugars and in particular glucose. Studies of glucose with Raman spectroscopy have primarily been geared towards medical applications and in particular the monitoring of blood glucose levels as a potential tool for monitoring diabetes (Shao et al., 2012). In terms of sugars there is research to suggest that glucose is the most common sugar found in samples of organic carbon and therefore that is one of the sugars analysed in this study (Routh et al., 2001).

Amino acids are another key component of organic matter and the use of Raman spectroscopy and in particular SERS to analyse amino acids is well established (Kim et al., 1987, Achterberg et al., 2001, Kang et al., 2013) with papers on glycine (Suh and Moskovits, 1986, Podstawka et al., 2004) and the other three amino acids measured in this project (tyrosine, tryptophan and phenylalanine) (Stewart and Fredericks, 1999). Some of these papers, such as Kang are focussed on protein research, while others are dealing with cancer treatment (Vendrell et al., 2013). As these tests are well-established a decision was made on this project not to spend much time analysing amino acids using Raman. However, some tests were conducted on amino acids in solutions in concentrations close to what we would expect to see in a fluvial environment, but the results indicated that the limits of detection were significantly below what would be measured in a fluvial environment. As there is material that describes the measurement of amino acids using Raman spectroscopy and the limits of detection are lower than desired little time was given to analysing these substances in order to focus more on humic and fulvic acids. Thus, the results from amino acids are not presented here (and also to prevent the chapter from becoming too cluttered with graphs).

Humic substances are extremely important in determining the quality of DOC samples because they account for a large proportion (approximately 60-70%) of the carbon in soils (Mobed et al., 1996). The difficulty with humic substances as discussed in the introduction chapter is that it is itself an umbrella term used to describe a huge number of different molecules. Humic substances are formed from many sources including microbial activity and the degradation of plants and animals (Moran and Hodson, 1990). Additionally, they consist of complex carbon structures that often contain complex aromatic and aliphatic structures (Ferreira et al., 2014).

The indicators from organic carbon research are primarily that those organic molecules such as glucose and amino acids that are linked to health studies have been (understandably) targeted by larger numbers of researchers and the subject of studies aimed at their detection. However, humic and fulvic acids are still important for public health reasons as they can lead to water colouration (Worrall et al., 2003, Armstrong et

al., 2010) but their removal from water incorrectly can lead to the production of disinfectant by-products (DBPs) (Li et al., 1998, Chow, 2006). Therefore, it is important to be able to distinguish between humic acids and fulvic acids. As humic and fulvic acids can contain such a wide range of molecules then Raman is a promising technique as it may be able to identify component fingerprints that may indicate a particular treatment for removal is required. Chemically humic acids are insoluble at pHs below 2, whereas fulvic acids are soluble at all pHs (Sutton and Sposito, 2005). It is also the case that fluorescence spectroscopy can be used to distinguish between humic and fulvic acids (Baker, 2001, Chen et al., 2003) and therefore if Raman spectroscopy is going to be seen as an alternative method to fluorescence spectroscopy then it is important that it is also able to distinguish between humic and fulvic acids.

One of the most important developments in the evolution of Raman spectroscopic analysis has been the discovery and development of surface enhanced Raman spectroscopy (SERS) which enables the enhancement of Raman signals (Fleischmann et al., 1974). SERS is achieved by having the target substance adsorbed onto a metallic surface (usually though not always copper, silver or gold) where that surface is rough at a nano-scale level (Campion and Kambhampati, 1998). Initial experiments used roughened silver electrodes but later approaches used colloids and custom-designed substrates.

A variation on standard Raman spectroscopy is Surface Enhanced Raman Spectroscopy (SERS), which enables the collection of enhanced Raman signals. One of the major drawbacks of Raman spectroscopy is that the signals produced are weak and can be easily obscured by background noise or fluorescence. SERS is a powerful technique for probing many materials that produce Raman spectra that are too weak to observe and distinguish from background noise. Some Raman enhancements have been reported as great as 10^{14}-10^{15} when compared to the strength of standard Raman spectra (Campion and Kambhampati, 1998, Kneipp et al., 1999). However SERS relies on substances to bond to the surface of the metal used (Campion and Kambhampati, 1998). If adsorption does not occur then surface enhancement of the Raman signals will not occur.

It has been shown that there is seasonal variation in [DOC] at the Drumtee catchment (Waldron et al., 2009, Murray, 2012). From this an intriguing question is whether or not the composition of the DOC is varying in addition to the concentration. Raman spectroscopy measuring at a similar temporal resolution similar to the Spectro::lyser™ and TROLL® offers the opportunity of being able to identify compositional changes in DOC over, not only over an annual period, but also potentially with regards to variation during events. However, to achieve this it is critical to establish whether key components of

DOC, such as humic substances and sugars, can be identified under ideal conditions in the laboratory.

One of the motivations for the testing Raman spectroscopy in this project is that portable Raman spectrometers are available on the market, for example the SafeInspec-0TM Handheld Raman analyser from BaySpec's, the ReporterR from Sci-Aps, the FirstGuard™ and the Xantus™ from Ragaku. With portable Raman spectrometers, the creation of a field-deployable sensor to collect data on DOC composition, as the Spectro::lyser™ does with [DOC], may be possible. The advantage that Raman spectroscopy would have over a UV-visible light system is that it can be more detailed about molecular structures (Kneipp et al., 1999). As DOC comprises an extremely large variety of different molecules, a system that can differentiate between different organic carbon types could reveal how catchment DOC varies seasonally.

One potential long term goal of this project is to develop a field-based sensor capable of measuring variations in the composition within DOC in real time and that sensor may be based on Raman Spectroscopy. However, before construction can begin on a field-deployable sensor for measuring or analysing DOC it would be prudent to establish further how well Raman spectroscopy can characterise DOC in a laboratory setting, where test conditions are more ideal and can be controlled with fewer unexpected variables such as damage to casings from debris being washed downstream or the effects of frost. The aims of this chapter are therefore to:

1. establish baseline background signals for the experiments and thereby reduce the risks of falsely attributing Raman bands to target materials (sugars and humic substances)
2. make Raman measurements of sugars
3. make Raman measurements of humic and fulvic acids
4. compare the results of the humic and fulvic acids to those found in the literature and determine whether my interpretation of those results is the same as that found in the literature.

One topic discussed in chapter 2 was that there are two types Raman scattering– one is Stokes Raman scattering and the other is anti-Stokes Raman scattering. The detail of this can be found in chapter 2 but the important point to note is that Stokes scattering occurs when light is absorbed at one wavelength but emitted at a longer wavelength and that the vast majority of Raman studies make measurements of the Stokes Raman scattered lines because the Stokes signals are stronger than the corresponding anti-Stokes lines. In this chapter all the Raman measurements discussed (both from the experimentation and the references to other literature) are from using Stokes scattering.

There was not a prior comparable experimental setup to build on, so a template for working had to be produced. The major question was would a water sample from the field site produce a distinctive Raman spectrum and if so could individual components of organic carbon be tested to determine what those Raman bands related to. This thought process is detailed in Figure 6.1 and this formed the key question that needed to be addressed in this chapter – is Raman spectroscopy a suitable tool for analysing organic carbon and if so would it be suitable for field deployment?

In summary, the aim of the Raman experimentation in this project is to produce an evaluation of Raman spectroscopy to identify DOC molecules and humic substances in particular. In order to justify proposing Raman spectroscopy as a technique for analysing DOC it has to be at least able to offer improved performance in at least one aspect (such as sensitivity or more structural information) over other approaches such as the use of resins (Aiken et al., 1992, Chow et al., 2006), fluorescence spectroscopy (Baker, 2001, Baker and Spencer, 2004, Carstea et al., 2010) and NMR (Al-Faiyz, 2012). Each of these techniques is already well established for analysing humic substances and so there are many more years backing up this research. If Raman spectroscopy can be shown to have some tangible benefit over these other techniques then it may be a viable method for analysing DOC.

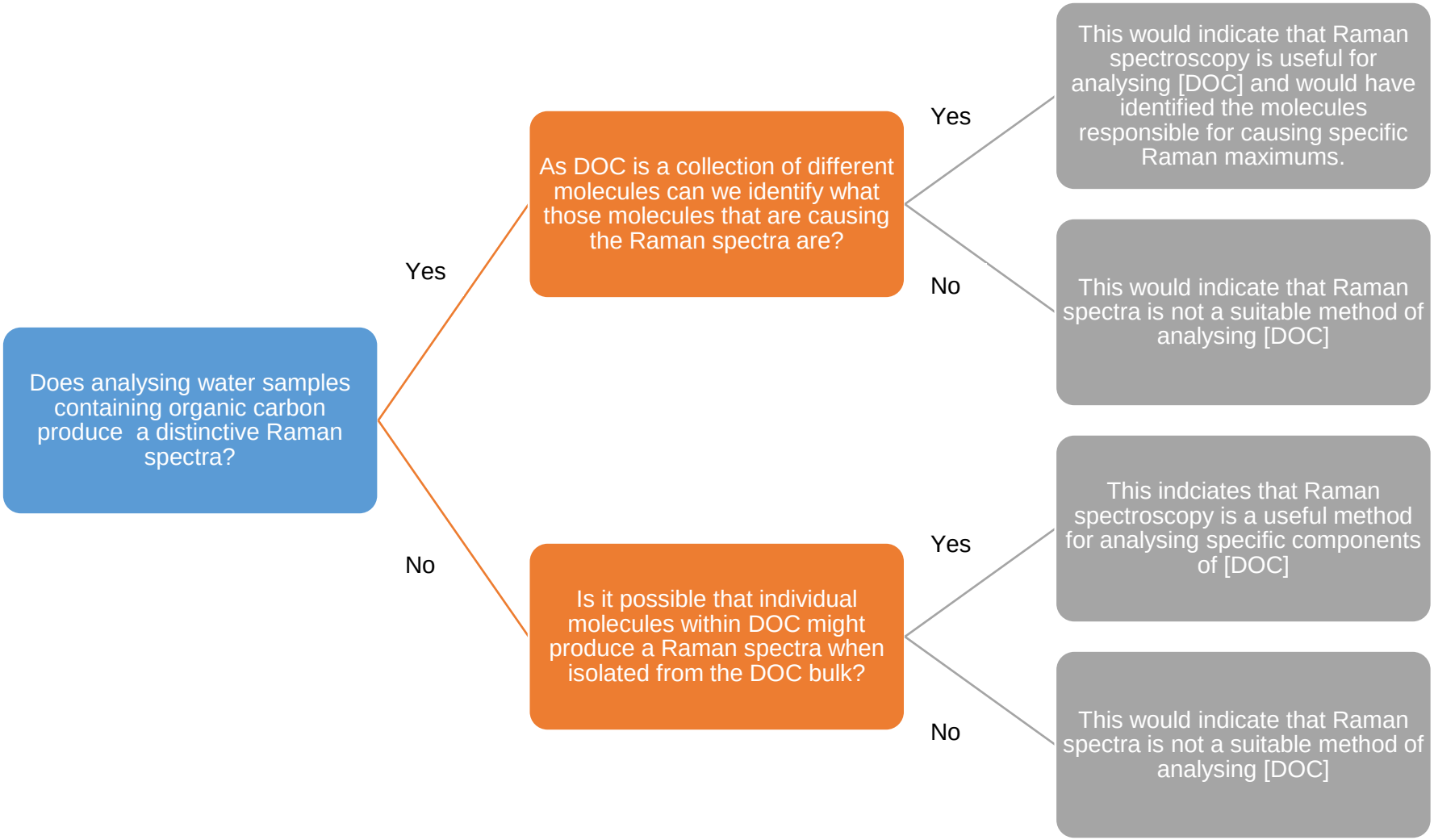


Figure 6.1 – Flow diagram of initial Raman experimentation

6.3 Methods

Two aspects are of primary importance when considering methodology are: the setup of the equipment and the preparation of the samples to be analysed.

6.3.1 Equipment set up

The Raman spectrometer used in these experiments was a Renishaw® In-Via Reflex Raman Spectrometer. The Renishaw In-Via Raman Spectrometer housed both a 514 nm edge filtered laser and a 785nm edge filtered laser. Both the 514 nm and the 785 nm lasers use edge filters in this setup but an alternative would have been to use a notch filter. The purpose of both edge and notch filters is to filter out Rayleigh scattered light, which is the light that is scattered at the same wavelength it is absorbed at (Collette and Williams, 2002, Zepp et al., 2004) and accounts for the majority of scattered light and produces a much stronger signal than Raman signals do. For example, in nitrogen the Rayleigh signal is approximately 1000 time stronger than the Raman signal (Hahn, 2007). In practice, this would produce a strong signal on the graph at close 0 cm^{-1}. Therefore, an edge filter is applied to wavenumber values up to 100 cm^{-1} for both the 514 nm and 785 nm lasers. Edge filters are more durable and provide better filtration performance but notch filters do have the ability to filter out the anti-stokes component of Rayleigh scattering as well. Fortunately, in chapter 7 even though anti-stokes experimentation is conducted on DOC without a notch filter the wavenumbers are sufficiently far away from 0 cm^{-1} that this was not an issue.

The 785 nm laser was used to measure all of the spectra presented In this project. Generally, the 514nm laser can produce a better Raman signal due to superior signal to noise ratio (SNR). However, the 514 nm laser resulted in greater fluorescence and background signals and the 514 nm laser could not be used with the SERS substrates because gold strongly absorbs light at this wavelength (which makes 514 nm lasers unsuitable for gold SERS substrates). That the 514 nm laser generated greater fluorescence than the 785 nm laser is unsurprising since fluorescence is commonly used to analyse DOC (Baker and Spencer, 2004, Carstea et al., 2010).

The 785 nm edge laser is air cooled, with a plasma filter and rated at 300 mW but where the power is adjustable and the power setting is controlled using the computer software WiRE 3.0. Power settings were reduced in some circumstances (discussed in section 6.3.4) when setting to laser to 300 mW was damaging the target sample. The laser light was focussed onto the sample using a microscope using a x50 long objective lens. The 50 part of the x50 long objective lens refers to the image magnification by a factor of 50, while the word long refers to the objective lens being further away from the surface being

analysed than a standard x50 objective lens would be. Use of the x50 long objective was advantageous as it enabled the lens to focus without having to worry about the objective lens coming into direct contact with the droplet of liquid being analysed.

After the light is passed through the microscope the light that is scattered from the target material is passed through a grating with 1,200 lines per mm. This grating splits up light into its constituent wavenumbers. These values are then recorded by the detector, which is typically a charged couple device (CCD) camera. In this setup it is a Master: Renishaw CCD camera.

Experiments are carried out at room temperature (20-25 °C) with no artificial heating source except for the laser itself. To prevent too much laser heating most samples were measured using an exposure time of 1 second. Although temperature has been shown to influence the results of Raman spectroscopy these differences are noticeable when comparing across ranges such as -150 °C to 75 °C (Fantini et al., 2004). All the samples analysed are kept in a separate laboratory in fridges, but these were removed from the fridge thirty plus minutes prior to analysis.

The substrates are positioned through the use of a High Speed Encoding Stage (HSES) platform, which is built into the Raman spectrometer. The HSES is used to help focus the target material under the objective lens and move the analyte into position. The HSES platform is controlled in the x, y and z planes using the WiRE 3.0 software using a proprietary controller, consisting of a tracker ball.

In summary, a sample of a target material is analysed on the HSES platform using a 785 nm laser with an edge filter, which is focussed onto the target material using a x50 long objective lens on a microscope. The scattered light is then passed through a 1200 l/mm grating and subsequently the Raman spectra is recorded by the Master: Renishaw CCD camera. All these operations are controlled using the WiRE^{TM} 3.0 software on a desktop PC.

6.3.2 Substrate and setup for standard Raman experimentation

Samples to be analysed are put onto a plate, which had to serve for both solid and liquid samples. Here two aluminium slides measuring approximately 25 x 50 mm were prepared with the samples. Onto the first of these slides, two holes capable of holding a water droplet were punched and onto the second plate nine such holes were punched (Figure 6.2). The first plate was used primarily as a backup plate for analysing powders. The second plate had less space available for depositing powders onto a flat surface due to

the punch holes, but was used for liquids. The first plate with 2 holes was also used for liquids when the first 9 on the second plate had all been used before a return to a different laboratory for cleaning. Liquids were deposited into the dimples using a water dropper. The dimples held the solutions in place, especially when the plate was moved around. A different water dropper was used for every different sample used.

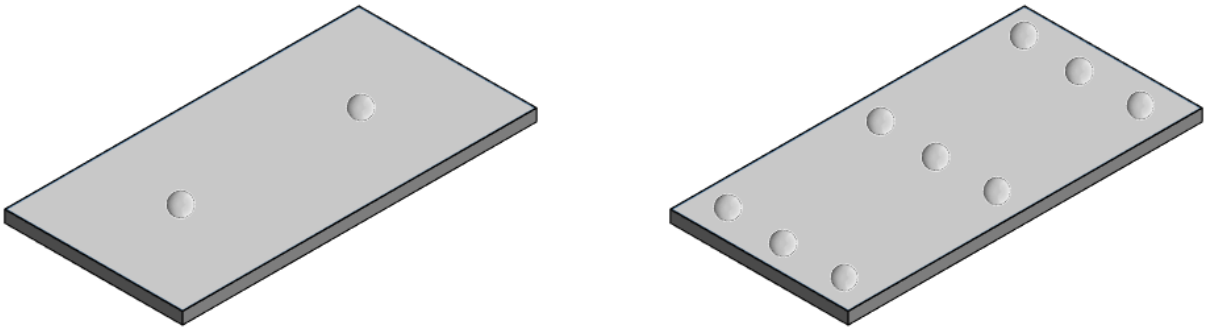


Figure 6.2 – Illustrations of metal plates
Metal plates are about 50 mm by 25 mm in cross sectional area. The first plate manufactured had two dimples with a larger flat surface area so was used to deposit powder samples onto. The second plate had nine dimples to make holding larger number of water samples before cleaning was required.

Magnifying lenses were used to focus the laser beam to be positioned on the sample. There were five lenses available to focus on the sample: x5, 20, 50, 50 long and 100 objectives.

For these experiments the 50 long x objective lens was used as this enabled a strong optical zoom for targeting the solid particles, but without coming into contact with the sample, preventing moisture from being deposited on the objective lens. The 20x, 50x and 100x objective lenses had to be positioned so close to the target sample that on occasion they would come into contact with the solution being analysed which could potentially deposit water, dirt or other substances onto the lens. The use of the long lens was particularly important for the samples measured using standard Raman as the aluminium plate did not have a completely flat surface. Using the 5x or 50 long x objective lenses enabled the focussing of the laser without risking the objective lens, however the 5x lens did not focus on a sufficiently specific point for this project. The spectrometer is equipped with both an eye piece and a camera. During the analysis, the camera was used for focussing the lens. The position of the samples under the objective can be controlled from the computer either up/down, forward/back or left/right.

The objective lens was simply focussed using the computer controls until the image on the screen came into focus. Solids were easier to focus on; liquids would appear as black on the screen and therefore had no direct method of focussing. To focus on liquid solutions (illustrated in Figure 6.3), the laser was focussed on the aluminium as close to the water

droplet as possible. The aluminium plate that samples are sitting on is then positioned using computer controls so that the water droplet is directly beneath the objective lens. When the SERS analysis of liquids was conducted the objective lens was primarily focussed on the structure of the SERS substrate, although attempts at obtaining data were also made by moving the sample plate vertically up and down out of focus to identify whether better Raman spectra were obtained by doing so.

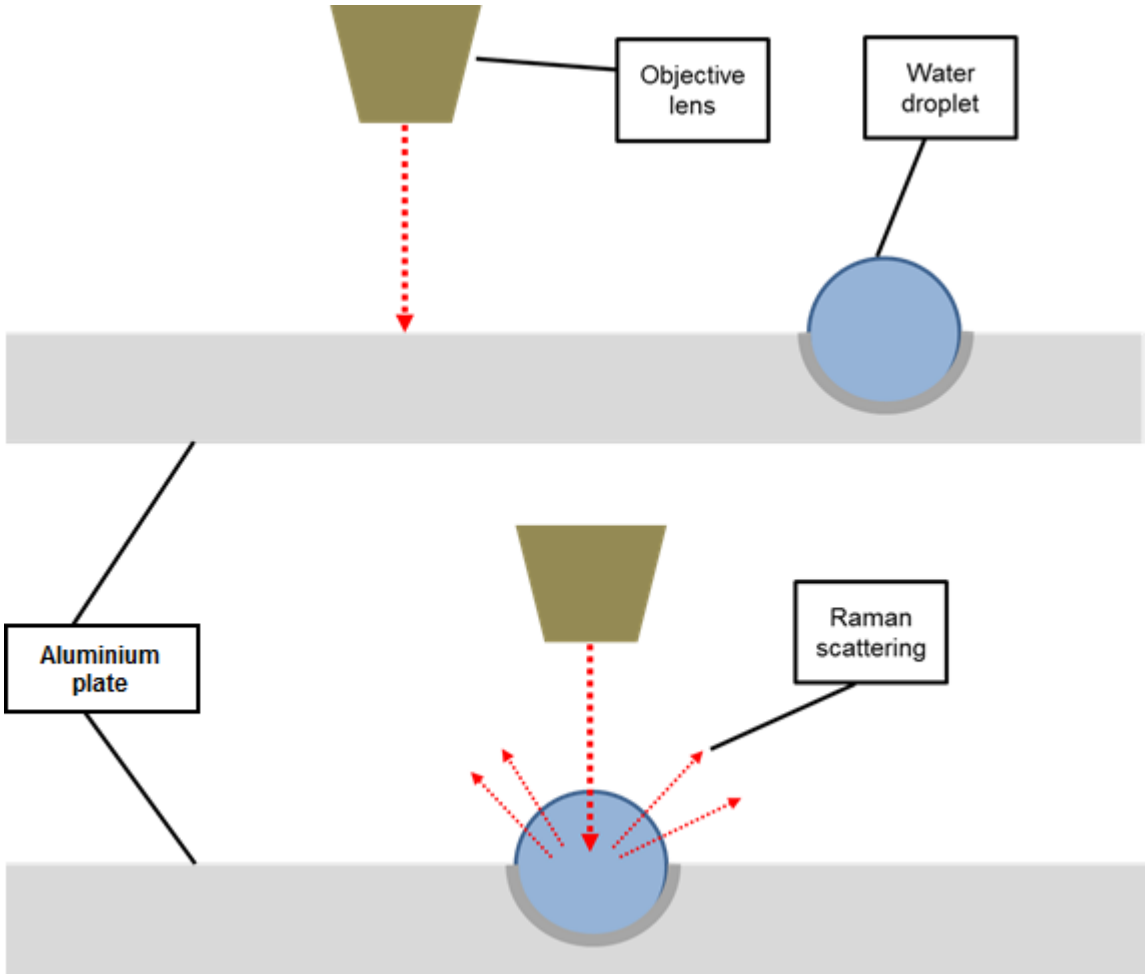


Figure 6.3 – Illustration of focussing objective lens for analysis of water droplet The lens is focussed onto aluminium near to the water droplet (top diagram) and then the water droplet moved into position. The diagram is not to scale.

6.3.3 Substrate and setup for standard Raman experimentation

For SERS analysis, substrates purchased from diagnostic AnSers were used. These substrates are printed onto paper with the substrate comprised of gold manifest in a rough pattern. The manufacturer's notes from Diagnostic AnSers indicated that the gold SERS substrates that they manufactured are designed for use with a 785 nm laser and so this was the laser used for all SERS experimentation. When the substrates were delivered, a

sample of BPE (1,2-Di(4-pyridyl)ethylene) was delivered as a standard solution with a known Raman spectra. BPE is a commonly-used standard for testing SERS substrates (Félidj et al., 2002, Chaney et al., 2005, Billot et al., 2006). The BPE produced a clear Raman spectra consistent with that supplied with the substrate and this was clearly differentiable from the background measurements made extensively throughout this project. Raman bands were recorded with peaks being measured at 1023.1 cm^{-1}, 1201.5 cm^{-1}, 1341.6 cm^{-1}, 1608.7 cm^{-1} and 1639.4 cm^{-1} (Figure 6.4). The Raman bands identified using the standard compared well with references in the literature of expected with Raman bands for BPE at 1200 cm^{-1}, 1340 cm^{-1}, 1610 cm^{-1} and 1640 cm^{-1} (Félidj et al., 2002). This test was conducted on 18 August 2014, 13 November 2014 and 17 June 2015 with no significant differences in the recorded spectra measured over this period of time, indicating that there was no significant degradation in the quality of the substrates over the period of time that they were used.

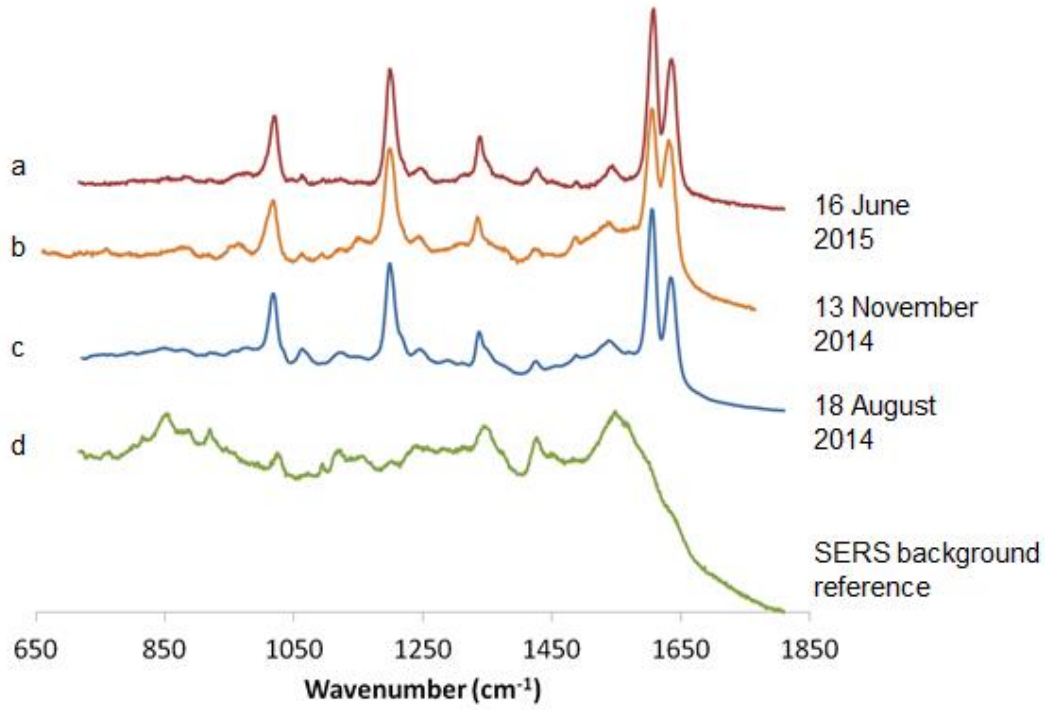


Figure 6.4 Comparison of BPE SERS results to test degradation of the substrate
The difference between the BPE SERS signals and the SERS background reference is noticeable with strong Raman bands.

The samples of solution used for the Raman experimentation were stored in the same way as described for the laboratory measurements of [DOC] that were measured by the Thermalox® (Chapter 3). After removal from the fridge the samples would be moved to the laboratory housing the Raman spectrometer and by this time would have warmed to close to room temperature compared with the 5°C they were stored at.

6.3.4 Laser settings: power, exposure time and number of accumulations

The intensity of the Raman spectra measured is dependent not only on the target material and how well the laser is focussed, but also the laser's settings. There are three main settings that could be varied on the spectrometer:

1. Laser power
2. Length of exposure time
3. Number of accumulations

The laser power could be set as high as 300 mW, but using the computer this could be reduced using a pre-set list of percentages. The laser power is not controlled directly but the light is passed through a motorised filter wheel with neutral density filters, which can be used to adjust the power settings using the WiRE™ 3.0 software. There are 16 available laser power settings, which range from 0.00005% to 100% of the power rating. Therefore, it was not possible to identify a precise maximum power value before the SERS substrate started to exhibit visible spectral signs that it was being damaged by overly intense light. SERS substrates and humic powders would burn if the laser power was too high so these had to be analysed at lower power settings. Figure 6.5 summarises how the laser power was set depending on the substance being analysed during experimentation. The laser was set to the highest power possible without damaging the sample (or the SERS substrate where one was used). For example, humic and fulvic acid powders burned at high laser powers, which is likely occurring because the excitation intensities are too high resulting in thermal decomposition (Vankeirsbilck et al., 2002a).

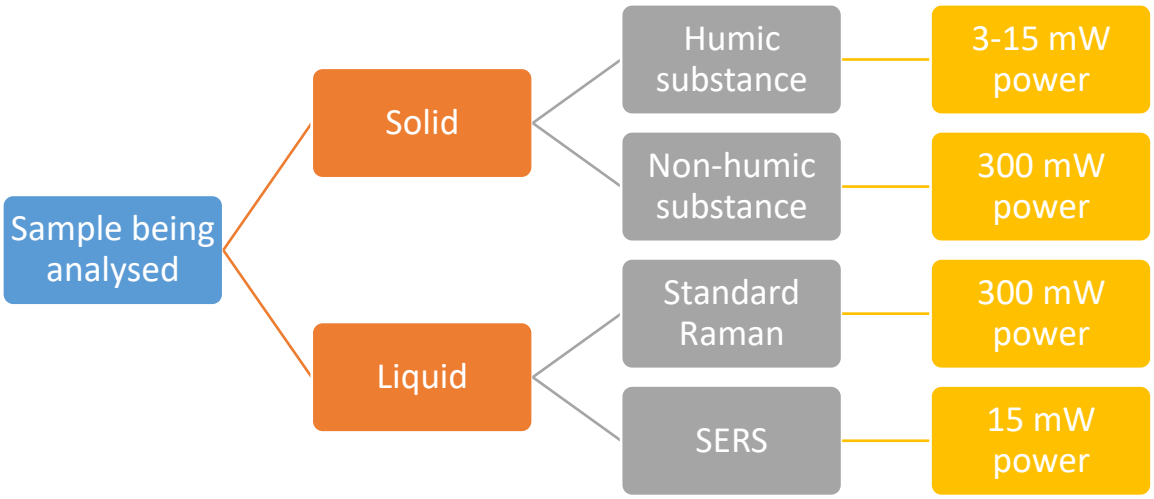


Figure 6.5 – Determination of power settings for different experiments
To improve results the amount of information collected from the Raman spectrometer either the exposure time or the number of accumulations can be increased. A single accumulation refers to a single measurement of the spectra. Five accumulations mean

that the spectrum is measured five times, which has the effect of averaging out noise or other unwanted signals.

When long exposures are used it enables weaker Raman bands to be identified but it can unfortunately also heat up a target material causing it to burn if the substance is volatile. Consequently, an exposure time of one second was used in almost all cases as the SERS substrates unfortunately burn if the exposure time was too long. The number of accumulations describes the number of times that a laser is fired at a target. A larger number of accumulations is useful for reducing the signal to noise ratio (SNR) by allowing many scans of the same object to be averaged and this will remove random artefacts of noise. For SERS experimentations the instructions issued by Diagnostic anSERS$_{Inc}$ recommended 5 accumulations. However, usually 20 accumulations were used in this project in order to reduce the SNR, and no problems were encountered while doing this. There are periods where a target is not being exposed to the laser light, and so the increased number of accumulations did not appear to result in the same problems of overheating substances caused by long exposure times.

6.3.5 Preparation of samples

The samples analysed were deposited onto either the aluminium slide or the SERS substrates.

1) Blank measurements were taken of the aluminium and SERS substrates as well as the DI water used in the sample preparations. These background measurements were made to prevent any Raman peaks being falsely attributed to organic carbon, which might in fact be background signals. The aluminium slide was scanned by the Spectrometer using 300 mW and the SERS substrate was scanned using 15 mW power. After this DI water was dropped onto the aluminium slide and the SERS substrate and the resulting spectra was also measured.
2) The BPE standard was measured.
3) Sugars (glucose and sucrose) were measured in three steps:
    a. As solid powder. The sugars were deposited onto the flat of the aluminium slide and the objective lens was focussed on one of the granules.
    b. Droplets were deposited onto the aluminium slide to determine if the sugars could be detected in solution. In the case of glucose and sucrose limit of detection tests were performed.
    c. The sugars were tested on the SERS substrates to determine if we could measure enhanced Raman signals. If enhanced signals could be detected then a new limit of detection test could be performed.
4) Humic acids and fulvic acids were tested in the same three phases as the sugars.

a. As solid powders, deposited onto the aluminium slide.
b. Droplets were deposited onto the aluminium slide to determine if the sugars could be detected in solution and if so at what concentration?
c. Droplets were deposited onto the SERS substrate to determine if we could measure enhanced Raman signals and if so at what concentration?

It was easier to focus the objective lens and therefore the laser on the lighter-coloured fulvic acid, especially as this material was looser and seemed to form a more even surface than the humic acid, which was more 'clumpy'.

If the above experimentation was successful then the sample from Drumtee could be analysed to determine if any of the Raman bands from this sample would match any Raman bands measured in the sugar or humic substances.

Glucose and sucrose solutions were produced to concentrations of >100,000 mg/l but doing this for humic and fulvic acid was not practical. Firstly, humic and fulvic acids were expensive and only available in only small quantities so how concentrated a solution could be made was limited Secondly the low quantities available meant a high concentration solution would require only a few ml of water and there was difficulty dissolving the powders: the humic and fulvic acid powders were quite bulky and of low density and tended to settle out of solution when the concentration was too high. The highest concentration of humic acid produced (excluding the analysis of the powder) was approximately 26,000 mg/l.

6.3.6 Preparation of humic substances with increased pH and subsequent reduction

It has been proposed that clearer Raman spectra of humics may be obtained using SERS when the pH for the solution is increased (Vogel et al., 1999, Francioso et al., 2001, Leyton et al., 2005, Corrado et al., 2008). Therefore, an experiment was set up where solutions of humic acid were prepared and then the pH was adjusted using sodium hydroxide until the solution was measured to a target value of about 10. Due to the small volumes it was hard to be too precise in the target pH, but based on the studies by Leyton and Corrado the important issue seemed to be to make sure that the pH was increased to an alkaline value, so close to 10 was fine. Solutions of humic acid were prepared to 2500 mg/l using standards obtained from the International Humic Substance Society (IHSS) and deionised using UV-filtered water. To prepare the 2500 mg/l solution 0.5 g of this solution was weighed on a scale and then diluted using pH 7 buffer until ~ 25 g of solution had been added. The concentration of the undiluted humic acid could not be measured

because the volume of liquid was too small. The pH of the solution with humic acid and pH buffer was measured to be 7.1 and the pH of the solution with humic acid, pH7 buffer and sodium hydroxide was 12.8.

Recently it has been proposed that chemically-reducing a solution of humic acid after the pH has been increased can improve the SERS spectra observed in experimentation (Heighton, 2013). As this was a time-consuming procedure both the humic acid and fulvic acid were prepared simultaneously between 28 January and 30 January 2015. The experimental setup was based on the experiment described by Heighton and also a 2010 paper by Ma (Ma et al., 2010) where the humic acid solution is reduced after adding 5 mg of borohydride to the solution and purging with nitrogen gas.

The procedure started by measuring 0.0025g humic acid powder into a centrifuge tube. Subsequently 50.0013g of DI-UV-filtered water was added to create a solution of 50.0 mg/l humic acid. An aliquot of 15.3 g of solution was retained in a centrifuge tube for analysis. Heighton's thesis states that prior to the reduction process the pH of the humic acid was increased to an alkaline value in a manner similar to that described before (Heighton, 2013) although no such instruction is given in the Ma paper (Ma et al., 2010). A decision was made to increase the pH to a target value of 10 to achieve this. To the humic acid solution, 6.0 g of sodium hydroxide was added, which increased the pH to 10.5, but reduced the concentration of humic acid to 35.9 mg/l, lower than the target solution as listed in the Ma paper. In the event that the reduction reaction produced additional Raman spectra, the sodium hydroxide could have been tested to ensure that this was not the source of these additional spectra (but this was not needed).

The same procedure was subsequently conducted for the fulvic acid, using 0.0027 g of powder to make a solution that was calculated to be 52.7 mg/l of humic acid. To a 16.2 g of solution aliquot, 7.1 g of sodium hydroxide was added, creating a solution of 23.4g, with a pH measured at 10.6. The resulting concentration was 36.6 mg/l.

The second phase of the reduction reaction was conducted by sealing approximately 3 ml of each solution in a sealed plastic cell and purging with nitrogen gas until the solution had degassed. Then 5 mg of sodium borohydride was added to each of the 3ml solutions and after mixing the containers were resealed and purged with nitrogen gas overnight. This concluded the preparation phase of the reduced humic and fulvic acids. Subsequently the SERS spectra of the reduced humic and reduced fulvic acids were obtained using the 785 nm laser.

6.3.7 Stokes experimentation using standard Raman spectroscopy

Stokes-based Raman Spectroscopy was conducted on stream samples obtained from Drumtee water and samples made up in the laboratory. Initial experimentation revealed that the 785 nm was the most effective laser for analysing humic substances (regardless of whether they were in a powder or solution phase) because humic substances are fluorescent and the fluorescence was more intense when using the 514 nm laser. As these fluorescence signals would potentially obscure any Raman peaks produced it was important to minimise their prevalence and hence the use of the 785 nm laser.

Measurements were made of samples obtained from Drumtee, humic and fulvic acids obtained from the International Humic Substance Society (IHSS), and two sugars (glucose and sucrose) from Sigma Aldrich. During analysis both solid phase and solutions were tested. For any substance that produced clear Raman spectra when in high concentration solutions, further solutions were prepared at different concentrations to find the limit of detection.

To increase the chances of measuring and identifying Raman bands related to the DOC samples collected from Drumtee, a concentrated sample was prepared. Typically, a stream sample would have a concentration of 5 – 60 mg/l. More concentrated solutions generate stronger Raman signals in substances that produce Raman spectra and a solid extract not dissolved in solution will usually produce the strongest Raman signals. Using rotary evaporation and freeze drying I extracted a powder sample from a 5 litre water sample collected from Drumtee in November 2013.

Most Raman studies focus on wavenumbers between 0 cm^{-1} and 1800 cm^{-1}. The region of approximately 600 cm^{-1} up to 1800 cm^{-1} is where preliminary analysis coupled with reading of studies in the literature (Leyton et al., 2005, Heighton, 2013) indicated might yield the most interesting results with respect to humic substances. However, samples were initially analysed using an extended scan from 130 cm^{-1} up to 3200 cm^{-1} to ensure nothing was being missed and identify which regions to focus on. This width of scan was not used in the full analysis as the data returned were noisy. The Renishaw In-via Raman spectrometer allowed the selection of a target wavenumber and it would scan wavenumbers above and below this in a range. For example, the most commonly selected wavenumber was 1250 cm^{-1} and so the scan was from 660 cm^{-1} to 1765 cm^{-1}.

Most of the spectra were recorded between 716.8 cm^{-1} to 1810.4 cm^{-1}, a region commonly considered in the literature reviewed for organic carbon. Leyton measured between 300 cm^{-1} and 1800 cm^{-1} with most of the Raman spectra identified as being between the

middle and larger wavenumbers (Leyton et al., 2005). Heighton measured between 750 cm^{-1} and 1800 cm^{-1} (Heighton, 2013). Corrado used the range from 200 cm^{-1} to 1800 cm^{-1} to identify the Raman bands relevant to humic acid and a broad scan that included a 3420 cm^{-1} wavenumber, which was used to calibrate the spectra (Corrado et al., 2008).

6.4 How to read Raman graphs

Graphs displaying Raman information have two axes. The x-axis displays the wavenumber, which is defined as 1/wavelength and the y-axis displays the magnitude of the wavenumber. There are several "peaks" on most Raman spectra but to prevent confusion between these different peaks and the absolute maximum value on a spectrum, each of these peaks is referred to as a "band". Each band is identified by the wavenumber at which the maximum intensity within that band is recorded.

Raman graphs do not always list the numerical value of the intensity on the y-axis, for example spectra have been presented for mine waters (Sobron et al., 2009), carbon nanotubes (Al-Attar et al., 2012), graphene (Ferrari and Basko, 2013), glucose and anthrax (Zhang et al., 2006), cocaine (Garrido et al., 2007) and exploration of surface-enhanced Raman scattering (SERS) (Campion and Kambhampati, 1998) with no y-axis intensity. Some studies use the absolute intensity measured of the Raman spectra on the y-axis, while others normalise the minimum and maximum values recorded to values between 0 and 1. The reason for this is that Raman is sometimes used for quantitative purposes such as the study of nitrate impurities in water where higher concentrations did yield greater Raman intensities at the target wavenumbers of 1056 cm^{-1} and 1326 cm^{-1} (Gajaraj et al., 2013), but it is the relative intensity of the bands that are of interest not the absolute values.

In this dissertation a mixture of graphs with and without intensities have been rendered. For example, where multiple traces are stacked, it is easier to enable comparison of wavenumbers if an intensity scale is not used. Where intensities are used, all graphs have been normalised between values of 0 and 1.

6.5 Results

Results are presented here for the Stokes Raman spectra of the different substances that were analysed, including the background signals of the aluminium, SERS substrates and water blanks. Following the presentation of the blank and background results are the results from glucose and sucrose Raman experiments, which were measured in both powder and solution form. Only the standard Raman results are presented for sugars

because the SERS analysis produced no enhanced signals and given an already large number of graphs and all that was observed was the background signal, the graphs have not been included. Finally, the results from humic and fulvic acids are presented including solid and solution measurements. Solutions of humic substances included humic and fulvic acids dissolved in water, the same solutions but with pH artificially increased, and finally those solutions measured after undergoing a reduction reaction.

Sometimes research results are presented with a correction to remove fluorescence signals, but in the data presented here no background correction has been applied. Removing the background involves manipulating the results and it was determined that for this project that as humic substances are strongly fluorescent (Baker and Spencer, 2004, Esteves et al., 2009) that it is important to determine how easy it is to distinguish between any overlapping Raman signals and fluorescence signals. It was also decided not to filter out the background signal as removing the background signal when the Raman signals are very weak can result in random noise appearing to look like Raman signals.

6.5.1 Blanks, background and standards

The measurement of aluminium (Figure 6.4) from 716.8 cm^{-1} to 1810.4 cm^{-1}, produced a largely featureless spectrum with broad spectral features that are at maximum of 716.8 cm^{-1} within this range and then a broad spectral feature appearing as a “hump” between 1200 cm^{-1} and 1810.4 cm^{-1}.

To identify DOC Raman spectra, it was important to assess what features of the background signal were significant and if adding water to the aluminium slide or SERS substrate yielded a stronger background signal. The Raman spectrum measured for aluminium alone (Figure 6.6) contains a slope that decreases as the wavenumber increases between approximately 700 cm^{-1} and 1750 cm^{-1}. When water is measured on the aluminium a weak Raman spectrum is observed at approximately 1620 cm^{-1} (Figure 6.7), but otherwise the spectroscopic trace measured is similar to that for the aluminium slide. The water spectra (Figure 6.7) shows two traces taken with the 785 nm laser of water on the aluminium slide. The two traces are similar but not identical and one of them exhibits a weak Raman band at 987.4 cm^{-1}.

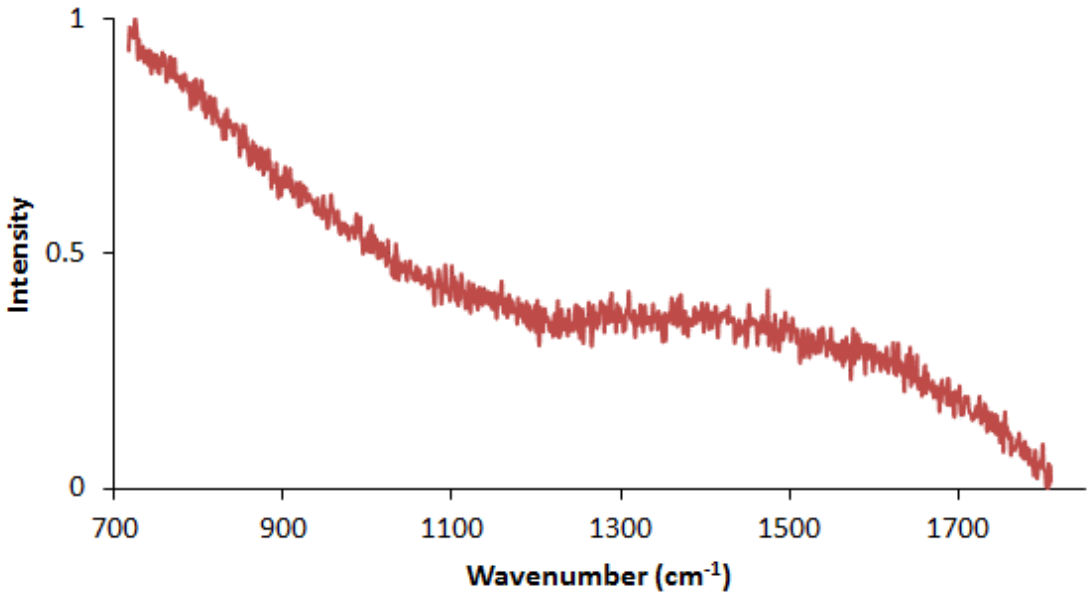


Figure 6.6 – Raman spectra of aluminium

Raman spectrum with no liquid or other target material deposited on it. The spectrum is relatively featureless and noisy because of the low signal strength of the measurements made.

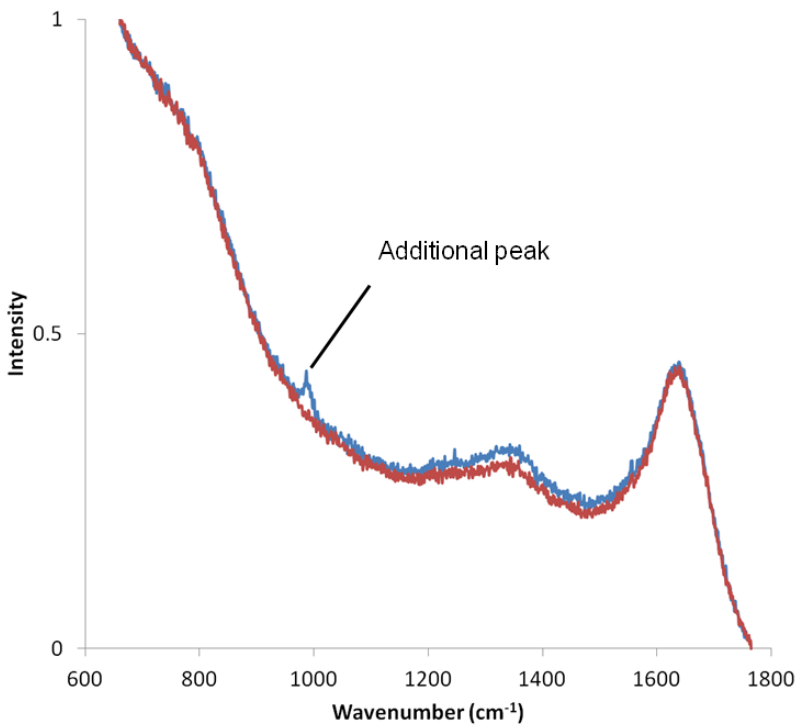


Figure 6.7 – Raman spectra of droplet of water
A water droplet was placed onto aluminium plate used for using standard Raman spectral analysis. Two spectra are displayed, being taken a few minutes apart. to show that although there are slight differences and inconsistency in measurements the two spectral lines are similar. The different colours are used to make the different spectra easier to distinguish from one another.

In addition to measuring the aluminium background spectrum, it was also important to measure the background spectra of the SERS substrates, to ensure that Raman spectra were not falsely attributed to the target material (e.g., sugars, humic acids etc). The

background spectrum of the SERS substrate (Figure 6.8) is somewhat inconsistent and does not contain any particularly strong Raman bands. These spectra were measured using three SERS substrates.

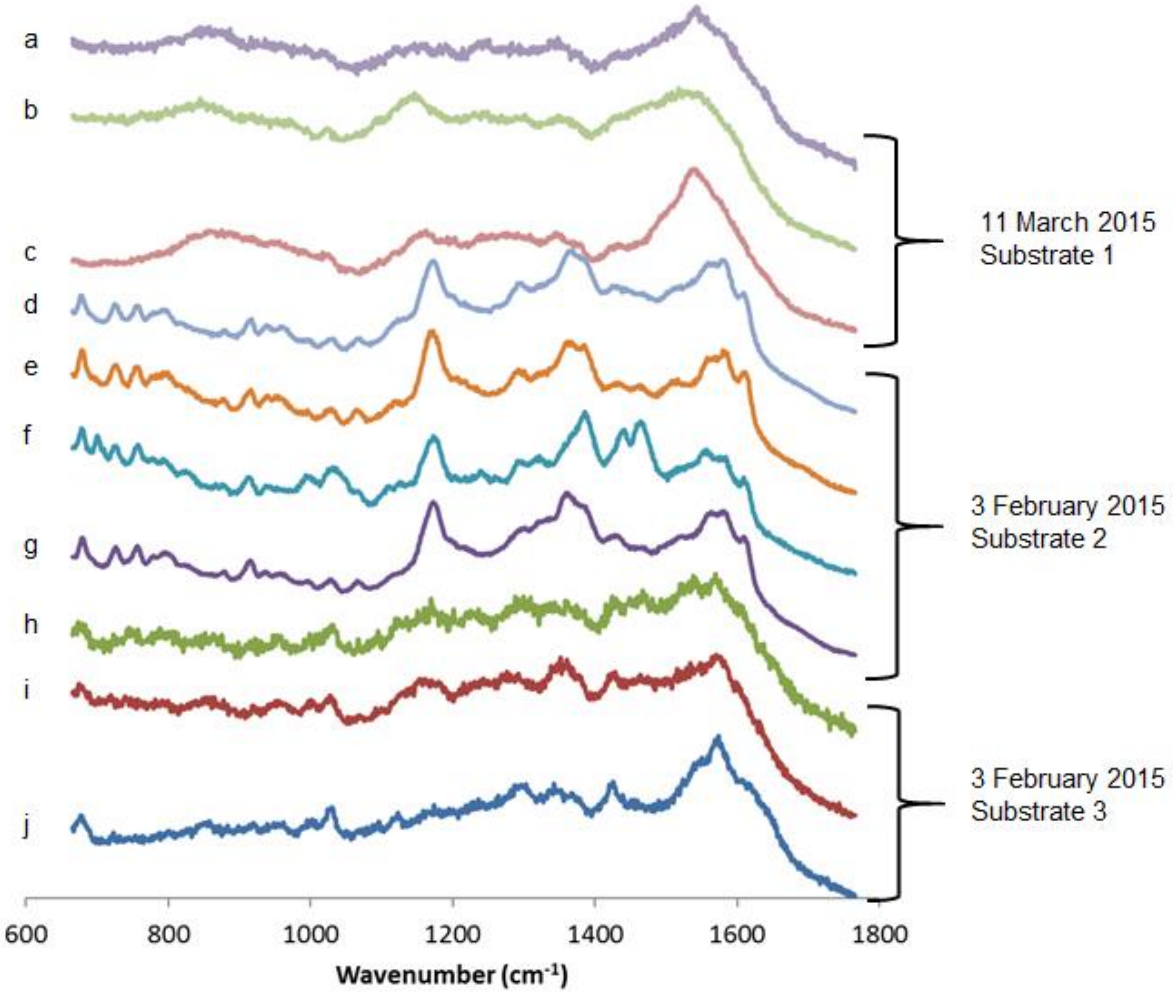


Figure 6.8 – 10 gold SERS background spectra
10 gold SERS background spectra measured from 3 different SERS substrates. The power, exposure and accumulation settings were not recorded for the measurements of substrate 1 on 11 March 2015. Substrates 2 and 3 (two separate substrates), recorded on 3 February 2015, were measured using power setting of 15 mW and 5 accumulations with a 1 second exposure time on a single substrate.

When DI water was added to the SERS substrate (Figure 6.9), the spectra remained very similar to the SERS spectra without water. One difference between the two is that the SERS with added water provided a more consistent signal than did the SERS spectra shown on Figure 6.7. However, neither the SERS reference nor the SERS of water produced strong spectral bands when compared to SERS active substances such as BPE (Figure 6.4).

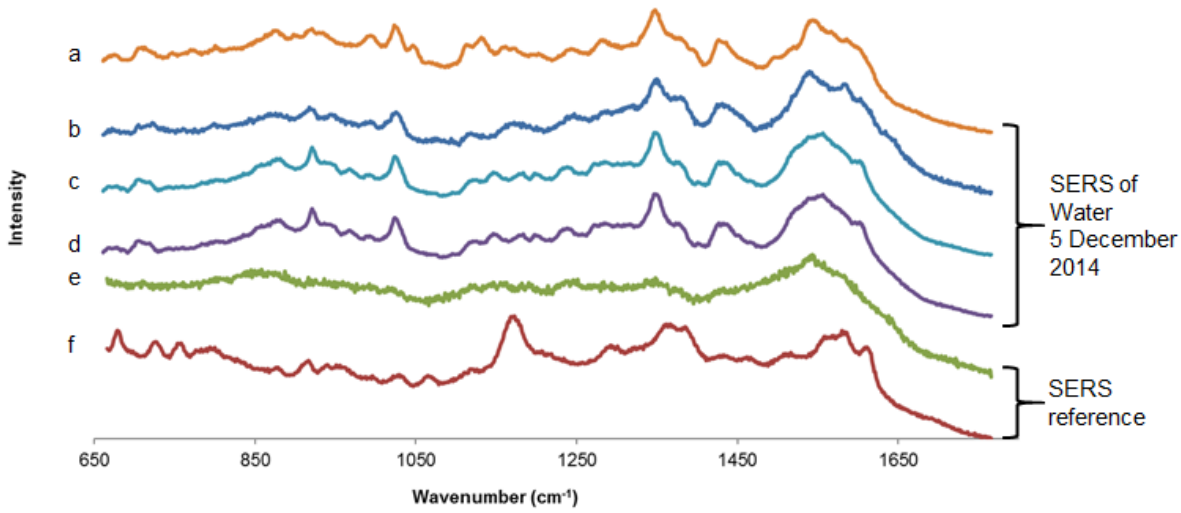


Figure 6.9 – Four SERS spectra of DI water
Four SERS spectra of DI water recorded on 5 December 2014. The 785 nm laser was used at 3 mW power with a 3 second exposure for 5 accumulations.

Not all of the Raman bands identified were necessarily a true representation of the substances being analysed as errors could arise. There were four anomalous results conducted when measuring the Raman spectrum of water on SERS (Figure 6.10). Three of these anomalous spectra were recorded on 5 December 2014 and the other was recorded on 18 March 2015. The anomalous Raman spectra observed all returned different observed maximum values (Table 6.1). The values extracted had Raman bands that were clearly differentiable from background noise. However, although these were consistent across the 20 accumulations measured these strong Raman bands did not show up when water was tested on other substrates (see Figure 6.10). Signal (d), taken on 18 March 2015 showed similarities to one of the SERS substrate reference signals, which had been recorded on 7 February 2015 (Figure 6.7). The results in Figure 6.8 and Figure 6.9 show that for whatever reason anomalous results could be encountered in this analysis and that any experiment that appeared to show an interesting Raman band would require retesting to prevent any Raman signals being falsely attributed to the target substance being analysed.

Date of anomalous water result and the wavenumbers of the bands recorded			
5 December 2014 Wavenumber (cm^{-1})	5 December 2014 Wavenumber (cm^{-1})	18 March 2015 Wavenumber (cm^{-1})	5 December 2014 Wavenumber (cm^{-1})
1095.9	925.3, 1027.3, 1113.8, 1224.9, 1351.6, 1429.4	739.8, 757.9, 802.1, 918.3, 946.1, 1109.3, 1172.4, 1303.7 1372.8, 1385.4, 1584.5, 1615.7	923.0, 1623.9, 1346.3

Table 6.1 – Wavenumbers of anomalyous water SERS spectra
List of the four anomalous water results from the SERS experimentation. Wavenumbers are given in cm^{-1}. There were 1, 6, 13 and 3 Raman bands observed in the four respective spectra.

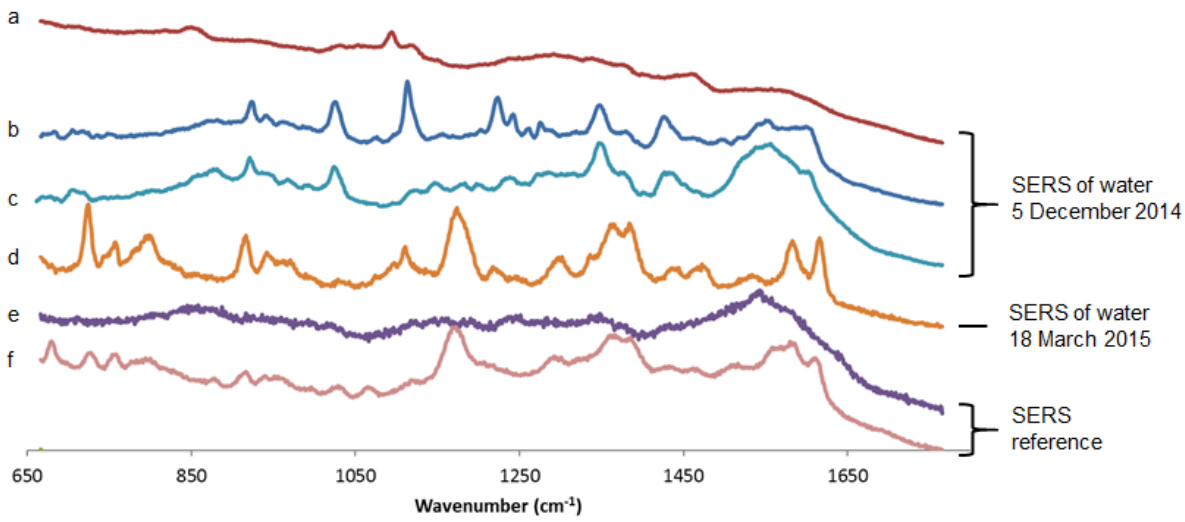


Figure 6.10 – Anomalous water results on SERS substrates
Three of the samples ((a) – (c)) were measured on 5 December 2014 and the other sample (d) was measured on 18 March 2015. Two background signals measured from the SERS substrate are given as a reference at the bottom of the plot.

A final set of background results was obtained of the four reagents used to prepare the various humic and fulvic acids when increasing their pH or subjecting them to reduction reactions (Figure 6.11). Sodium borohydride was not tested for health and safety reasons and because it was added in a very small mass. Additionally, the reduced humic and fulvic acids did not return additional spectral information (as will be shown in section 6.5.4) and although the spectrum of sodium borohydride was generated this was not problematic as spectra of neither the reduced humic or fulvic acid were generated. Had these generated results the sodium borohydride would have been tested once additional safety constraints has been identified, but as it transpired this was not necessary. The hydrochloric acid and sodium hydroxide produced Raman spectra that were differentiable from the background. The two pH buffers did not generate any intense Raman bands.

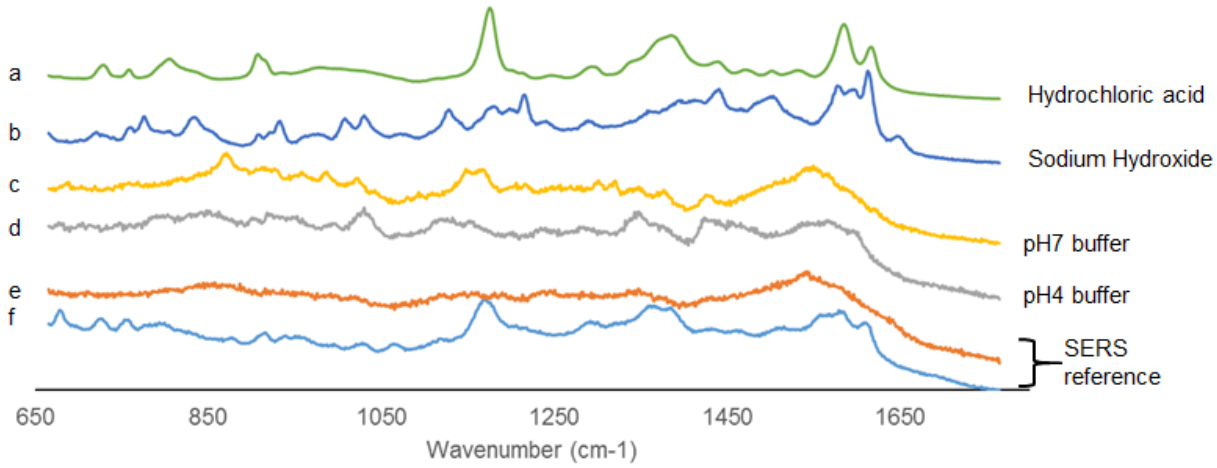


Figure 6.11 – SERS spectra of preparation reagents
The spectra generated by preparation reagents used to manipulate the pH of the humic and fulvic acids (which were also used in reduction reactions for humic and fulvic acids). The SERS substrates spectra are provided as a reference signal as signals (e) and (f).

6.5.2 Sugars

Glucose, fructose and sucrose were first tested in a powder form to obtain the true Raman spectra of these substances. Each of these experiments produced a clear spectrum (Figure 6.12 for Glucose and Figure 6.13 for Sucrose), similar to that documented in the literature (e.g.(Shao et al., 2012) for glucose and (Mathlouthi and Luu, 1980) for sucrose). The specific Raman bands are clearest in powder, but in solution these Raman bands merge with one another and produce a smoother spectral profile. Limit of detection tests on glucose and sucrose established this concentration to be 50,000 gm/l and 20,000 mg/l respectively. It is possible to possibly identify Raman bands in the sucrose spectrum below 50,000 mg/l due to the better SNR for sucrose than glucose.

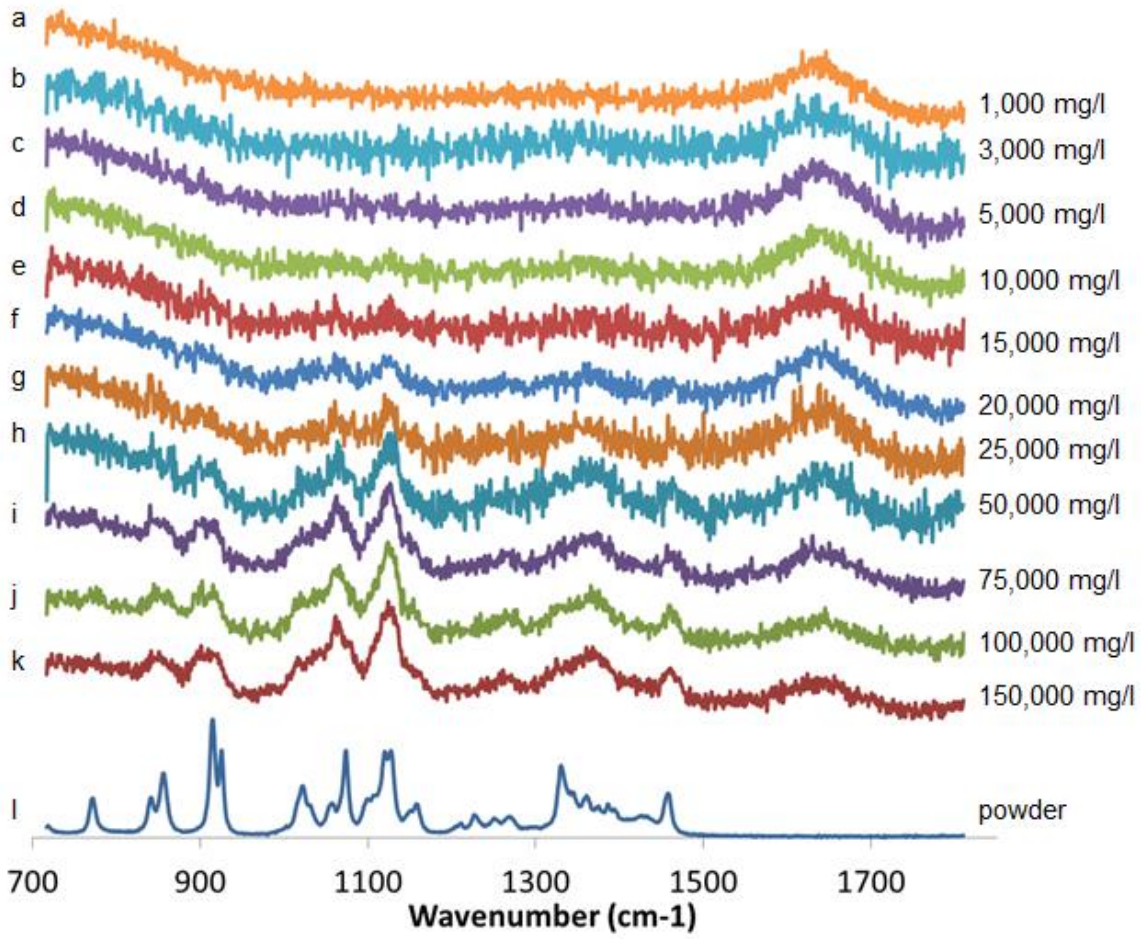


Figure 6.12 - Glucose limit of detection experimentation spectra
As samples become more concentrated going from top to bottom (a – l) then the Raman bands become clearer and easier to distinguish from the background. Raman bands are observable about 50,000 mg/l and debatably at 25,000 mg/l. Measurements were made using a 785 nm laser, with 20 accumulations at 300 mW, with a 1 second exposure time.

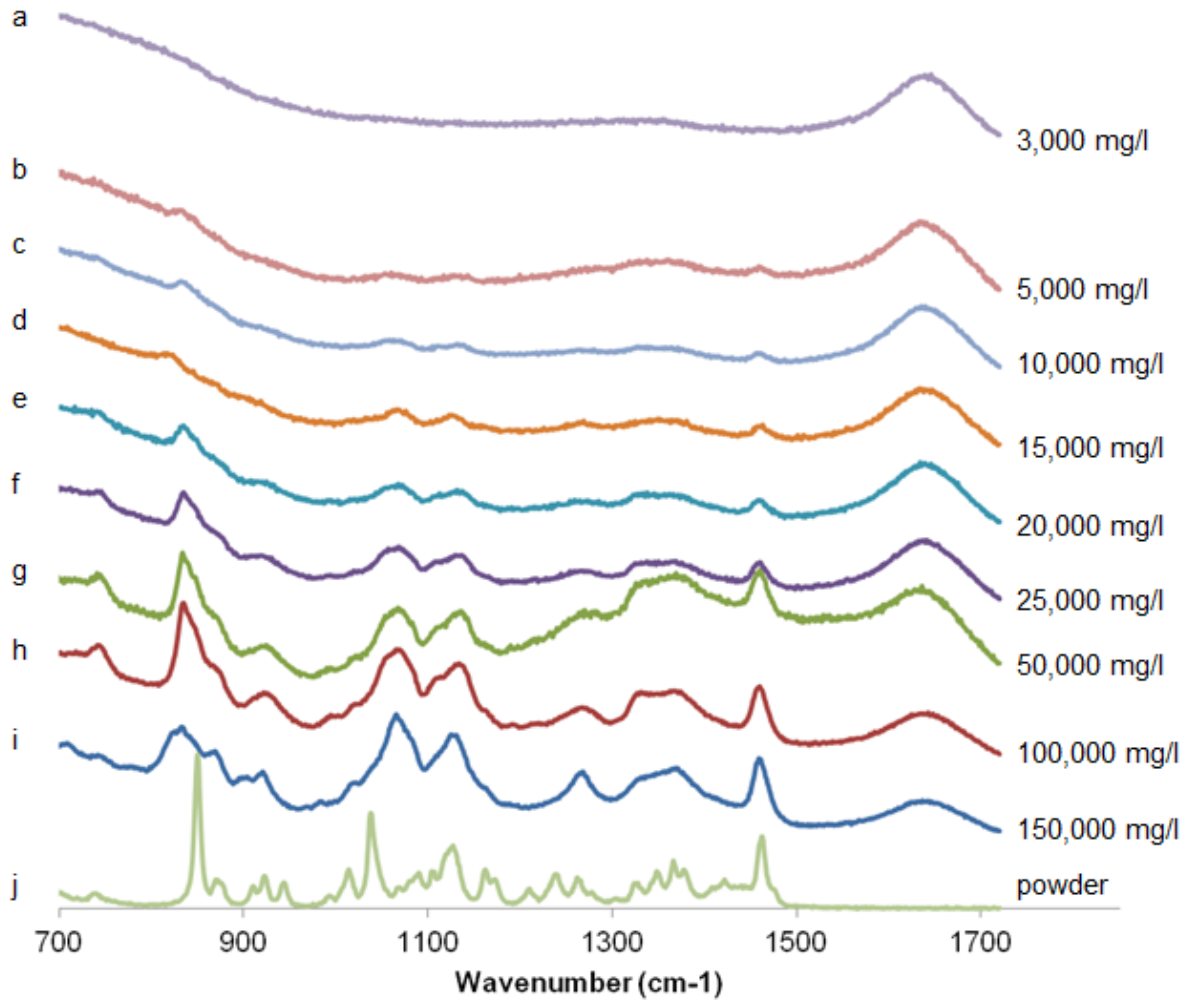


Figure 6.13 – Sucrose limit of detection Raman spectra
As samples become more concentrated going from top to bottom (a – l) then the Raman bands become clearer and easier to distinguish from the background. Raman bands are observable about 50,000 mg/l and debatably at 25,000 mg/l. Measurements were made using a 785 nm laser, with 20 accumulations at 300 mW, with a 1 second exposure time.

6.5.3 Humic and fulvic acids

Humic and fulvic acids were analysed using Raman spectroscopy in four stages:

1. the powders of humic and fulvic acid as they were supplied from the IHSS
2. comparison of two solutions of humic acid (one dilute at approximately 9 mg/l and the other concentrated at approximately 2,500 mg/l) on both an aluminium slide and SERS substrates
3. comparison of humic acid at different pH values in an attempt to mimic the results from Leyton and Corrado (Leyton et al., 2005, Corrado et al., 2008).
4. measurement of humic and fulvic acids subsequent to a reduction reaction to compare with results from Heighton (Heighton, 2013).

Humic and fulvic acid powders both exhibit strong “wide band” signals but no obvious Raman bands (Figure 6.14). The source of these wide bands is likely from fluorescence spectroscopy as humic material has been noted as being strongly fluorescent (Matthews et al., 1996, Baker and Spencer, 2004, Saadi et al., 2006). The maximum intensity of the humic and fulvic acid spectra is measured at wavenumbers of 361.6 cm^{-1} and 251.0 cm^{-1} respectively. The spectra illustrated in Figure 6.13 shows a broad spectrum with no narrow Raman bands, which is particularly apparent if we compare these powder results with the results of the Raman measured of glucose and sucrose powders above (Figure 6.12 L and Figure 6.13 J). For glucose and sucrose there are narrow Raman hands (approximately 100 cm^{-1} across at the bases but there is not an equivalent of this in the humic or fulvic acid powders.

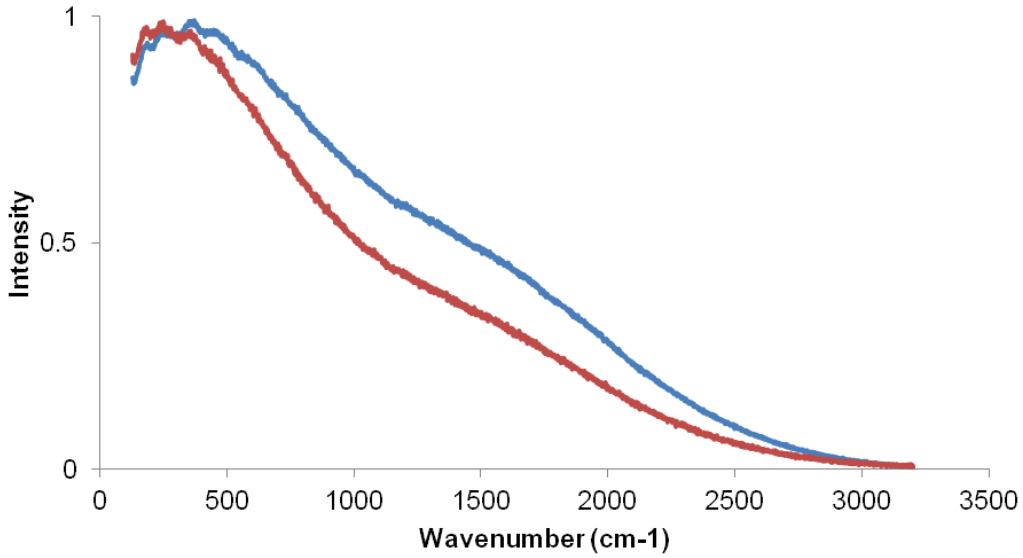


Figure 6.14 - Comparison of humic acid and fulvic acid powders.
The blue line is the humic acid powder and the red line is the fulvic acid powder.

The first set of measurements of humic acids in this project compared the results of both concentrated and dilute humic acid spectra on both aluminium and a SERS substrate (Figure 6.15):

a) 2,500 mg/l of humic acid deposited onto the SERS substrate compared with 2,500 mg/l of humic acid deposited onto the aluminium slide. The spectrum of the 2500 mg/l humic acid on aluminium is essentially featureless with no distinct Raman bands. In contrast, the SERS spectrum of 2500 mg/l humic acid produced two Raman bands that merged at about 1400 cm^{-1}, with the Raman band attributed to 1520 cm^{-1}.

b) 2,500 mg/l of humic acid deposited onto the aluminium slide compared with 9 mg/l of humic acid also deposited onto the aluminium slide. The Raman spectrum of the 2500 mg/l sample is relatively featureless, but in the 9 mg/l humic acid there is a Raman band about 1640 cm^{-1}. This is the same band observed in Figure 6.7, where the 1640 cm^{-1} was attributed to water. The 9 mg/l humic acid solution produces a Raman spectrum with poor SNR, resulting from a very week Raman signal.

c) 9 mg/l of humic acid deposited onto the SERS substrate compared with 2,500 mg/l of humic acid deposited onto the SERS substrate. The 9 mg/l humic acid has clearer Raman bands than the broader, less distinct bands generated from the 2500 mg/l solution. Both solutions show spectral features between 1000 cm^{-1} and 1600 cm^{-1} but the 9 mg/l spectrum has more distinct Raman bands at 1100 cm^{-1}, 1200 cm^{-1} and 1350 cm^{-1}, whereas the 2500 mg/l solution has a broad signal (distinct from the solution on aluminium – comparison (A)) but that does not have a distinctive peak.

d) 9 mg/l of humic acid deposited onto the SERS substrate compared with 9 mg/l of humic acid deposited onto the aluminium slide. The solution on the aluminium slide has the water band at 1600 cm^{-1}, whereas the solution on the SERS substrate has bands at 1100 cm^{-1}, 1200 cm^{-1}, 1350 cm^{-1} and about 1550 cm^{-1} so there is little similarity between the two signals.

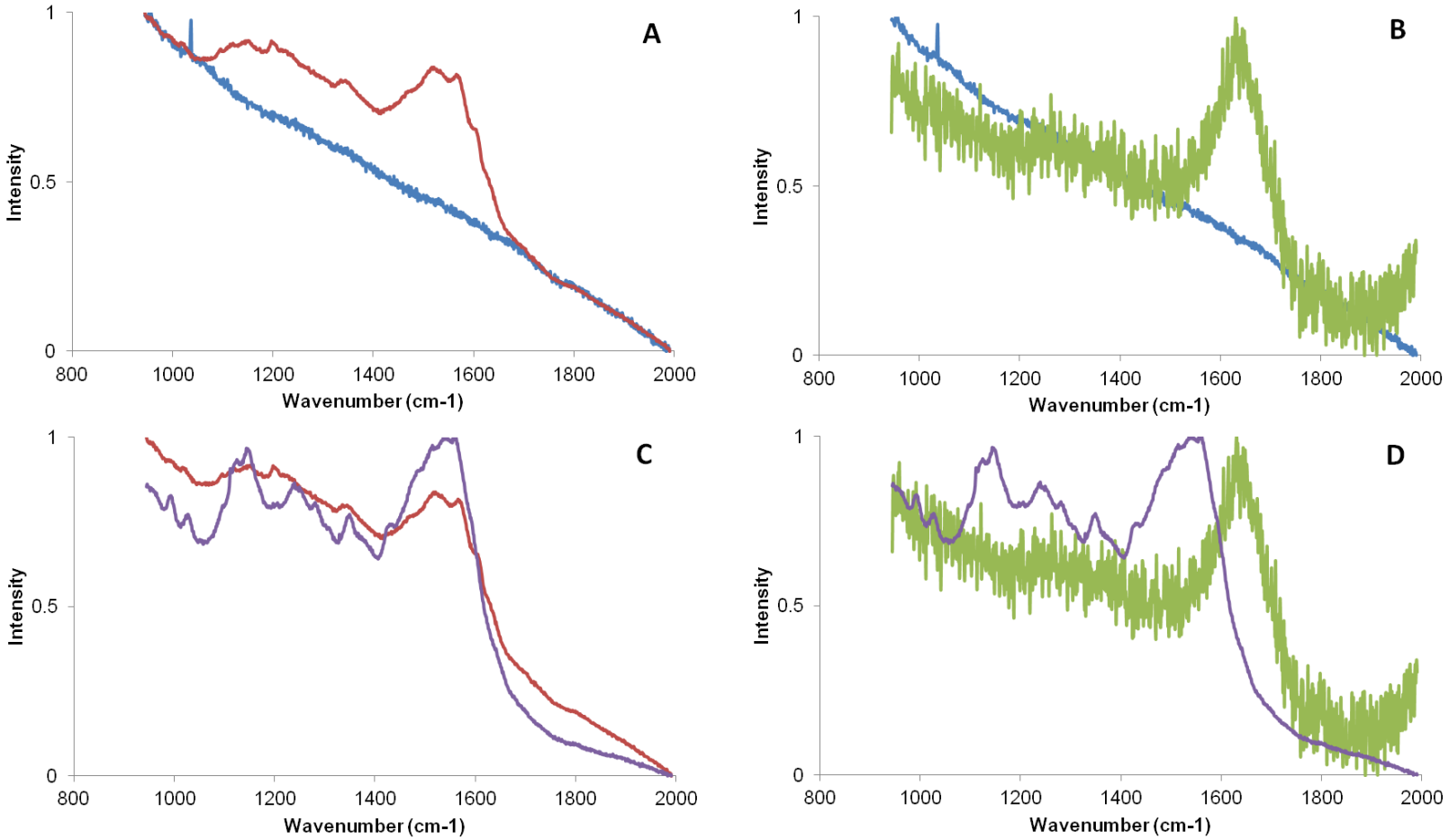


Figure 6.15 - Comparison of 4 sets of results

(A) 2500 mg/l humic acid on SERS substrate (red) and 2500 mg/l humic acid on aluminium (blue)

(B) 9 mg/l humic acid on aluminium (green) and 2500 mg/l humic acid on aluminium (blue)

(C) 2500 mg/l humic acid on SERS substrate (red) and 9 mg/l humic acid on SERS substrate (purple)

(D) 9 mg/l humic acid on SERS substrate (purple) and 9 mg/l humic acid on aluminium (green)

Subsequent to these initial experiments further tests were conducted on humic and fulvic acids using the SERS substrates. Examples of typical responses were recorded (many more spectra were obtained but not saved as they returned essentially the same results). Raman bands were measured on one a single SERS measurement of humic acid solution on 23 September 2014 at wavenumbers 620 cm^{-1}, 1450 cm^{-1} and 1600 cm^{-1} but these were not repeated in the other measurements taken (Figure 6.16).

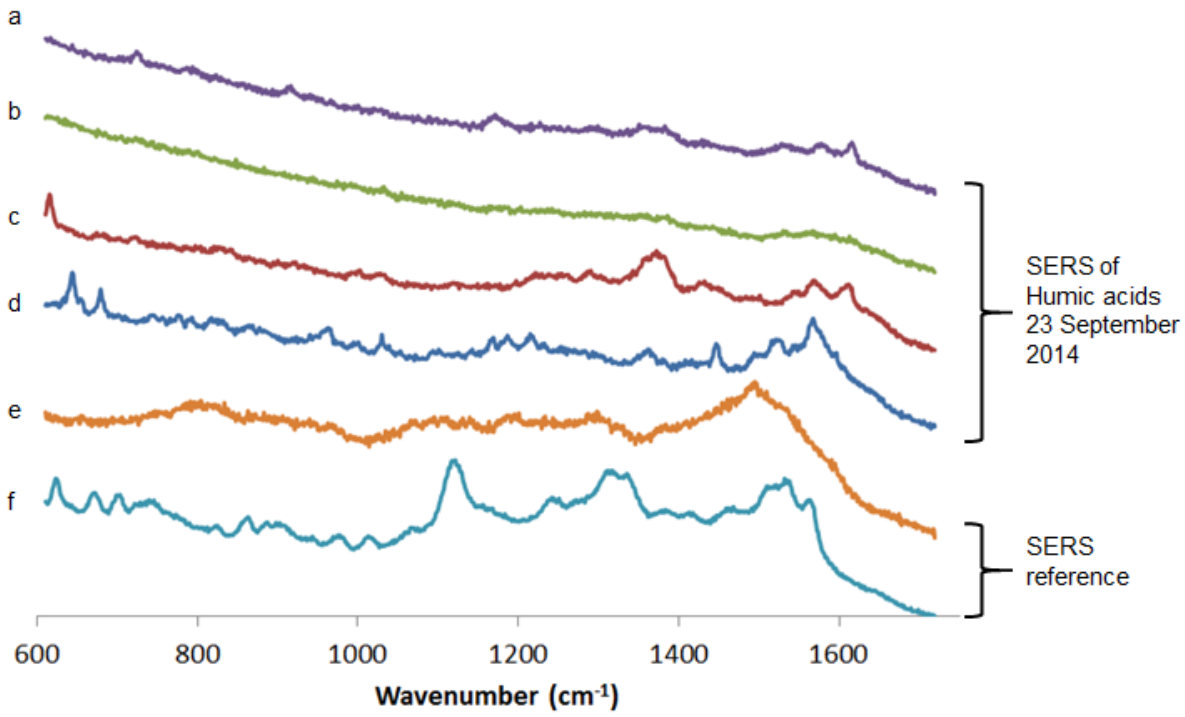


Figure 6.16 – SERS of 2,500 mg/l humic acid solutions
SERS of 2,500 mg/l humic acid solutions on 23 September 2014. Samples were measured on SERS substrates using the 785 nm laser with a 0.15 mW power setting, 1 second exposure times and 20 accumulations. Spectrum (d) shows some Raman bands not measured on the other four spectra at wavenumbers 620 cm^{-1}, 1450 cm^{-1} and 1600 cm^{-1}.

As with the water samples on the SERS substrates there were some anomalous SERS spectra of the humic acid in solution. On 3 October 2014, there were several spectra returned that deviated from the normal results discussed in the previous paragraph - these measured samples contained identifiable Raman bands although these bands were returned at inconsistent wavenumbers and could not be replicated (Figure 6.17). When compared to the reference background signal of water on SERS (bottom trace on Figure 6.16) these spectra from the humic acid solution samples do not correspond to Raman bands associated with the substrate but they cannot be linked to any other Raman bands measured in this project either. As a comparison, the spectral bands measured when analysing BPE (Figure 6.3) were stable in both terms of wavenumber and relative magnitude.

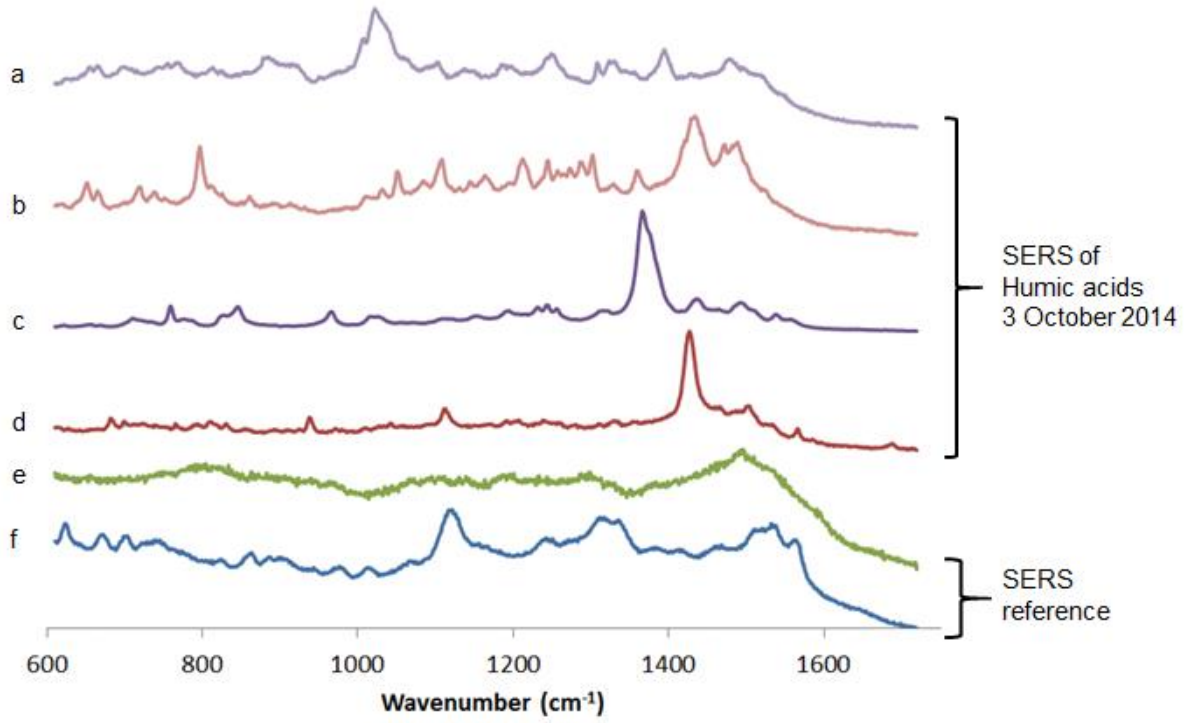


Figure 6.17 – Anomalous humic acid on SERS
The top four spectra on this figure are from anomalous SERS results returned for solutions of humic acid. The bottom of the five traces is a reference spectrum of gold SERS substrate. Humic acid solutions were measured on 3 October 2014 using the 785 nm laser with a power setting of 0.3 mW and 20 accumulations with a 1 second exposure time.

However, the day when the most Raman bands were measured on a single SERS substrate when measuring a 2,500 mg/l humic acid solution on 13 October 2014 (Figure 6.18). When the substrate dried out more of the same humic acid solution was applied using a water dropper. The cause of these Raman bands was not clear but several strong Raman bands were measured at 1000 cm^{-1}, 1200 cm^{-1}, 1350 cm^{-1}, 1450 cm^{-1}, 1600 cm^{-1} and 1650 cm^{-1}. The substrate has a cross sectional area much bigger than the area the laser is focussed on and so the laser was fired at multiple points to determine how consistent the results were. There were 121 measurements taken of this substrate on 13 October 2014 and to display all this information on a graph would lose clarity, so only examples of the different line shapes have been extracted as many of them are similar.

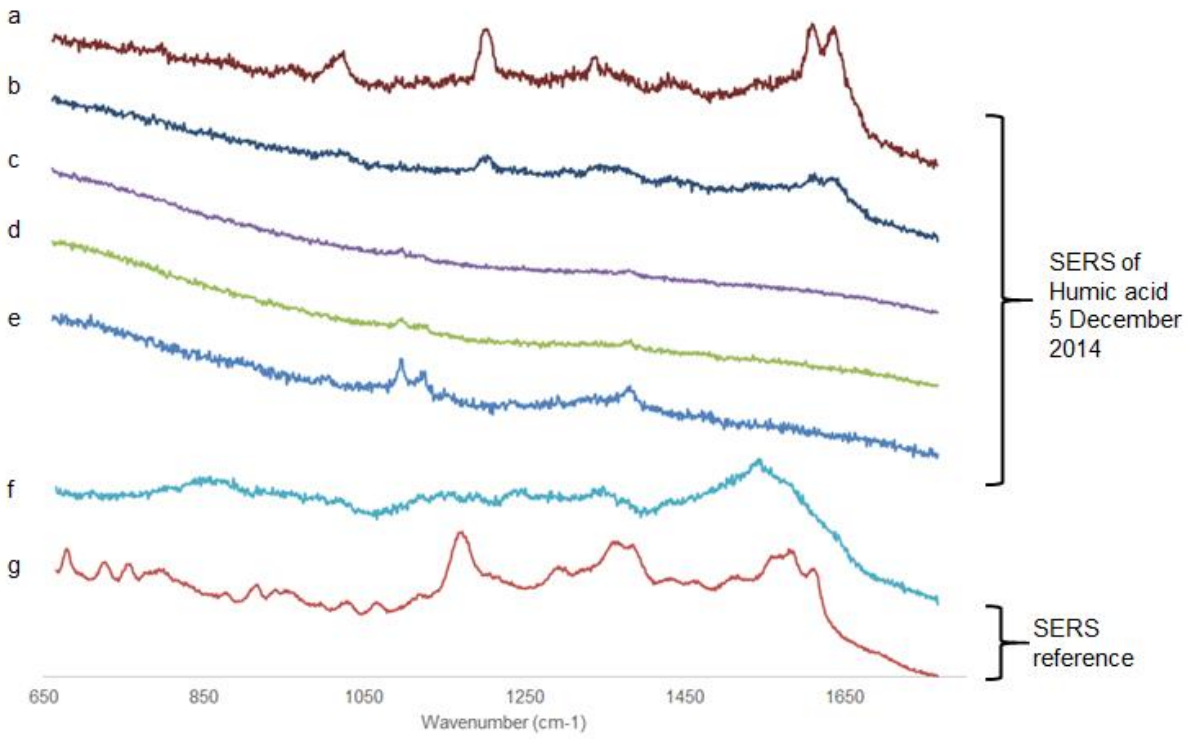


Figure 6.18 – Humic acid on SERS measured on 13 October 2014
An overlay of the Raman spectra of humic acid measured on a single SERS substrate on 13 October 2014. Raman bands at 1000 cm^{-1}, 1200 cm^{-1}, 1350 cm^{-1}, 1450 cm^{-1}, 1600 cm^{-1} and 1650 cm^{-1} are identifiable but these are not common to all spectra measured. Unfortunately, a note of power settings was not taken on this day, although by this point in the project,15Mw with an exposure time of 1 second and 20 accumulations for a measurement was typical.

The next step in the analysis was to determine if the pH had any influence on the SERS spectra. On 5 December 2014, a solution of approximately 2,500 mg/l humic acid was prepared and analysed. The four reagents used to control the pH were all analysed as well (hydrochloric acid, sodium hydroxide and the pH 4 and 7 buffers). Four separate measurements of humic acid solution at an increased pH were taken, although only two are shown in the plot as they had very consistent spectra. However, the spectra of the humic acid repeats seem to share a spectral peak at around 1550 cm^{-1} with the pH 7 buffer and a spectral peak at around 1300 cm^{-1} with the pH 4 buffer. One of the SERS reference signals also contained a spectral band peaking at around 1550 cm^{-1}. None of the Raman bands identified with HCl or sodium hydroxide were identified in the manipulated humic acid solution. Also, none of the spectra were as strong as the one associated with BPE in the test solution (Figure 6.4).

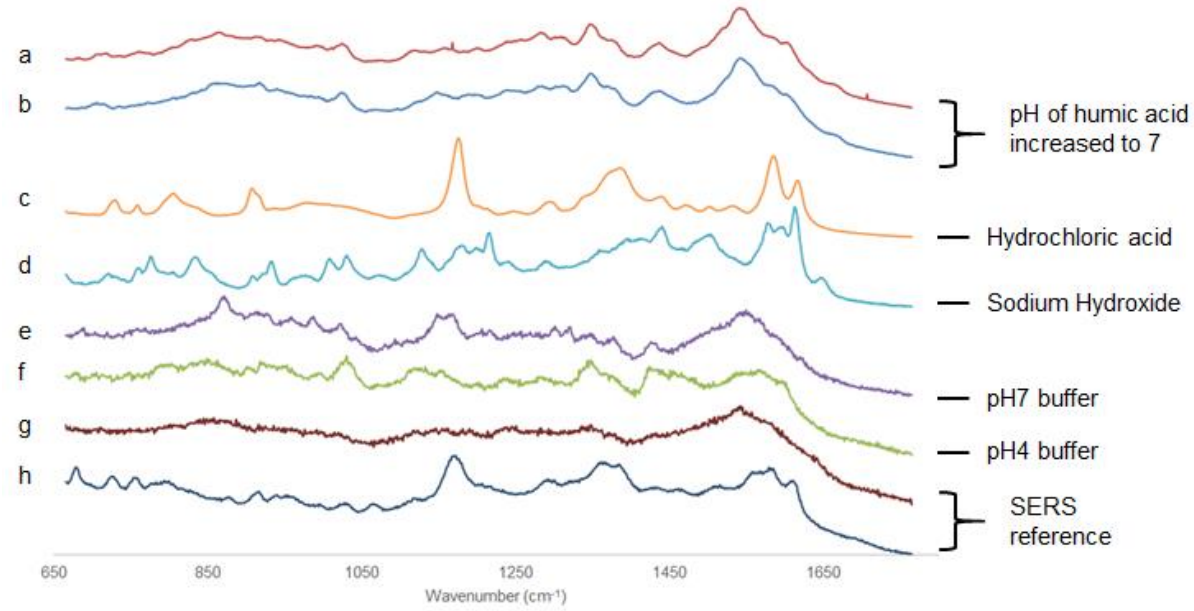


Figure 6.19 – pH adjusted humic acid solutions on SERS
Showing the results of the 2500 mg/l humic acid solution with the pH increased to a value of about 12 using the sodium hydroxide and pH7 buffer solution. The pH4 buffer and HCl were used to correct for overshooting when preparing samples as the volumes were very small.

The final set of results was obtained from measuring the SERS Raman spectra of the reduced humic acids and the reduced fulvic acids. However, when the reduced humic and fulvic acids were tested they generally did not produce appear to produce Raman bands that could be differentiated from the background SERS signal (Figures 6.19 and 6.20).

The overall results indicate that no additional Raman bands have been obtained when comparing the results of the non-reduced humic substances with the reduced humic substances.

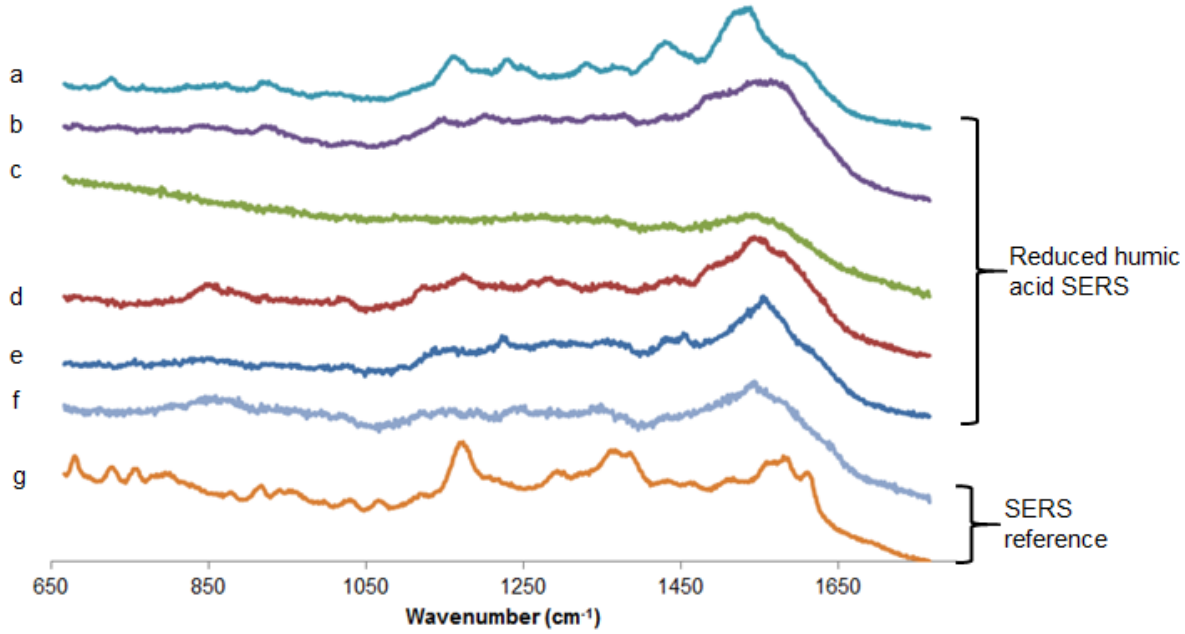


Figure 6.20 – Reduced humic acid solution
Humic acid solution are about 50 mg/l, measured on 3 February 2015 using the 785 nm laser with settings of 5 mW power and 1 second exposure times with 5 accumulations. There are no consistently clear differences between the SERS reference signals at the bottom and the humic acid.

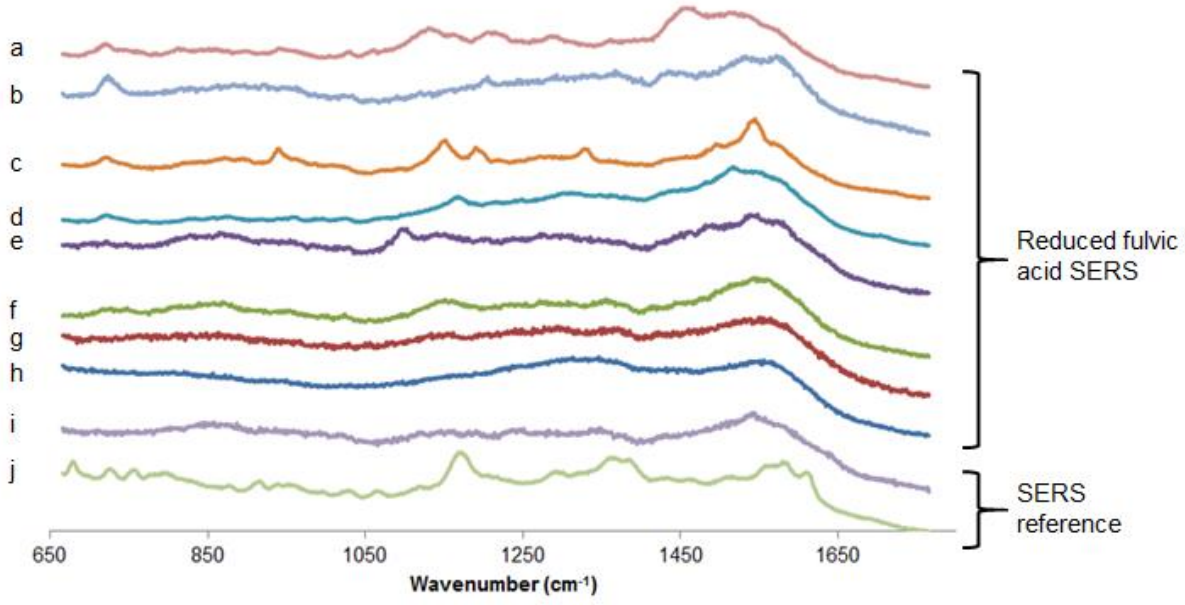


Figure 6.21 – Reduced fulvic acid solution
Fulvic acid solution are about 50 mg/l, measured on 3 February 2015 using the 785 nm laser with settings of 5 mW power and 1 second exposure times with 5 accumulations. Trace (c) exhibits peaks at 700 cm^{-1}, 900 cm^{-1}, 1150 cm^{-1}, 1200 cm^{-1}, 1325 cm^{-1} and 1550 cm^{-1}. There are no consistently clear differences between the SERS reference signals at the bottom and the fulvic acid.

6.6 Discussion of Results

6.6.1 Clarity of Raman spectra

The measurement of glucose and sucrose solutions produce visible and distinctive Raman spectra but only in concentrations of thousands of mg/l. The high limit of detection corresponds with other studies that have measured sugars using Raman spectroscopy (Soderholm et al., 1999, Shafer-Peltier et al., 2003). Natural water systems will only rarely exceed 100 mg/l of mixed DOC and therefore developing an in-situ field based sensor does not appear of value for detecting sugars using Raman spectroscopy. SERS spectra of sugars at lower concentrations have been measured in other work using custom-built substrates that are not available on the commercial market at the time of writing and that at present would require specialist knowledge and in-house manufacturing (Kreno et al., 2014).

Unlike glucose and sucrose, the humic and fulvic acids solids did not produce any tall, narrow Raman bands. It would seem that direct measurement of the humic acid when there is no water or other substance to dilute its concentration would coincide with the strongest Raman signals, except perhaps where enhanced Raman signals could be recorded in a solution. However, measurements of humic and fulvic acids in solution on SERS substrates (regardless of concentration) appeared to produce no definable Raman bands. Even after increasing the pH of the humic and fulvic acid solutions and performing the reduction reaction that was advocated by others (Heighton, 2013, Leyton et al., 2005, Corrado et al., 2008), the humic substances do not generate unique Raman spectra. There is significant variation in the results but the humic substance solutions on the SERS substrate does not produce any consistently strong Raman bands. However, there is a signal of a broad spectral peak that appears in most of the Raman analysis conducted between about 1500 cm^{-1} and 1700 cm^{-1}, which appears in the blank water spectrum when measured on aluminium (Figure 6.5), the background spectrum of the SERS substrate (Figure 6.6) and again with water on a SERS substrate (Figure 6.7). A PDF published by Horiba was helpful in identifying that there is a broad and weak Raman band corresponding to water centred around 1640 cm^{-1} (HoribaScientific). There is little published research on water using Raman spectroscopy because Raman is a weak water scatterer and therefore it is hard to find published sources identifying this band.

6.6.2 Comparison of increased pH measurements of humics and reduction of humics using SERS with blank water

Much of the experimental work here considered replication of approaches published in other work that increase the pH of a humic acid from an acidic to alkaline state to improve the measured SERS (Leyton et al., 2005, Corrado et al., 2008) and that performing a reduction reaction of the humic acid prior to analysis could also improve the SERS spectra observed (Heighton, 2013). However, my standard Raman measurements are indistinguishable from my background measurements and the stronger the concentration of humic acid used then the weaker the results. The large double band (Leyton et al., 2005) is where I find water in my experiments. Below I have copied one of the graphs from Leyton's 2005 paper (Figure 6.21). There are two main bands at around 1400 cm^{-1} and 1600 cm^{-1}. Leyton attributed these two spectral bands to aromatic CC stretching modes within the humic acid and attributed the smaller sharper peaks that appear at higher pHs to different functional groups. Compared with Figure 6.7 in this chapter, the 1600 cm^{-1} band, which corresponds to water, is at the same wavenumber as the largest band in Figure 6.21. The band that Leyton presents at 1400 cm^{-1} is similar to the one shown in Figure 6.15A at about 1350 cm^{-1} but this is a big enough difference in wavenumber question if these two signals are the same.

One difference between the results presented in this project compared with Leyton's (Figure 6.21) is that Leyton's results have greater number of sharper spectral features as the pH increases, whereas my measurements were more random in terms of when spectral features were measured.

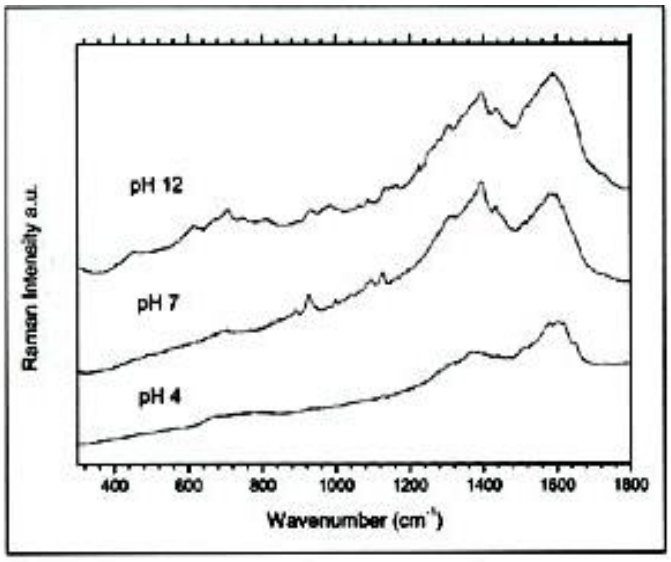


Fig.4. SERS spectra of humic acid at different pH.

Figure 6.22 – Copy of SERS reported by Leyton (Leyton et al., 2005)
Shows SERS of humic acid at different pH.

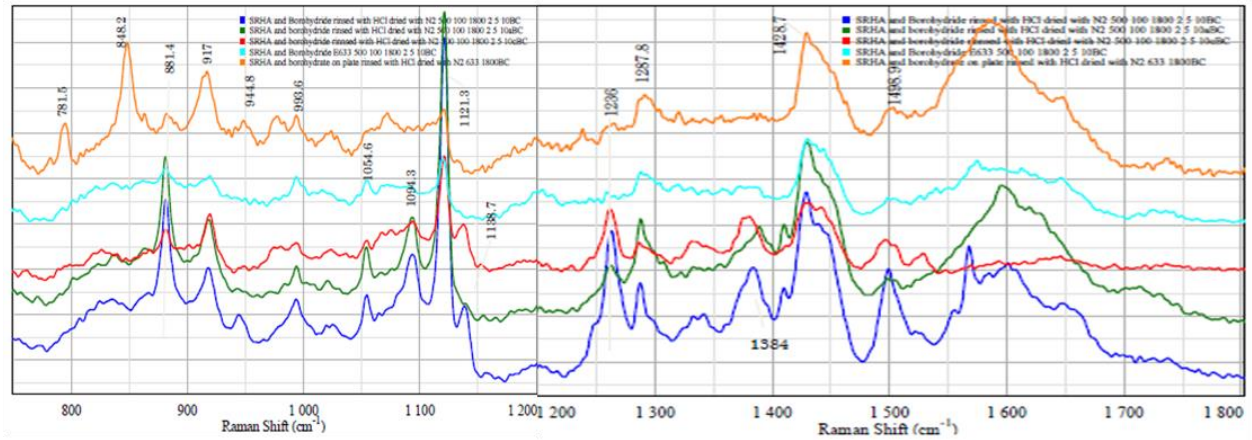


Figure 6.23: Copy of results reported by Heighton (Heighton, 2013)
Spliced results. I have spliced the two graphs that appear on pages 175 and 176 of the thesis (Heighton, 2013) together so that I can more clearly observe the entire spectra recorded. Comparison of five different reduced humic acid solutions have been colour-coded. In this chapter I will refer to spectra (a), (b), (c), (d) and (e).

A major experiment was to perform a reduction reaction on the humic and fulvic acids prior to analysis. The SERS results here for reduced humic and fulvic acids still do not allow identification of Raman bands. Unfortunately this means that the conclusion from this experiment differs from the experimental results that were being replicated (Heighton, 2013) – (see Figure 6.23). It is possible that some of these differences could be the result of differences in experimental setup – discussed later (section 6.6.3). Noteworthy for this discussion is that there is some inconsistency between the five Heighton Raman spectra (although they are more consistent than the results presented in this project). There are strongly identified Raman bands at 1121.3 cm^{-1} and 1428.7 cm^{-1}, which are consistent in terms of wavenumber, if not magnitude across all five traces. However, in four of the five spectra, the trace around 1600 cm^{-1} varies in relative intensity and wavenumber. Some of the five spectra have a strong Raman band at 1236 cm^{-1} but not all of the spectra. There are other Raman bands that appear at certain wavenumbers (e.g., 781.5 cm^{-1} and 944.8 cm^{-1}) but not in all. One possibility is that because humic substances are complex they will not always return the same spectra as spectra is dependent on the structure and the structure of humic and fulvic acids is not 100% consistent.

6.6.3 Concentration and light penetration

The simplest interpretation of my humic and fulvic acid analysis is that the spectra reflect that of the SERS substrate. Altering the concentration of solutions containing humic and fulvic acids can have two separate influences. With most substances increasing the concentration of the solute will usually lead to Raman signals of greater intensity: with more target molecules the frequency of the Stokes-scattering should be greater and result in more intense Raman spectra. However, humic and fulvic acids colour water to the

extent the solutions are essentially opaque at 2500 mg/l. The opacity of these solutions at high concentrations could influence laser light penetration, and spectra, including of the SERS substrate would be weaker. So even this spectrum was not clear due to lack of light penetration, exacerbated by a reduced laser power setting is being used because the fragility of the SERS substrates.

Such a mechanism could explain why the 9 mg/l humic acid produced a Raman spectrum of greater clarity than the Raman spectrum produced by the 2500 mg/l SERS analysis. When analysing the 9 mg/l solution the laser light penetrated the liquid more clearly and this might have enabled greater illumination of the gold SERS substrate and therefore the background signal caused by the gold would have been measured in the Raman spectra more clearly. In the 2500 mg/l solution much of the laser light is likely to have been absorbed or reflected by the humic acid and therefore the spectra for the gold substrate would be weaker. The lack of light penetration at high concentrations could account for the seemingly counter-intuitive results that the diluted solution was resulting in a more intense Raman spectrum – in reality it may have been that there wasn't a stronger Raman signal being produced but rather the background signal was clearer. In essence by increasing the pH, the concentration of the humic acid would be decreased and this would actually increase the apparent background signals, which in this case would be the SERS substrate and the water diluting it.

If dilution was an issue in the results presented here, then this needs to be considered when addressing the results from experiments conducted by others who detected an increase in Raman intensity when the pH is manipulated (Leyton et al., 2005, Corrado et al., 2008). However, the Corrado experiments seem to have been conducted, while maintaining a 10 mg/l concentration of humic acid (Figure 6.23), so dilution should not be an issue. It is not completely clear in the Leyton paper what the concentration of the humic acid tested was. Leyton also notes that although the SERS results do contain variation (for individual accumulations) the results seemed to be stable between different batches tested.

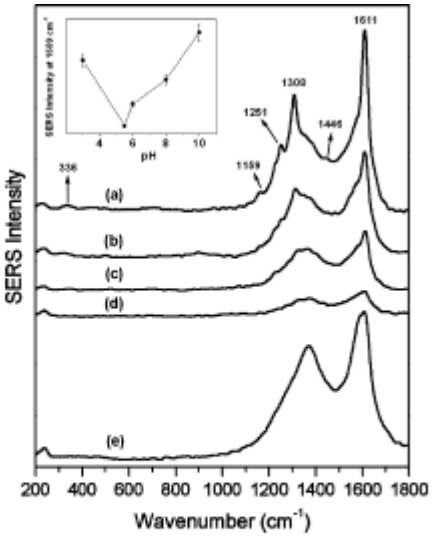


Fig. 3 – SERS spectra of C_{30} sample (10 μg mL^{-1}) after fluorescence correction at the following pH: (a) 10.0, (b) 8.0, (c) 6.0, (d) 5.5 and (e) 3.0. Inset: variation of the 1609 cm^{-1} SERS band on changing the pH.

Figure 6.24 - Copy of SERS reported by Corrado (Corrado et al., 2008) Shows SERS of humic acid at different pH.

Recently methods using advanced techniques having been developed for enabling the SERS detection of glucose by integrating a metal organic framework (MOF) within a SERS substrate (Kreno et al., 2014). MOFs are metallic structures designed to have huge surface areas and because of this they have extremely high adsorption properties and can enable materials that would not normally adsorb to the surface of a SERS substrate to do so. However, they are specialist materials and not available commercially. Although during the PhD I undertook some training in the James Watt Nanofabrication Centre in anticipation of making analytical components, there was insufficient time remaining to develop such substances.

6.6.4 Why do the results presented here not corroborate with other studies?

I have been unable to corroborate the results of (Leyton et al., 2005) and (Heighton, 2013) who identified Raman spectra of humic acids. One question that needs to be addressed is whether this is because the results generated here are different or whether the different conclusions are as a result of different interpretations of the results? There are some differences in the results and the reasons behind these differences are discussed in the following paragraphs. The other possibility that needs to be considered is whether I am interpreting the results obtained differently from other researchers.

A simple question is: was there an issue with the SERS substrates used? I did have confidence in the substrates used as the BPE standard produced the spectra expected and therefore the lack of strong spectral bands associated with the sugars and humic

substances is unlikely due to the substrates themselves. I have already discussed how glucose does not adsorb onto most SERS surfaces (Shafer-Peltier et al., 2003, Shao et al., 2012) so this was a challenging substance to obtain Raman spectra from solutions. It may also be the case that other sugars such as sucrose also do not adsorb readily onto most SERS substrates – but existing understanding to support this could not be found since there is much more research tied to glucose. This research has also shown humic and fulvic acids do not adsorb onto the surface of gold and therefore using standard methods for producing Raman spectra from SERS will not work with humic substances. The results shown in Figure 6.4 also showed that there was no significant degradation of the SERS substrates over time as BPE was still measurable on the substrates purchased at the same time for the duration of the experimental period.

In this project, the Raman spectra of target samples of humic substances was too similar to that of the background signals to attribute Raman bands to humic substances as the similarity meant the spectra of the target material and background could not be readily differentiated. I suggest two possible reason why I have interpreted these results differently from other researchers: firstly, that dilution increases the background noise and this can happen when the pH is increased; secondly that what has previously been interpreted as Raman bands signifying humic substances might not actually be true Raman signals of humic materials. However, if I assume that dilution is not an issue and that the previous experiments have been interpreted correctly, then there may be experimental setup differences that may also contribute to the lack of reproducibility: both Leyton and Heighton used silver nano-particles in a colloidal solution as substrate, whereas gold-plated SERS substrates have been used here (Leyton et al., 2005, Heighton, 2013).

The key reason for choosing gold as the SERS substrate was that it is easier to find market off-the-shelf solutions for SERS analysis using gold than silver. The preparation and use of gold or silver nano-particles in colloidal solutions seemed to require more specialist knowledge and due to time limitations available for this experimental work an off the shelf solution was selected. It should also be noted that while silver is generally considered a superior SERS substrate to gold (Aikens and Schatz, 2006, Sharma et al., 2013) it does not mean that gold is poor at producing SERS Raman spectra – it will produce the same SERS spectra – just usually with a poor SNR. Published Raman spectra comparing gold and silver could not be found (as such traces might be considered too generic by those that work regularly with Raman measurements), but there are databases, such as the RRUFF^{TM} database (RRUFF), that show the background radiation to be expected. The referenced reference spectra indicate that gold does have a stronger

background signal and thus would results in a poorer SNR. However, when using a 785 nm laser, gold may produce better enhancement than silver in some cases (Smith, 2008).

The reason why silver produces greater enhancement of the Raman signal over the background generally is as a consequence of the minimisation of Plasmon resonance damping (Rycenga et al., 2009, Fernanda Cardinal et al., 2010). Plasmon resonance describes electromagnetic surface waves that are distributed along the interface between a conducting material and a dielectric (Liz-Marzán, 2006) and so can interfere with or reduce the Raman signal. One of the mechanisms that can lead to Plasmon resonance is the incidence of photons onto metal surfaces (Salamon et al., 1997). When Plasmon resonance is damped the energy distributed in this manner is reduced and therefore a better SERS effect is obtained.

The choice of laser may also have influenced the spectra generation. As a general 'rule-of-thumb' lasers with lower wavelengths produce Raman spectra with greater SNR, but they are also more susceptible to fluorescence interference - wavelengths of fluorescence light tend to be stronger with an excitation at shorter wavelengths compared to longer wavelengths (Vankeirsbilck et al., 2002a, Thorley et al., 2006). It may be that the use of the 785 nm laser produced too much noise to identify specific Raman bands when compared with the 514 nm laser (Corrado et al., 2008) or the 633 nm lasers (Leyton et al., 2005, Heighton, 2013). I also had access to a 514 nm laser but this overlaps with an absorption band of the gold SERS substrate, limiting the experimentation to use of the 785 nm laser (Smith, 2008). However, Raman research into the analysis of red blood cells using different wavelengths of incidence lasers shows differences in the magnitude of certain Raman bands depending on what laser was selected (Wood et al., 2007), but the most prominent Raman bands appear to be identifiable across all lasers with only smaller peaks appearing on some but not all of the traces.

Having considered the different experimental setups, I remain unconvinced that the setup is the reason why I am not able to corroborate past experimental results. I think that the major difference is one of interpretation of the results rather than the results themselves being too different.

6.6.5 Reproducibility of the results

It is critical that measurements of Raman spectra are consistent and repeatable as this will enable the creation of models and this will provide confidence in the results generated in the laboratory and towards the potential development of a field based sensor. A simple observation is that when really strong Raman signals are measured this greatly improves

the SNR. When there are strong Raman spectra, as was the case with the BPE SERS spectra, the signals returned are very consistent (Figure 6.4). There are still some small variations in the relative magnitude of the Raman bands but in general these Raman bands are consistent in that they are associated with the same wavenumbers and are visibly distinguishable from the background. From a very simple perspective the more consistently a spectral pattern can be repeated then the more confidence can be placed on the result and subsequent interpretation.

Knowledge and understanding of humic substances and their structure is improving each year: it is now considered that humic acids are supramolecular structures consisting of small molecules linked together by processes such as hydrogen bonding, rather than as large polymeric structures (Piccolo, 2002), and tools such as NMR are able to identify specific extracted groups within humic substances such as lignite groups (Cao et al., 2016). The lack of reproducibility of the SERS spectra for humic and fulvic acids could arise as they are chemically so complex (Kuwatsuka et al., 1978, De Paolis and Kukkonen, 1997, Rodríguez et al., 2016), or what was being measured were contaminants. However, this may contribute to lack of consistent spectra and the core issue was a lack of spectra generated.

6.7 Conclusions and future work

This chapter has shown that the use of Raman to identify the composition of DOC from fluvial systems is challenging and there is difficult in detecting and identifying components in the laboratory and between users. Thus, extending portable spectrometers to in-situ sensors requires first considerable testing and development.

The tests conducted in the experimentation presented in this chapter have identified potential difficulties in using Stokes-scattered Raman signals as a method of analysing DOC composition. Sugars did produce clearly defined and reproducible spectral maximums that corroborated well with the spectral data found in online databases and other studies. In concentrations of several tens of thousands of mg/l or greater the sugars continued to produce identifiable spectral maximums although below this concentration the spectra maximums were obscured by noise. In the concentration range found in natural water systems at Drumtee (~ 8 mg/l to 60 mg/l) then sugars would not be observable using Raman spectroscopy or the regular type SERS analysis as performed here.

The experimental set-up I used (514 nm or a 785 nm laser, SERS, chemical alteration of the test sample) did not enable identification of humic substances - I was unable to

reproduce or corroborate the results indicated in the two studies I attempted to replicate (Leyton et al., 2005, Heighton, 2013). Whilst my experimental set-up differed from Leighton and Heighton it seems would more likely influence the strength of Raman signal than the shape. Their experimental results do not illustrate the strength of Raman signals as found in other substances such as graphene (Ferrari and Basko, 2013) or that were measured in this project for BPE, and they have traces similar to those from water. Thus, there is the unavoidable query of whether their tracers are humic signals or a background signal.

If laboratory analysis of isolated humic and fulvic acids cannot reliably be used to identify these substances, then constructing a working sensor for field deployment is more challenging because of the environmental variables and the humic and fulvic acids will be mixed in with other substances. However, during these experiments I observed that the anti-Stokes section of the Raman spectra may enable a clearer visualisation of humic substances using Raman spectroscopy and this is detailed in the next chapter.

7.0 – Anti-Stokes Raman analysis of dissolved organic carbon

7.1 Abstract

Here I present a reproducible method for measuring concentrated humic and fulvic acids using a 785 nm laser and measuring the anti-Stokes scattered Raman signals generated. Results are presented for powdered humic and fulvic acids from the Suwannee River (International Humic Substance Society. Tests were also conducted using the solutions of humic acid, fulvic acid and mixtures of the two acids in different proportions. The limit of detection for both the humic and fulvic acids was measured to be approximately 500 mg/l. When tested at concentrations of 10,000 mg/l, graded mixtures of humic and fulvic acid produced similar spectra, but with sufficient differences to enable the construction of a linear predictive model for identifying the proportion of humic acids to fulvic acids in a DOC mixture. Further, the linear relationship varied depending upon the concentration of the humic acid and fulvic acid solution and the stronger the concentration, the more significant the linear relationship. The results showed that the greater the concentration then the easier it was to determine whether a humic solution was dominated more by humic acids or fulvic acids. Once the solution's concentrations were diluted down to about 1,000 mg/l then it became difficult to distinguish between humic and fulvic acids.

One of the advantages of Raman spectroscopy over many other spectral methods is that samples usually require little to no preparation prior to analysis. Unfortunately, the method for detection of humic substances presented here using anti-Stokes measurements requires highly concentrated solutions or powder extracts and this necessitates an undesirable amount of preparation work at the present time if the long term goal is sensor development. Despite a consistent Raman signal being identified when analysing humic substances a very similar signal was identified when the SERS substrates purchased for this project were burned. The SERS substrates were printed onto paper and the burned paper is likely the source of a similar signal. Despite the similarity when comparing the humic substances to the burned paper, the signal of the humic substances is distinct from other tested background sources and this signal is consistent and reproducible enough to suggest that the anti-Stokes measurement of humic substances is a valid and viable method of detection.

Although the limit of detection measured would be insufficient for measuring DOC at concentrations typically measured in streams it is likely that the use of anti-Stokes Raman spectroscopy represents a promising opportunity to study and analyse humic substances. As such Raman spectroscopy, even if there are potential challenges to be overcome, could be a potential in-situ sensor for measuring DOC but only if the identified issue of limit of detection could be overcome. Furthermore, it is not clear from these

experiments whether there is a benefit to using Raman spectroscopy over fluorescence spectroscopy for analysing humic substances.

7.2 Introduction

As discussed in chapter 6 Raman spectroscopy has been suggested as a potential tool for analysing humic acids, which are a major component of DOC (Leyton et al., 2005, Corrado et al., 2008). However, the methods discussed in the literature do not produce strong, distinctive, Raman signals that are easily replicated. The techniques discussed rely on manipulation of the pH of the acid and in one case conducting a reduction reaction on the humic substances prior to analysis. These methods have produced results which have lacked distinctive Raman bands that are narrow and clearly defined as we would hope to observe in an ideal sample. Instead the Raman bands identified tended to be broad or in some cases sporadic and not clearly replicable.

An area of research that has remained largely unexplored is to identify humic substances using anti-Stokes scattering Raman spectroscopy. Anti-Stokes scattering Raman spectroscopy is a natural phenomenon that occurs as a result of an electron being promoted to a higher, virtual energy state before decaying into an energy level that is slightly lower in energy than the original energy level (Marshall and Verdun, 2016). In theory all of the substances that produce Stokes scattering also produce anti-Stokes scattering – indeed if the Stokes Raman and anti-Stokes Raman signals of a substance that produces strong spectra (such as sulphur) are shown on the same graph, the two will appear as mirror images of one another, except that the anti-Stokes signals are weaker than the Stokes Raman spectra (Carriere and Havermeyer, 2012). However, in most cases researchers don't consider the anti-Stokes measurements - although they are essentially always produced when Stokes scattering occurs because the intensity of the anti-Stokes signals is significantly weaker than the Stokes Raman effect at room temperature (Karamehmedovic, 2006, Singh et al., 2007).

For many materials the issue of poor Raman signal intensity can be overcome through the process of surface enhanced Raman spectroscopy (SERS). The use of a SERS substrate can enhance the Raman signals that would otherwise be obscured by background noise (as discussed in chapter 2 – Literature review). For SERS to work the target material needs to be in close proximity to the substrate, which is most easily achieved by the target material being adsorbed onto the surface of the substrate (or colloid) as this absorption enables the enhanced signals to be produced. Unfortunately, not all materials adsorb onto metallic substrates. For example even although glucose is generally a reactive substance it does not adsorb onto the surface of most SERS

substrates (Sugikawa et al., 2011, Kreno et al., 2014). As was shown in chapter 6 it seems that the humic substances fall into the category of being substances that do not readily adsorb onto metallic substrates.

One property of humic substances is that they contain many components that are fluorescent, called fluorophores, although it should be noted that there may be non-fluorescent humics acids contained within a humic acid sample (Matthews et al., 1996). Consequently, fluorescence spectroscopy has been used for many studies into humic substances, soil and dissolved organic carbon studies such as (Coble et al., 1990, Senesi et al., 1991, Miano and Senesi, 1992, Mopper and Schultz, 1993, Coble, 1996, McKnight et al., 2001, Hudson et al., 2007, Kowalczuk et al., 2010, Herzsprung et al., 2012). There are slight differences in the fluorescence spectra of humic acid and fulvic acids (Baker, 2001) and consequently fluorescence has been used for analysing changes in chemistry at different points along rivers (Hudson et al., 2007). A major drawback when trying to use Raman spectroscopy to analyse many substances is that fluorescence signals are stronger than Raman signals (Nie and Emory, 1997, Martyshkin et al., 2004) and so they will dominate a spectrum that contains both Raman and fluorescence signals. In the case of the results of humic substances presented in chapter 6 the fluorescence signals appeared to obscure the Raman spectra and background signals. However, in the anti-Stokes region there are no fluorescence signals and so anti-Stokes spectra may allow study of materials that produce fluorescence signals, especially if those materials do not readily bond to the surface of a SERS substrate and consequently there would be no prospect of using SERS to enhance the recorded signals.

Sample analysis was carried out as discussed in chapter 6, incorporating the following five major components to the anti-Stokes Raman spectroscopy experiments:

1. measurement of the background signal in the anti-Stokes region of the aluminium slide, the SERS substrates and the water used to make the solutions. It is necessary to measure all the potential background signals to ensure that background anti-Stokes signals are characterised and corrected for, and this not misattributed to organic carbon.
2. measurement of sugar powders (glucose, fructose and sucrose).
3. measurements of amino acid powders (glycene, tyrosine, phenylalanine and tryptophan).
4. measurement of humic and fulvic acids in isolation and in mixtures using both their powders and measuring them in solution.
5. if any anti-Stokes Raman spectra were identified related to more than one substance then differentiating the spectra of these solutions.

The major motivation of these five tests is to determine if anti-Stokes Raman spectroscopy can be used to identify components of organic carbon and in particular humic and fulvic acids, particularly given that testing using the Stokes scattering method had not yielded consistent results. As before, in order to be able to identify the components of a mixed substance like DOC, it was important, firstly, to establish whether similar but different Raman spectra could be identified and secondly, to assess the limit of detection was on any substance that produced a definable Raman spectrum and whether the limit of detection was comparable to [DOC] in a natural fluvial environment.

7.3 Methodology for the measurement of anti-Stokes measurements

Here the Renishaw In-Via Raman Spectrometer was again used to conduct the experiments. The spectrometer had both a 514 nm laser and a 785 nm laser. In this case the 785 nm laser was selected as the 514 nm laser produced no spectral lines in the Anti-Stokes region. The laboratory set-up was the same as detailed in chapter 6 relating to the Stokes measurements using Raman spectroscopy, including the use of the aluminium slide for holding water droplets or depositing powders onto and exploring the use of SERS. However, it transpired that the SERS substrates did not generate enhanced signals and therefore the aluminium was used for all but a few tests.

The sugars and amino acids were stored in plastic centrifuge tubes, while the humic substances from which small volume solutions (~ 1 – 3 ml) were made within a few days of any analysis and were stored in small glass containers with plastic lids. These analytes were stored in smaller glass containers with plastic seals, and as with the solid samples of humic substances, sugars, amino acids and the extract of DOC, were refrigerated at 5 ^{o}C when not being used. The reason for preparing the humic substances differently from the sugars and amino acids was that they were relatively expensive and only a small quantity of the humic and fulvic acid was available for analysis.

The same criteria described in chapter 6 were used to determine the laser power selected for each test carried out. The 785 nm laser was set to 300 mW when analysing all solutions and for the sugar and amino acid powders. Unfortunately, as was identified in chapter 6, the powders of humic acid and fulvic acid had a tendency to burn when the laser power was too high and so a power setting of 5 mW for analysing humic acid powder and 15 mW was used for analysing fulvic acid powder.

However, as the SERS substrates did not enable enhanced signals on the organic carbon, the aluminium plate was used for analysing all the substances. For the solids, it

was inconsequential whether the aluminium plate or SERS substrate was used as the powders merely required a surface to sit on. For the solutions, the aluminium plate was used because the laser power could be set up to 300 mW, whereas the power could only be set to a maximum of 15 mW when using the SERS substrates, which would result in a poorer signal to noise ratio (SNR).

7.3.1 Normalisation

When the Raman spectra are generated they are not automatically scaled to a common reference point. The spectrometer essentially calculates the intensity of a signal at a particular wavenumber and puts it onto a graph. If an experiment uses four accumulations, then all these four accumulations will be added to one another (Figure 7.1). Therefore, the maximum value on the intensity scale after one accumulation might be different dependent upon the material being analysed, laser power, the number of accumulations or the exposure time. As most signals are compared visually on a graph it is preferable to scale all the signals so that they can be seen on the same graph as otherwise some traces on a graph may be several orders of magnitude greater than another, making the smaller trace difficult to see. Therefore, to make the graphs readable they are normalised. In the majority of cases this was achieved by setting the largest value recorded to one and the smallest value recorded to zero and scaling everything linearly in-between. The exception was for the Raman traces of humic substance powders as a strong background signal was present in these and the signals being discussed were compressed too much, so instead the value recorded at a wavenumber of -1477.6cm^{-1} since this value was not associated with background signals and was identified as the strongest signal on a sample of 10,000 mg/l of humic acid.

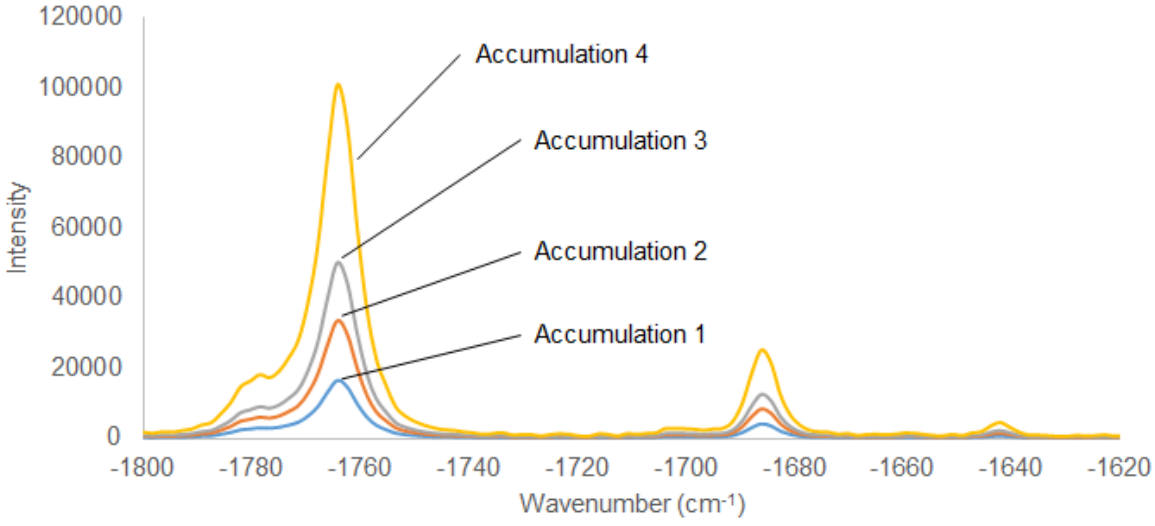


Figure 7.1 – How accumulations work
Illustrating how the intensity of a signal will increase for every accumulation carried out. If the spectrometer settings differ for different measurements this can make visualising different signals on the same graph difficult and therefore they are normalised.

7.4 Results

Here the results of the anti-Stokes Raman spectroscopy experimentation are presented. However, first I present the measurements of potential background sources and all the spectra arising from background sources are labelled. Clearly labelling background signals enables a comparison with the spectra generated using the sugars, amino acids and humic substances, where any generated spectra are labelled differently.

7.4.1 - Measuring the background signal in the anti-Stokes region of the aluminium slide and the SERS substrate.

The background signal of the aluminium plate (Figure 7.2) and the SERS substrate (Figure 7.3) were measured to identify background signals. The signal of aluminium is clearer than that for the gold SERS substrate because the laser could be set to its full power of 300 mW, Power above 15mW risks damaging the SERS substrate, and results in a poorer SNR.

Both the aluminium and SERS substrate have two strong absorption peaks at -1684.3 cm^{-1} and -1762.4 cm^{-1}. The aluminium also shows peaks at -1642.1 cm^{-1}, -2169.5 cm^{-1}, -2257.5 cm^{-1}, -2338.9 cm^{-1} and -2413.4 cm^{-1}. These five signals are present in the SERS substrate signal but are more difficult to observe because of the poorer SNR. These background Raman bands seemed to appear in samples regardless of whether water was present or not. Identifying the root cause of these background signals was not determined but they were important to note of ensure that they were not attributed to organic carbon and because they did interfere with later results.

As the use of anti-Stokes Raman has been little explored with respect to humic substances, there were no reference points for these results in the literature and consequently, I have not found names or notation for these peaks. However, there are a lot of spectral bands that are of potential interest and therefore I have used my own nomenclature notation to label these bands and make the graphs easier to read. I ascribed different notation to background signals from those I have interpret in the following to be attributable to organic carbon.

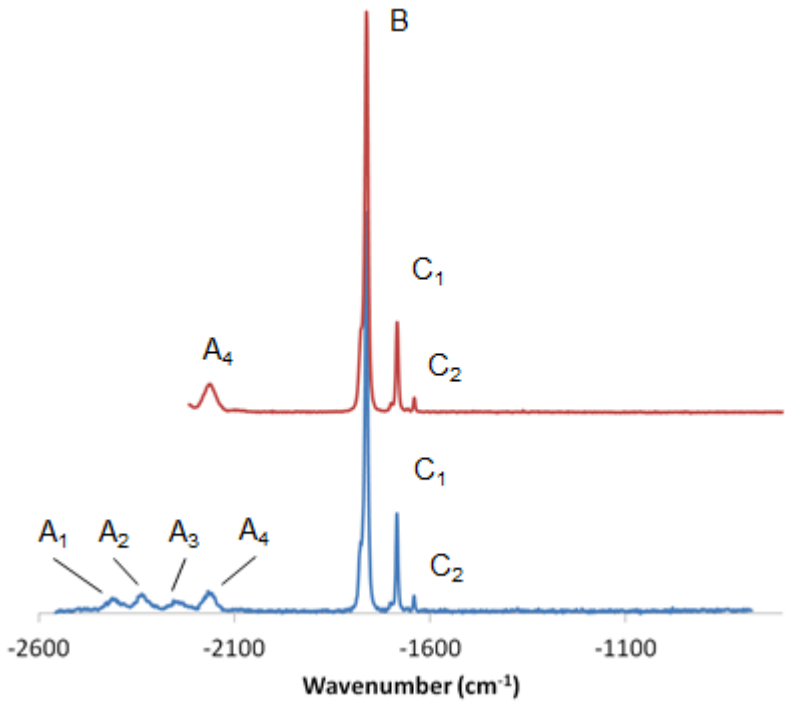


Figure 7.2 – Aluminium anti-Stokes background signal
A 785 nm laser was used producing power of 300 Mw, with 20 accumulations and a 1 second exposure time. Raman bands A1-A4, B, C_1 and C_2 were measured and can be attributed as a background signal. There were two traces because during initial analysis it was not entirely clear what wavenumbers should be focussed on.

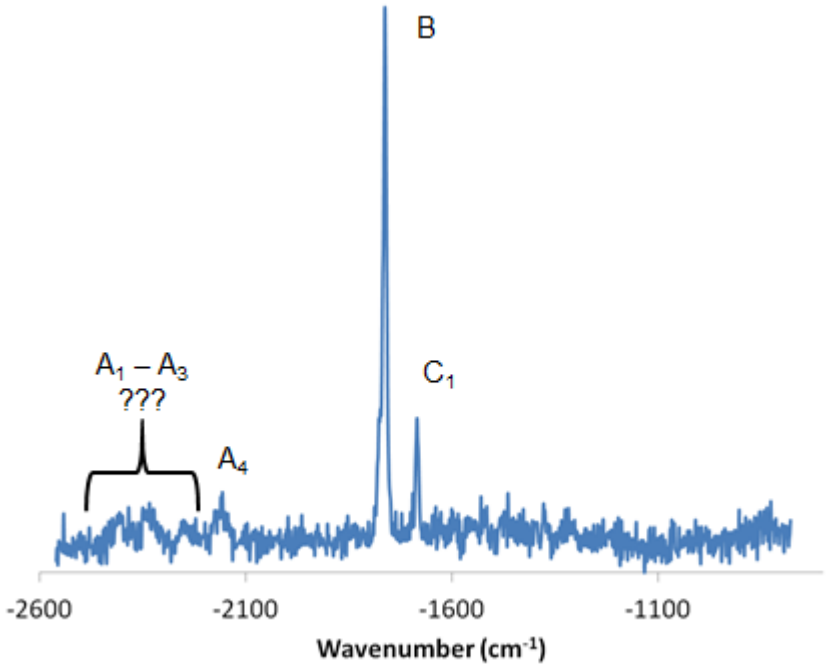


Figure 7.3 – SERS substrate anti-Stokes background signal
Raman bands A4, B and C_1 were measured and can be attributed as a background signal. It is debatable whether we can see A_1, A_2 and A_3 or whether these are merely noise. Signal C_2 is definitely not visible.

A final background spectrum was obtained of a burned SERS substrate, which occurred when the laser power was set too high. The laser power was set to 300mW rather than the usual 15 mW. The resultant spectrum is shown (Figure 7.4) exhibiting a large number of Raman bands. The several bands within the signal have been labelled x_1 to x_{19} from

left to right, except for the last two bands which are labelled y_1 and y_2 as these were narrow single peaks rather than the x bands, which show either a broad single peak or more than one peak. Raman bands x_1 to x_{19} are characterised by relatively broad bands with a wide, shallow peak or multiple narrow peaks. For example, x_{14}, which has the most intense peak after the B peak consists of two separate Raman peaks at -1450.3 cm^{-1} and -1477.6 cm^{-1}. Band x_{16} has 3 spectral peaks at wavenumbers -1305.5 cm^{-1}, -1330.5 cm^{-1} and -1340.6 cm^{-1}. Band x_{10} overlaps with band B (as shown on Figure 7.2 and Figure 7.3) but Raman bands A_{1-4} or D cannot be identified on this graph.

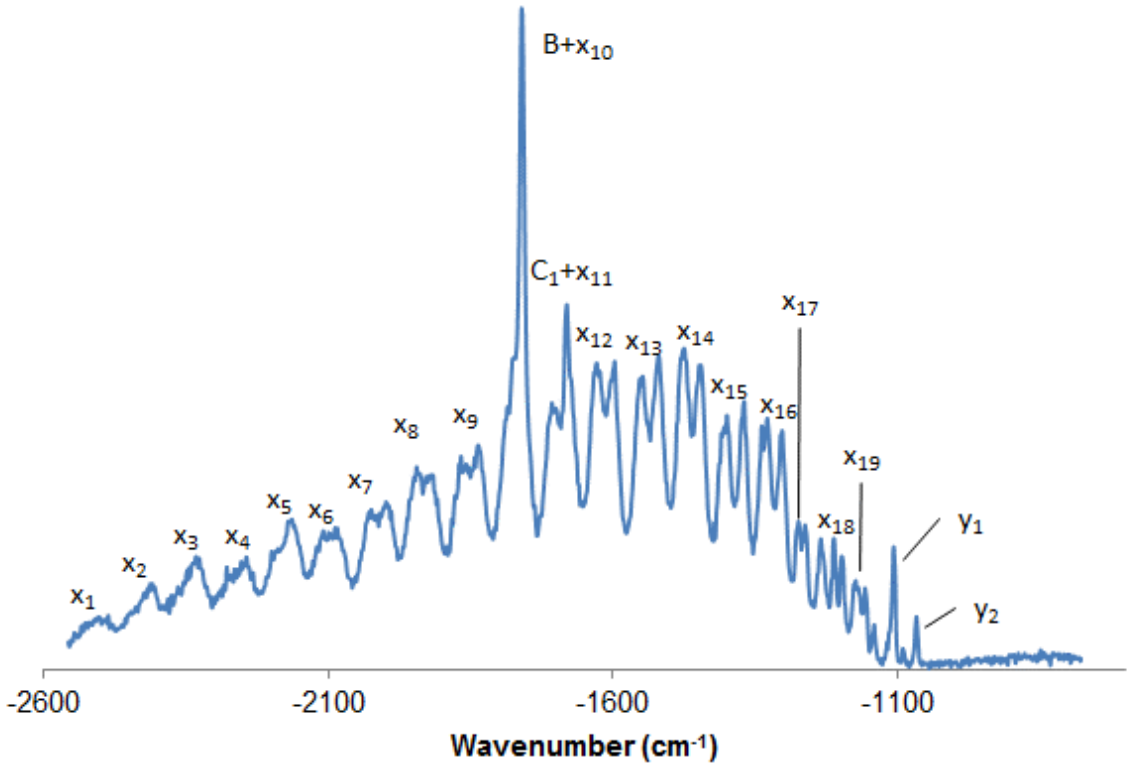


Figure 7.4 – Spectrum of burned SERS substrate
Bands B and C_1 (Figures 7.2 and 7.3) are still visible but there are many other spectral bands now visible. Two narrow single bands on the right-hand side of the graph are shown as y_1 and y_2.

7.4.2 - Humic and fulvic acids detected

Between 17 June 2015 and 29 June 2015 there were 60 measurements made of humic and fulvic acids. Two of these measurements were taken from the condensed Drumtee DOC sample; 21 measurements were made of humic acids; 20 measurements were made of fulvic acids in powder and solution form; 17 measurements were made of solutions that contained mixtures of humic acids and fulvic acids.

Clear Raman spectra were obtained for the analysis of humic and fulvic acid powders and the Drumtee sample using the 785 nm laser and standard Raman spectroscopy on the

aluminium plate (Figure 7.5). The background signal B (at -1762 cm^{-1}) is visible in three samples, however the acid and DOC Raman spectra are significantly different to the background sample of water on aluminium. The fulvic and humic acids exhibit Raman bands at the same wavenumbers, which also are the same wavenumbers identified from the burned SERS substrate (Figure 7.4). The powdered fulvic acid samples had a better SNR than the powdered humic acids, primarily because the fulvic acids could be measured with a greater laser power. A power setting of 15 mW could be used to analyse fulvic acids but this power setting was found to burn the humic acid and so a setting of 5 mW was used to analyse humic acid powders. It was also easier to focus the laser on the paler coloured fulvic acid, which also potentially assisted in improving the SNR. When observing the results, the bands x_2 – x_5 are very close in terms of wavenumber to the background Raman bands A_1 – A_4.

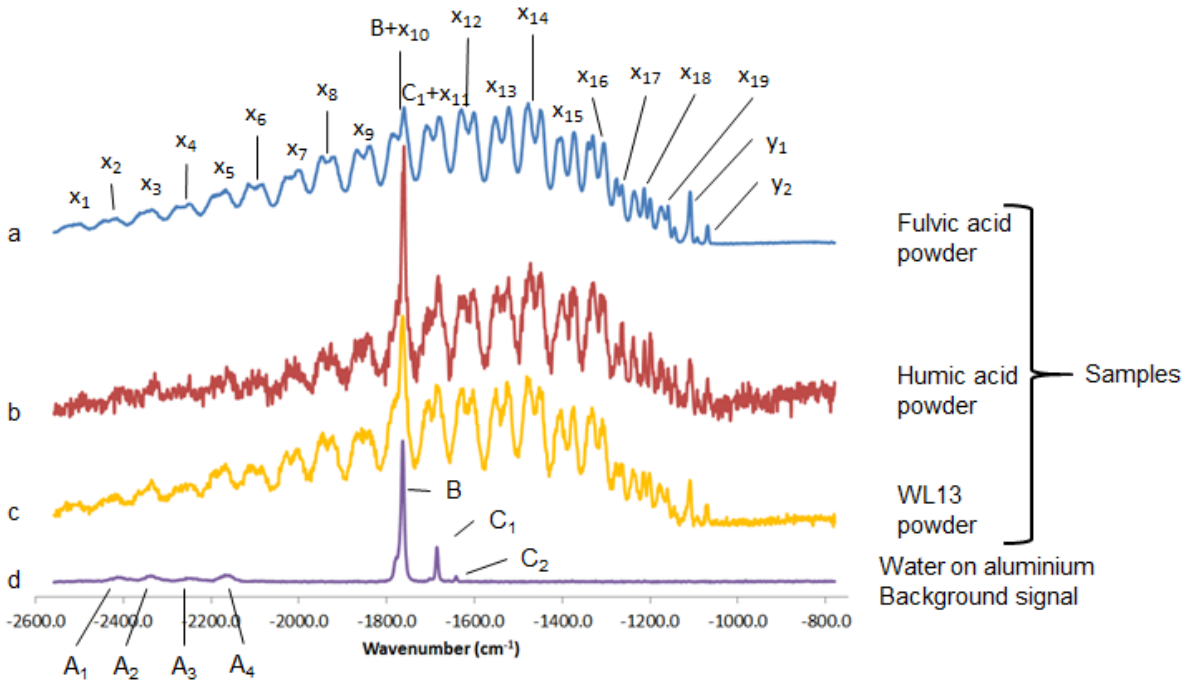


Figure 7.5 – Anti-Stokes spectra of humic acid, fulvic acid and WL13 powder
Traces of humic acid, fulvic acid and Drumtee powders analysed on aluminium plate with the (d) background reference for comparison. The spectral bands produced by the background (A-C2) are listed in the background reference and the bands attributed to humic substances are shown on the top trace (Fulvic acid). The humic substances are normalised between the minimum value and -1477.6cm^{-1}.

7.4.3 – Limit of Detection

Starting from a concentration of 10,000 mg/l, solutions of both humic acid (Figure 7.6) and fulvic acid (Figure 7.7) (though no mixtures at this stage) were diluted in increments to identify the limit of detection. For both humic and fulvic acids the limit of detection is approximately 500 mg/l. The fulvic acid powder gives a clearer spectrum than the humic

acid powder. The main reason for this was that fulvic acid powder could be analysed at a greater laser power as it didn't combust as easily and as it was slightly easier to focus the objective lens it meant it was more likely the laser would be better focussed.

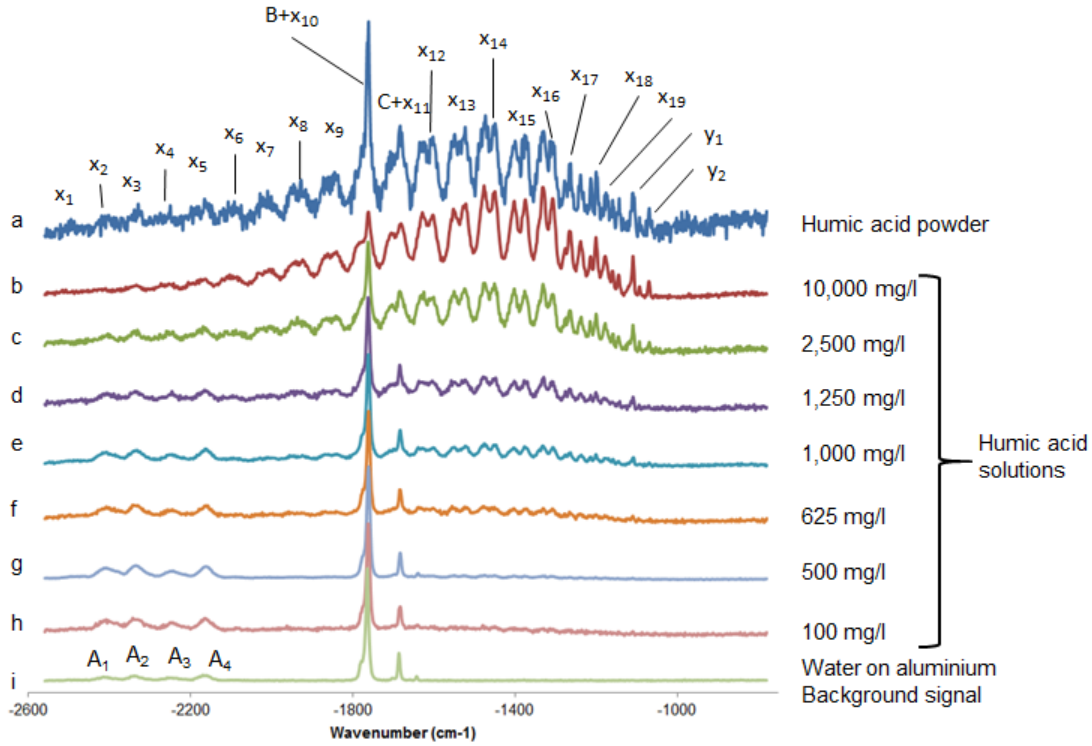


Figure 7.6 – Humic acid in a powder form and solutions
Humic acid in a powder form and in different concentrations in solution, with a blank water sample shown at the bottom of the graph for reference. The B peak is seen prominently in all sample from (a) to (i).

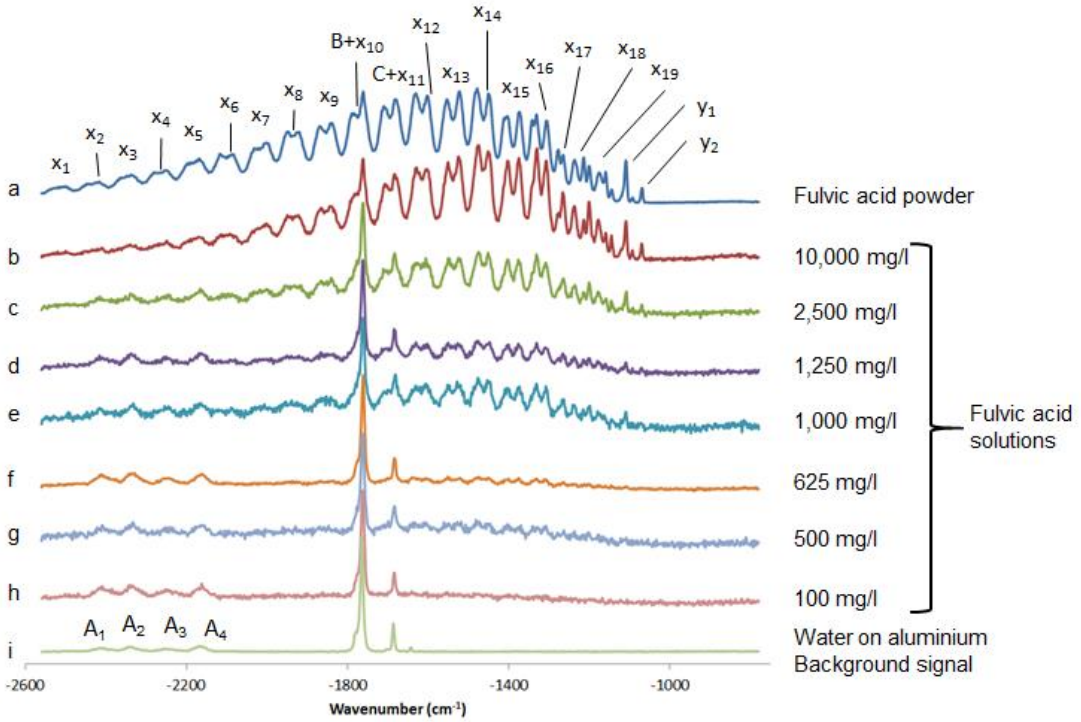


Figure 7.7 – Fulvic acid in a powder form and solutions
Fulvic acid in a powder form and in different concentrations in solution, with a blank water sample shown at the bottom of the graph for reference. The B peak is seen prominently in all sample from (a) to (i).

It should be noted that the limit of detection established for humic substances was for solutions containing only this mixture. A sample collected from the field likely contains additional organic carbon components, such as sugars and amino acids and so the concentration of humic substances would be reduced.

7.4.4 – Differentiating humic and fulvic acids

The results from a comparison of the powdered humic acid and fulvic acid indicated that, after normalisation, the humic acids produced more intense Raman spectra at smaller negative wavenumbers, whilst the fulvic acids appeared to produce more intense Raman spectra at larger negative wavenumbers (Figure 7.8). The difference supports exploring whether a ratio produced from the mean intensity at shorter wavenumbers (-1100 cm^{-1} to -1400 cm^{-1}, designated S1), and the mean intensity at longer wavenumbers (-1800 cm^{-1} to -2000 cm^{-1}, designated S2), could identity humic and fulvic acids end-members. Thus S1:S2 for the humic acid powder was calculated to be 1.186 and S1:S2 for the fulvic acid was calculated to be 0.658 and so were considerably different. Therefore, it appears possible that this ratio, if end-members generate consistent spectra over time, could be used to determine the proportion of humic to fulvic acid in a field sample.

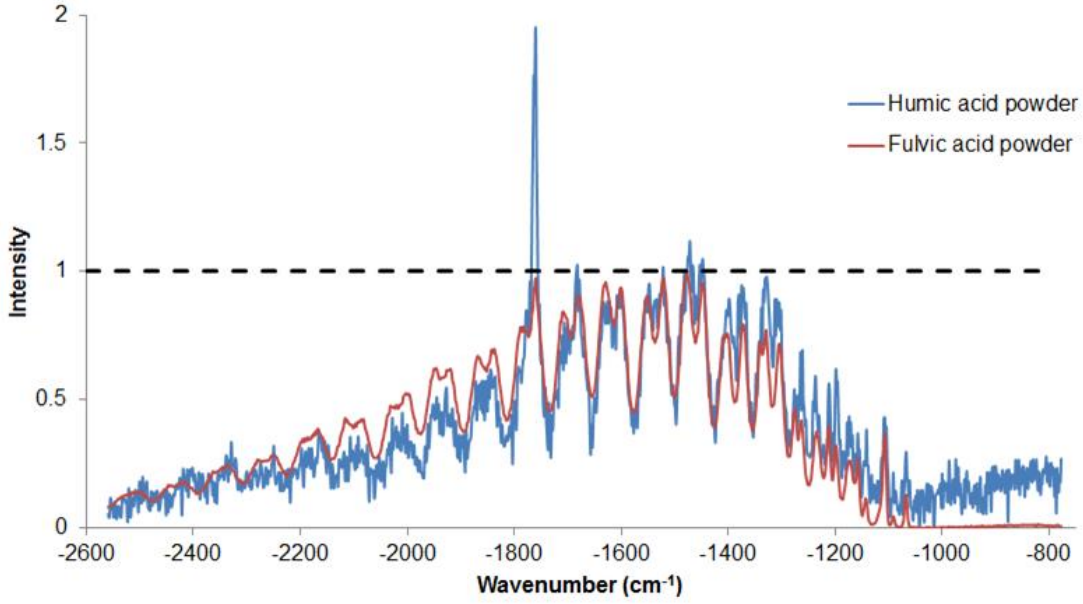


Figure 7.8 – Comparison of humic and fulvic acids in powder form
Comparison of humic acid and fulvic acid in powder form using anti-Stokes Raman scattering with the two figures normalised to the wavenumber -1477.6 cm^{-1} to illustrate sections of the spectra where one substance is producing a more intense spectrum than the other. Unlike most plots presented here the intensity axis is shown numerically, to show where the numbers for S1:S2 come from.

S1:S2 was found to vary depending upon the concentration of the solution. When the solutions of humic and fulvic acids are lower in concentrations then the similarity of S1:S2

increased (Figure 7.9). However, at very high concentrations (e.g., 10,000 mg/l) then S1:S2 is measured to be greater than at lower concentrations. At concentrations lower than 2500 mg/l, S1:S2 for humic acid and fulvic acid cannot be differentiated. However, the separation still allowed exploration of whether different proportions of humic acid and fulvic acid in the same solution would show a relationship in S1:S2 that reflected these proportions.

The spectra produced from 11 solutions containing different proportions of humic acid and fulvic acid were analysed and the S1:S2 calculated (Figure 7.10). The results indicate that when a humic mixture is dominated by humic acid then S1:S2 ratio is greater, but when the solution is dominated by fulvic acid then S1:S2 is smaller. The humic acid powder S1:S2 of 1.186 and the fulvic acid powder S1:S2 of 0.658 fit with this mode in that it may be expected that the S1:S2 would be greatest for powdered humic acid is there is no dilution but this may be a result of not being able to use 300 mW laser power and instead being restricted to 5 mW.

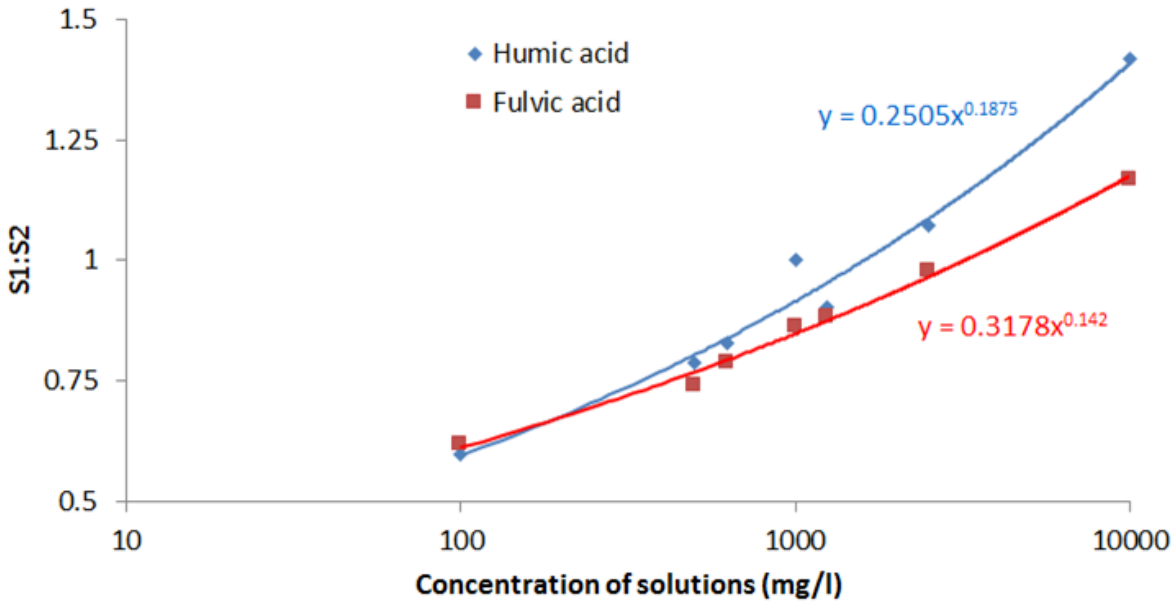


Figure 7.9 – A comparison of S1:S2 for humic acid and fulvic acid at different concentrations.
Solutions contained either humic acid or fulvic acid but not a mixture of the two. Results were collected on 29 June 2015 using a 300 mW laser, with 20 accumulations and 3 second exposures. Prediction and confidence intervals are not shown to make the graph easier to read.

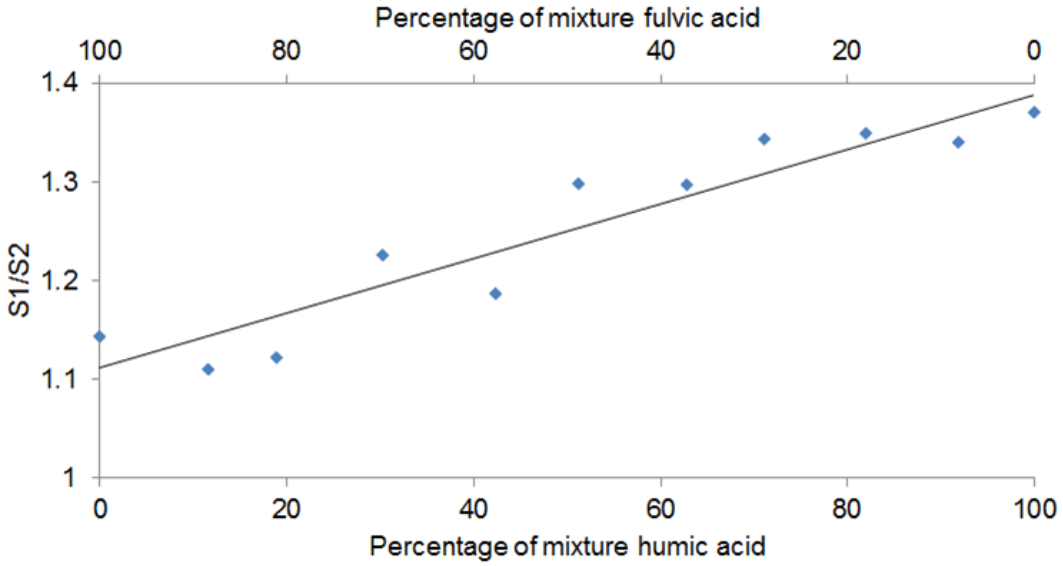


Figure 7.10 – Comparison of S1:S2 for 10,000 mg/l solutions that contained different proportions of humic acid and fulvic acid.
Results (except for the 100% fulvic acid and 100% humic acid solutions) were collected on 24 June 2015. The 100% fulvic acid and 100% humic acid solutions were collected on 29 June 2015. All measurements were made at 300 mW with 25 accumulations and 3 second exposure times.

7.4.5 Anti-Stokes spectra of sugars and amino acids

The anti-Stokes spectra for the solid form of three sugars (glucose, fructose and sucrose) and amino acids (glycine, tryptophan, tyrosine, phenylalanine) were generated. However, only glucose produced spectra different from the background signals (Figure 7.11) and the other six substances (fructose, sucrose, glycine, phenylalanine, phenylalanine and tyrosine) only showed the background peaks of $A_{1\text{-}4}$, B, C_1 and C_2. The glucose signals, designated here as Z1-12, indicate a different set of Raman bands than that observed either as background signals or within the humic substances.

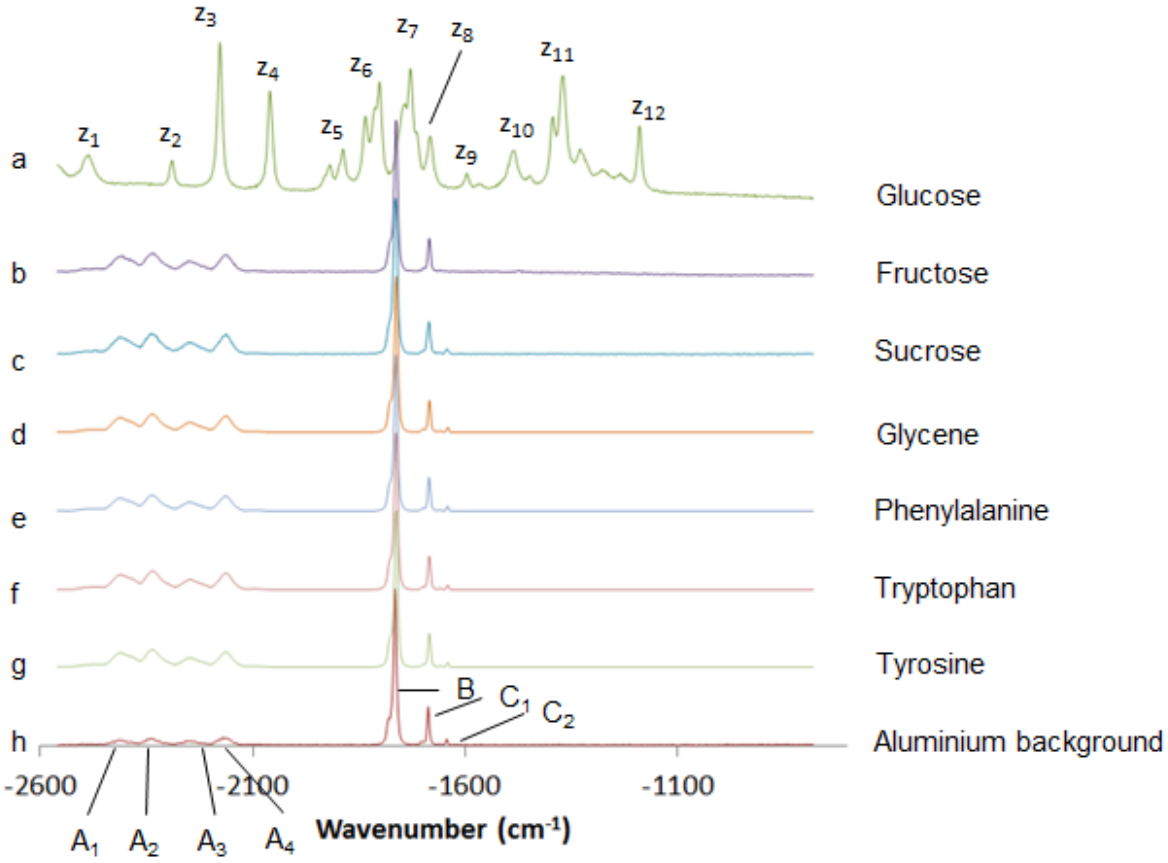


Figure 7.11 – Sugars and amino acids measured using anti-Stokes Raman spectroscopy.
Solids of sugars and amino acids were analysed using a Raman spectrometer to determine if they produced any anti-Stokes Raman scattering. All samples were measured using the 785 nm laser at 300 mW, using a 1 second exposure for 20 accumulations. The aluminium background is shown for reference.

7.5 Discussion

The analysis of several organic carbon components has been shown to be feasible using anti-Stokes Raman scattering. In particular spectra were identified for glucose, humic acid and fulvic acid. While components of DOC have been measured using Raman Stokes scattering - glucose and sucrose (Mathlouthi and Luu, 1980, Soderholm et al., 1999, Shafer-Peltier et al., 2003, Shao et al., 2012); several amino acids (Suh and Moskovits, 1986, Kim et al., 1987, Stewart and Fredericks, 1999, Podstawka et al., 2004, Kang et al., 2013); humic substances (Leyton et al., 2005, Corrado et al., 2008, Heighton, 2013) – no literature could be found on the use of anti-Stokes Raman spectroscopy to analyse humic substances. The anti-Stokes Raman scattering results are promising because they show consistent Raman spectra for humic substances, especially when compared to the results presented in chapter 6. Measuring humic substances using anti-

Stokes scattering is feasible because there is no fluorescence in this region, and they are exhibiting strong and consistent Raman signals.

7.5.1 Correcting for background contributions to the spectra and explaining the burned SERS substrate signals

The main interference in the anti-Stokes Raman spectra for the humic and fulvic acids were the A_{1-4}, B, C_1 and C_2 bands. These signals appeared in all the spectra, unless the spectra were strong enough to drown these Raman bands out, as was the case with the glucose.

One of the observations made during the laboratory testing was that the burned SERS substrates caused the same Raman spectra as did the humic substances (compare Figure 7.4 with Figure 7.5). The reason for this similarity is unclear but the spectral bands measured on the burned SERS substrate were recorded when no other substance had been added to them. The SERS substrates being used from Diagnostic AnSers are printed on paper and this result was produced by accident when the laser power was set too high when the background signals were being analysed. One possibility is that when the substrate is burned it is generating enough thermal energy to burn the paper and this is what is being measured by the spectrometer. It has been identified that when Raman power is too high it can cause an ignition of a target material (Vankeirsbilck et al., 2002a). However, the segment of paper without substrate could not be ignited by direct exposure to the laser (I did try this). However, as the gold substrate consists of extremely thin strands it (as SERS substrates are roughened at a microscopic scale) the energy is not dissipated across a wide surface area, and this seemed to lead to burning. In effect, this is the opposite of a heat sink in a circuit: –the heat sink dissipates the heat generated over a large surface area to prevent any one area becoming too hot, but here the strands of gold melting may have provided sufficient energy to burn the paper. As paper is carbon this could produce a reaction that is producing the x1-x19 and y_1-y_2 bands. The S1:S2 ratio of the burned substrate was 0.69, which is very similar to the same ratio for fulvic acid, which was 0.65 and less similar to the ratio for humic acid of 1.19.

The analysis of humic and fulvic acids were conducted on the aluminium slides rather than the SERS substrate, but these substrates were not the source of the x_1-x_{19} and y_1-y_2 bands observed in the anti-Stokes analysis.

7.5.2 Differentiating humic and fulvic acids

Identifying the ideal ratio to determine the proportion of humic acids to fulvic acids from the results presented in section 7.4 is complicated in part due to the B band, which is

more prominent at lower concentrations. At high concentrations (10,000 mg/l) the B band has minimums influence on band x10 but at lower concentrations the B peak starts to dominate and obscure the x10 band and is significantly stronger than any Raman band from x_1-x_{19} or y_1-y_2. The ratio produced in this project (S1:S2) circumnavigates the interference caused by the B peak by selecting the wavenumbers that do not overlap with this. A wide range of wavelengths is used (S1 being the mean value between -1100 cm^{-1} and -1400 cm^{-1} and S2 being the mean value between -1800 cm^{-1} and -2000 cm^{-1}) so that noise in the signal does not cause issues for the S1:S2 ratio that could arise if a single wavenumber was selected. It may be possible to produce a more sophisticated ratio for differentiating between humic and fulvic acids but the S1:S2 ratio should be quick and easy to implement.

7.5.3 Lack of spectra from sugars and amino acids

Of the three sugars and four amino acids, only one, the glucose produced an identifiable Raman spectrum. That only the glucose produced a clear Raman spectrum was disappointing as Raman spectroscopy has been applied to a wide variety of analytical applications. As discussed in the literature review some of the applications of Raman spectroscopy today include astrobiology (Villar and Edwards, 2006, Hutchinson et al., 2014), drug detection (Garrido et al., 2007, Hargreaves et al., 2008), diabetes measurements (Shafer-Peltier et al., 2003, Robertson and Harmon, 2006, Shao et al., 2012) and graphene studies (Frank et al., 2011, Ferrari and Basko, 2013, Nanda et al., 2016). The difficulties in detecting components of DOC such as sugars and amino acids would currently restrict the usefulness of Raman spectroscopy for analysing DOC.

There are, however, both benefits and drawbacks to the lack of anti-Stokes spectra from sugars and amino acids. The largest benefit is that there is going to be little to no interference with the results returned from the humic substances. As the gradient used to identify the proportions of humic acids to fulvic acids is dependent upon very subtle differences in the spectra this is beneficial: if there were Raman bands in the same region from sugars or amino acids then there would be interference from the spectra of these other substance, but because they do not produce spectra in this region we can study the humic substances in isolation as if these other materials were not even present. The drawback is that sugars (with the exception of glucose) and amino acids do not produce clear Raman spectra in Stokes scattering either except in extremely high concentrations and therefore there isn't an easy way to identify the presence of these molecules using anti-Stokes Raman.

7.5.4 Limit of detection limitations

The results presented demonstrate that anti-Stokes Raman measurements of humic and fulvic acids are possible at sufficiently high concentrations. However, the limit of detection analyses indicates that waters must have a minimum concentration of 500 mg/l of humic or fulvic acid or a mixture of these two components. This information is useful in terms of identifying the technical limitations of using Raman in its simplest form to identify humic substances. However, obtaining clear spectra requires the humic substances to be of 1000 mg/l or greater and for natural samples this presents a challenge as this is well above the values typically found in most natural fluvial environments. Furthermore, it was found that to effectively determine whether a humic mixed solution was predominantly comprised of humic acid or fulvic acid, a concentration of 2500 mg/l or greater was required, with better results as the solution concentration increased.

Natural waters by contrast are very unlikely to ever reach the concentrations required in the limit of detection discussed. The [DOC] measured in some porewater samples has been measured at over 100 mg/l and up to about 150 mg/l (Fiedler et al., 2008, Armstrong et al., 2012, Calmano and Förstner, 2012), but in fluvial systems it can be less than this. For example, the maximum [DOC] at Drumtee was 81.1 mg/l (chapter 4) and this is at the higher end of fluvial concentrations. The value of 81.1 mg/l is still considerably below the limit of detection listed here of about 500 mg/l and even further away from the recommended minimum concentration of 2,500 mg/l needed to differentiate mixtures. However, this could be produced by concentrating the sample by rotary evaporation e.g. one litre of a 60 mg/l C could be reduced to a 25 ml sample would have a concentration of 3000 mg/l C. However, this is very time consuming and may not be desirable in a field deployable sensor with a high throughput.

As this method has just been identified at the tail end of this project there is room for improvement. Four possible methods for generating signals from lower concentrations than were obtained in the experiments presented here would be:

1) conducting the experiments at a greater temperature. Anti-Stokes Raman spectroscopy is temperature dependent and yields more intense spectra at higher temperatures (Karamehmedovic, 2006, Singh et al., 2007). One of the difficulties with this approach might be that the humic acids were shown to be volatile in this project and burned under intense laser powers. To make this a practical application it may be that the solutions would need to be analysed in a chamber with no oxygen and under a more inert gas such as nitrogen that would not produce reduce C compounds.

2) combing SERS substrates with a metal organic framework (MOF). In recent years developments in some labs have successfully combined these two technologies to enable adsorbance of target materials that wouldn't normally adsorb onto the surface of a SERS substrate (Sugikawa et al., 2011, Kreno et al., 2014).
3) the testing of surface enhanced resonance Raman spectroscopy (SERRS). Resonance Raman spectroscopy occurs when a material absorbs some of the incident light at a real, rather than a virtual energy state and this can result in enhanced Raman bands (Merlin, 1985). However, it is not necessarily necessary for the absorption band to be the target material itself and it is possible to add a dye to the target material, which absorbs at the desired wavelength (Faulds et al., 2007). SERRS combines SERS methodology with resonance Raman and has been used to obtain spectra from low concentration solutions.
4) the use of coherent anti-Stokes Raman scattering (CARS). CARS is where two separate lasers are used and set up so that their wavenumbers differences resonantly excite a particular vibrational mode within the target material (Bergner et al., 2011).

Each of these techniques would require additional time and likely investment, which due to time constraints I did not have as the results obtained in this chapter were collected at the end of my research period.

7.5.5 Further consideration

One of the outcomes of these experiments is that it appears that fulvic acids generally have better SNR than do humic acids when in a powder form due to being able to make use of a higher laser power. The difference of fulvic acid having a better SNR than humic acid seems to disappear when the humic substances are in solution as the laser can be set to full power of 300 mW for both samples. Counterintuitively, it may be better to analyse the humic substances in solution as a means to prevent burning it, as this will enable setting the laser to a higher power.

It is interesting to consider why studies measuring organic carbon using anti-Stokes lines have not been published before. Firstly, there is limited research into the measurement of humic substances using Raman spectroscopy in general, largely because fluorescence spectroscopy has been used for the past few decades. As such studies have focussed on the Stokes scattering because this is where the strongest Raman signals are typically found, and the anti-Stokes has not been explored as poor responses with Stokes

scattering may have left the community to consider there was little value in pursuing this further.

Given that fluorescence spectroscopy is well-established and also yields information about the composition of DOC (Coble et al., 1990, Baker, 2001, Esteves et al., 2009) it is questionable the value of considering further experiments using anti-Stokes Raman scattering. However, Raman spectroscopy does have some advantages over fluorescence that are worth considering. For example, it is possible to identify more detailed structural information about some molecules compared to fluorescence (Vankeirsbilck et al., 2002a) and it is possible to perform label-free analysis of substances using Raman spectroscopy (Huh et al., 2009). Here 'labels' refer to a molecule that can be attached to a target molecule to make tracking and identifying the target molecule easier. Where labels are required for Raman spectroscopy there are a greater number of options available (Faulds et al., 2004, Keren et al., 2008). Therefore, there may be value in exploring further the application of anti-Stokes Raman spectroscopy.

One of the normal advantages of Raman spectroscopy is that samples often require little to no preparation and that this allows for rapid analysis of many samples (Hargreaves et al., 2008). To an extent this advantage is lost if samples have to be concentrated by rotary evaporation. However, this is still quicker (~ 2 hours) that fulvic and humic acid extraction (a few days in total). Thus, if a method can be developed to analyse the humic substances using Raman spectroscopy at lower concentrations then this would alleviate this hurdle and make this a potentially very powerful technique for analysing DOC samples. Unlike some other experimental techniques suggested in the literature for measuring humic substances using Raman, there is no need to manipulate the pH (Vogel et al., 1999, Sanchez-Cortes et al., 2006, Corrado et al., 2008) or conduct reduction reactions (Heighton, 2013). Once the sample has been concentrated it can simply be measured by a Raman spectrometer.

7.5.6 Viability of Raman spectroscopy for analysing DOC

In section 6.2.1 it was stated that for Raman spectroscopy to be a useful technique for analysing DOC then it would need to be shown to have some advantage over currently employed techniques. Based on the results presented in chapters 6 and 7 of this there is little justification for recommending the use of Raman Spectroscopy as a method for analysing DOC generally or humic substances specifically.

In terms of being able to characterise molecular structure the experiments based on anti-stokes Raman measurements have shown some success at being able to distinguish between humic acids and fulvic acids. However, it is also possible to distinguish between

humic and fulvic acids using fluorescence spectroscopy (Chen et al., 2003, Esteves et al., 2009). The paper by (Chen et al., 2003) also indicates that it is possible to identify amino acids using fluorescence, whereas the Raman experimentation did not produce clear results in the instances presented in this project. Therefore, the structural information obtained from the Raman spectroscopy is not greater than that obtainable from fluorescence spectroscopy or NMR (Esteves et al., 2009) and consequently the use of Raman to analyse DOC in a laboratory setting is not recommended over fluorescence or NMR.

The second issue is one of portability when regarding how suitable Raman spectroscopy would be as the basis of a field based sensor. Raman spectrometers have been miniaturised into portable sensors by various companies. Although the anti-stokes Raman approach did produce a ratio that could be used to calculate the proportion of humic acids to fulvic acids (Figure 7.10) in solution this relies on the solutions being at concentrations greater than 1000 mg/l (Figure 7.9). In terms of producing a field deployable sensor this introduces the additional complexity of concentrating samples to a value of high concentrations such as 10,000 mg/l. It would present a significant technological challenge to automate the concentrating process of a field deployable sensor without sacrificing the rate of throughput. Additionally, sensors in the field are usually subject to greater margins of error than measurements made in a laboratory. In order to make this approach viable a method would have to be identified to reduce the limit of detection to a value that would mean that concentrating solutions is not required.

7.6 Conclusions

These experiments have shown that anti-Stokes Raman spectroscopy is a viable method of measuring humic and fulvic acids. The experimental results presented here show the following achievements:

1) measurements of anti-Stokes Raman measurements of humic and fulvic acids are possible.
2) the spectra generated here are consistent and produce extremely clear Raman spectra that are easy to differentiate from any background signals when concentrations were sufficiently high (>2,500 mg/l).
3) the detection of humic substances using the anti-Stokes measurements has been achieved using both purchased standards and a sample extracted from a local fieldsite. Unfortunately, at present the limit of detection is greater than what would be found in natural fluvial systems.

4) a numerical approach for differentiating humic acid from fulvic acid through differences in spectra intensity, 'the S1:S2 ratio', has been established.

At present, there are three issues that need to be addressed:

1) poor limit of detection. [DOC] at Drumtee is usually between 10 mg/l and 60 mg/l but the limit of detection in these experiments was measured to be about 1,000 mg/l and as discussed, I would recommend a concentration of about 2,500 mg/l or more.
2) concentrating each sample to 10,000 mg/l or to a powder is a time-consuming process.
3) why do the humic substances and the burned SERS substrates produce similar spectra? The similarity of the spectra produced from these separate sources could not be explained satisfactorily and the number of Raman bands suggests that it can't be a coincidence that they are so similar. Why was this, and are there other substances not analysed here that give the same or very similar results?

Unfortunately, the full usefulness of Raman spectroscopy for analysing DOC has not been fully explored as this breakthrough has come at the end of the PhD with little time to consider further. It is possible that with more time to explore the use of Anti-Stokes Raman scattering as a method of probing humic substances that some, or indeed all, of the concerns listed above could be addressed. However, without these improvements this approach does not give any clear advantages over using a fluorescence based spectrometer.

8.0 Conclusions and future work

In this chapter, I first summarise the key findings of the time series analysis, generated to understand DOC export dynamics, and then of the Raman spectroscopy, undertaken to consider DOC composition. Thereafter I consider how future research could develop further the novel research findings summarised here.

8.1 – Summary of key findings

DOC is a challenging substance to analyse because it is compositionally complex largely due to its heterogeneity (McKnight et al., 1992, Sondergaard and Middelboe, 1995, Lutz et al., 2012) and because its export from terrestrial sources such as soils is influenced by more than one variable including pH (Kennedy et al., 1996), discharge (Raymond and Saiers, 2010) and atmospheric CO_2 levels (Freeman et al., 2004). This dissertation focussed on advancing the methodology for measuring and analysing DOC with two key aims. The first aim of my research was to use high resolution data to help better understand the drivers of [DOC] and if possible to establish a model between the [DOC] and other hydrological variables that are comparatively easier to measure. The second aim of my research was to use Raman spectroscopy to determine if this is a suitable tool for analysing the composition of DOC.

8.1.1 – Time series outcomes

More than two years of high resolution hydrochemical data were collected from the Drumtee catchment, generating near-continuous time series for the period between May 2012 and December 2014, incorporating both the 2012/13 and 2013/14 hydrological years. Fortuitously this time period offered an interesting contrast and opportunity for comparison - the summer of 2012 was unusually wet with a large number of events (which were defined as increases in discharge brought on by precipitation) and the summer of 2013 was drier with fewer rainfall events, where the discharge would be measured to significantly increase for a short period and then recover to baseflow. The summer of 2012 exhibited greater maximum [DOC] than the summer of 2013. In both years, the pH and SC measured at Drumtee decreased rapidly during events and then increased between events more slowly.

One of the interesting outcomes of this research was how similar the C export budget calculated using the Spectro::lyser™ was to the budget calculated using the laboratory data despite several hundred times more data points used in constructing the budget for

the former. The similarity of budgets calculated using field based data and laboratory based data contrasted with a German study that showed a much larger difference between export budgets generated using low and high resolution data sets (Strohmeier et al., 2013).

The data showed that the top 10% of flows were accounting for 44.2% of C exports, which is in line with other similar research (Raymond and Saiers, 2010). More detailed analysis using the high resolution [DOC] data revealed that [DOC] would frequently decrease during events and that the minimum [DOC] would frequently coincide with the time period when the discharge value was increasing most quickly, indicating that dilution of the DOC entering the stream was occurring. During these time periods, although the concentration was decreasing the total magnitude of C being exported was increasing.

Multiple contemporaneous hydrochemistry time series allowed exploration of whether commonly measured parameters could be used to describe [DOC] (an advantage if possible, as time series construction of [DOC] requires more complex sensor logistics). However, there was limited predictive power of any one variable for estimating [DOC] in the catchment at any particular time. Temperature was a reasonable proxy, especially if the mean temperature during the previous 200 hours was considered. However, a multi-variate model, in particular a generalised additive model (GAM), that combined time series measurements of pH, SC, temperature, the time between events, and discharge offered a much more successful approach in constructing a model that described [DOC]. This approach may also be of benefit because it is possible the hydrological relationships with [DOC] are non-linear.

Wavelet analysis was identified as a useful exploratory analysis tool for handling the large volumes of data generated. In this project, I made use of both continuous wavelet transforms (CWTs), maximal overlap discrete wavelet transforms (MODWTs) and wavelet coherence analysis (WTCs). Although wavelet analysis was not suitable to pinpoint specific details it was useful as a tool to inform which way the analysis should proceed. Wavelets were valuable for identifying patterns that may exist within the data and enabling the analysis to be focussed in particular directions such as showing transient, short term relationships of a few weeks between pH and [DOC], even if that relationship did not extend over a full hydrological year.

Scatter plots for all of the data sets collected were generated to explore linear interactions between different variables and these were shown to have no clear pattern. However, when the periods of data identified by the wavelet coherence analysis were extracted for a scatter plot those same variables showed a much stronger correlation. It would be

extremely unlikely that the timing of this type of relationship could be identified without the use of wavelet analyses. One of the goals of this project has been to determine suitable analytical tools for analysing high resolution time series data sets from [DOC]. Wavelet analysis appears to be a powerful method of exploratory analysis that can help direct other aspects of the analysis.

8.1.2 – Raman outcomes

At the outset of the project there was a desire to determine if Raman spectroscopy could be used to understand the composition of DOC, particularly considering humic and fulvic acids. Even narrowing the analysis down to humic and fulvic acids presents a challenge, for although humic and fulvic acids are only one component of DOC they are a heterogeneous mixture of chemicals (McKnight et al., 1992) and therefore, they are difficult to analyse.

The experimentation based on existing studies using Raman spectroscopy to identify humic substances was unable to replicate the Raman spectra that had been reported in other studies (Leyton et al., 2005, Sanchez-Cortes et al., 2006, Corrado et al., 2008, Roldán et al., 2011, Heighton, 2013). The Leyton, Sanchez-Cortes, Corrado and Roldán papers relied on increasing pH to make the solution alkaline, with their papers showing attenuation of the signals. In the experiments that I conducted I was unable to differentiate my signals from the background fluorescence. Sometimes a Raman signal would be prominent in a test but when these did occur they could not be replicated consistently and therefore I could not definitively state that they were the product of humic substances rather than another analytical artefact, for example contamination. I too performed a reduction reaction on the humic and fulvic acids (as per Heighton 2013) but this seemed to have little influence on the spectral lines measured. Again, sometimes a prominent spectral line would be observed but this seemed no more or less common than the random spectral lines just discussed. Therefore, I was unable to attribute these spectral lines to humic or fulvic acid.

The Raman spectrometer used in these experiments (Renishaw In-Via Raman Spectrophotometer) was equipped with both a 785 nm and a 514 nm laser. The 785 nm laser was selected in this project as it produces less intense fluorescence signals than does a 514 nm laser as its wavelength range is in the near infrared (Viskari and Landers, 2006, Qin et al., 2010). In general the longer a laser wavelength the less fluorescence is produced (Vítek et al., 2012). Additionally, some of the experiments made use of gold SERS substrates but one of the absorption bands of gold overlaps close to the 514 nm

laser, making the use of gold substrates in conjunction with the 514 nm laser is not particularly effective (Yarwood et al., 2009).

This component of the research has revealed that great care and attention must be taken to identify all possible sources of background signals, particularly when (as is the case with humic substances) the substances are strongly fluorescent (Mobed et al., 1996, Baker, 2001). Even if there are Raman spectra in the sample these can be overwhelmed by the stronger fluorescence signals (Porterfield and Campion, 1988). In this project, I have taken considerable care to identify background signals including measurements of substrates and the aluminium slide, both when wet and when dry.

A different method than that listed in the literature for potentially measuring humic and fulvic acids using Raman spectroscopy was identified. The approach of measuring humic and fulvic acids presented in this project differed from previous methods reported for identifying humic and fulvic acid using Raman spectroscopy by measuring the anti-Stokes Raman spectra. Often this part of the Raman spectra is ignored because it tends to produce weaker signals compared to Stokes scattering at room temperature (Fantini et al., 2004).

8.2 – Future research directions

Although considerable insight has been presented on the drivers of DOC export from Drumtee, and so other headwater catchments, all controls cannot be identified from the data presented here. Additionally, the use of Raman spectroscopy to measure DOC was exploratory and so there is scope to expand on this research. I have the following suggestions on future research needs arising from this research.

1) Continued measurements of [DOC] from fluvial catchments such as Drumtee or other peatland catchments using high resolution sensors to understand better long- and short-term trends which are complex and difficult to predict. In particular, the reasons for the dilution of [DOC] at the start of some but not all events are not entirely clear. Why does this happen sometimes but not all the time? The collection of data could also be improved with either more concurrent data being collected such as sensors to measure soil moisture, water table depth, soil temperature etc.
   Additionally, this research considered only one catchment scale (one point on the river). It would be insightful to deploy a second Spectro::lyser™ (or similar sensor), or even a suite of sensors, either upstream or downstream to determine DOC

attenuation of input and consider synchronicity in DOC behaviour across the catchment with changing scale. Simultaneous measurements at different locations would be important to understand how much DOC is lost or gained, through for example respiration, or change in land use influencing influx. Related to this would be to use the DOC time series to understand other aspects of the fluvial C cycle. For example, an additional C pool that would be interesting to characterise is the degassing of CO_2 from the river, particularly considering hydrological controls. The release of free CO_2 from the Drumtee catchment has been measured and analysed (Long et al., 2015) but spot samples only and whether there is a relationship with DOC export has not been explored. The sensor data could enable a clearer picture on the controls on CO_2 efflux especially as pH is a potential control on DOC export (Kennedy et al., 1996) and influences CO_2 efflux. It has been known for some time that many catchments can be very responsive to hydrological events (Clark et al., 2007a, Bass et al., 2011). Perhaps most fundamentally to the understanding of the short-term drivers of variations within the [DOC] time series measured at Drumtee is the question of what is the root cause of the surprising dilutions during many but not all events. One approach to answering this question would be the deployment of synchronised sensors placed to measure different aspects of the catchment at different locations. For example, sensors could be used to continuously measure the water table depth, peat moisture levels, ground temperature and potentially relate this back into the [DOC] time series to determine if there are any commonalities in these other variables that could determine whether [DOC] would decrease or not at the beginning of an event.

2) Refine the GAM model presented here or alternately explore the use of other statistical modelling tools that could generate a useful model for calculating [DOC] using other hydrological variables. A better [DOC] model could enable the Spectro::lyser™ to be moved to different sites over time until it enough data is generated to create a model at that different site and then it could be moved again. Moving the sensor to different sites until a model is generated would be an approach that would enable high resolution data to be calculated at multiple sites without the need to invest in another [DOC] sensor. At present the GAM model can calculate [DOC] with a reasonable degree of accuracy but there is room for improvement with further data collection and greater insight into the data collected.
3) The new approach for using Raman spectroscopy for analysing humic substances using anti-Stokes spectra leaves significant scope for future work and improvements. However, given I had great difficulty replicating the work of others (Leyton et al., 2005, Corrado et al., 2008, Heighton, 2013), or at the least was reaching different conclusions about my data, I suggest first the anti-Stokes

findings here are replicated in another laboratory. If this approach could be replicated in a different laboratory then it would lend credence to the anti-Stokes methodology.

4) Assuming the results can be replicated, then future research should focus on how could this technique be developed to offer a practical method of measuring DOC in a field environment. A key issue is that, a field sensor designed around anti-Stokes Raman spectroscopy seems to need a significantly concentrated solution (>1000 mg/l) for analysis. Automating the concentration process and still being able to maintain a 30 minute resolution (such as that recorded by the Spectro::lyser^{TM}) would not be likely to be practical. The more realistic recommendation would be to find a method of improving the limit of detection. One methods that could not be tested in this project but that may provide a possible way forward would be coherent anti-Stokes Raman scattering (CARS) spectroscopy. CARS requires a more complex set up as it utilises more than one laser wavelength for analysing a sample but enables the stronger measurement of Raman signals (Bergner et al., 2011) and so the concentration may be lowered.
5) Raman spectroscopic analysis of humic and fulvic acids using a 1064 nm laser. As has been described (Vítek et al., 2012) longer wavelength lasers produce less fluorescence despite the trade-off of weaker Raman signals. A laser with this wavelength may be capable of detecting humic and fulvic acids in the Stokes region of the spectrum.

8.3 – Contributions made from this research

I will finish by summarising the contribution this research has made to understanding fluvial DOC dynamics. However, first I will acknowledge that the research has been more effective as it builds on understanding generated from research previously carried out at this site (Waldron et al., 2009, Murray, 2012), and other research using sensors, including projects that have worked with the Spectro::lyser^{TM} (Müller et al., 2014) (Sandford et al., 2010, Grayson and Holden, 2012, Strohmeier et al., 2013), or with high resolution hydrology data (Kirchner et al., 2004, Jeong et al., 2012, Neal et al., 2012) or simply is of interest as we now know that [DOC] is increasing (Evans et al., 2001, Worrall et al., 2004b) in UK waters and internationally (Monteith et al., 2007).

However, this research has furthered understanding. The ability to measure [DOC] at high resolutions using sensor technology has only really been possible in the last few years (Sandford et al., 2010, Grayson and Holden, 2012, Strohmeier et al., 2013, Bass et al., 2014, Jones et al., 2014, Müller et al., 2014) and this is one of the longest high

resolution [DOC] time series produced to date. Consequently, the number of studies that have been able to generate the type of long term high resolution [DOC] time series presented here is extremely limited and some of the research presented here does not appear in any known published literature:

1. The measurement of [DOC] at a high resolution in concentrations as high as those presented here. From this and accompanying hydrochemistry time series, the amount of C exported, the processing that controls it and the dynamics of export have been considered. Producing this sensor-generated DOC time series was not straightforward and required method development beyond the manufacturer's algorithm, and considerable tenacity in field work. The research published by (Müller et al., 2014) was from a lake and measured maximum [DOC] during the period of study as 26.1 mg/l. The paper published by (Strohmeier et al., 2013) measured a maximum [DOC] of 33.8 mg/l. Furthermore, the Strohmeier paper also exhibits a very different [DOC] profile, where it has a consistent baseline and responds more predictably to events with [DOC] increase as the discharge increased and then decreased to baseflow after the event. At Drumtee the response was significantly more complicated due to the dilutions that occurred during events.
2. The use of wavelet analysis to analyse the time series. Frequency domain measurements have been made by (Müller et al., 2014) but these were made using Fourier analysis. For the first time wavelet analysis has been applied to consider DOC dynamics and possible controls on this variation.
3. For the first time, anti-Stokes Raman spectroscopy has been explored to identify humic acids within fluvial DOC, prompted by the inability to produce humic acid spectra from existing Raman spectroscopic approaches, which in itself prompts focus on the importance of reproducibility in research. I have identified a method using Raman spectroscopy to measure humic acids and fulvic acids using the anti-Stokes section of the spectrum. I have also determined a limit of detection. If methods to improve the limit of detection could be identified, then this would be a potentially very powerful technique for identifying humic substances.
4. Identified a ratio (S1:S2) that can be used to determine the relative proportion of humic acid to fulvic acid in a mixture of the two substances when using anti-Stokes Raman spectroscopy. This ratio showed that there is not only the possibility of identifying humic substances within fluvial systems but that the relative proportion of fulvic acid to humic acid can be calculated using the simple approach identified using S1:S2.

All these advances are important as whilst they provide insight, they show that there is still much to be explored in understanding fluvial C dynamics from sensor-generated time series. As the climate changes the need for field-based measurement at higher frequencies will grow, for example to understand tipping points, or in remote

environments, now accessible, but still logistically challenging for intensive measurements. Research such as this is needed to ensure that the understanding that these systems generate is accurate and that appropriate tools for interpretation are identified, whether they be statistical models fit-for-purpose or the development of new hardware to provide a new and improved understanding.

References

ACHTERBERG, E. P., BRAUNGARDT, C. B., SANDFORD, R. C. & WORSFOLD, P. J. 2001. Uv digestion of seawater samples prior to the determination of copper using flow injection with chemiluminescence detection. *Analytica Chimica Acta,* 440**,** 27-36.

ADAMS, J. M. & FAURE, H. 1998. A new estimate of changing carbon storage on land since the last glacial maximum, based on global land ecosystem reconstruction. *Global and Planetary Change,* 16**,** 3-24.

ÅGREN, A., BUFFAM, I., COOPER, D., TIWARI, T., EVANS, C. & LAUDON, H. 2013. Can the heterogeneity in stream dissolved organic carbon be explained by contributing landscape elements? *Biogeosciences Discussions,* 10.

AHIPATHY, M. & PUTTAIAH, E. 2006. Ecological characteristics of vrishabhavathy river in bangalore (india). *Environmental geology,* 49**,** 1217-1222.

AIKEN, G. R., MCKNIGHT, D. M., THORN, K. A. & THURMAN, E. M. 1992. Isolation of hydrophilic organic-acids from water using nonionic macroporous resins. *Organic Geochemistry,* 18**,** 567-573.

AIKEN, G. R., THURMAN, E. M., MALCOLM, R. L. & WALTON, H. F. 1979. Comparison of xad macroporous resins for the concentration of fulvic-acid from aqueous-solution. *Analytical Chemistry,* 51**,** 1799-1803.

AIKENS, C. M. & SCHATZ, G. C. 2006. Tddft studies of absorption and sers spectra of pyridine interacting with au20. *The Journal of Physical Chemistry A,* 110**,** 13317-13324.

AL-ATTAR, N., KOPF, I., KENNEDY, E., FLAVIN, K., GIORDANI, S. & RICE, J. H. 2012. Surface-enhanced raman scattering from small numbers of purified and oxidised single-walled carbon nanotubes. *Chemical Physics Letters,* 535**,** 146-151.

AL-FAIYZ, Y. S. 2012. Cpmas< sup> 13</sup> c nmr characterization of humic acids from composted agricultural saudi waste. *Arabian Journal of Chemistry*.

ALAM, M. M., REHMAN, S., AL-HADHRAMI, L. M. & MEYER, J. P. 2014. Extraction of the inherent nature of wind speed using wavelets and fft. *Energy for Sustainable Development,* 22**,** 34-47.

ALLEN, H. L. 1973. Production and utilization of dissolved organic carbon during in-situ phyto plankton photosynthesis measurements. *Internationale Revue der Gesamten Hydrobiologie,* 58**,** 843-849.

ANEMÜLLER, J., SEJNOWSKI, T. J. & MAKEIG, S. 2003. Complex independent component analysis of frequency-domain electroencephalographic data. *Neural Networks,* 16**,** 1311-1323.

ANESIO, A. M. & GRANELI, W. 2003. Increased photoreactivity of doc by acidification: Implications for the carbon cycle in humic lakes. *Limnology and Oceanography,* 48**,** 735-744.

APPELS, J., KUESTER, E. & VAN DEN BROEKE, J. 2007. Combination of an on-line biomonitor using light emitting bacteria and a uv spectrophotometer probe for homeland security and drinking water safety - art. No. 67391e. *In:* KAMERMAN, G. W., STEINVALL, O. K., LEWIS, K. L., KRAPELS, K. A., CARRANO, J. C. & ZUKAUSKAS, A. (eds.) *Electro-optical remote sensing, detection, and photonic technologies and their applications.*

ARMSTRONG, A., HOLDEN, J., KAY, P., FRANCIS, B., FOULGER, M., GLEDHILL, S., MCDONALD, A. & WALKER, A. 2010. The impact of peatland drain-blocking on dissolved organic carbon loss and discolouration of water; results from a national survey. *Journal of Hydrology,* 381**,** 112-120.

ARMSTRONG, A., HOLDEN, J., LUXTON, K. & QUINTON, J. 2012. Multi-scale relationship between peatland vegetation type and dissolved organic carbon concentration. *Ecological Engineering,* 47**,** 182-188.

BAKER, A. 2001. Fluorescence excitation-emission matrix characterization of some sewage-impacted rivers. *Environmental Science & Technology,* 35**,** 948-953.

BAKER, A. & SPENCER, R. G. M. 2004. Characterization of dissolved organic matter from source to sea using fluorescence and absorbance spectroscopy. *Science of the Total Environment,* 333**,** 217-232.
BALL, D. 1964. Loss-on-ignition as an estimate of organic matter and organic carbon in non-calcareous soils. *Journal of Soil Science,* 15**,** 84-92.
BARRENETXEA, G., INGELREST, F., LU, Y. M. & VETTERLI, M. Assessing the challenges of environmental signal processing through the sensorscope project. 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. IEEE, 5149-5152.
BASILE, A., CIOLLARO, G. & COPPOLA, A. 2003. Hysteresis in soil water characteristics as a key to interpreting comparisons of laboratory and field measured hydraulic properties. *Water Resources Research,* 39.
BASS, A. M., BIRD, M. I., LIDDELL, M. J. & NELSON, P. N. 2011. Fluvial dynamics of dissolved and particulate organic carbon during periodic discharge events in a steep tropical rainforest catchment. *Limnology and Oceanography,* 56**,** 2282-2292.
BASS, A. M., MUNKSGAARD, N., LEBLANC, M., TWEED, S. & BIRD, M. 2014. Contrasting carbon export dynamics of human impacted and pristine tropical catchments in response to a short-lived discharge event. *Hydrological Processes,* 28**,** 1835-1843.
BASS, A. M., O' GRADY, D., BERKIN, C., LEBLANC, M., TWEED, S., NELSON, P. N. & BIRD, M. I. 2013. High diurnal variation in dissolved inorganic c, delta c-13 values and surface efflux of co2 in a seasonal tropical floodplain. *Environmental Chemistry Letters,* 11**,** 399-405.
BEGGS, K. M. & SUMMERS, R. S. 2011. Character and chlorine reactivity of dissolved organic matter from a mountain pine beetle impacted watershed. *Environmental science & technology,* 45**,** 5717-5724.
BERGNER, G., ALBERT, C. R., SCHILLER, M., BRINGMANN, G., SCHIRMEISTER, T., DIETZEK, B., NIEBLING, S., SCHLÜCKER, S. & POPP, J. 2011. Quantitative detection of c-deuterated drugs by cars microscopy and raman microspectroscopy. *Analyst,* 136**,** 3686-3693.
BERNAL, D. & YOUSSEF, A. 1998. A hybrid time frequency domain formulation for non-linear soil–structure interaction. *Earthquake engineering & structural dynamics,* 27**,** 673-685.
BEVEN, K. 2006. Searching for the holy grail of scientific hydrology: Q t=(s, r,? T) a as closure. *Hydrology and Earth System Sciences Discussions,* 10**,** 609-618.
BILLETT, M., CHARMAN, D., CLARK, J., EVANS, C., EVANS, M., OSTLE, N., WORRALL, F., BURDEN, A., DINSMORE, K. & JONES, T. 2010. Carbon balance of uk peatlands: Current state of knowledge and future research challenges. *Climate Research,* 45**,** 13-29.
BILLETT, M., PALMER, S., HOPE, D., DEACON, C., STORETON-WEST, R., HARGREAVES, K., FLECHARD, C. & FOWLER, D. 2004. Linking land-atmosphere-stream carbon fluxes in a lowland peatland system. *Global Biogeochemical Cycles,* 18.
BILLOT, L., DE LA CHAPELLE, M. L., GRIMAULT, A.-S., VIAL, A., BARCHIESI, D., BIJEON, J.-L., ADAM, P.-M. & ROYER, P. 2006. Surface enhanced raman scattering on gold nanowire arrays: Evidence of strong multipolar surface plasmon resonance enhancement. *Chemical physics letters,* 422**,** 303-307.
BLOOMFIELD, P. 2004. *Fourier analysis of time series: An introduction*, John Wiley & Sons.
BOGNER, K. & KALAS, M. 2008. Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the upper danube catchment. *Atmospheric Science Letters,* 9**,** 95-102.
BOLTON, L. 2003. *The application of excitation-emission fluorescence spectrophotometry to the monitoring of dissolved organic matter in upland catchments in the united kingdom.*
BRADSHAW, G. & MCINTOSH, B. 1994. Detecting climate-induced patterns using wavelet analysis. *Environmental Pollution,* 83**,** 135-142.

BRIGHT, J., LANGSTON, R., BULLMAN, R., EVANS, R., GARDNER, S. & PEARCE-HIGGINS, J. 2008. Map of bird sensitivities to wind farms in scotland: A tool to aid planning and conservation. *Biological Conservation,* 141**,** 2342-2356.
BROADBENT, F. 1953. The soil organic fraction. *Advances in agronomy,* 5**,** 153-183.
BUFFLE, J. 1990. The analytical challenge posed by fulvic and humic compounds. *Analytica Chimica Acta,* 232**,** 1-2.
BUJOSA, M., GARCÍA-FERRER, A. & YOUNG, P. C. 2007. Linear dynamic harmonic regression. *Computational statistics & data analysis,* 52**,** 999-1024.
BURT, R. 1992. *Soil survey laboratory methods manual*, USDA.
CALMANO, W. & FÖRSTNER, U. 2012. *Sediments and toxic substances: Environmental effects and ecotoxicity*, Springer Science & Business Media.
CAMPION, A. & KAMBHAMPATI, P. 1998. Surface-enhanced raman scattering. *Chem. Soc. Rev.,* 27**,** 241-250.
CAO, X., DROSOS, M., LEENHEER, J. A. & MAO, J. 2016. Secondary structures in a freeze-dried lignite humic acid fraction caused by hydrogen-bonding of acidic protons with aromatic rings. *Environmental science & technology,* 50**,** 1663-1669.
CARLSON, C. A. & DUCKLOW, H. W. 1996. Growth of bacterioplankton and consumption of dissolved organic carbon in the sargasso sea. *Aquatic Microbial Ecology,* 10**,** 69-85.
CARRIERE, J. T. & HAVERMEYER, F. Ultra-low frequency stokes and anti-stokes raman spectroscopy at 785nm with volume holographic grating filters. SPIE BiOS, 2012. International Society for Optics and Photonics, 821905-821905-9.
CARSTEA, E. M., BAKER, A., BIEROZA, M. & REYNOLDS, D. 2010. Continuous fluorescence excitation-emission matrix monitoring of river organic matter. *Water Research,* 44**,** 5356-5366.
CAUWET, G. 1994. Htco method for dissolved organic-carbon analysis in seawater - influence of catalyst on blank estimation. *Marine Chemistry,* 47**,** 55-64.
CAZELLES, B., CHAVEZ, M., BERTEAUX, D., MÉNARD, F., VIK, J. O., JENOUVRIER, S. & STENSETH, N. C. 2008. Wavelet analysis of ecological time series. *Oecologia,* 156**,** 287-304.
CHANEY, S. B., SHANMUKH, S., DLUHY, R. A. & ZHAO, Y.-P. 2005. Aligned silver nanorod arrays produce high sensitivity surface-enhanced raman spectroscopy substrates. *Applied Physics Letters,* 87**,** 031908.
CHATHAM, J. C. & BLACKBAND, S. J. 2001. Nuclear magnetic resonance spectroscopy and imaging in animal research. *IlAR Journal,* 42**,** 189-208.
CHEN, C. & WANG, X. 2007. Sorption of th (iv) to silica as a function of ph, humic/fulvic acid, ionic strength, electrolyte type. *Applied Radiation and Isotopes,* 65**,** 155-163.
CHEN, W., WESTERHOFF, P., LEENHEER, J. A. & BOOKSH, K. 2003. Fluorescence excitation - emission matrix regional integration to quantify spectra for dissolved organic matter. *Environmental Science & Technology,* 37**,** 5701-5710.
CHEN, W. H. & WANGERSKY, P. J. 1993. A high-temperature catalytic-oxidation method for the determination of marine dissolved organic-carbon and its comparison with the uv photooxidation method. *Marine Chemistry,* 42**,** 95-106.
CHERUKURU, N., FORD, P. W., MATEAR, R. J., OUBELKHEIR, K., CLEMENTSON, L. A., SUBER, K. & STEVEN, A. D. 2016. Estimating dissolved organic carbon concentration in turbid coastal waters using optical remote sensing observations. *International Journal of Applied Earth Observation and Geoinformation,* 52**,** 149-154.
CHOW, A. T. 2006. Comparison of dax-8 and xad-8 resins for isolating disinfection byproduct precursors. *Journal of Water Supply Research and Technology-Aqua,* 55**,** 45-55.
CHOW, A. T., GUO, F., GAO, S., BREUER, R. & DAHLGREN, R. A. 2005. Filter pore size selection for characterizing dissolved organic carbon and trihalomethane precursors from soils. *Water Research,* 39**,** 1255-1264.
CHOW, A. T., GUO, F. M., GAO, S. D. & BREUER, R. S. 2006. Size and xad fractionations of trihalomethane precursors from soils. *Chemosphere,* 62**,** 1636-1646.

CLARK, J. M., LANE, S. N., CHAPMAN, P. J. & ADAMSON, J. K. 2007a. Export of dissolved organic carbon from an upland peatland during storm events: Implications for flux estimates. *Journal of Hydrology,* 347**,** 438-447.
CLARK, J. M., LANE, S. N., CHAPMAN, P. J. & ADAMSON, J. K. 2007b. Export of dissolved organic carbon from an upland peatland during storm events: Implications for flux estimates. *Journal of Hydrology,* 347**,** 438-447.
COBLE, P. G. 1996. Characterization of marine and terrestrial dom in seawater using excitation emission matrix spectroscopy. *Marine Chemistry,* 51**,** 325-346.
COBLE, P. G., GREEN, S. A., BLOUGH, N. V. & GAGOSIAN, R. B. 1990. Characterization of dissolved organic-matter in the black-sea by fluorescence spectroscopy. *Nature,* 348**,** 432-435.
COLE, J. J., PRAIRIE, Y. T., CARACO, N. F., MCDOWELL, W. H., TRANVIK, L. J., STRIEGL, R. G., DUARTE, C. M., KORTELAINEN, P., DOWNING, J. A. & MIDDELBURG, J. J. 2007. Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. *Ecosystems,* 10**,** 172-185.
COLLETTE, T. W. & WILLIAMS, T. L. 2002. The role of raman spectroscopy in the analytical chemistry of potable water. *Journal of Environmental Monitoring,* 4**,** 27-34.
CORNISH, C. R., BRETHERTON, C. S. & PERCIVAL, D. B. 2006. Maximal overlap wavelet statistical analysis with application to atmospheric turbulence. *Boundary-Layer Meteorology,* 119**,** 339-374.
CORRADO, G., SANCHEZ-CORTES, S., FRANDOSO, O. & GARCIA-RAMOS, J. V. 2008. Surface-enhanced raman and fluorescence a joint analysis of soil humic acids. *Analytica Chimica Acta,* 616**,** 69-77.
COX, P. M., BETTS, R. A., JONES, C. D., SPALL, S. A. & TOTTERDELL, I. J. 2000. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model (vol 408, pg 184, 2000). *Nature,* 408**,** 750-750.
CROWLEY, P. M. 2007. A guide to wavelets for economists. *Journal of Economic Surveys,* 21**,** 207-267.
DANIEL, M. H., MONTEBELO, A. A., BERNARDES, M. C., OMETTO, J. P., DE CAMARGO, P. B., KRUSCHE, A. V., BALLESTER, M. V., VICTORIA, R. L. & MARTINELLI, L. A. 2002. Effects of urban sewage on dissolved oxygen, dissolved inorganic and organic carbon, and electrical conductivity of small streams along a gradient of urbanization in the piracicaba river basin. *Water, Air, and Soil Pollution,* 136**,** 189-206.
DARBRE, G. R. & WOLF, J. P. 1988. Criterion of stability and implementation issues of hybrid frequency-time-domain procedure for non-linear dynamic analysis. *Earthquake engineering & structural dynamics,* 16**,** 569-581.
DAVIDSON, E. A. & JANSSENS, I. A. 2006. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature,* 440**,** 165-173.
DAWSON, J., BILLETT, M., NEAL, C. & HILL, S. 2002. A comparison of particulate, dissolved and gaseous carbon in two contrasting upland streams in the uk. *Journal of Hydrology,* 257**,** 226-246.
DAWSON, J. J., BILLETT, M. F., HOPE, D., PALMER, S. M. & DEACON, C. M. 2004. Sources and sinks of aquatic carbon in a peatland stream continuum. *Biogeochemistry,* 70**,** 71-92.
DE LANGE, H. J., MORRIS, D. P. & WILLIAMSON, C. E. 2003. Solar ultraviolet photodegradation of doc may stimulate freshwater food webs. *Journal of Plankton Research,* 25**,** 111-117.
DE MONTETY, V., MARTIN, J., COHEN, M., FOSTER, C. & KURZ, M. 2011. Influence of diel biogeochemical cycles on carbonate equilibrium in a karst river. *Chemical Geology,* 283**,** 31-43.
DE PAOLIS, F. & KUKKONEN, J. 1997. Binding of organic pollutants to humic and fulvic acids: Influence of ph and the structure of humic material. *Chemosphere,* 34**,** 1693-1704.
DETWILER, R. P. & HALL, C. A. 1986. *Tropical forests and the global carbon cycle.* Cornell University.

DINSMORE, K., BILLETT, M. & DYSON, K. 2013. Temperature and precipitation drive temporal variability in aquatic carbon and ghg concentrations and fluxes in a peatland catchment. *Global change biology,* 19**,** 2133-2148.
DOGLIOLI, A., BLANKE, B., SPEICH, S. & LAPEYRE, G. 2007. Tracking coherent structures in a regional ocean model with wavelet analysis: Application to cape basin eddies. *Journal of Geophysical Research: Oceans,* 112.
DULKEITH, E., RINGLER, M., KLAR, T., FELDMANN, J., MUNOZ JAVIER, A. & PARAK, W. 2005. Gold nanoparticles quench fluorescence by phase induced radiative rate suppression. *Nano Letters,* 5**,** 585-589.
EBEL, B. A., LOAGUE, K. & BORJA, R. I. 2010. The impacts of hysteresis on variably saturated hydrologic response and slope failure. *Environmental Earth Sciences,* 61**,** 1215-1225.
ECKHARDT, B. & MOORE, T. 1990. Controls on dissolved organic carbon concentrations in streams, southern quebec. *Canadian Journal of Fisheries and Aquatic Sciences,* 47**,** 1537-1544.
ESTEVES, V. I., OTERO, M. & DUARTE, A. C. 2009. Comparative characterization of humic substances from the open ocean, estuarine water and fresh water. *Organic Geochemistry,* 40**,** 942-950.
EVANS, C. & MONTEITH, D. 2001. Chemical trends at lakes and streams in the uk acid waters monitoring network, 1988-2000: Evidence for recent recovery at a national scale. *Hydrology and Earth System Sciences Discussions,* 5**,** 351-366.
EVANS, C. D., CHAPMAN, P. J., CLARK, J. M., MONTEITH, D. T. & CRESSER, M. S. 2006. Alternative explanations for rising dissolved organic carbon export from organic soils. *Global Change Biology,* 12**,** 2044-2053.
EVANS, C. D., MONTEITH, D. T. & COOPER, D. M. 2005. Long-term increases in surface water dissolved organic carbon: Observations, possible causes and environmental impacts. *Environmental Pollution,* 137**,** 55-71.
EVANS, C. D., MONTEITH, D. T. & HARRIMAN, R. 2001. Long-term variability in the deposition of marine ions at west coast sites in the uk acid waters monitoring network: Impacts on surface water chemistry and significance for trend determination. *Science of the Total Environment,* 265**,** 115-129.
EVERALL, N., HAHN, T., MATOUSEK, P., PARKER, A. W. & TOWRIE, M. 2001. Picosecond time-resolved raman spectroscopy of solids: Capabilities and limitations for fluorescence rejection and the influence of diffuse reflectance. *Applied spectroscopy,* 55**,** 1701-1708.
FANTINI, C., JORIO, A., SOUZA, M., STRANO, M., DRESSELHAUS, M. & PIMENTA, M. 2004. Optical transition energies for carbon nanotubes from resonant raman spectroscopy: Environment and temperature effects. *Physical review letters,* 93**,** 147406.
FARQUHARSON, S., SMITH, W. W., CARANGELO, R. M. & BROUILLETTE, C. R. Industrial raman: Providing easy, immediate, cost-effective chemical analysis anywhere. Photonics East'99, 1999. International Society for Optics and Photonics, 14-25.
FASCHING, C. & BATTIN, T. J. 2012. Exposure of dissolved organic matter to uv-radiation increases bacterial growth efficiency in a clear-water alpine stream and its adjacent groundwater. *Aquatic Sciences,* 74**,** 143-153.
FAULDS, K., BARBAGALLO, R. P., KEER, J. T., SMITH, W. E. & GRAHAM, D. 2004. Serrs as a more sensitive technique for the detection of labelled oligonucleotides compared to fluorescence. *Analyst,* 129**,** 567-568.
FAULDS, K., MCKENZIE, F., SMITH, W. E. & GRAHAM, D. 2007. Quantitative simultaneous multianalyte detection of DNA by dual-wavelength surface-enhanced resonance raman scattering. *Angewandte Chemie,* 119**,** 1861-1863.
FÉLIDJ, N., AUBARD, J., LÉVI, G., KRENN, J., SALERNO, M., SCHIDER, G., LAMPRECHT, B., LEITNER, A. & AUSSENEGG, F. 2002. Controlling the optical response of regular arrays of gold particles for surface-enhanced raman scattering. *Physical Review B,* 65**,** 075419.

FELLMAN, J. B., D'AMORE, D. V. & HOOD, E. 2008. An evaluation of freezing as a preservation technique for analyzing dissolved organic c, n and p in surface water samples. *Science of the total environment,* 392**,** 305-312.
FELLMAN, J. B., HOOD, E. & SPENCER, R. G. 2010. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review. *Limnology and Oceanography,* 55**,** 2452-2462.
FENNER, N., FREEMAN, C., LOCK, M. A., HARMENS, H., REYNOLDS, B. & SPARKS, T. 2007. Interactions between elevated co2 and warming could amplify doc exports from peatland catchments. *Environmental science & technology,* 41**,** 3146-3152.
FENNER, N., WILLIAMS, R., TOBERMAN, H., HUGHES, S., REYNOLDS, B. & FREEMAN, C. 2011. Decomposition 'hotspots' in a rewetted peatland: Implications for water quality and carbon cycling. *Hydrobiologia,* 674**,** 51-66.
FERNANDA CARDINAL, M., RODRÍGUEZ-GONZÁLEZ, B., ALVAREZ-PUEBLA, R. A., PÉREZ-JUSTE, J. & LIZ-MARZÁN, L. M. 2010. Modulation of localized surface plasmons and sers response in gold dumbbells through silver coating. *The Journal of Physical Chemistry C,* 114**,** 10417-10423.
FERRARI, A. C. & BASKO, D. M. 2013. Raman spectroscopy as a versatile tool for studying the properties of graphene. *Nature nanotechnology,* 8**,** 235-246.
FERREIRA, E. C., FERREIRA, E. J., VILLAS-BOAS, P. R., SENESI, G. S., CARVALHO, C. M., ROMANO, R. A., MARTIN-NETO, L. & MILORI, D. M. B. P. 2014. Novel estimation of the humification degree of soil organic matter by laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy,* 99**,** 76-81.
FIEDLER, S., HÖLL, B., FREIBAUER, A., STAHR, K., DRÖSLER, M., SCHLOTER, M. & JUNGKUNST, H. 2008. Particulate organic carbon (poc) in relation to other pore water carbon fractions in drained and rewetted fens in southern germany. *Biogeosciences,* 5**,** 1615-1623.
FLEISCHMANN, M., HENDRA, P. & MCQUILLAN, A. 1974. Raman spectra of pyridine adsorbed at a silver electrode. *Chemical Physics Letters,* 26**,** 163-166.
FOVET, O., RUIZ, L., HRACHOWITZ, M., FAUCHEUX, M. & GASCUEL-ODOUX, C. 2015. Hydrological hysteresis and its value for assessing process consistency in catchment conceptual models.
FRAEDRICH, K., JIANG, J., GERSTENGARBE, F. W. & WERNER, P. C. 1997. Multiscale detection of abrupt climate changes: Application to river nile flood levels. *International Journal of Climatology,* 17**,** 1301-1315.
FRANCIOSO, O., SANCHEZ-CORTES, S., TUGNOLI, V., MARZADORI, C. & CIAVATTA, C. 2001. Spectroscopic study (drift, sers and h-1 nmr) of peat, leonardite and lignite humic substances. *Journal of Molecular Structure,* 565**,** 481-485.
FRANK, H., PATRICK, S., PETER, W. & HANNES, F. 2000. Export of dissolved organic carbon and nitrogen from gleysol dominated catchments–the significance of water flow paths. *Biogeochemistry,* 50**,** 137-161.
FRANK, O., MOHR, M., MAULTZSCH, J., THOMSEN, C., RIAZ, I., JALIL, R., NOVOSELOV, K. S., TSOUKLERI, G., PARTHENIOS, J. & PAPAGELIS, K. 2011. Raman 2d-band splitting in graphene: Theory and experiment. *ACS nano,* 5**,** 2231-2239.
FREEMAN, C., FENNER, N., OSTLE, N. J., KANG, H., DOWRICK, D. J., REYNOLDS, B., LOCK, M. A., SLEEP, D., HUGHES, S. & HUDSON, J. 2004. Export of dissolved organic carbon from peatlands under elevated carbon dioxide levels. *Nature,* 430**,** 195-198.
FRIMMEL, F. H. 1998. Characterization of natural organic matter as major constituents in aquatic systems. *Journal of Contaminant Hydrology,* 35**,** 201-216.
FURUKAWA, H., KO, N., GO, Y. B., ARATANI, N., CHOI, S. B., CHOI, E., YAZAYDIN, A. Ö., SNURR, R. Q., O'KEEFFE, M. & KIM, J. 2010. Ultrahigh porosity in metal-organic frameworks. *Science,* 329**,** 424-428.

GAJARAJ, S., FAN, C., LIN, M. & HU, Z. 2013. Quantitative detection of nitrate in water and wastewater by surface-enhanced raman spectroscopy. *Environmental Monitoring and Assessment,* 185**,** 5673-5681.
GARCÍA-BAQUERO, G., SILVERTOWN, J., GOWING, D. J. & VALLE, C. J. 2016. Dissecting the hydrological niche: Soil moisture, space and lifespan. *Journal of vegetation science,* 27**,** 219-226.
GARRIDO, J. M., MARQUES, M. P. M., SILVA, A. M., MACEDO, T. R., OLIVEIRA-BRETT, A. M. & BORGES, F. 2007. Spectroscopic and electrochemical studies of cocaine–opioid interactions. *Analytical and Bioanalytical Chemistry,* 388**,** 1799-1808.
GDEISAT, M. A., BURTON, D. R. & LALOR, M. J. 2006. Spatial carrier fringe pattern demodulation by use of a two-dimensional continuous wavelet transform. *Applied optics,* 45**,** 8722-8732.
GENÇAY, R., SELÇUK, F. & WHITCHER, B. J. 2001. *An introduction to wavelets and other filtering methods in finance and economics*, Academic press.
GORHAM, E. 1990. Biotic impoverishment in northern peatlands. *The earth in transition: patterns and processes of biotic impoverishment. Cambridge University Press, Cambridge, UK***,** 65-98.
GORHAM, E. 1991. Northern peatlands: Role in the carbon cycle and probable responses to climatic warming. *Ecological applications,* 1**,** 182-195.
GRAPS, A. 1995. An introduction to wavelets. *Ieee Computational Science & Engineering,* 2**,** 50-61.
GRAYSON, R. & HOLDEN, J. 2012. Continuous measurement of spectrophotometric absorbance in peatland streamwater in northern england: Implications for understanding fluvial carbon fluxes. *Hydrological Processes,* 26**,** 27-39.
GRIEVE, I. & GILVEAR, D. 2008. Effects of wind farm construction on concentrations and fluxes of dissolved organic carbon and suspended sediment from peat catchments at braes of doune, central scotland. *Mires and Peat,* 4.
GRINHUT, T., HADAR, Y. & CHEN, Y. 2007. Degradation and transformation of humic substances by saprotrophic fungi: Processes and mechanisms. *Fungal Biology Reviews,* 21**,** 179-189.
GRINSTED, A., MOORE, J. C. & JEVREJEVA, S. 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear processes in geophysics,* 11**,** 561-566.
GROSCH, S., MONAKHOVA, Y. B., KUBALLA, T., RUGE, W., KIMMICH, R. & LACHENMEIER, D. W. 2013. Comparison of gc/ms and nmr for quantification of methyleugenol in food. *European Food Research and Technology,* 236**,** 267-275.
GROSSMANN, A. & MORLET, J. 1984. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis,* 15**,** 723-736.
GU, Y. H. & BOLLEN, M. H. 2000. Time-frequency and time-scale domain analysis of voltage disturbances. *IEEE Transactions on Power Delivery,* 15**,** 1279-1284.
GUISAN, A., EDWARDS, T. C. & HASTIE, T. 2002. Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological modelling,* 157**,** 89-100.
HAEI, M., OQUIST, M. G., ILSTEDT, U. & LAUDON, H. 2012. The influence of soil frost on the quality of dissolved organic carbon in a boreal forest soil: Combining field and laboratory experiments. *Biogeochemistry,* 107**,** 95-106.
HAHN, D. W. 2007. Raman scattering theory. *Department of Mechanical and Aerospace Engineering, University of Florida*.
HAN, X. X., ZHAO, B. & OZAKI, Y. 2009. Surface-enhanced raman scattering for protein detection. *Analytical and bioanalytical chemistry,* 394**,** 1719-1727.
HANSON, R., NEWHOUSE, M. & DETTINGER, M. 2004. A methodology to asess relations between climatic variability and variations in hydrologic time series in the southwestern united states. *Journal of Hydrology,* 287**,** 252-269.
HARGREAVES, M. D., PAGE, K., MUNSHI, T., TOMSETT, R., LYNCH, G. & EDWARDS, H. G. 2008. Analysis of seized drugs using portable raman spectroscopy in an

airport environment—a proof of principle study. *Journal of Raman Spectroscopy,* 39**,** 873-880.
HART, J. K. & MARTINEZ, K. 2006. Environmental sensor networks: A revolution in the earth system science? *Earth-Science Reviews,* 78**,** 177-191.
HASTIE, T. & TIBSHIRANI, R. 1990. *Generalized additive models*, Wiley Online Library.
HAUTALA, K., PEURAVUORI, J. & PIHLAJA, K. 2000. Measurement of aquatic humus content by spectroscopic analyses. *Water Research,* 34**,** 246-258.
HEIGHTON, L. P. 2013. Modeling the ph dependent optical properties of aquatic, terrestrial and microbial humic substances.
HEIL, C. E. & WALNUT, D. F. 1989. Continuous and discrete wavelet transforms. *SIAM review,* 31**,** 628-666.
HERRERO-HERNÁNDEZ, E., MARÍN-BENITO, J., ANDRADES, M., SÁNCHEZ-MARTÍN, M. & RODRÍGUEZ-CRUZ, M. 2015. Field versus laboratory experiments to evaluate the fate of azoxystrobin in an amended vineyard soil. *Journal of environmental management,* 163**,** 78-86.
HERZSPRUNG, P., VON TUEMPLING, W., HERTKORN, N., HARIR, M., BUETTNER, O., BRAVIDOR, J., FRIESE, K. & SCHMITT-KOPPLIN, P. 2012. Variations of dom quality in inflows of a drinking water reservoir: Linking of van krevelen diagrams with eemf spectra by rank correlation. *Environmental Science & Technology,* 46**,** 5511-5518.
HILL, T. & NEAL, C. 1997. Spatial and temporal variation in ph, alkalinity and conductivity in surface runoff and groundwater for the upper river severn catchment. *Hydrology and Earth System Sciences Discussions,* 1**,** 697-715.
HOLDEN, J. & BURT, T. 2002. Infiltration, runoff and sediment production in blanket peat catchments: Implications of field rainfall simulation experiments. *Hydrological Processes,* 16**,** 2537-2557.
HONGVE, D. 1999. Production of dissolved organic carbon in forested catchments. *Journal of Hydrology,* 224**,** 91-99.
HOPE, D., BILLETT, M. F. & CRESSER, M. S. 1997. Exports of organic carbon in two river systems in ne scotland. *Journal of Hydrology,* 193**,** 61-82.
HOPE, D., KRATZ, T. K. & RIERA, J. L. 1996. Relationship between pco2 and dissolved organic carbon in northern wisconsin lakes. *Journal of environmental quality,* 25**,** 1442-1445.
HORIBASCIENTIFIC. *Raman data and analysis* [Online]. Available: http://www.horiba.com/fileadmin/uploads/Scientific/Documents/Raman/bands.pdf [Accessed 7 June 2016 2016].
HUANG, N. E., SHEN, Z., LONG, S. R., WU, M. C., SHIH, H. H., ZHENG, Q., YEN, N.-C., TUNG, C. C. & LIU, H. H. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 1998. The Royal Society, 903-995.
HUDSON, N., BAKER, A. & REYNOLDS, D. 2007. Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters - a review. *River Research and Applications,* 23**,** 631-649.
HUH, Y. S., CHUNG, A. J. & ERICKSON, D. 2009. Surface enhanced raman spectroscopy and its application to molecular and cellular analysis. *Microfluidics and nanofluidics,* 6**,** 285-297.
HUTCHINSON, I. B., INGLEY, R., EDWARDS, H. G., HARRIS, L., MCHUGH, M., MALHERBE, C. & PARNELL, J. 2014. Raman spectroscopy on mars: Identification of geological and bio-geological signatures in martian analogues using miniaturized raman spectrometers. *Phil. Trans. R. Soc. A,* 372**,** 20140204.
JAMESHUTTONINSTITUTE 2014. Whitelee catchment. 1.0 ed.
JEBB, A. T., TAY, L., WANG, W. & HUANG, Q. 2015. Time series analysis for psychological research: Examining and forecasting change. *Frontiers in psychology,* 6.
JENNINGS, E., JÄRVINEN, M., ALLOTT, N., ARVOLA, L., MOORE, K., NADEN, P., AONGHUSA, C. N., NÕGES, T. & WEYHENMEYER, G. A. 2009. Impacts of

climate on the flux of dissolved organic carbon from catchments. *The impact of climate change on european lakes.* Springer.
JEONG, J. J., BARTSCH, S., FLECKENSTEIN, J. H., MATZNER, E., TENHUNEN, J. D., LEE, S. D., PARK, S. K. & PARK, J. H. 2012. Differential storm responses of dissolved and particulate organic carbon in a mountainous headwater stream, investigated by high-frequency, in situ optical measurements. *Journal of Geophysical Research: Biogeosciences,* 117.
JEVREJEVA, S., MOORE, J. & GRINSTED, A. 2003. Influence of the arctic oscillation and el niño-southern oscillation (enso) on ice conditions in the baltic sea: The wavelet approach. *Journal of Geophysical Research: Atmospheres,* 108.
JOHNSON, M. S., COUTO, E. G., ABDO, M. & LEHMANN, J. 2011. Fluorescence index as an indicator of dissolved organic carbon quality in hydrologic flowpaths of forested tropical watersheds. *Biogeochemistry,* 105**,** 149-157.
JONATHAN H, S. 1973. Total organic carbon in seawater — comparison of measurements using persulfate oxidation and high temperature combustion. *Marine Chemistry,* 1**,** 211-229.
JONES, T. D., CHAPPELL, N. A. & TYCH, W. 2014. First dynamic model of dissolved organic carbon derived directly from high-frequency observations through contiguous storms. *Environmental science & technology,* 48**,** 13289-13297.
JORGE-ARAÚJO, P., QUIQUAMPOIX, H., MATUMOTO-PINTRO, P. & STAUNTON, S. 2015. Glomalin-related soil protein in french temperate forest soils: Interference in the bradford assay caused by co-extracted humic substances. *European Journal of Soil Science,* 66**,** 311-319.
JORGENSEN, C. B. 1976. Putter,a, krough,a and modern ideas on use of dissolved organic-matter in aquatic environments. *Biological Reviews of the Cambridge Philosophical Society,* 51**,** 291-328.
KAISER, K., GUGGENBERGER, G., HAUMAIER, L. & ZECH, W. 2001. Seasonal variations in the chemical composition of dissolved organic matter in organic forest floor layer leachates of old-growth scots pine (pinus sylvestris l.) and european beech (fagus sylvatica l.) stands in northeastern bavaria, germany. *Biogeochemistry,* 55**,** 103-143.
KAISLER, S., ARMOUR, F., ESPINOSA, J. A. & MONEY, W. Big data: Issues and challenges moving forward. System Sciences (HICSS), 2013 46th Hawaii International Conference on, 2013. IEEE, 995-1004.
KALBITZ, K. & GEYER, W. 2001. Humification indices of water-soluble fulvic acids derived from synchronous fluorescence spectra—effects of spectrometer type and concentration. *Journal of Plant Nutrition and Soil Science,* 164**,** 259-265.
KALBITZ, K., SCHMERWITZ, J., SCHWESIG, D. & MATZNER, E. 2003. Biodegradation of soil-derived dissolved organic matter as related to its properties. *Geoderma,* 113**,** 273-291.
KALLACHE, M., RUST, H. & KROPP, J. 2005. Trend assessment: Applications for hydrology and climate research. *Nonlinear Processes in Geophysics,* 12**,** 201-210.
KANG, L., XU, P., CHEN, D., ZHANG, B., DU, Y., HAN, X., LI, Q. & WANG, H.-L. 2013. Amino acid-assisted synthesis of hierarchical silver microspheres for single particle surface-enhanced raman spectroscopy. *The Journal of Physical Chemistry C,* 117**,** 10007-10012.
KARAMEHMEDOVIC, E. 2006. Incoherent optical frequency domain reflectometry for distributed thermal sensing. *Technical University of Denmark, Department of Electromagnetic System*.
KAYRANLI, B., SCHOLZ, M., MUSTAFA, A. & HEDMARK, A. 2010. Carbon storage and fluxes within freshwater wetlands: A critical review. *Wetlands,* 30**,** 111-124.
KEERY, J., BINLEY, A., CROOK, N. & SMITH, J. W. 2007. Temporal and spatial variability of groundwater–surface water fluxes: Development and application of an analytical method using temperature time series. *Journal of Hydrology,* 336**,** 1-16.
KEIM, D. A. 2002. Information visualization and visual data mining. *IEEE transactions on Visualization and Computer Graphics,* 8**,** 1-8.

KEIM, D. A., MANSMANN, F., SCHNEIDEWIND, J. & ZIEGLER, H. Challenges in visual data analysis. Tenth International Conference on Information Visualisation (IV'06), 2006. IEEE, 9-16.
KENNEDY, J., BILLETT, M. F., DUTHIE, D., FRASER, A. R. & HARRISON, A. F. 1996. Organic matter retention in an upland humic podzol; the effects of ph and solute type. *European Journal of Soil Science,* 47**,** 615-625.
KEREN, S., ZAVALETA, C., CHENG, Z., DE LA ZERDA, A., GHEYSENS, O. & GAMBHIR, S. 2008. Noninvasive molecular imaging of small living subjects using raman spectroscopy. *Proceedings of the National Academy of Sciences,* 105**,** 5844-5849.
KERNAN, M., BATTARBEE, R., CURTIS, C., MONTEITH, D. & SHILLAND, E. 2010. Uk acid waters monitoring network 20 year interpretative report.
KIM, S. K., KIM, M. S. & SUH, S. W. 1987. Surface-enhanced raman scattering (sers) of aromatic amino acids and their glycyl dipeptides in silver sol. *Journal of Raman spectroscopy,* 18**,** 171-175.
KINGSBURY, N. The dual-tree complex wavelet transform: A new efficient tool for image restoration and enhancement. Signal Processing Conference (EUSIPCO 1998), 9th European, 1998. IEEE, 1-4.
KIRCHNER, J. W., FENG, X. H., NEAL, C. & ROBSON, A. J. 2004. The fine structure of water-quality dynamics: The (high-frequency) wave of the future. *Hydrological Processes,* 18**,** 1353-1359.
KITAGAWA, T. & NOMURA, T. 2003. A wavelet-based method to generate artificial wind fluctuation data. *Journal of Wind Engineering and Industrial Aerodynamics,* 91**,** 943-964.
KITIS, M., KARANFIL, T., KILDUFF, J. E. & WIGTON, A. 2001. The reactivity of natural organic matter to disinfection byproducts formation and its relation to specific ultraviolet absorbance. *Water Science and Technology,* 43**,** 9-16.
KITIS, M., KARANFIL, T., WIGTON, A. & KILDUFF, J. E. 2002. Probing reactivity of dissolved organic matter for disinfection by-product formation using xad-8 resin adsorption and ultrafiltration fractionation. *Water Research,* 36**,** 3834-3848.
KNEIPP, K., KNEIPP, H., ITZKAN, I., DASARI, R. R. & FELD, M. S. 1999. Ultrasensitive chemical analysis by raman spectroscopy. *Chemical Reviews,* 99**,** 2957-2976.
KOEHLER, A.-K., MURPHY, K., KIELY, G. & SOTTOCORNOLA, M. 2009. Seasonal variation of doc concentration and annual loss of doc from an atlantic blanket bog in south western ireland. *Biogeochemistry,* 95**,** 231-242.
KONONOVA, M. A. M. 1966. Soil organic matter; its nature, its role in soil formation and in soil fertility. Pergamon Press.
KOWALCZUK, P., ZABLOCKA, M., SAGAN, S. & KULINSKI, K. 2010. Fluorescence measured in situ as a proxy of cdom absorption and doc concentration in the baltic sea. *Oceanologia,* 52**,** 431-471.
KRENO, L. E., GREENELTCH, N. G., FARHA, O. K., HUPP, J. T. & VAN DUYNE, R. P. 2014. Sers of molecules that do not adsorb on ag surfaces: A metal–organic framework-based functionalization strategy. *Analyst,* 139**,** 4073-4080.
KROGH, A. & KEYS, A. 1934. Methods for the determination of dissolved organic carbon and nitrogen in sea water. *Biological Bulletin,* 67**,** 132-144.
KUMAR, D., SINGH, B., BAUDDH, K. & KORSTAD, J. 2015. Bio-oil and biodiesel as biofuels derived from microalgal oil and their characterization by using instrumental techniques. *Algae and environmental sustainability.* Springer.
KUMAR, P. & FOUFOULA-GEORGIOU, E. 1997. Wavelet analysis for geophysical applications. *Reviews of geophysics,* 35**,** 385-412.
KUWATSUKA, S., TSUTSUKI, K. & KUMADA, K. 1978. Chemical studies on soil humic acids: 1. Elementary composition of humic acids. *Soil Science and Plant Nutrition,* 24**,** 337-347.
LACHAUX, J. P., LUTZ, A., RUDRAUF, D., COSMELLI, D., LE VAN QUYEN, M., MARTINERIE, J. & VARELA, F. 2002. Estimating the time-course of coherence between single-trial brain signals: An introduction to wavelet coherence. *Neurophysiologie Clinique-Clinical Neurophysiology,* 32**,** 157-174.

LAKOWICZ, J. R. 2013. *Principles of fluorescence spectroscopy*, Springer Science & Business Media.
LAKSHMANAN, M. K. & NIKOOKAR, H. 2006. A review of wavelets for digital wireless communication. *Wireless Personal Communications,* 37**,** 387-420.
LANE, S. N. 2007. Assessment of rainfall-runoff models based upon wavelet analysis. *Hydrological processes,* 21**,** 586-607.
LANGERGRABER, G., FLEISCHMANN, N. & HOFSTADTER, F. 2003. A multivariate calibration procedure for uv/vis spectrometric quantification of organic matter and nitrate in wastewater. *Water Science and Technology,* 47**,** 63-71.
LARK, R. 2006. The representation of complex soil variation on wavelet packet bases. *European journal of soil science,* 57**,** 868-882.
LEENHEER, J. A. 1981. Comprehensive approach to preparative isolation and fractionation of dissolved organic-carbon from natural-waters and wastewaters. *Environmental Science & Technology,* 15**,** 578-587.
LEENHEER, J. A. & CROUÉ, J.-P. 2003. Peer reviewed: Characterizing aquatic dissolved organic matter. *Environmental science & technology,* 37**,** 18A-26A.
LEHAITRE, M., DELAUNEY, L. & COMPÈRE, C. 2008. Biofouling and underwater measurements. *Real-time observation systems for ecosystem dynamics and harmful algal blooms: Theory, instrumentation and modelling. Oceanographic Methodology Series. UNESCO, Paris***,** 463-493.
LEYTON, P., LIZAMA-VERGARA, P., CAMPOS-VALLETTE, M., BECKER, M. I., CLAVIJO, E., CORDOVA REYES, I., VERA, M. & JEREZ, C. A. 2005. Surface enhanced raman spectrum of nanometric molecular systems. *Journal of the Chilean Chemical Society,* 50**,** 725-730.
LI, C. W., KORSHIN, G. V. & BENJAMIN, M. M. 1998. Monitoring dbp formation with differential uv spectroscopy. *Journal American Water Works Association,* 90**,** 88-100.
LIMPENS, J., BERENDSE, F., BLODAU, C., CANADELL, J. G., FREEMAN, C., HOLDEN, J., ROULET, N., RYDIN, H. & SCHAEPMAN-STRUB, G. 2008. Peatlands and the carbon cycle: From local processes to global implications - a synthesis. *Biogeosciences,* 5**,** 1475-1491.
LIU, C.-H., DAS, B., GLASSMAN, W. S., TANG, G., YOO, K., ZHU, H., AKINS, D., LUBICZ, S., CLEARY, J. & PRUDENTE, R. 1992. Raman, fluorescence, and time-resolved light scattering as optical diagnostic techniques to separate diseased and normal biomedical media. *Journal of Photochemistry and Photobiology B: Biology,* 16**,** 187-209.
LIU, X., WANG, Z. A., BYRNE, R. H., KALTENBACHER, E. A. & BERNSTEIN, R. E. 2006. Spectrophotometric measurements of ph in-situ: Laboratory and field evaluations of instrumental performance. *Environmental science & technology,* 40**,** 5036-5044.
LIZ-MARZÁN, L. M. 2006. Tailoring surface plasmons through the morphology and assembly of metal nanoparticles. *Langmuir,* 22**,** 32-41.
LONG, H., VIHERMAA, L., WALDRON, S., HOEY, T., QUEMIN, S. & NEWTON, J. 2015. Hydraulics are a first-order control on co2 efflux from fluvial systems. *Journal of Geophysical Research: Biogeosciences,* 120**,** 1912-1922.
LUTZ, B. D., BERNHARDT, E. S., ROBERTS, B. J., CORY, R. M. & MULHOLLAND, P. J. 2012. Distinguishing dynamics of dissolved organic matter components in a forested stream using kinetic enrichments. *Limnology and Oceanography,* 57**,** 76-89.
MA, J., DEL VECCHIO, R., GOLANOSKI, K. S., BOYLE, E. S. & BLOUGH, N. V. 2010. Optical properties of humic substances and cdom: Effects of borohydride reduction. *Environmental science & technology,* 44**,** 5395-5402.
MALCOLM, R. L. & MACCARTHY, P. 1992. Quantitative-evaluation of xad-8 and xad-4 resins used in tandem for removing organic solutes from water. *Environment International,* 18**,** 597-607.
MALLAT, S. & HWANG, W. L. 1992. Singularity detection and processing with wavelets. *IEEE transactions on information theory,* 38**,** 617-643.

MARSHALL, A. G. & VERDUN, F. R. 2016. *Fourier transforms in nmr, optical, and mass spectrometry: A user's handbook*, Elsevier.
MARTINI, I. P., CORTIZAS, A. M. & CHESWORTH, W. 2007. *Peatlands: Evolution and records of environmental and climate changes*, Elsevier.
MARTYSHKIN, D. V., AHUJA, R. C., KUDRIAVTSEV, A. & MIROV, S. B. 2004. Effective suppression of fluorescence light in raman measurements using ultrafast time gated charge coupled device camera. *Review of Scientific Instruments,* 75**,** 630-635.
MATHLOUTHI, M. & LUU, D. V. 1980. Laser-raman spectra of d-glucose and sucrose in aqueous solution. *Carbohydrate Research,* 81**,** 203-212.
MATTHEWS, B. J. H., JONES, A. C., THEODOROU, N. K. & TUDHOPE, A. W. 1996. Excitation-emission-matrix fluorescence spectroscopy applied to humic acid bands in coral reefs. *Marine Chemistry,* 55**,** 317-332.
MCCARTHY, J. F., CZERWINSKI, K. R., SANFORD, W. E., JARDINE, P. M. & MARSH, J. D. 1998. Mobilization of transuranic radionuclides from disposal trenches by natural organic matter. *Journal of Contaminant Hydrology,* 30**,** 49-77.
MCCREERY, R. L. 2005. *Raman spectroscopy for chemical analysis*, John Wiley & Sons.
MCKNIGHT, D. M., BENCALA, K. E., ZELLWEGER, G. W., AIKEN, G. R., FEDER, G. L. & THORN, K. A. 1992. Sorption of dissolved organic carbon by hydrous aluminum and iron oxides occurring at the confluence of deer creek with the snake river, summit county, colorado. *Environmental Science & Technology,* 26**,** 1388-1396.
MCKNIGHT, D. M., BOYER, E. W., WESTERHOFF, P. K., DORAN, P. T., KULBE, T. & ANDERSEN, D. T. 2001. Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnology and Oceanography,* 46**,** 38-48.
MENZEL, D. W. & VACCARO, R. F. 1964. The measurement of dissolved organic and particulate carbon in seawater. *Limnology and Oceanography,* 9**,** 138-142.
MERLIN, J. C. 1985. Resonance raman spectroscopy of carotenoids and carotenoid-containing systems. *Pure and Applied Chemistry,* 57**,** 785-792.
MERRY, R. & STEINBUCH, M. 2005. Wavelet theory and applications. *Literature Study, Eindhoven University of Technology, Department of Mechanical Engineering, Control Systems Technology Group*.
MIANO, T. & SENESI, N. 1992. Synchronous excitation fluorescence spectroscopy applied to soil humic substances chemistry. *Science of the total environment,* 117**,** 41-51.
MOBED, J. J., HEMMINGSEN, S. L., AUTRY, J. L. & MCGOWN, L. B. 1996. Fluorescence characterization of ihss humic substances: Total luminescence spectra with absorbance correction. *Environmental Science & Technology,* 30**,** 3061-3065.
MONTEITH, D. T., STODDARD, J. L., EVANS, C. D., DE WIT, H. A., FORSIUS, M., HOGASEN, T., WILANDER, A., SKJELKVALE, B. L., JEFFRIES, D. S., VUORENMAA, J., KELLER, B., KOPACEK, J. & VESELY, J. 2007. Dissolved organic carbon trends resulting from changes in atmospheric deposition chemistry. *Nature,* 450**,** 537-U9.
MOORE, S., EVANS, C. D., PAGE, S. E., GARNETT, M. H., JONES, T. G., FREEMAN, C., HOOIJER, A., WILTSHIRE, A. J., LIMIN, S. H. & GAUCI, V. 2013. Deep instability of deforested tropical peatlands revealed by fluvial organic carbon fluxes. *Nature,* 493**,** 660-+.
MOPPER, K. & SCHULTZ, C. A. 1993. Fluorescence as a possible tool for studying the nature and water column distribution of doc components. *Marine Chemistry,* 41**,** 229-238.
MOPPER, K., ZHOU, X., KIEBER, R. J., KIEBER, D. J., SIKORSKI, R. J. & JONES, R. D. 1991. Photochemical degradation of dissolved organic carbon and its impact on the oceanic carbon cycle. *Nature,* 353**,** 60-62.
MORAN, M. A. & HODSON, R. E. 1990. Bacterial production on humic and nonhumic components of dissolved organic-carbon. *Limnology and Oceanography,* 35**,** 1744-1756.

MORAN, M. A. & ZEPP, R. G. 1997. Invited review role of photoreactions in the formation of biologically labile compounds from dissolved organic matter. *Oceanography,* 42.
MORLET, J., ARENS, G., FOURGEAU, E. & GLARD, D. 1982. Wave propagation and sampling theory-part i: Complex signal and scattering in multilayered media. *Geophysics,* 47**,** 203-221.
MORRIS, D. P., ZAGARESE, H., WILLIAMSON, C. E., BALSEIRO, E. G., HARGREAVES, B. R., MODENUTTI, B., MOELLER, R. & QUEIMALINOS, C. 1995. The attentuation of solar uv radiation in lakes and the role of dissolved organic carbon. *Limnology and Oceanography,* 40**,** 1381-1391.
MULLER, E., GUNNEMANN, S., FARBER, I. & SEIDL, T. Discovering multiple clustering solutions: Grouping objects in different views of the data. 2012 IEEE 28th International Conference on Data Engineering, 2012. IEEE, 1207-1210.
MULLER, F. L. & TANKÉRÉ-MULLER, S. P. 2012. Seasonal variations in surface water chemistry at disturbed and pristine peatland sites in the flow country of northern scotland. *Science of the Total Environment,* 435**,** 351-362.
MÜLLER, R. A., KOTHAWALA, D. N., PODGRAJSEK, E., SAHLÉE, E., KÖHLER, B., TRANVIK, L. J. & WEYHENMEYER, G. A. 2014. Hourly, daily, and seasonal variability in the absorption spectra of chromophoric dissolved organic matter in a eutrophic, humic lake. *Journal of Geophysical Research: Biogeosciences,* 119**,** 1985-1998.
MUNCH, J. M., TOTSCHE, K. U. & KAISER, K. 2002. Physicochemical factors controlling the release of dissolved organic carbon from columns of forest subsoils. *European Journal of Soil Science,* 53**,** 311-320.
MURRAY, H. 2012. *Assessing the impact of wind-farm related disturbance on streamwater carbon, phosphorus and nitrogen dynamics: A case study of the whitelee catchments. .* University of Glasgow.
NAKKEN, M. 1999. Wavelet analysis of rainfall–runoff variability isolating climatic from anthropogenic patterns. *Environmental Modelling & Software,* 14**,** 283-295.
NANDA, S. S., KIM, M. J., YEOM, K. S., AN, S. S. A., JU, H. & YI, D. K. 2016. Raman spectrum of graphene with its versatile future perspectives. *TrAC Trends in Analytical Chemistry,* 80**,** 125-131.
NASON, G. 2006a. Stationary and non-stationary time series. *Statistics in Volcanology. Special Publications of IAVCEI,* 1**,** 000-000.
NASON, G. P. 2006b. Stationary and non-stationary time series. *In:* MADER H, C. S. C. (ed.) *Statistics in volcanology. Special publications of iavcei.* The Geological Society.
NAYAK, D., MILLER, D., NOLAN, A., SMITH, P. & SMITH, J. 2010. Calculating carbon budgets of wind farms on scottish peatlands. *Mires and Peat,* 4**,** 1-23.
NAYAK, D. R., MILLER, D., NOLAN, A., SMITH, P. & SMITH, J. 2008. Calculating carbon savings from wind farms on scottish peat lands–a new approach. *Final Report (June)***,** 1-83.
NEAL, C., HILL, T., ALEXANDER, S., REYNOLDS, B., HILL, S., DIXON, A. J., HARROW, M., NEAL, M. & SMITH, C. J. 1997. Stream water quality in acid sensitive uk upland areas; an example of potential water quality remediation based on groundwater manipulation. *Hydrology and Earth System Sciences Discussions,* 1**,** 185-196.
NEAL, C., REYNOLDS, B., ROWLAND, P., NORRIS, D., KIRCHNER, J. W., NEAL, M., SLEEP, D., LAWLOR, A., WOODS, C., THACKER, S., GUYATT, H., VINCENT, C., HOCKENHULL, K., WICKHAM, H., HARMAN, S. & ARMSTRONG, L. 2012. High-frequency water quality time series in precipitation and streamflow: From fragmentary signals to scientific challenge. *Science of the Total Environment,* 434**,** 3-12.
NEAL, C. & ROBSON, A. J. 2000. A summary of river water quality data collected within the land-ocean interaction study: Core data for eastern uk rivers draining to the north sea. *Science of the Total Environment,* 251**,** 585-665.
NELSON, D. W. & SOMMERS, L. E. 1996. Total carbon, organic carbon, and organic matter. *Methods of soil analysis part 3—chemical methods***,** 961-1010.

NIE, S. & EMORY, S. R. 1997. Probing single molecules and single nanoparticles by surface-enhanced raman scattering. *science,* 275**,** 1102-1106.
OGAWA, H. & OGURA, N. 1992. Comparison of 2 methods for measuring dissolved organic-carbon in sea-water. *Nature,* 356**,** 696-698.
OSBURN, C. L. & ST-JEAN, G. 2007. The use of wet chemical oxidation with high-amplification isotope ratio mass spectrometry (wco-irms) to measure stable isotope values of dissolved organic carbon in seawater. *Limnology and Oceanography-Methods,* 5**,** 296-308.
PAGE, S. E., SIEGERT, F., RIELEY, J. O., BOEHM, H.-D. V., JAYA, A. & LIMIN, S. 2002. The amount of carbon released from peat and forest fires in indonesia during 1997. *Nature,* 420**,** 61-65.
PATTERSON, H., CRONAN, C., LAKSHMAN, S., PLANKEY, B. & TAYLOR, T. 1992. Comparison of soil fulvic acids using synchronous scan fluorescence spectroscopy, ftir, titration and metal complexation kinetics. *Science of the total environment,* 113**,** 179-196.
PAWSON, R., LORD, D., EVANS, M. & ALLOTT, T. 2008. Fluvial organic carbon flux from an eroding peatland catchment, southern pennines, uk. *Hydrology and Earth System Sciences Discussions,* 12**,** 625-634.
PEARCE-HIGGINS, J., STEPHEN, L., LANGSTON, R. & BRIGHT, J. 2008. Assessing the cumulative impacts of wind farms on peatland birds: A case study of golden plover pluvialis apricaria in scotland. *Mires and Peat,* 4**,** 1-13.
PERCIVAL, D. B. & MOFJELD, H. O. 1997. Analysis of subtidal coastal sea level fluctuations using wavelets. *Journal of the American Statistical Association,* 92**,** 868-880.
PERCIVAL, D. B. & WALDEN, A. T. 1993. *Spectral analysis for physical applications*, Cambridge University Press.
PERCIVAL, D. B. & WALDEN, A. T. 2006. *Wavelet methods for time series analysis*, Cambridge university press.
PEURAVUORI, J., LEHTONEN, T. & PIHLAJA, K. 2002. Sorption of aquatic humic matter by dax-8 and xad-8 resins - comparative study using pyrolysis gas chromatography. *Analytica Chimica Acta,* 471**,** 219-226.
PEURAVUORI, J. & PIHLAJA, K. 1998. Multimethod characterization of lake aquatic humic matter isolated with sorbing solid and tangential membrane filtration. *Analytica chimica acta,* 364**,** 203-221.
PICCOLO, A. 2002. The supramolecular structure of humic substances: A novel understanding of humus chemistry and implications in soil science. *Advances in Agronomy,* 75**,** 57-134.
PICKUP, J. C., HUSSAIN, F., EVANS, N. D., ROLINSKI, O. J. & BIRCH, D. J. 2005. Fluorescence-based glucose sensors. *Biosensors and Bioelectronics,* 20**,** 2555-2565.
PODSTAWKA, E., OZAKI, Y. & PRONIEWICZ, L. M. 2004. Part i: Surface-enhanced raman spectroscopy investigation of amino acids and their homodipeptides adsorbed on colloidal silver. *Applied spectroscopy,* 58**,** 570-580.
POLANSKY, L., WITTEMYER, G., CROSS, P. C., TAMBLING, C. J. & GETZ, W. M. 2010. From moonlight to movement and synchronized randomness: Fourier and wavelet analyses of animal location time series data. *Ecology,* 91**,** 1506-1518.
PORTERFIELD, D. R. & CAMPION, A. 1988. Fluorescence-free scanning raman spectroscopy. *Journal of the American Chemical Society,* 110**,** 408-410.
PROVENZANO, M. R., GIGLIOTTI, G., CILENTI, A., ERRIQUENS, F. & SENESI, N. 2006. Spectroscopic and thermal investigation of hydrophobic and hydrophilic fractions of dissolved organic matter. *Compost Science & Utilization,* 14**,** 191-200.
QIN, J., CHAO, K. & KIM, M. 2010. Raman chemical imaging system for food safety and quality inspection. *Transactions of the ASABE,* 53**,** 1873-1882.
RAMAN, C. V. & KRISHNAN, K. S. 1928. A new type of secondary radiation. *Nature,* 121**,** 501-502.
RAYMOND, P. A. & SAIERS, J. E. 2010. Event controlled doc export from forested watersheds. *Biogeochemistry,* 100**,** 197-209.

RIEGER, L., LANGERGRABER, G., THOMANN, M., FLEISCHMANN, N. & SIEGRIST, H. 2004. Spectral in-situ analysis of no2, no3, cod, doc and tss in the effluent of a wwtp. *Water Science and Technology,* 50**,** 143-152.
RINKE, K., KUEHN, B., BOCANIOV, S., WENDT-POTTHOFF, K., BÜTTNER, O., TITTEL, J., SCHULTZE, M., HERZSPRUNG, P., RÖNICKE, H. & RINK, K. 2013. Reservoirs as sentinels of catchments: The rappbode reservoir observatory (harz mountains, germany). *Environmental earth sciences,* 69**,** 523-536.
RIOUL, O. & DUHAMEL, P. 1992. Fast algorithms for discrete and continuous wavelet transforms. *IEEE transactions on information theory,* 38**,** 569-586.
RIVERA, D., ARUMÍ, M. L. Y. J. & DEMİREL, M. C. Wavelets transform applied to hydrological time series.
ROBERTSON, R. P. & HARMON, J. S. 2006. Diabetes, glucose toxicity, and oxidative stress: A case of double jeopardy for the pancreatic islet β cell. *Free Radical Biology and Medicine,* 41**,** 177-184.
RODRÍGUEZ, F. J., SCHLENGER, P. & GARCÍA-VALVERDE, M. 2016. Monitoring changes in the structure and properties of humic substances following ozonation using uv–vis, ftir and 1 h nmr techniques. *Science of the Total Environment,* 541**,** 623-637.
ROLDÁN, M. L., CORRADO, G., FRANCIOSO, O. & SÁNCHEZ-CORTÉS, S. 2011. Interaction of soil humic acids with herbicide paraquat analyzed by surface-enhanced raman scattering and fluorescence spectroscopy on silver plasmonic nanoparticles. *Analytica Chimica Acta,* 699**,** 87-95.
ROULET, N. & MOORE, T. R. 2006. Environmental chemistry: Browning the waters. *Nature,* 444**,** 283-284.
ROUTH, J., GROSSMAN, E. L., MURPHY, E. M. & BENNER, R. 2001. Characterization and origin of dissolved organic carbon in yegua ground water in brazos county, texas. *Ground Water,* 39**,** 760-767.
RRUFF. *Rruffdatabase* [Online]. Available: http://rruff.info/ [Accessed 20 August 2016 2016].
RUA, A. 2012. Wavelets in economics. *Economic Bulletin and Financial Stability Report Articles***,** 71-79.
RUCKA, M. & WILDE, K. 2006. Application of continuous wavelet transform in vibration based damage detection method for beams and plates. *Journal of Sound and Vibration,* 297**,** 536-550.
RYCENGA, M., HOU, K. K., COBLEY, C. M., SCHWARTZ, A. G., CAMARGO, P. H. & XIA, Y. 2009. Probing the surface-enhanced raman scattering properties of au–ag nanocages at two different excitation wavelengths. *Physical Chemistry Chemical Physics,* 11**,** 5903-5908.
SAADI, I., BORISOVER, M., ARMON, R. & LAOR, Y. 2006. Monitoring of effluent dom biodegradation using fluorescence, uv and doc measurements. *Chemosphere,* 63**,** 530-539.
SALAMON, Z., MACLEOD, H. A. & TOLLIN, G. 1997. Surface plasmon resonance spectroscopy as a tool for investigating the biochemical and biophysical properties of membrane protein systems. Ii: Applications to biological systems. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes,* 1331**,** 131-152.
SANCHEZ-CORTES, S., CORRADO, G., TRUBETSKAYA, O. E., TRUBETSKOJ, O. A., HERMOSIN, B. & SAIZ-JIMENEZ, C. 2006. Surface-enhanced raman spectroscopy of chernozem humic acid and their fractions obtained by coupled size exclusion chromatography-polyacrylamide gel electrophoresis (sec-page). *Applied Spectroscopy,* 60**,** 48-53.
SANDFORD, R. C., BOL, R. & WORSFOLD, P. J. 2010. In situ determination of dissolved organic carbon in freshwaters using a reagentless uv sensor. *Journal of Environmental Monitoring,* 12**,** 1678-1683.
SCHAEFLI, B., MARAUN, D. & HOLSCHNEIDER, M. 2007. What drives high flow events in the swiss alps? Recent developments in wavelet spectral analysis and their application to hydrology. *Advances in Water Resources,* 30**,** 2511-2525.

SCHIFF, S., ARAVENA, R., TRUMBORE, S. E. & DILLON, P. 1990. Dissolved organic carbon cycling in forested watersheds: A carbon isotope approach. *Water Resources Research,* 26**,** 2949-2957.
SCHIMEL, D. S. 1995. Terrestrial ecosystems and the carbon cycle. *Global change biology,* 1**,** 77-91.
SCHLEICHER, C. 2002. *An introduction to wavelets for economists*, Bank of Canada.
SCHLESINGER, W. H. 1977. Carbon balance in terrestrial detritus. *Annual review of ecology and systematics***,** 51-81.
SCHUMACHER, B. A. 2002. Methods for the determination of total organic carbon (toc) in soils and sediments.
SCOTTISHPOWER 2002. Cre energy: Whitelee windfarm environmental statement, chapters 1, 4, 5, 18.
SCOTTISHPOWER. 2008. Two golden eagle chicks hatched at scottishpower renewables' windfarm site. *Press Release - Scottish Power*, p.1.
SCOTTISHPOWER. 2009. Switch-on at europe's largest onshore windfarm powering 180,000 homes, as planning consent is granted to power a further 70,000. *Press Release - Scottish Power*, p.1.
SCOTTISHPOWER. 2012. Europe's largest wind farm sets another record as new turbines at whitelee extension are connected to grid. *Press Release - Scottish Power*, p.1.
SCOTTISHPOWER. 2013. Scottishpower chairman ignacio galán calls for greater clarity on uk energy reform. *Press Release - Scottish Power*, p.1.
SEEGER, M., ERREA, M.-P., BEGUERıA, S., ARNÁEZ, J., MARTı, C. & GARCıA-RUIZ, J. 2004. Catchment soil moisture and rainfall characteristics as determinant factors for discharge/suspended sediment hysteretic loops in a small headwater catchment in the spanish pyrenees. *Journal of Hydrology,* 288**,** 299-311.
SENESI, N., MIANO, T. M., PROVENZANO, M. R. & BRUNETTI, G. 1991. Characterization, differentiation, and classification of humic substances by fluorescence spectroscopy. *Soil Science,* 152**,** 259-271.
SERROUKH, A., WALDEN, A. & PERCIVAL, D. B. 2000. Statistical properties and uses of the wavelet variance estimator for the scale analysis of time series. *Journal of the American Statistical Association,* 95**,** 184-196.
SHAFER-PELTIER, K. E., HAYNES, C. L., GLUCKSBERG, M. R. & VAN DUYNE, R. P. 2003. Toward a glucose biosensor based on surface-enhanced raman scattering. *Journal of the American Chemical Society,* 125**,** 588-593.
SHAO, J., LIN, M., LI, Y., LI, X., LIU, J., LIANG, J. & YAO, H. 2012. In vivo blood glucose quantification using raman spectroscopy. *PloS one,* 7**,** e48127.
SHARMA, B., CARDINAL, M. F., KLEINMAN, S. L., GREENELTCH, N. G., FRONTIERA, R. R., BLABER, M. G., SCHATZ, G. C. & VAN DUYNE, R. P. 2013. High-performance sers substrates: Advances and challenges. *MRS bulletin,* 38**,** 615-624.
SHARP, J., PELTZER, E., ALPERIN, M. J., CAUWET, G., FARRINGTON, J., FRY, B., KARL, D., MARTIN, J., SPITZY, A. & TUGRUL, S. 1993. Procedures subgroup report. *Marine Chemistry,* 41**,** 37-49.
SHARP, J. H. 1973. Size classes of organic carbon in seawater. *Limnology and Oceanography,* 18**,** 441-447.
SHILLAND, E., MONTEITH, D., MILLIDINE, K. & MALCOLM, I. 2014. The united kingdom upland waters monitoring network data report for 2012–2013 (year 25). *Report to the Department for Environment, Food and Rural Affairs (DEFRA)*.
SIEBECK, O., VAIL, T., WILLIAMSON, C., VETTER, R., HESSEN, D., ZAGARESE, H., LITTLE, E., BALSEIRO, E. & MODENUTTI, B. 1994. Impact of uv-b radiation on zooplankton and fish in pelagic freshwater ecosystems. *Advances in limnology[ERGEB. LIMNOL./ADV. LIMNOL.]. 1994.*
SIESLER, H. W. 2009. *Infrared and raman spectroscopic imaging*, John Wiley & Sons.
SIFUZZAMAN, M., ISLAM, M. & ALI, M. 2009. Application of wavelet transform and its advantages compared to fourier transform.

SIMONSEN, J. & HARREMOËS, P. 1978. Oxygen and ph fluctuations in rivers. *Water Research,* 12**,** 477-489.
SIMPSON, A., BURDON, J., GRAHAM, C., HAYES, M., SPENCER, N. & KINGERY, W. 2001. Interpretation of heteronuclear and multidimensional nmr spectroscopy of humic substances. *European Journal of Soil Science,* 52**,** 495-509.
SINGH, S. P., GANGWAR, R. & SINGH, N. 2007. Nonlinear scattering effects in optical fibers. *Progress In Electromagnetics Research,* 74**,** 379-405.
SKOOG, A., THOMAS, D., LARA, R. & RICHTER, K. U. 1997. Methodological investigations on doc determinations by the htco method. *Marine Chemistry,* 56**,** 39-44.
ŠMEJKALOVÁ, D. & PICCOLO, A. 2007. Aggregation and disaggregation of humic supramolecular assemblies by nmr diffusion ordered spectroscopy (dosy-nmr). *Environmental Science & Technology,* 42**,** 699-706.
SMEKAL, A. 1923. Zur quantentheorie der dispersion. *Naturwissenschaften,* 11**,** 873-875.
SMITH, B. A. V. 2016. *Assessment of carbon and nutrient export from a peatland windfarm construction site.* University of Glasgow.
SMITH, J., NAYAK, D. R. & SMITH, P. 2012. Renewable energy: Avoid constructing wind farms on peat. *Nature,* 489**,** 33-33.
SMITH, M. B., KOREN, V. I., ZHANG, Z., REED, S. M., PAN, J.-J. & MOREDA, F. 2004. Runoff response to spatial variability in precipitation: An analysis of observed data. *Journal of hydrology,* 298**,** 267-286.
SMITH, W. 2008. Practical understanding and use of surface enhanced raman scattering/surface enhanced resonance raman scattering in chemical and biological analysis. *Chemical Society Reviews,* 37**,** 955-964.
SOBRON, P., SANZ, A., ACOSTA, T. & RULL, F. 2009. A raman spectral study of stream waters and efflorescent salts in rio tinto, spain. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy,* 71**,** 1678-1682.
SODERHOLM, S., ROOS, Y. H., MEINANDER, N. & HOTOKKA, M. 1999. Raman spectra of fructose and glucose in the amorphous and crystalline states. *Journal of Raman Spectroscopy,* 30**,** 1009-1018.
SONDERGAARD, M., HANSEN, B. & MARKAGER, S. 1995. Dynamics of dissolved organic carbon lability in a eutrophic lake. *Limnology and Oceanography,* 40**,** 46-54.
SONDERGAARD, M. & MIDDELBOE, M. 1995. A cross-system analysis of labile dissolved organic-carbon. *Marine Ecology-Progress Series,* 118**,** 283-294.
SPENCER, R. G. M., PELLERIN, B. A., BERGAMASCHI, B. A., DOWNING, B. D., KRAUS, T. E. C., SMART, D. R., DAHGREN, R. A. & HERNES, P. J. 2007. Diurnal variability in riverine dissolved organic matter composition determined by in situ optical measurement in the san joaquin river (california, USA). *Hydrological Processes,* 21**,** 3181-3189.
SPYRES, G., NIMMO, M., WORSFOLD, P. J., ACHTERBERG, E. P. & MILLER, A. E. J. 2000. Determination of dissolved organic carbon in seawater using high temperature catalytic oxidation techniques. *Trac-Trends in Analytical Chemistry,* 19**,** 498-506.
STEVENSON, F. J. 1994. *Humus chemistry: Genesis, composition, reactions*, John Wiley & Sons.
STEWART, S. & FREDERICKS, P. 1999. Surface-enhanced raman spectroscopy of amino acids adsorbed on an electrochemically prepared silver surface. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* 55**,** 1641-1660.
STOCK, J. H. & WATSON, M. W. 1988. Variable trends in economic time series. *The Journal of Economic Perspectives,* 2**,** 147-174.
STOCKWELL, R. G., MANSINHA, L. & LOWE, R. 1996. Localization of the complex spectrum: The s transform. *IEEE transactions on signal processing,* 44**,** 998-1001.
STRICKLAND, J. 1972. Research on the marine planktonic food web at the institute of marine resources: A review of the past seven years of work. *Oceanography and Marine Biology Annual Review,* 10**,** 349-414.

STROHMEIER, S., KNORR, K. H., REICHERT, M., FREI, S., FLECKENSTEIN, J. H., PEIFFER, S. & MATZNER, E. 2013. Concentrations and fluxes of dissolved organic carbon in runoff from a forested catchment: Insights from high frequency measurements. *Biogeosciences,* 10**,** 905-916.
STUTTER, M., LANGAN, S. & COOPER, R. 2008. Spatial and temporal dynamics of stream water particulate and dissolved n, p and c forms along a catchment transect, ne scotland. *Journal of Hydrology,* 350**,** 187-202.
SUGIKAWA, K., FURUKAWA, Y. & SADA, K. 2011. Sers-active metal–organic frameworks embedding gold nanorods. *Chemistry of Materials,* 23**,** 3132-3134.
SUGIMURA, Y. & SUZUKI, Y. 1988. A high-temperature catalytic-oxidation method for the determination of non-volatile dissolved organic-carbon in seawater by direct injection of a liquid sample. *Marine Chemistry,* 24**,** 105-131.
SUH, J. & MOSKOVITS, M. 1986. Surface-enhanced raman spectroscopy of amino acids and nucleotide bases adsorbed on silver. *Journal of the American Chemical Society,* 108**,** 4711-4718.
SUNDEKILDE, U. K., LARSEN, L. B. & BERTRAM, H. C. 2013. Nmr-based milk metabolomics. *Metabolites,* 3**,** 204-222.
SUTTON, R. & SPOSITO, G. 2005. Molecular structure in soil humic substances: The new view. *Environmental Science & Technology,* 39**,** 9009-9015.
SUZUKI, Y. 1993. On the measurement of doc and don in seawater. *Marine Chemistry,* 41**,** 287-288.
TADINI, A. M., MOREIRA, A. B. & BISINOTI, M. C. 2014. Fractionation of aquatic humic substances and dynamic of chromium species in an aquatic body influenced by sugarcane cultivation. *Journal of the Brazilian Chemical Society,* 25**,** 119-125.
TAILLEFERT, M., LUTHER III, G. W. & NUZZIO, D. B. 2000. The application of electrochemical tools for in situ measurements in aquatic systems. *Electroanalysis,* 12**,** 401-412.
TERCIER-WAEBER, M.-L., HEZARD, T., MASSON, M. & SCHÄFER, J. R. 2009. In situ monitoring of the diurnal cycling of dynamic metal species in a stream under contrasting photobenthic biofilm activity and hydrological conditions. *Environmental science & technology,* 43**,** 7237-7244.
TETZLAFF, D., WALDRON, S., BREWER, M. J. & SOULSBY, C. 2007. Assessing nested hydrological and hydrochemical behaviour of a mesoscale catchment using continuous tracer data. *Journal of Hydrology,* 336**,** 430-443.
THOMAS, J. D. 1997. The role of dissolved organic matter, particularly free amino acids and humic substances, in freshwater ecosystems. *Freshwater Biology,* 38**,** 1-36.
THORLEY, F. C., BALDWIN, K. J., LEE, D. C. & BATCHELDER, D. N. 2006. Dependence of the raman spectra of drug substances upon laser excitation wavelength. *Journal of Raman Spectroscopy,* 37**,** 335-341.
TIPPING, E., CORBISHLEY, H. T., KOPRIVNJAK, J. F., LAPWORTH, D. J., MILLER, M. P., VINCENT, C. D. & HAMILTON-TAYLOR, J. 2009. Quantification of natural dom from uv absorption at two wavelengths. *Environmental Chemistry,* 6**,** 472-476.
TIPPING, E., WOOF, C., RIGG, E., HARRISON, A. F., INESON, P., TAYLOR, K., BENHAM, D., POSKITT, J., ROWLAND, A. P., BOL, R. & HARKNESS, D. D. 1999. Climatic influences on the leaching of dissolved organic matter from upland uk moorland soils, investigated by a field manipulation experiment. *Environment International,* 25**,** 83-95.
TOMASBARBERAN, F. A., BLAZQUEZ, M. A., GARCIAVIGUERA, C., FERRERES, F. & TOMASLORENTE, F. 1992. A comparative-study of different amberlite xad resins in flavonoid analysis. *Phytochemical Analysis,* 3**,** 178-181.
TORRENCE, C. & COMPO, G. P. 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society,* 79**,** 61-78.
TOTSCHE, K., DANZER, J. & KÖGEL-KNABNER, I. 1997. Dissolved organic matter-enhanced retention of polycyclic aromatic hydrocarbons in soil miscible displacement experiments. *Journal of Environmental Quality,* 26**,** 1090-1100.
TRAPERO, J. R., KOURENTZES, N. & MARTIN, A. 2015. Short-term solar irradiation forecasting based on dynamic harmonic regression. *Energy,* 84**,** 289-295.

TURETSKY, M., WIEDER, K., HALSEY, L. & VITT, D. 2002. Current disturbance and the diminishing peatland carbon sink. *Geophysical Research Letters,* 29.
VAN DEN BERG, L. J., SHOTBOLT, L. & ASHMORE, M. R. 2012. Dissolved organic carbon (doc) concentrations in uk soils and the influence of soil, vegetation type and seasonality. *Science of the Total Environment,* 427**,** 269-276.
VAN HALL, C. E., SAFRANKO, J. & STENGER, V. A. 1963. Rapid combustion method for the determination of organic substances in aqueous solutions. *Analytical Chemistry,* 35**,** 315-319.
VANKEIRSBILCK, T., VERCAUTEREN, A., BAEYENS, W., VAN DER WEKEN, G., VERPOORT, F., VERGOTE, G. & REMON, J. P. 2002a. Applications of raman spectroscopy in pharmaceutical analysis. *Trac-Trends in Analytical Chemistry,* 21**,** 869-877.
VANKEIRSBILCK, T., VERCAUTEREN, A., BAEYENS, W., VAN DER WEKEN, G., VERPOORT, F., VERGOTE, G. & REMON, J. P. 2002b. Applications of raman spectroscopy in pharmaceutical analysis. *TrAC trends in analytical chemistry,* 21**,** 869-877.
VARGAS, R., DETTO, M., BALDOCCHI, D. D. & ALLEN, M. F. 2010. Multiscale analysis of temporal variability of soil co2 production as influenced by weather and vegetation. *Global Change Biology,* 16**,** 1589-1605.
VENDRELL, M., MAITI, K. K., DHALIWAL, K. & CHANG, Y.-T. 2013. Surface-enhanced raman scattering in cancer detection and imaging. *Trends in biotechnology,* 31**,** 249-257.
VILLAR, S. E. J. & EDWARDS, H. G. M. 2006. Raman spectroscopy in astrobiology. *Analytical and Bioanalytical Chemistry,* 384**,** 100-113.
VISKARI, P. J. & LANDERS, J. P. 2006. Unconventional detection methods for microfluidic devices. *Electrophoresis,* 27**,** 1797-1810.
VÍTEK, P., ALI, E. M., EDWARDS, H. G., JEHLIČKA, J., COX, R. & PAGE, K. 2012. Evaluation of portable raman spectrometer with 1064nm excitation for geological and forensic applications. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* 86**,** 320-327.
VITT, D. H., BAYLEY, S. E. & JIN, T.-L. 1995. Seasonal variation in water chemistry over a bog-rich fen gradient in continental western canada. *Canadian Journal of Fisheries and Aquatic Sciences,* 52**,** 587-606.
VOGEL, E., GESSNER, R., HAYES, M. H. B. & KIEFER, W. 1999. Characterisation of humic acid by means of sers. *Journal of Molecular Structure,* 482**,** 195-199.
WALDRON, S., FLOWERS, H., ARLAUD, C., BRYANT, C. & MCFARLANE, S. 2009. The significance of organic carbon and nutrient export from peatland-dominated landscapes subject to disturbance, a stoichiometric perspective. *Biogeosciences,* 6**,** 363-374.
WALKER, J. P., WILLGOOSE, G. R. & KALMA, J. D. 2004. In situ measurement of soil moisture: A comparison of techniques. *Journal of Hydrology,* 293**,** 85-99.
WALLAGE, Z. E., HOLDEN, J. & MCDONALD, A. T. 2006. Drain blocking: An effective treatment for reducing dissolved organic carbon loss and water discolouration in a drained peatland. *Science of the total environment,* 367**,** 811-821.
WALLING, D. & WEBB, B. 1985. Estimating the discharge of contaminants to coastal waters by rivers: Some cautionary comments. *Marine Pollution Bulletin,* 16**,** 488-492.
WARK, T., HU, W., CORKE, P., HODGE, J., KETO, A., MACKEY, B., FOLEY, G., SIKKA, P. & BRUNIG, M. Springbrook: Challenges in developing a long-term, rainforest wireless sensor network. Intelligent Sensors, Sensor Networks and Information Processing, 2008. ISSNIP 2008. International Conference on, 2008. IEEE, 599-604.
WATERLOO, M. J., OLIVEIRA, S. M., DRUCKER, D. P., NOBRE, A. D., CUARTAS, L. A., HODNETT, M. G., LANGEDIJK, I., JANS, W. W. P., TOMASELLA, J., DE ARAUJO, A. C., PIMENTEL, T. P. & MUNERA ESTRADA, J. C. 2006. Export of organic carbon in run-off from an amazonian rainforest blackwater catchment. *Hydrological Processes,* 20**,** 2581-2597.

WELLINGTON, B. I. & DRISCOLL, C. T. 2004. The episodic acidification of a stream with elevated concentrations of dissolved organic carbon. *Hydrological Processes,* 18**,** 2663-2680.
WERNEKINCK, E., VALENZUELA, H. & ANFOSSI, I. On the analysis of power electronics circuits waveforms with wavelets. Power Conversion Conference, 1993. Yokohama 1993., Conference Record of the, 1993. IEEE, 544-549.
WEST, T. R. 2010. *Geology applied to engineering*, Waveland Press.
WHITNEY, A. V., VAN DUYNE, R. P. & CASADIO, F. Silver island films as substrate for surface-enhanced raman spectroscopy (sers): A methodological study on their application to artists' red dyestuffs. Optics East 2005, 2005. International Society for Optics and Photonics, 59930K-59930K-10.
WIEBINGA, C. J. & DE BAAR, H. J. W. 1998. Determination of the distribution of dissolved organic carbon in the indian sector of the southern ocean. *Marine Chemistry,* 61**,** 185-201.
WILLIAMS, P. 1971. The distribution and cycling of organic matter in the ocean. *Organic compounds in aquatic environments***,** 145-163.
WILLIAMS, P. M. & DRUFFEL, E. R. 1987. Radiocarbon in dissolved organic matter in the central north pacific ocean. *Nature,* 330.
WILLIAMSON, C. E., HARGREAVES, B. R., ORR, P. S. & LOVERA, P. A. 1999a. Does uv play a role in changes in predation and zooplankton community structure in acidified lakes? *Limnology and Oceanography,* 44**,** 774-783.
WILLIAMSON, C. E., MORRIS, D. P., PACE, M. L. & OLSON, A. G. 1999b. Dissolved organic carbon and nutrients as regulators of lake ecosystems: Resurrection of a more integrated paradigm. *Limnology and Oceanography,* 44**,** 795-803.
WILSON, H. F., SAIERS, J. E., RAYMOND, P. A. & SOBCZAK, W. V. 2013. Hydrologic drivers and seasonality of dissolved organic carbon concentration, nitrogen content, bioavailability, and export in a forested new england stream. *Ecosystems,* 16**,** 604-616.
WINKLER, S., SARACEVIC, E., BERTRAND-KRAJEWSKI, J.-L. & TORRES, A. 2008. Benefits, limitations and uncertainty of in situ spectrometry. *Water science and Technology,* 57**,** 1651-1658.
WOLF, J. P. 1987. Soil-structure-interaction analysis in time domain. *Structural mechanics in reactor technology.*
WOOD, B. R., CASPERS, P., PUPPELS, G. J., PANDIANCHERRI, S. & MCNAUGHTON, D. 2007. Resonance raman spectroscopy of red blood cells using near-infrared laser excitation. *Analytical and bioanalytical chemistry,* 387**,** 1691-1703.
WORRALL, F., BURT, T. & ADAMSON, J. 2004a. Can climate change explain increases in doc flux from upland peat catchments? *Science of the Total Environment,* 326**,** 95-112.
WORRALL, F., BURT, T., JAEBAN, R., WARBURTON, J. & SHEDDEN, R. 2002. Release of dissolved organic carbon from upland peat. *Hydrological Processes,* 16**,** 3487-3504.
WORRALL, F., BURT, T. & SHEDDEN, R. 2003. Long term records of riverine dissolved organic matter. *Biogeochemistry,* 64**,** 165-178.
WORRALL, F., HARRIMAN, R., EVANS, C. D., WATTS, C. D., ADAMSON, J., NEAL, C., TIPPING, E., BURT, T., GRIEVE, I., MONTEITH, D., NADEN, P. S., NISBET, T., REYNOLDS, B. & STEVENS, P. 2004b. Trends in dissolved organic carbon in uk rivers and lakes. *Biogeochemistry,* 70**,** 369-402.
YAMADA, M. & OHKITANI, K. 1991. Orthonormal wavelet analysis of turbulence. *Fluid Dynamics Research,* 8**,** 101-115.
YARWOOD, J., DOUTHWAITE, R. & DUCKETT, S. 2009. *Spectroscopic properties of inorganic and organometallic compounds*, Royal Society of Chemistry.
YU, Z. 2012. Northern peatland carbon stocks and dynamics: A review. *Biogeosciences,* 9**,** 4071-4085.
ZALBA, P., AMIOTTI, N. M., GALANTINI, J. A. & PISTOLA, S. 2016. Soil humic and fulvic acids from different land use systems evaluated by e4/e6 ratios. *Communications in Soil Science and Plant Analysis*.

ZEPP, R. G., SHELDON, W. M. & MORAN, M. A. 2004. Dissolved organic fluorophores in southeastern us coastal waters: Correction method for eliminating rayleigh and raman scattering peaks in excitation-emission matrices. *Marine Chemistry,* 89**,** 15-36.

ZHANG, X., SHAH, N. C. & VAN DUYNE, R. P. 2006. Sensitive and selective chem/bio sensing based on surface-enhanced raman spectroscopy (sers). *Vibrational Spectroscopy,* 42**,** 2-8.