Comparison and development of advanced techniques for organic matter characterisation in water and wastewater processing

A thesis submitted to the University of Adelaide
in fulfilment of the requirements for the degree of Doctor of Philosophy

EMMA LOUISE PLANT

School of Agriculture, Food and Wine
University of Adelaide

October 2013

TABLE OF CONTENTS

CHAPTER THREE: CAN ORGANIC RESIDUES FROM A PULP AND PAPER MILL BE IDENTIFIED IN THE SEDIMENTS OF THE RECEIVING ENVIRONMENT USING NMR AND PYROLYSIS TECHNIQUES?

CHAPTER FOUR: CHANGES IN THE NATURE OF THE DISSOLVED ORGANICS DURING PULP AND PAPER MILL WASTEWATER TREATMENT: A MULTIVARIATE STATISTICAL STUDY COMBINING DATA FROM THREE ANALYTICAL TECHNIQUES

CHAPTER FIVE: CHARACTERISATION OF ORGANIC MATTER IN SEWAGE EFFLUENT: COMBINING AND COMPARING NMR, PYROLYSIS AND HPSEC ANALYTICAL APPROACHES

CHAPTER SIX: SYNOPSIS

ABSTRACT

Organic matter (OM) is one of the most complex natural mixtures on Earth. It is ubiquitous in all natural environments and plays an important role in a number of natural processes. However, despite the rich literature on its role and function in many environments, key aspects of OM composition remain unclear. What is now appreciated, is that OM exists as a complex mixture rather than as a single, defined material, whose structure depends on the source of the OM, the environment in which it was produced or transported to, and its stage of degradation.

There are different analytical techniques currently available that are used to study the composition and structure of OM. The more sophisticated techniques, including spectroscopic and chromatographic techniques, can provide valuable chemical or structural information on a sample, and have been widely applied to the study of OM. Characterisation of OM now commonly involves the combined use of techniques for an informative analysis. However, when they are used together, it is unclear how much information provided by the different techniques is the same, and how much is unique.

The focus of this study was to develop a protocol to quantitatively compare the information provided by different techniques when used in combination to characterise OM. This approach to assess the extent of technique complementarity relies on multivariate statistics. It involves the use of ordination plots to assess the information qualitatively, and the Spearman Rank Correlation method to assess the information quantitatively. The sophisticated analytical techniques chosen for the multi-technique approach are at the forefront of OM characterisation and include (i) solid-state ^{13}C nuclear magnetic resonance (NMR) spectroscopy, (ii) flash pyrolysis-gas chromatography mass spectrometry (py-GCMS), and (iii) high performance size exclusion chromatography (HPSEC).

The protocol was initially developed for the combination and comparison of solid-state ^{13}C NMR spectroscopy and flash py-GCMS data. The approach was demonstrated on a set of well understood plant residues, where NMR and py-GCMS results could be compared to the literature. This methodology was further demonstrated on a different set of sediment organics. The approach was expanded to include HPSEC data in the analysis of pulp and paper mill water and wastewater (WW) organics. The three-technique approach was

then applied to the final set of samples in this thesis, consisting of partially and fully treated sewage effluent OM.

In each case, the ordination plots were able to help determine and compare how the different techniques differentiated between the organics in the sample sets. What is novel about the protocol developed is the quantitative comparison of this information. The Spearman Rank Correlation method was able to determine that two techniques in each study provided some complementary information to the analysis, and when the third technique was used, one technique provided unique information only. From this, it was determined that the most beneficial combination of techniques was when some complementary information and some unique information were provided. There was little benefit to the analysis when a large degree of complementary information was provided by two techniques. With the ability to determine how much complementary information is provided, analytical techniques can be more appropriately applied to OM characterisation therefore improving the allocation of resources including time and money.

Importantly, the degree of technique complementarity varied with each study. This was a promising result, as the complexity and variability of OM was therefore reflected in the analysis. The limits to the protocol were thought to have been reached in the final study of sewage effluents, as the ordination plots were thought to reflect primarily random variation due to the high degree of similarity between the spectral results. However, these results were put into context by the addition of reference organics to the ordination plots.

DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

8/10/13.

Emma Plant   Date

ACKNOWLEDGEMENTS

PUBLICATIONS ARRISING FROM THIS THESIS

Plant EL, Smernik RJ, van Leeuwen J. 2011. Comparison and development of advanced techniques for organic matter characterisation in water and wastewater processing. Platform presentation at: Water Quality Research Australia Limited's Members Meeting, Gold Coast, Australia.

Plant EL, Smernik RJ, van Leeuwen J, Macdonald LM, Dewy R, Nothrop S, Everson A. 2012. Lake Bonney: how have mill wastewater treatment practices influenced organics? Poster presentation at: Australian Water Association's OzWater '12 Conference, Sydney, Australia.

Plant EL, Smernik R, Greenwood P, van Leeuwen J, Macdonald LM, Everson A. 2013. Stirring-up sediment in Lake Bonney: multivariate comparison of organic profiling. Poster presentation at: Australian Water Association's OzWater '13 Conference, Perth, Australia.

Plant EL, Smernik R, Greenwood P, Macdonald LM, van Leeuwen J (2013) The organic chemistry of plant residues: comparison of NMR and pyrolysis data using multivariate statistical approaches. *Current Organic Chemistry*. 17:3006-3012.

Plant EL, Smernik RJ, van Leeuwen J, Greenwood P, Macdonald LM (submitted) Can organic residues from a pulp and paper mill be identified in the sediments of the receiving environment using NMR and pyrolysis techniques? *Environmental Science and Pollution Research.*

Plant EL, Smernik RJ, van Leeuwen J, Greenwood P, Macdonald LM (2013) Changes in the nature of dissolved organics during pulp and paper mill wastewater treatment: a multivariate statistical study combining data from three analytical techniques. *Environmental Science and Pollution Research*. DOI 10.1007/s11356-013-2351-0.

Plant EL, Smernik RJ, van Leeuwen J, Greenwood P, Macdonald LM, Leverett D. 2013. Pushing the limits of organic matter characterisation in wastewater: a quantitative approach to technique complementarity. Poster presentation at: SETAC North America 34th Annual Meeting, Nashville, Tennessee.

STRUCTURE OF THIS THESIS

This thesis is presented as a combination of papers that have been accepted for publishing, submitted for publication or have been prepared for submission.

Chapter One provides an overview of the literature on the nature, behaviour and importance of OM in different environments and the analytical techniques used to study it. It is in this chapter that the analytical challenges facing characterisation of OM are discussed. This chapter includes the proposed objective of this research. Introductory material relevant to the accepted, submitted and prepared papers is not presented in detail in the literature review because it appears in the introduction of each chapter.

Chapter Two comprises a paper published in *Current Organic Chemistry*. It describes the development and demonstration of a multivariate statistical method to combine and compare data from solid-state ^{13}C NMR spectroscopy and flash py-GCMS.

Chapter Three comprises a paper that has been submitted to *Environmental Science and Pollution Research*. It describes a demonstration of the method in Chapter Two to the characterisation of a set of sediment organic samples.

Chapter Four comprises a paper published in *Environmental Science and Pollution Research*. It describes the expansion of the method to include HPSEC for the characterisation of a set of pulp and paper mill WW organics.

Chapter Five comprises a paper that has been prepared for submission to *Environmental Monitoring and Assessment*. It describes a demonstration of the expanded method to the characterisation of a set of partially and fully treated sewage effluent organics.

Chapter Six provides a synthesis of the findings contained in this thesis and includes recommendation for future research.

CHAPTER ONE

A REVIEW OF THE LITERATURE

*

Organic matter (OM) is a term that describes all organic degradation products of plant and animal material (. It is ubiquitous in the environment, being found in soils, sediments, natural waters and the atmosphere (Hanninen 2010; Bolan et al. 2011; Simpson et al. 2011). Organic matter is chemically heterogeneous and highly complex (Drikas 2003; Simpson and Simpson 2012; Nebbioso and Piccolo 2013), and has been described as the most complex natural mixture on Earth (Simpson and Simpson 2012). Having a presence in many natural environments, OM can form associations with different materials, especially those of soil and water, thereby influencing their properties (Kordel et al. 1997; Bolan et al. 2011). Anthropogenic organic materials are becoming more common in the natural environment, this includes the release of organics in industrial and domestic wastewaters (Liu et al. 2011; Lange et al. 2012; Zuloaga et al. 2012). These anthropogenic organic materials can further influence the properties of different materials in the natural environment, potentially causing environmental stress (Nebbioso and Piccolo 2013). Importantly, OM represents a large environmental pool of carbon (C) and plays a key role in the global carbon cycle (Simpson et al. 2011; Simpson and Simpson 2012; Nebbioso and Piccolo 2013).

THE CARBON CYCLE

The carbon cycle describes the exchange of C within and between four major reservoirs: the atmosphere, land, the oceans and fossil fuels (Houghton 2003; Jurado et al. 2008). The time taken for C to transfer from one reservoir to the next can vary from seconds to millennia. Some of the faster transfers include photosynthesis, which is the process by which atmospheric carbon dioxide (CO_2) is fixed into sugar by plants, while the longer transfers include the accumulation of fossil C including oil, gas and coal, through the deposition, diagenesis and thermal maturation of OM (Houghton 2003). Carbon in these reservoirs is exchanged by processes that are biological, chemical, geological and physical. A simplified model of the carbon cycle showing the reservoirs and some of the exchange processes is shown in Figure 1.
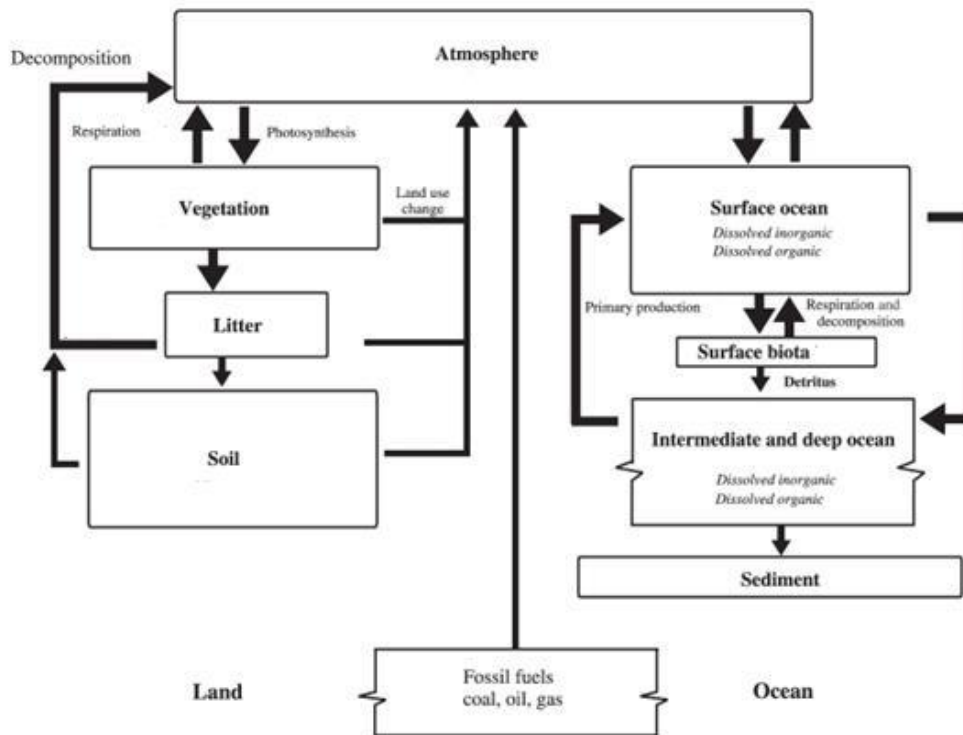
Figure 1 The carbon cycle, showing the reservoirs and some exchange processes between these reserviors. Adapted from (Houghton 2003).

ORGANIC MATTER IN THE CARBON CYCLE

Organic matter is an important part of the carbon cycle associated with the geological, terrestrial and oceanic components (Templier et al. 2005; Bolan et al. 2011; Nebbioso and Piccolo 2013). Organic matter can be transported between the four main reservoirs by interacting chemically and physically with its surrounding environment. Due to this movement, there are terms used to describe the origin of OM: autochthonous and allochthonous. Organic matter can originate from where it is located (autochthonous), or be from a different origin to its present location (allochthonous).

In the sections below, OM from each environmental reservoir (soils, sediments, waters and the atmosphere) are discussed in terms of their environmental importance; their known chemical composition; the processes by which they degrade and are transformed in their environment; how they interact and move through the environment; their influence on the fate of pollutants; and some examples of their study.

ORGANIC MATTER IN SOILS

Soil OM (SOM) plays a key role in plant, animal and microbial life in soils. It stabilises the entire soil matrix by buffering the soil, has exchange capacity with other components of the soil matrix and retains water and nutrients which are central to the maintenance of soil fertility (Kordel et al. 1997; Bolan et al. 2011; Kunhi Mouvenchery et al. 2012; Simpson and Simpson 2012). Organic matter exists in the entire soil column from the surface, where most terrestrial organics enter the soil (Pignatello 2012), to deeper in the profile, which receives inputs from the translocation of surface OM and from living and dead roots (Lynch 1991; Rumpel and Kogel-Knabner 2011). In soil, OM acts as both a sink and source of atmospheric CO_2 thereby playing an important role in regulating climate (Bolan et al. 2011).

Soils contain a variety of different chemical and physical forms of OM. Chemical forms include bio-macromolecules (such as cellulose, lignin, proteins and lipids) (Derenne and Largeau 2001; Li et al. 2012) as well as humic substances (polymeric OM that is not recognisable as source biopolymers) (Sutton and Sposito 2005; Hanninen 2010), kerogen (the insoluble organic fraction of sedimentary source rocks) and bitumen (the soluble organic fraction of sedimentary source rocks) (Brady and Weil 2008; Ahangar 2010; Cao et al. 2013). The overall composition of SOM depends on the origin of the material and the environment in which it has undergone degradation (Gaffney et al. 1996). In soils, OM exists as both particulate organic matter (POM) and dissolved organic matter (DOM) (Hepplewhite and Newcombe 2001; Bolan et al. 2011; Nebbioso and Piccolo 2013). The distinction between these forms is operationally defined, with DOM describing organics that pass through a filter pore size of 0.45 μm, and POM being the fractions that do not (Thurman 1985; Zsolnay 2003; Bolan et al. 2011; Nebbioso and Piccolo 2013). Dissolved OM typically represents a small fraction of SOM; however, it is acknowledged as influencing many biogeochemical processes as it is the most mobile and actively cycling fraction of OM (Akagi et al. 2007; Sanderman et al. 2008; Sanderman and Amundson 2009; Sanderman et al. 2009; Bolan et al. 2011; Xue et al. 2013).

A distinction is often made within OM between non-humic and humic substances. Non-humic OM includes recognisable bio-macromolecules (e.g. cellulose, lignin, protein and lipids) both inside and outside living organisms (Hanninen 2010). Humic substances are the complex macromolecules of reconstructed plant and microbial products that have

been transformed to various degrees through biotic and abiotic degradation (Hanninen 2010; Simpson et al. 2011; Simpson and Simpson 2012). They are characteristically aromatic and can make up to 80% of SOM (Simpson et al. 2011). Clearly, "humic substances" is a term that describes a complex mix of materials with no defined chemical structure or stoichiometry. Humic substances are traditionally divided into three types, fulvic acids, humic acids and humin, based on their solubility in acids and bases (Stevenson 1994; Gaffney et al. 1996; Kordel et al. 1997; Sutton and Sposito 2005; Hanninen 2010; Nebbioso and Piccolo 2013). Fulvic acids are soluble in water at any pH; humic acids are soluble in alkaline solution but precipitate at low pH(<2); and humin is insoluble at any pH (Gaffney et al. 1996). This distinction, based on solubility in different pH solutions, is operationally defined and reflects traditional isolation practices. Recent research indicates that compositional differences between the humic fractions are often subtle (Thomsen et al. 2002; Sutton and Sposito 2005).

Soil OM has been said to have a typical elemental composition consisting of approximately 40-60% carbon, 30-50% oxygen, 5% hydrogen, 1-4% nitrogen, 1-2% sulphur, and ≤0.03% phosphorus (Gaffney et al. 1996; Bolan et al. 2011). A popular model for the structure of humic substances is based on an alkyl/aromatic skeleton cross-linked by nitrogen and oxygen groups, with carboxylic acids, alcoholic and phenolic hydroxyls, quinones and ketones as major functional groups (Gaffney et al. 1996; Hayes and Clapp 2001; Hanninen 2010; Bolan et al. 2011). Solubility differences between fulvic and humic acids result from differences in structure and functional groups, with fulvic acids containing more aliphatic structures and more carboxylic acid, phenol and ketone groups compared to humic acids, which results in its high solubility in water at all pH values (Gaffney et al. 1996; Yakimenko 2001).

There are various processes by which OM is transformed in the soil, including diagenesis, humification and microbial degradation. Diagenesis is a process that occurs when OM is initially buried (Henrichs 1992). Following diagenesis, catagenesis starts with the increase in pressure and temperature after burial of OM and facilitates the transformation of discrete bio-macromolecules (e.g. lignin and cellulose) to immature humic materials (humic and fulvic acids and humus), then to mature coals and kerogens (Hedges and Keil 1999; Golding et al. 2004; Schimmelmann et al. 2006). During diagenesis and more-so catagenesis, oxygen-containing functional groups are lost, resulting in an increase in

aromatic structures compared to those in aliphatic linkages, and an increase in molecular weight (Golding et al. 2004).

Humification is another process by which plant and animal OM is broken down in the top-most soil to a point of stability, resulting in the preservation of refractory material known as humus (Wershaw 1993; Orlov and Sadovnikova 2005; Hanninen 2010). This is a basic step in the carbon cycle and this form of OM degradation decreases with depth in the soil (Kordel et al. 1997). The amount of humus in soils controls a number of soil properties, including buffering capacity, water-holding capacity, stabilisation of soil aggregates and metal-binding capacity (Wershaw 1993; Yao et al. 2011).

There are various microorganisms in the soil that contribute to the breakdown of OM. These include invertebrates, bacteria and fungi (Swift 1979; Kalbitz et al. 2000; Dilly et al. 2004; Bolan et al. 2011). These microorganisms provide a dual role, as a decomposer of OM and as a carbon sink (Bolan et al. 2011). Microbial biomass represents an important reservoir of OM, and soil fauna (earthworms etc) and microflora (e.g., bacteria) facilitate the turnover of this biomass (Bolan et al. 2011). As SOM is mainly present in the mineral topsoil horizon in temperate soils, microbial activity is limited at depth due to the reduced bioavailability of organic C (Ghiorse and Wilson 1988; Rodriguez-Zaraqoza et al. 2008; Celerier et al. 2009) as well as reduced supply of other essential nutrients, such as nitrogen and phosphorus (Grandy et al. 2008; Enowashu et al. 2009; Bolan et al. 2011). As with all forms of decomposition, including diagenesis and humification, the nature of the organic source material and environmental conditions are important factors that control the rate of decomposition by microbes (Kalbitz et al. 2000; Bolan et al. 2011).

The fate of SOM is influenced by the extent and strength of interactions with many soil components, especially clays. Clays provide surface area for the sorption of DOM and other solutes in the soil, and can directly interact with microbes thus affecting the microbial degradation of SOM (Bolan et al. 2011). There are several processes by which organics can sorb to clay surfaces including anion/cation exchange, hydrogen bonding, water bridging and van der Waals forces (Gu et al. 1994; Kordel et al. 1997; Bolan et al. 2011). Sorption of DOM to clays increases with increasing amounts of aluminium and iron oxides in soils (Bolan et al. 2011). Those DOM constituents with high aromaticity or those rich in organic nitrogen and acidic groups are preferentially sorbed (Bolan et al. 2011).

Generally OM moves downward due to leaching by seepage or through the action of animals, including earthworms and burrowing animals (Kordel et al. 1997; Rumpel and Kogel-Knabner 2011). The movement of DOM in soils influences the bioavailability of nutrients (Bolan et al. 2011). This is especially true for metals and metalloids, whose transport and bioavailability can be strongly influenced by complexation with DOM (Bolan et al. 2011). These interactions can alter the chemical speciation of metals and metalloids, influencing their affinity for sorption to the soil matrix, their accumulation, uptake, and toxicity to organisms (Arnold et al. 2010; Bolan et al. 2011; Yao et al. 2011). The occurrence and mobility of DOM in soils can also influence the sorption, degradation, and mobility of contaminants including non-ionic hydrocarbons, organic pesticides and other hydrophobic organic compounds (HOCs) (Wershaw 1993; Kordel et al. 1997; Kopinke et al. 2001; Orlov and Sadovnikova 2005; Chefetz and Xing 2009; Bolan et al. 2011). Hydrophobic organic compounds can sorb to soil organics that have both aromatic and aliphatic moieties; however, the degree of aromaticity or aliphaticity of SOM cannot be used to predict the affinity of sorbent sorption (Chefetz and Xing 2009). Many studies have shown that SOM is the most important soil component for the sorption of PAHs, therefore SOM has the greatest effect on the environmental persistence and bioavailability of these compounds (Orlov and Sadovnikova 2005; Ahangar 2010; Pignatello 2012).

Determination of the chemical composition of SOM often requires it to be separated from the soil matrix. Separation can be achieved by one of a number of solvent based extraction techniques (Chen et al. 1978; Schnitzer and Schulten 1992; Northcott and Jones 2000). Extraction generally results in the fractionation of OM. Alkali extraction results in fractionation into three organic types, as mentioned above: humic acid, fulvic acid, and humin. Extraction with organic solvents can be used to isolate the lipid fraction of SOM (Otto and Simpson 2007; de Blas et al. 2013). Treatment with hydrofluoric acid can also be used to isolate SOM, but in this case the mineral fraction is dissolved leaving behind a solid residue that is primarily organic (Schmidt et al. 1997; Smernik et al. 2003). Solvent extraction of soil DOM can affect the characteristics of the DOM especially if it is strongly influenced by the water content of the soil (Akagi et al. 2007). Extraction methods provide OM in solution-form thereby making it available for characterisation using wet chemical techniques.

Our understanding of SOM chemistry and the way it plays a role in the environment has been enriched by the diverse types of characterisation techniques available to study SOM.

Some of the more sophisticated techniques include isotopic analysis, nuclear magnetic resonance (NMR) spectroscopy and analytical pyrolysis. Isotopic analyses have included the study of hydrogen isotope compositions during diagenesis and C turn over in the soil (Schimmelmann et al. 2006; Rumpel and Kogel-Knabner 2011); pyrolysis has been used extensively to characterise humic structures in the soil (Hanninen 2010; Mao et al. 2011a); and NMR has become an important research tool for SOM analysis (Plante et al. 2009; Leinweber et al. 2013).

ORGANIC MATTER IN SEDIMENTS

Sedimentary OM (SeOM) is increasingly recognised as playing key biogeochemical roles. For example, SeOM contributes to the early stages of fossil fuel development through diagenesis and catagenesis (Henrichs 1992; Golding et al. 2004). For this reason SeOM has a significant influence on the carbon cycle. The movement of OM in sediments and its interaction with the aqueous environment also influences the mobility, bioavailability and persistence of organic pollutants in the environment (Alexander 2000; Lueking et al. 2000; Golding et al. 2004).

Sedimentary OM covers a range of environments from deeply earth buried, lakes and rivers, to estuaries and the ocean. As a consequence, SeOM can comprise different combinations of OM of terrestrial and aquatic origin. The SeOM composition of terrestrial OM is reflective of that of vascular plants, being characteristically aromatic rich and nitrogen poor (Hedges and Oades 1997; Golding et al. 2004; Simpson et al. 2011). This is due to the contribution of lignin and tannin, which are unique to terrestrial plants and relatively resistant to degradation. In contrast, aquatic OM is nitrogen rich due to a primarily planktonic origin (Golding et al. 2004; Zehr 2011; Zehr and Kudela 2011). As with SOM, SeOM can exist as both POM and DOM and its chemical structure and nature depends on the source material and the environment in which it is degraded.

Sedimentary OM can be transformed in a number of ways which are similar to the degradation processes of SOM. These processes include degradation by bacteria and other microorganisms and diagenesis (Hedges and Oades 1997; Golding et al. 2004; Zehr 2011; Zehr and Kudela 2011). Most decomposition of OM takes place near the sediment-water interface. The primary degraders of SeOM are bacteria, which are most prevalent in the upper 10 cm of the sediment (Hedges and Oades 1997). The microorganisms responsible

for SeOM degradation can be nitrogen-rich, therefore contributing a biomass residue that is rich in nitrogen (Golding et al. 2004; Zehr 2011; Zehr and Kudela 2011). Diagenesis in marine sediments occurs in the upper several hundred meters of the sediment column (Henrichs 1992). During diagenesis, oxic decomposition of OM is significant in most marine environments, although stratification can lead to anoxic bottom waters where anaerobic process will be more significant. The rates of OM oxidation in both oxic and anoxic marine sediments increase with increasing temperature (Henrichs 1992).

Burial efficiency influences the rate of diagenesis and is low in oxic deep-sea sediments where OM inputs (< 10 g C m^{-2}.year^{-1}) and sediment accumulation rates (< 1cm per 1000 years) are low (Henrichs 1992). High efficiencies are associated with sediments with high OM inputs (> 100 g C m^{-2}.year^{-1}) and high sedimentation rates (> 0.1cm.year^{-1}). In the early stages of diagenesis, decomposition of fresh fungal detritus is generally faster than SeOM, with the highest decomposition rates being measured for small, soluble organics including glucose, amino acids and carboxylic acids (Henrichs 1992).

The mobility and bioavailability of organic compounds in sediments, like soils, is regulated by the process of sorption (Pignatello 2012). Adsorption to sediment particles is a mechanism by which solutes (DOM) are removed from porewaters; however, only some adsorbed organics are inhibited from decomposing in this way (Henrichs 1992). The movement of porewater through sediment facilitates the exchange of solutes and particles, including OM, between the sediment and water (Santos et al. 2012). The circulation of seawater through permeable sand is thought to be a controlling factor of the biogeochemistry of the sediment and the overlying waters.

The fate of anthropogenic pollutants in sediments, including HOCs and trace metals (Cu, Zn, Mn and Pb), is influenced by their chemical structure and size. These properties influence the ability of pollutants to interact with their environment. Sediments are known as transport pathways for a range of pollutants, including anthropogenic trace elements and antibiotic-resistant bacteria (Thevenon and Pote 2012). As with SOM, movement is facilitated by the pollutant's solubility, ability to interact with POM and ability to sorb to the sediment (Lueking et al. 2000). Hydrophobic pollutants will be preferentially adsorbed to sediments, while those that are hydrophilic and soluble have the potential to be transported with the porewater and DOM and exchanged with an overlying water column.

Due to the broad range of sedimentary environments and OM sources, there are countless possibilities of sedimentary organic structures. The typically ill-defined structural composition of SeOM composition was thought to represent a significant knowledge gap in organic environmental studies in the 90's (Henrichs 1992). The major analytical challenge of SeOM includes difficulty with sample collection, which has been discussed by Henrichs (1992). Of particular difficulty still today is the collection of samples reflecting the steep compositional gradients at the sediment-water interface which, as discussed above, is the point at which solutes are exchanged between the two environments.

Sedimentary OM structure has been probed at a detailed level by sophisticated solid-based techniques such as ^{13}C NMR spectroscopy, pyrolysis and thermochemolysis (Deshmukh et al. 2001; Krull et al. 2009), as well as carbon isotope techniques (Loh et al. 2012; Thevenon and Pote 2012). Deshmukh et al. (2001) applied ^{13}C NMR and analytical pyrolysis to compare the contribution of terrestrial OM and anthropogenic HOCs to marine SeOM of a contaminated site and a pristine site; and Loh et al. (2012) and Thevenon and Pote (2012) demonstrated the value of carbon isotope analysis to study the mobilisation and fate of anthropogenic OM in sediments.

ORGANIC MATTER IN WATER

In natural waters, OM can exist as both particulate (POM) and dissolved (DOM) forms, both serving a number of important roles. In particular, POM provides a vehicle for the redistribution of bioactive elements and provides the ecosystem with a source of nutrients. From the surface waters of the ocean, POM rains down to the deeper waters transferring carbon and nutrients to depth and supporting deep-sea life (Wakeham et al. 1997; Mao et al. 2011b). These sinking organics, which are partially and selectively degraded on transit to the seafloor, record into the underlying sediments an imprint of the water column processes (Kordel et al. 1997; Wakeham et al. 1997). DOM is the most mobile and actively cycling fraction and is therefore recognised as influencing key environmental parameters and numerous biogeochemical processes in aquatic, and terrestrial, environments (McDonald et al. 2004; Bolan et al. 2011).

Both POM and DOM can have autochthonous (e.g., the breakdown of aquatic biota) or allochthonous (plant litter from runoff events and anthropogenic chemicals) sources. This

is particularly true for rivers, which integrate biogeochemical processes occurring across the entire basin and therefore contain OM of terrestrial and aquatic origins (Kordel et al. 1997). Due to the natural flow of waters from rivers to the ocean, fresh water and oceanic OM share some similarities. However, as water very rarely flows upstream, there are also some considerable differences.

Fresh water DOM is believed to have a composition similar to that of soil DOM (Nebbioso and Piccolo 2013), as a result of the transport of DOM from soil to water during run-off events. The humic substances found in fresh water therefore contain compounds derived from lignin degradation which are more aromatic than oceanic humic substances (Kordel et al. 1997; Esteves et al. 2009). Humic substances in the ocean are a result of microbial action and are compositionally rich in nitrogen, and often contain branched aliphatic structures (Esteves et al. 2009), as well as protein and carbohydrate residues from marine algae (Kordel et al. 1997).

Oceanic, or marine, OM contains primarily autochthonous OM and is therefore largely planktonic in origin (Wakeham et al. 1997; Golding et al. 2004). POM in the ocean includes both detritus and living organisms, with the relative proportions of these two components varying with location and depth (Kordel et al. 1997). Generally, phytoplankton makes up more than a quarter of the POM in the euphotic zone (Kordel et al. 1997), the surface layer of water where light can penetrate to support photosynthesis. Below this zone, the concentration of POM rapidly decreases to a low constant level at depths below approximately 200 m.

During the transport of POM from surface to deeper waters, most (>99%) of the POM is selectively transformed by diagenetic and mineralisation processes (Wakeham and Lee 1993; Hedges and Oades 1997; Kordel et al. 1997; Wakeham et al. 1997; Mao et al. 2011b). Mineralisation by bacteria is the principle mechanism by which POM is removed from the water column (Kordel et al. 1997). Microorganisms are also major degraders and producers of aquatic OM (Hedges and Oades 1997). Microbial biomass is nitrogen-rich, therefore during the recycling of this material, metabolites are produced which are compositionally rich in nitrogen (Esteves et al. 2009).

As has been discussed above for SOM and SeOM, aquatic OM has the ability to interact with environmental pollutants (Kordel et al. 1997). Pollutants including HOCs and heavy

metals are capable of interacting with the hydrophobic components of aquatic OM (Kordel et al. 1997). These associations facilitate the movement of pollutants within the aquatic environment as well as between aquatic and sedimentary environments (Kordel et al. 1997; Lueking et al. 2000). The degradation, toxicity and bioavailability of pollutants are also strongly influenced by their association with OM, as has been discussed in the sections above. Pollutants enter the food web when they are associated with POM, which when ingested by aquatic life, is an important part of the food chain (Kordel et al. 1997).

Many methods have been used to study OM from different water sources, and many different issues addressed. In addition to enriching our understanding of its chemistry and the roles it plays in the environment, OM has been studied to optimise its removal from water sources for potable purposes, and for removal of organics from wastewater (WW) before release to the environment. Organic matter is removed from drinking water sources not because it necessarily poses a risk when ingested, but mainly because its presence impacts on water quality parameters associated with production and distribution of potable water (Drikas 2003; Allpike et al. 2010a; Matilainen et al. 2011; Uyguner-Demirel et al. 2013). As WW from municipals and industry are generally released to a receiving environment, OM must be removed to prevent the release of high loads of anthropogenic OM and any sorbed contaminants which can disrupt the receiving ecosystem (Liu et al. 2011; Kharayat 2012; Zuloaga et al. 2012; Feitosa et al. 2013).

A range of techniques used to study aqueous OM from a variety of different source waters have been discussed and compared in a review by Matilainen et al. (2011). Here, a variety of wet and solid-based analytical techniques providing information on the basic parameters of the OM as well as more detailed chemical and molecular structural features are discussed. One popular solid state technique is ^{13}C NMR spectroscopy (Simpson and Simpson 2009; Simpson et al. 2011; Simpson et al. 2012). This spectroscopic technique was used, for example, to study the structural changes and fate of sinking POM, as discussed above (Mao et al. 2011b).

ORGANIC MATTER IN THE ATMOSPHERE

An important but previously under-appreciated form of OM that is drawing increasing interest due to its relevance in global carbon dynamics is atmospheric OM (Nebbioso and Piccolo 2013). Water soluble organic compounds (WSOCs) in the atmosphere are thought

to play an important role in the carbon cycle by providing a temporal source of organic carbon to surface waters (Duarte and Duarte 2011). These WSOCs have also been recognised as potentially altering the properties of atmospheric particles (Dinar et al. 2007). For example, WSOCs can limit the ability of particles to act as cloud-condensation nuclei and thus provide an indirect aerosol effect (Dinar et al. 2007; Hallquist et al. 2009; Fors et al. 2010; Duarte and Duarte 2011). These WSOCs also absorb light in the ultraviolet and visible spectral regions, thereby contributing to atmospheric heating and thus directly contributing to the effect of aerosols on climate (Hallquist et al. 2009; Fors et al. 2010; Duarte and Duarte 2011). Anthropogenic aerosols in the atmosphere are also of concern as they can affect human health, specifically the respiratory and cardiovascular systems (Volkamer et al. 2006; Hallquist et al. 2009). However the true impact of atmospheric aerosols to climate and human life is uncertain as knowledge regarding their formation, composition and reactivity is limited (Hallquist et al. 2009; Duarte and Duarte 2011).

Atmospheric OM is the organic C fraction of atmospheric aerosols, which can exist as solid or liquid-phase particles (Hallquist et al. 2009). Humic-like substances (HULIS) have been identified as major contributors of atmospheric carbon (Fors et al. 2010), contributing 15-60% of aerosol particle mass (Dinar et al. 2007). The HULIS found in aerosol particles are thought to resemble the humic material found in soils, rivers and the ocean (Dinar et al. 2007; Fors et al. 2010). Some of this HULIS can enter the atmosphere from the bubble-bursting processes at the surface of the ocean, or by the transformation of organics that reside in the condensation phase (Carpenter et al. 2012). The heterogeneous mixture of HULIS contains substituted aromatic and aliphatic structures with acidic, ester and phenolic groups (Dinar et al. 2007).

Atmospheric OM is often dominated by WSOCs, which can represent between 10 and 80% of OM in atmospheric samples (Duarte and Duarte 2011). A major component of atmospheric WSOCs is secondary organic aerosols (SOAs). In particular, SOAs are significant in the troposphere (Hallquist et al. 2009). These secondary particles are formed in the atmosphere by gas particle conversion processes of condensation and nucleation, as well as chemical reactions that are heterogeneous or multiphase (Hallquist et al. 2009). These processes can occur through biogenic or anthropogenic sources, with recent studies revealing that 70-88% of WSOC C is formed by oxidative processes that involve biogenic volatile organics (Duarte and Duarte 2011). However, the chemical and physical processes associated with the formation of SOAs are varied and complex and understanding them

has become a major focus for research in atmospheric science (Hallquist et al. 2009). In urban areas, large occurrences of SOAs (and nitrogen oxides) can be the cause of atmospheric pollution known as photochemical smog (Volkamer et al. 2006).

Organic aerosols (OAs) have been shown to mature in the atmosphere. This maturing process involves the oxygenation of OAs, making them less volatile and more hygroscopic (Jimenez et al. 2009; Duarte and Duarte 2011). The result of maturing is the formation of oxygenated organic aerosols (OOAs), and atomic O:C ratios as high as 1 have been measured for very mature OA (Jimenez et al. 2009).

Organic matter is extremely mobile in the atmosphere compared to OM transport in the ocean or soils, with organics capable of redistribution on a regional or even global scale (Jurado et al. 2008). Aside from transport by air, OM can be deposited in the ocean by three processes. Firstly OM can enter the ocean by exchange across the air-ocean interface (Jurado et al. 2008; Carpenter et al. 2012); secondly it can enter by the dry deposition of OM bound to aerosols; and thirdly it can enter by the wet deposition of precipitation (Jurado et al. 2008). Currently, few models of the carbon cycle take into account the exchange of organic C between the atmosphere and ocean (Jurado et al. 2008). This is more-so because of the scarcity of relevant chemical data due to the challenge of isolating and analysing atmospheric OM, and not because they have negligible impact (Jurado et al. 2008; Nebbioso and Piccolo 2013).

As with all analyses, the sample collection step is crucial to the analytical process and is the determinant of the representation of the sample (Krol et al. 2010). Aside from the challenge of isolating atmospheric OM as discussed by Jurado et al. (2008) and Nebbioso and Piccolo (2013), a large range of analytical procedures are available for the detection and monitoring of atmospheric OM (Krol et al. 2010). In a review of sampling strategies for analytical measurements, Krol et al. (2010) described three basic approaches to collect atmospheric samples: denuder, dynamic and passive. Denuder and dynamic approaches allow for greater sample pre-concentration than passive approaches, although the former requires an external power source. However passive sampling approaches are often the method of choice, especially in studies of air quality, due to their determination of long-term averaged concentrations of atmospheric organics (Krol et al. 2010).

Atmospheric OM has attracted several aspects of scientific interest. In terms of public health, the chemistry, reactivity, and fate of SOAs and other volatile OM in urban areas have been studied (Volkamer et al. 2006; Hallquist et al. 2009; Ebersviller et al. 2012). The deposition processes of atmospheric OM have also been studied so that they may be included in the carbon cycle (Jurado et al. 2008). Techniques that have been commonly used to study atmospheric organics to date include gas chromatography-mass spectrometry (GCMS) (Ebersviller et al. 2012), and ultra violet/visible (UV/VIS) spectroscopy (Dinar et al. 2006; Dinar et al. 2007).

CHARACTERISATION OF ORGANIC MATTER – AN INTRODUCTION

Despite the importance of OM and its role in the environment, many aspects of its chemistry and structure are still relatively unknown. Thus OM has been referred to as being largely "molecularly uncharacterised" (Hedges et al. 2000; Parsi et al. 2007; Simpson and Simpson 2012; Nebbioso and Piccolo 2013). The challenge of natural OM characterisation stems from it being a mixture of mostly large and complex molecules (Matilainen et al. 2011; Simpson et al. 2011; Greenwood et al. 2012; Simpson and Simpson 2012; Nebbioso and Piccolo 2013). The complexity of these mixtures provides difficulties for conventional analytical techniques (Simpson et al. 2011). Even the more sophisticated analytical tools used to study OM were mostly designed or developed for the analysis of synthetic organic compounds and are not always suited to studying mixtures of complex structures like OM without substantial modification. Full characterisation of OM at the molecular level would require an approach that is capable of identifying thousands of unknown structures (Simpson et al. 2011). This is an exceptionally difficult analytical problem and would require a multidisciplinary approach that pushes the boundaries of modern science (Simpson et al. 2011).

Analytical techniques that are currently used to study OM rely on a range of chemical or structural features for detection and measurement. Detection and measurement is therefore affected by the propensity of organics to interact with incoming energy, to fluoresce, or to fragment into recognisable substructures. Several quantitative methods are able to measure the amount of OM present in a sample and are mostly quick and easy to use. These techniques include the measurement of dissolved organic carbon (DOC), total organic carbon (TOC) and colour (Marquet et al. 1999; Chow et al. 2008; Matilainen et al. 2011). Whilst providing important quantitative information, they typically provide little or

no information on the nature of the OM (Chow et al. 2008). More sophisticated analytical techniques that have the ability to probe the chemical nature of OM include NMR spectroscopy (Simpson et al. 2011), degradative techniques (pyrolysis and thermochemolysis) (Shadkami and Helleur 2010) and chromatography techniques (Matilainen et al. 2011). The increased level of qualitative detail provided by these techniques comes at a cost of increased time and expense needed to prepare the OM, undertake the analysis and interpret the results (Chow et al. 2008; Matilainen et al. 2011). However, with an improved understanding of how analytical methods function, there will be an improvement to our understanding of OM across all environmental settings discussed above.

Three sophisticated analytical techniques commonly used to study OM are solid-state ^{13}C NMR spectroscopy, flash py-GCMS, and HPSEC. In the following section, these techniques are discussed in terms of their importance, how they work, typical conditions or protocols for these analytical approaches, how qualitative and quantitative information can be obtained, the analytical challenges of these techniques, and some examples of their application to study OM.

SOLID-STATE ^{13}C NMR SPECTROSCOPY

Nuclear magnetic resonance spectroscopy is one of the most powerful contemporary analytical tools for studying the structure of organic materials (Smernik and Oades 1999; Smernik and Oades 2003; Simpson and Simpson 2009; Matilainen et al. 2011; Simpson et al. 2011; Simpson et al. 2012). It was first applied to the study of humic substances over 40 years ago (Barton and Schnitzer 1963; Simpson and Simpson 2009), although it wasn't until 1999 that POM was studied by solid-state NMR methods (Clark et al. 1999; Mao et al. 2011b). Since then the technique of solid-state ^{13}C NMR spectroscopy, in particular, has emerged as a powerful research tool for the study of OM (Smernik et al. 2003; Simpson et al. 2009; Matilainen et al. 2011; Simpson et al. 2011).

The basis of NMR is the difference in energy between spin states that arises when some nuclei are placed in a magnetic field. Nuclei that possess a spin (I) = ½ are the most easily observed by NMR techniques (Brauniger and Jansen 2013). Nuclei with I > ½ also possess a quadruple moment, and this complicates detection (Brauniger and Jansen 2013). When nuclei with I = ½ are placed in a magnetic field, they can absorb electromagnetic radiation

that corresponds to the difference in energy between the two spin states. The characteristic frequency of electromagnetic radiation corresponding to the difference is known as the Larmor frequency (Ito et al. 2011). This resonant frequency is slightly dependent on the chemical environment of the nucleus. The result is a spectrum, with one peak for each unique chemical environment (Simpson and Simpson 2009; Brauniger and Jansen 2013). The position of each peak in an NMR spectrum is termed its chemical shift (Simpson and Simpson 2009; Simpson et al. 2012), and since the absolute energy of the transitions (and hence the resonant frequency) is dependent on the strength of the applied magnetic field, it is usually expressed relative to that of a reference material. The small magnitude that differences in chemical environment have on the resonant frequency means that chemical shift is always expressed in parts per million (ppm), which is literally the difference in resonant frequency of the given peak divided by that of the reference material $\times$ 10^6. It is the chemical shift that provides most of the structural information about a sample (Simpson and Simpson 2009; Matilainen et al. 2011).

The most commonly studied nuclei in organic samples are ^{1}H and ^{13}C (Simpson and Simpson 2009). The most abundant isotope of hydrogen, ^{1}H, is detectable by NMR (I = ½); however, the most common carbon isotope, ^{12}C (98.9% isotopic abundance), is NMR silent (I = 0). Therefore NMR detection of carbon is carried out on the less abundant ^{13}C nucleus (1.1% isotopic abundance; I = ½). For this reason, ^{1}H NMR spectroscopy is inherently far more sensitive than ^{13}C NMR spectroscopy (Simpson and Simpson 2009). Other nuclei associated with OM that are studied by NMR techniques include ^{11}B, ^{15}N, ^{27}Al, ^{29}Si, and ^{31}P (Simpson and Simpson 2009; Matilainen et al. 2011; Brauniger and Jansen 2013).

Samples can be analysed by NMR either in solution or in the solid-state. Solution NMR is the more widely used method in chemistry, but requires the sample be soluble in an appropriate solvent. Solution NMR is most useful for the analysis of pure compounds, and generally requires only small amounts of sample (µg-mg). In contrast, solid-based techniques are often better for complex mixtures, especially for materials with little or no solubility. Larger sample sizes (~200 mg) are preferred as the large range of chemical environments present means that the concentration of each such environment is low and hence the signal is spread across a multitude of peaks. For this reason, solid-state NMR methods have been used more extensively to study OM (Simpson and Simpson 2009). Organic matter studies by solid-state NMR typically focus on the ^{13}C nucleus, even though

observing ^{13}C signals is compromised by the naturally low abundance of the ^{13}C isotope. Detection of ^{1}H nuclei, though attractive from a sensitivity perspective, is compromised by the broadness of signals and the relatively small range of variation in chemical shift (Simpson and Simpson 2009).

Solid-state ^{13}C NMR spectroscopy is commonly applied to the study of all types of organic samples (Smernik and Oades 1999; Simpson and Simpson 2009; Matilainen et al. 2011; Simpson et al. 2011; Cao et al. 2013; Nebbioso and Piccolo 2013). In a standard solid-state ^{13}C NMR experiment, differences in the chemical shift of the ^{13}C nuclei are used to identify and quantify broad carbon structural features in the sample (Matilainen et al. 2011). This is often done by dividing ^{13}C nuclei into the following broad chemical classes on the basis of chemical shift: carbonyl (200-160 ppm), aromatic (160-110 ppm), O-alkyl (110-45 ppm) and alkyl (45-0 ppm) (Almendros et al. 1996; Kögel-Knabner 1997; Smernik and Oades 2001; Smernik et al. 2004; Smith et al. 2008). However, solid-state ^{13}C NMR spectra contain a greater complexity of information than this and in most cases features that can be assigned to subsets within these broad classes can be identified, if not quantified. An example NMR spectrum showing the structural groups that can be identified for OM samples is shown in Figure 2, and Table 1 shows an alternative division of the chemical shift range into a larger number of carbon classes. A second advantage of solid-based techniques for OM characterisation is that they usually require minimal to no pretreatment (Smernik and Oades 1999). This means that the organic structures are not structurally or chemically compromised during isolation or purification steps. Finally, since solid-state techniques do not require a sample to be dissolved in a solvent, all of the organics in the sample, not just the soluble species, can be studied.
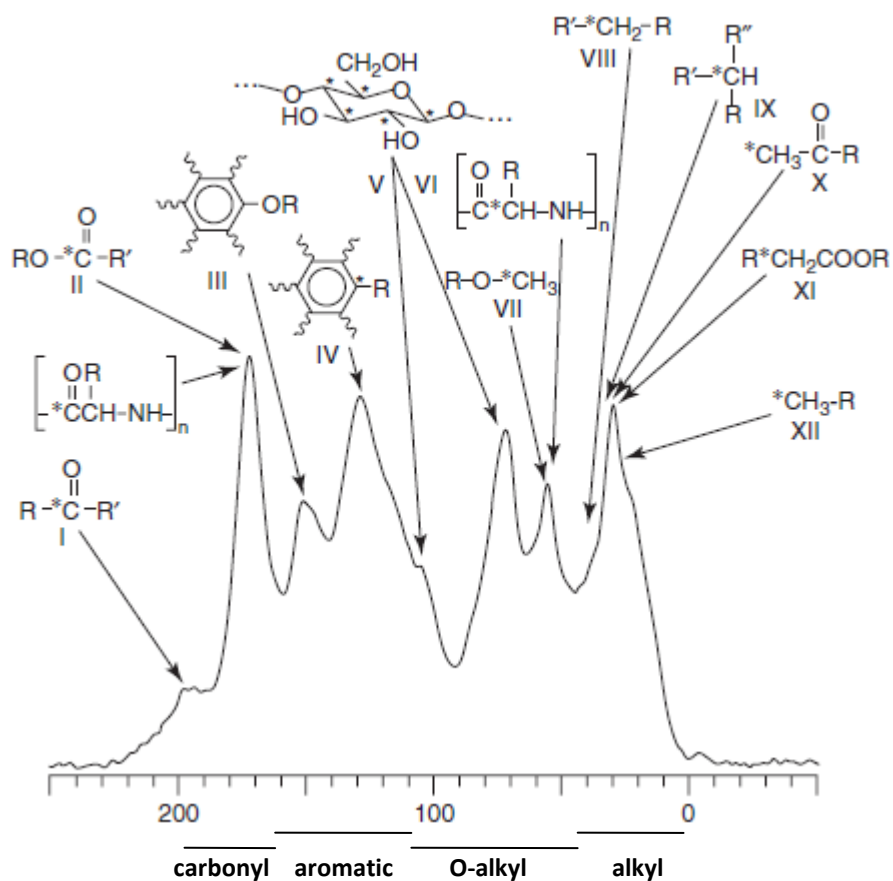
Figure 2 An example of the structural groups that can be identified in OM using solid-state ^{13}C NMR spectroscopy. The carbon atoms with the asterisk are the observed carbon functional groups, as highlighted in Table 1. Adapted from Simpson and Simpson (2009).

Table 1 The ^{13}C chemical shift ranges of major functional groups present in OM from soils and sediments. Adapted from Simpson and Simpson (2009).

[13]C Chemical Shift Range (ppm)	Assignment of Structures
0-45	Unsubstituted alkyl carbon: terminal methyl (15 ppm), straight-chain methylene carbon (30-34 ppm) and branched methylene carbon (35-45 ppm)
45-65	Substituted alkyl carbon in amines (45-46 ppm) and methoxyl groups (56 ppm)
65-95	Oxygen-substituted carbon, carbons in ethers, and ring carbons in carbohydrates
95-110	Anomeric carbon in carbohydrates (105 ppm), and di-oxygen-substituted aliphatic carbon
110-145	Aromatic carbon
145-160	Phenolic carbon
160-190	Carboxylic, ester and amide carbon
190-220	Carbonyl carbon

There are two methods that are commonly used to obtain solid-state ^{13}C NMR spectra of OM: cross-polarisation (CP) and direct-polarisation (DP); DP is also known as Bloch decay (Smernik 2005; Simpson and Simpson 2009; Mao et al. 2011a; Cao et al. 2013). Although DP is the simpler method, involving a simple 90° pulse and detect sequence, it is less often used. The CP method involves an initial 90° pulse to the more abundant adjacent ^{1}H nuclei in the sample followed by a controlled transfer of polarisation to the ^{13}C for detection (Cao et al. 2013). In doing this, the sensitivity of detection in increased, firstly because ^{13}C signal is enhanced by up to a factor of four and secondly because much less time is required between pulses to achieve complete relaxation between scans (Simpson and Simpson 2009). The disadvantage of CP is that it is the more complicated pathway of polarisation and more can go wrong, hence it can be quantitatively less reliable with some ^{13}C nuclei detected less efficiently or not at all (Smernik et al. 2006; Cao et al. 2013).

Quantification of broad chemical classes (alkyl, O-alkyl, aromatic and carbonyl) in an OM sample is possible with solid-state ^{13}C NMR spectroscopy. Under ideal conditions, each nucleus in a sample produces the same quantity of signal in a spectrum, whatever its chemical environment, so integration of spectral peaks provides a direct measure of the relative concentration of the species present (Smernik and Oades 2003; Simpson and Simpson 2009). As discussed above, DP relies only on the presence of ^{13}C in a sample in order to obtain a solid-state spectrum. This means that DP, if used appropriately, can provide the most accurate quantification in the solid-state (Smernik 2005; Simpson et al. 2011). Some ^{13}C nuclei in a sample may be underrepresented by the indirect method of CP, and debate has surrounded the degree of quantification achieved in solid-state CP spectra (Smernik 2005; Simpson and Simpson 2009). However, despite this, CP is considered semi-quantitative in terms of relative abundances of functional and structural groups (Smernik and Oades 2003; Simpson and Simpson 2009; Simpson et al. 2011).

Both CP and DP techniques require samples to be spun rapidly (in the kHz range) at an angle of 54.7° (Simpson and Simpson 2009; Simpson et al. 2011; Mao et al. 2012). This is known as magic angle spinning (MAS) and it is needed to overcome chemical shift anisotropy and dipolar interactions that would otherwise result in peaks being very broad. Chemical shift anisotropy broadening is a consequence of the different orientations that molecules can have in the solid phase relative to the applied magnetic field (Simpson and Simpson 2009). By spinning the sample at the "magic angle", these orientations are averaged resulting in much sharper resonances (Smernik 2005). When samples are spun at low speeds, spinning side bands (SSBs) may result in the spectral window (0-250 ppm for OM studies) and can interfere with qualitative and quantitative analyses (Simpson and Simpson 2009; Cao et al. 2013). The spinning speed required to overcome chemical shift anisotropy scales directly with magnetic field strength.

Dipolar interactions occur between chemically inequivalent nuclei and include heteronuclear interactions (e.g. between ^{13}C and ^{1}H) and homonuclear interactions (e.g. between ^{1}H and ^{1}H) (Simpson and Simpson 2009). Homonuclear ^{1}H-^{1}H dipolar interactions in the solid-state are hard to overcome and are a key reason why solid-state ^{1}H NMR experiments are difficult (Simpson et al. 2011; Brown 2012). On the other hand, heteronuclear ^{13}C-^{1}H interactions can be overcome through continuous irradiation of ^{1}H nuclei during signal acquisition, a process known as decoupling. The line shape of an

NMR signal, whether from solid-state or solution based techniques, is related to the relaxation time of the nuclei after they have absorbed the pulsed radiofrequency energy; more rapid relaxation in the solid-state contributes to the much broader line widths of solid-state spectra (Simpson and Simpson 2009).

When studying organics in whole soil or sediment samples, there are a number of possible challenges to overcome. One challenge is the presence of paramagnetic species, including iron and copper, which are naturally present in these samples. Paramagnetics can reduce the intensity of signal and result in broadening of peaks in a solid-state ^{13}C NMR spectrum (Smernik and Oades 2000; Simpson and Simpson 2009; Cao et al. 2013). The degree of signal loss in SOM samples has been shown to be dependent on the type of paramagnetic species present (Smernik and Oades 1999, 2000). Paramagnetic species also increase NMR relaxation rates and change the chemical shift of peaks (Smernik and Oades 2000; Smernik and Oades 2002). To remove the effects of paramagnetic species, samples can be treated with hydrofluoric acid (HF), which removes the mineral components of the sample (Simpson and Simpson 2009). By doing this, organics are also concentrated, therefore decreasing the time required to obtain a spectrum and improving the signal-to-noise ratio.

Solid-state ^{13}C NMR spectroscopy has been used extensively for structural investigations of OM (Simpson and Simpson 2009; Simpson et al. 2011; Nebbioso and Piccolo 2013) with developments to modern solid-state NMR techniques having improved the understanding of SOM structure significantly over the past 40 years (Kögel-Knabner 1997; Kogel-Knabner and Knicker 2001; Schaumann 2006; Simpson and Simpson 2009). Through the use of solid-state ^{13}C NMR spectroscopy it has been determined that humin has similar functional groups to the corresponding whole soil or sediment from which it is isolated (Simpson et al. 2011). In other studies, humin has been shown to contain more substituted-aliphatic carbon and less aromatic carbon compared to humic acids (Almendros et al. 1996; Bonin and Simpson 2007; Simpson et al. 2011). The research group of Mao et al. (2011a) has spent the past decade developing and modifying solid-state ^{13}C NMR techniques for exploring the structure of OM (Mao et al. 2000; Mao et al. 2001; Mao et al. 2002; Mao et al. 2007). From these developments, more structural moieties can be accurately quantified in contrast to the broad functional groups determined by CP-MAS (Mao et al. 2011a).

The non-destructive analyses provided by solid-state NMR techniques have become the method of choice for providing details on the chemical structure of kerogens (Cao et al. 2013). The most extensively used solid-state technique is that of ^{13}C CP-MAS (Zujovic et al. 1995; Lille et al. 2003; Wei et al. 2005; Werner-Zwanziger et al. 2005; Cao et al. 2013) which has led to significant advancements in the understanding of kerogen structures. In particular, NMR techniques have been able to identify functional groups as well as the connectivities between these groups, observe domains and heterogeneities, and provide structural information that is quantitative (Cao et al. 2013).

Improvements to NMR spectroscopy have facilitated developments in DOM research (Nebbioso and Piccolo 2013). Solid-state ^{13}C NMR methods have been used to elucidate the chemical and structural makeup of marine DOM (Templier et al. 2005; Koprivnjak et al. 2009; Mao et al. 2012), to study the nature and dynamics of DOM in soils (Sanderman et al. 2008; Sanderman et al. 2009), to investigate the structure of DOM in composts (Chefetz et al. 1998a; Chefetz et al. 1998b), and to characterise drinking water and wastewater (WW) DOM (Wong et al. 2002; Lewis et al. 2011). Many other NMR techniques, including solution and ^{1}H NMR, have been used to characterise the structure of DOM (Gigliotti et al. 2002; Esteves et al. 2009).

Aside from structural investigations, NMR has been crucial for developments in the understanding of the origins of organic materials and how they interact with their environment and with contaminants (Karathanasis 1999; Feng et al. 2006; Fomba et al. 2009; Simpson et al. 2011). Solid-state ^{13}C NMR spectroscopy has been applied over a wide range of OM research, and continues to advance the understanding of the structure and role of OM in the environment.

PYROLYSIS – GAS CHROMATOGRAPHY MASS SPECTROMETRY

Pyrolysis is a thermal degradation technique that can be used to study the molecular structure of organic samples (Hedges et al. 2000; Plante et al. 2009; Berwick et al. 2010a). Due to its convenience, speed and ability to provide complementary information to wet chemical degradation methods, pyrolysis has become a powerful research tool for the study of organic polymers and naturally occurring OM (Parsi et al. 2007; Silverio et al. 2008). Thermal analysis techniques involve monitoring a physical or chemical property of a sample against time or temperature and have been used in geosciences for over a century

(Plante et al. 2009). As early as 1887, Le Chatelier used thermal analysis to differentiate between clays based on their point of dehydration (Le Chatelier 1887 as cited by Plante et al. (2009)). However, it wasn't until 1935 that thermal analysis was applied to soil science (Mackenzie and Mitchell 1972; Plante et al. 2009).

When subject to pyrolysis, large complex molecules are broken down into more analytically amenable fragments by the application of heat under controlled atmospheric conditions (Baldock et al. 1991; Galletti and Bocchini 1995; Matilainen et al. 2011; Iwai et al. 2013). These fragments can be analysed when pyrolysis is used in combination with modern analytical techniques such as gas chromatography mass spectroscopy (py-GCMS) (Templier et al. 2005; Parsi et al. 2007; Iwai et al. 2013). With py-GCMS, the volatile fragments of the parent sample are swept onto the analytical column, where they are separated based on thermal volatility (related to size or molecular weight (MW)) and structural interaction with the activated phase of the analytical column. The separated analytes are then detected by mass spectroscopy (MS). A total ion chromatogram (TIC) is produced by the continual fast recording of mass spectra to capture the analytes as they elute from the gas chromatography (GC) column (Leinweber et al. 2013). The corresponding mass spectrum from full scan analysis of each peak can then be compared to the mass spectra of known compounds from standards or mass spectral libraries for assignment.

The quantitation of pyrolysates is possible via either the TIC or by the molecular, or another selected ion, representative of particular compound classes (Allpike et al. 2010b; Berwick et al. 2010b). Absolute analyte concentration can be determined with an internal standard, which involves spiking the samples with an organic standard of known concentration prior to py-GCMS (Ruiz et al. 2013). The combined signals from similar types of products, such as a common functional group, are often compared with other product groups (Greenwood et al. 2012; Ruiz et al. 2013).

There are a variety of pyrolysis techniques available, with the most significant variables including temperature, pyrolysis time, and type of atmosphere (hydrous or anhydrous, anoxic or oxic). Traditional flash pyrolysis techniques rapidly heat organic samples to high temperatures ($>500^{o}C$), held for several seconds, in an oxygen-limited environment (Berwick et al. 2010a). By rapidly heating the sample in an inert atmosphere, there is limited time for the many radicals produced to participate in secondary reactions (Galletti

and Bocchini 1995; Silverio et al. 2008), a certain proportion of which are inevitable. Secondary reactions, and also extensive degradation of the sample, can be problematic for structural interpretation (Berwick et al. 2010a). Flash pyrolysis techniques also traditionally struggle to effectively represent structural units of high polarity (Matilainen et al. 2011), though lower temperatures can overcome this to some extent (Greenwood et al., 2006; Berwick et al., 2010a). Despite its limitations, flash pyrolysis with GCMS detection is one of the most common methods for studying OM (Schulten and Gleixner 1999; Page 2006; Greenwood et al. 2012).

Many different pyrolysers have been used to analyse OM, including platinum filament and Curie point pyrolysers (Parsi et al. 2007). Platinum filament pyrolysers have been widely used and are probably the most commonly applied pyrolysis device. Quartz sample tubes are used to hold the sample and therefore the sample does not come into direct contact with the heated filament element (Parsi et al. 2007). This means that the exact temperature of pyrolysis is difficult to determine.

Curie point pyrolysis and double-shot pyrolysis are two pyrolysis techniques that use a specific pyrolysis temperature, known as the Curie point, of the ferromagnetic sample holder made of a wire alloy (Leinweber et al. 2013). The Curie point temperature is the temperature, between 300-800°C, at which the sample holder's permanent magnetism changes to induced magnetism. As Curie point pyrolysis is limited to only heating the sample to the Curie point temperature of the wire used to hold the sample (Mulder et al. 1992; Hatcher et al. 2001; Parsi et al. 2007), there is little flexibility with the applied thermal conditions. Double-shot pyrolysis uses two temperature steps in order to distinguish between thermally labile and more thermally stable compounds (Leinweber et al. 2013).

Several analytical pyrolysis techniques have also been developed in attempts to better control the fragmentation process. These include microscale sealed vessel pyrolysis (MSSVpy) and hydropyrolysis (HyPy) (Berwick et al. 2010a; Berwick et al. 2010b; Matilainen et al. 2011). These pyrolysis devices also couple to GCMS for the separation and detection of the thermally generated fragments (Berwick et al. 2010b). With MSSVpy, the sample is contained in a small glass tube, i.e., the microscale sealed vessel (MSSV), and typically heated over several days at a temperature in the range of 250-350°C. Once the pyrolysis has completed, the pyrolysates (volatile fragments) are released by cracking the

glass MSSV in a purpose built injector fitted to the GC. The pyrolysates are transferred to the capillary column of the GC with inert, typically Helium, carrier gas (Berwick et al. 2010b). The technique of MSSVpy can compliment flash pyrolysis techniques by providing additional molecular information as it reduces the occurrence of polar OM moieties and increases the yield of products amenable to GC detection (Matilainen et al. 2011; Greenwood et al. 2012). This approach has been shown to provide better control than traditional fast pyrolysis over the pyrolytic degradation of biochemicals (Berwick et al. 2010a). Previous studies have successfully applied this technique to the identification of bacterial biomarkers in OM samples (Greenwood et al. 2006) and the study of petroleum generation kinetics (Berwick et al. 2010a). It has been identified as a promising new approach for the characterisation of OM, although further developments with this technique are needed for more advanced characterisation of OM (Matilainen et al. 2011).

Hydropyrolysis is another method that employs a mild thermal regime, typically with a slow heating rate of 8 ^{o}C.min^{-1} from 300 to 500°C (Rocha et al. 1997b; Berwick et al. 2010b). High pyrolysis temperatures are unnecessary with the addition of a metal sulphide catalyst to the sample. The thermal reactor where pyrolysis occurs, is kept at a high pressure (typically ›10 MPa) using hydrogen gas, thus facilitating the rapid removal of pyrolysates from the system for analysis (Rocha et al. 1997a; Berwick et al. 2010b). These conditions allow for the "soft" release of pyrolysates with minimal structural rearrangement. A large yield of pyrolysates have been obtained in many HyPy studies, as documented by Berwick et al. (2010b).

The aim of pyrolysis is to generate molecular fragments which are characteristic of the original organic structure. However, there are a number of analytical challenges to the technique, including the extensive fragmentation, the secondary reaction of primary products or the inefficient transfer of products from the pyrolysis device to the GC, all of which can limit the representation of the parent material (Galletti and Bocchini 1995; Leinweber et al. 2013). Oxygen-containing macromolecules have been shown to be particularly vulnerable to de-functionalisation and rearrangement processes (Hatcher et al. 2001; Zang and Hatcher 2002; Leinweber et al. 2013). A number of common products from de-functionalisation and rearrangements have been listed by Leinweber et al. (2013); however, in some instances these products can also represent primary fragments from certain samples so their structural value cannot always be disregarded. Challenges such as

these have been the impetus for the development of "softer" pyrolysis techniques, such as MSSVpy and HyPy discussed above, which better control the fragmentation process.

Transfer of products from the pyrolyser to the analytical system (GCMS) can be hindered by cold spots along the interface or poorly swept areas, i.e., dead volumes, (Parsi et al. 2007), which can be more problematic for less volatile, high MW analytes. Consequently, this may result in a MW bias to detection.

Polar compounds are difficult to chromatographically separate through GC which is essentially a non-polar analytical method (Matilainen et al. 2011; Greenwood et al. 2012). Thus polar structures are often under-represented by traditional py-GCMS analysis. Several pyrolysis techniques have been developed to address this problem. Thermochemolysis was developed in the 1990's as an alternative for the study of lignin (Hatcher and Clifford 1994; Li et al. 2012). Thermochemolysis involves the addition of an alkylating reagent, most commonly tetramethylammonium hydroxide (TMAH), to a sample before heating (Shadkami and Helleur 2010). This results in the formation of methylated products, allowing acids in particular to be easily separated compared to their polar unmethylated counterparts (McKinney et al. 1995; Joll et al. 2003; Iwai et al. 2013). Furthermore, thermally promoted cyclisation and aromatisation reactions are reduced during thermochemolysis (Iwai et al. 2013).

Pyrolysis-GCMS has been used extensively for the study of various OM samples due to the detail provided at the molecular level (Greenwood et al. 2002; Greenwood et al. 2006; Chow et al. 2009; Leinweber et al. 2013). The extensive structural fragmentation of most condensed organic samples induced by py-GCMS represents a detailed molecular "fingerprint" of the organic material (Tas and van der Greef 1994; Galletti and Bocchini 1995). In particular, py-GCMS has been widely applied to SOM, DOM and SeOM (Galletti and Bocchini 1995; Hatcher et al. 2001; Zang and Hatcher 2002; Parsi et al. 2007; Alvarez et al. 2013). Its analytical application has been used to study structure, source and importance of OM in these environments. In a recent study by Alvarez et al. (2013), for example, humic acids were studied from different depths in the soil in the Mediterranean region. Here, pyrolysis results showed that water penetration was dependent on the OM content of the soil.

Organic matter in water and WW have been studied widely by analytical pyrolysis. Greenwood et al. (2012) used MSSV py-GCMS to study and compare DOM in WW effluents and several primary source waters. From these samples, MSSV py-GCMS was able to identify high concentrations of microbial metabolites, likely from the concentrated biota associated with WW treatment, as well as trace amounts of anthropogenics including pharmaceuticals and personal care products which can be a major issue for the toxicity and environmental fate of DOM (Ratola et al. 2012; Ruiz et al. 2013).

In another elegant example of the utility of MSSVpy, hopane biomarkers were generated from the bacteriohopanepolyols of bacterial isolate (Greenwood et al., 2006) as well as the bacterial populations of WW (Berwick et al. 2010b). Figure 3 shows a partial *m/z* 191 chromatogram highlighting the hopanoid biomarkers generated from *F. Aurntia* (Greenwood et al. 2006), as also listed in Table 2. In Figure 3, *m/z* 191 corresponds to the major A and B ring fragments diagnostic of the pentacylcic hopanoid structure.

Analytical pyrolysis has made an important contribution to our understanding of the structural make-up of naturally occurring OM. With a diverse range of pyrolysis approaches now available and which continue to evolve, it has the ability to continue to advance the molecular understanding of OM and the role it plays in different environments.
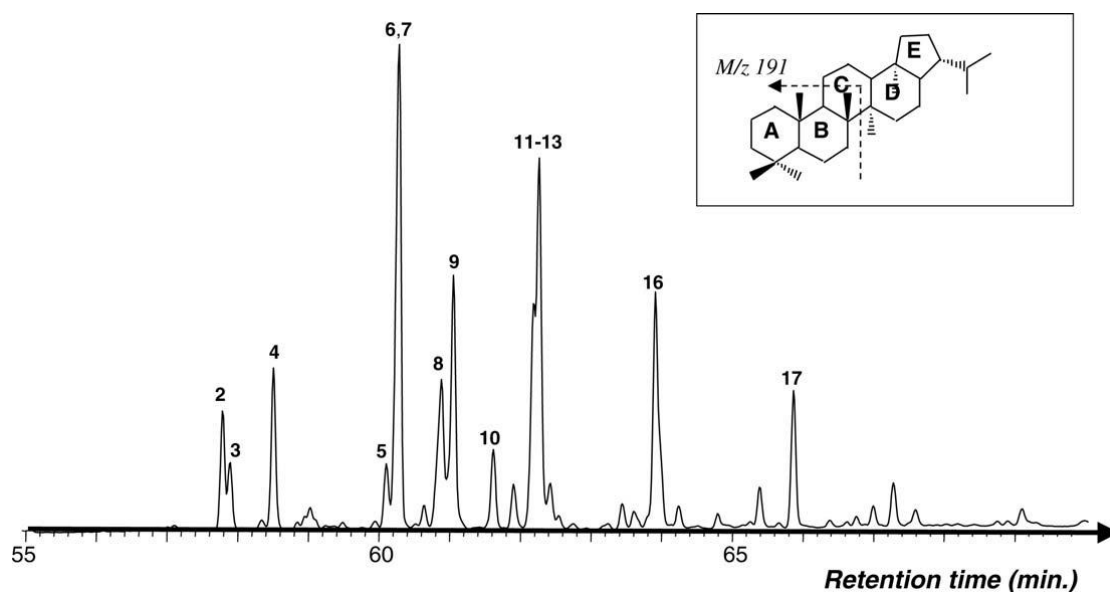
Figure 3 A partial *m/z* 191 chromatogram showing the hopanoid biomarker identified by MSSV py-GCMS of a bacterial isolate. Sourced from Greenwood et al. (2006). The peak assignments correspond to the products listed in Table 2.

Table 2 Bacterial biomarkers (hopanoid compounds) from the MSSV py-GCMS of a bacterial isolate. Sourced from Greenwood et al. (2006).

Peak nos.	Abbreviation	Compound
1	T_s	18α(H)-22,29,30-trisnorhopane
2	C_{27}H:1	Monounsaturated C_{27} Hopene
3	T_m	17α(H)-22,29,30-trisnorhopane
4	27β(H)	17β(H)-22,29,30-trisnorhopane
5	C_{29}H:1	Monounsaturated C_{29} hopene
6	C_{29}H	17α,21β-30-norhopane
7	C_{29}H:1	Monounsaturated C_{29} hopene
8	C_{30}H:1	Monounsaturated C_{30} hopene
9	C_{29}βα	17β,21α-30-normoretane
10	C_{30}H	17α,21β-hopane
11	C_{29}ββ	17β,21β-30-norhopane
12	C_{30}H:1	Monounsaturated C_{30} hopene
13	C_{30}βα	17β,21α-moretane
14	C_{31}HS	17α,21β-30-homohopane (22R)
15	C_{32}HS	17α,21β-30-homohopane (22R)
16	C_{30}ββ	17β,21β-hopane
17	C_{31}ββ	17β,21β-30-homohopane

HPSEC

High performance size exclusion chromatography (HPSEC) is not a characterisation technique per se, but rather a fractionation technique that can be coupled to another technique for sample characterisation. The earlier separation technique of size exclusion chromatography (SEC) was first applied to the separation of biomolecules (starch) in 1955 by Lindqvist and Storgards (1955) (Hong et al. 2012). By applying pressure to the SEC column, fractionation is enhanced and the method is referred to as HPSEC. The ability of HPSEC to separate and characterise DOM without pre-extraction makes it an attractive

approach for the study of OM (Wu et al. 2007). This technique is also rapid, sensitive and reproducible and has improved the understanding of the nature and role of OM (Vuorio et al. 1998; Wu et al. 2007; Chow et al. 2008). Being a solution technique it is only applicable to the soluble fraction of OM.

High performance size exclusion chromatography fractionates DOM based on molecular size (Vuorio et al. 1998; Chow et al. 2008; Matilainen et al. 2011). This is achieved by the elution of DOM through a column of porous solid phase material, commonly silica, using a buffer solution such as phosphate or sodium chloride (Vuorio et al. 1998; Chow et al. 2008; Woods et al. 2010; Lewis et al. 2012). The smaller OM molecules penetrate the pores of the solid phase, allowing the larger molecules to be eluted more quickly (Potschka 1993). Different solid phase particle sizes are available, which influences the movement of organics through the column. The apparent molecular weight (AMW) of each fraction that is eluted can then be determined by calibration of the elution time using standards of known MW.

The standards commonly used for calibration include various MWs of polyethylene glycol or polystyrene sulfonate (Chow et al. 2008; Lewis et al. 2011; Lewis et al. 2012). Polystyrene sulfonates are particularly suitable as standards as they can be produced as a homologous series that have been independently shown to have comparable MWs (Beckett et al. 1987) and similar ion exclusion properties (Berden and Berggren 1990) to humic substances (Alberts et al. 2002). Although these standards may not provide the most accurate representation of the true molecular size of the OM studied, they do provide an apparent measure of the MW that can be reproducible and provide consistency for comparisons between chromatograms collected at different times (Chow et al. 2008).

Elution time is decreased, and hence fractionation increased through the use of pressure, hence the alternative and equally appropriate expansion of the acronym HPSEC as high pressure size exclusion chromatography. When different pressures are used in a study, fractionation can be referred to as low pressure size exclusion chromatography to distinguish between the fractionation techniques; this terminology is used in the studies by Woods et al. (2010) and Comte et al. (2007). However to avoid confusion, HPSEC will be referred to as high performance size exclusion chromatography throughout this thesis.

High performance size exclusion chromatography must be coupled to a detection technique in order to detect or characterise the fractionated organics as they are eluted from the column. Detection techniques most commonly used are UV absorbance (at 254 nm), or chemical determination of DOC or TOC (Alberts et al. 2002; Wu et al. 2007; Chow et al. 2008; Lewis et al. 2012). These approaches are employed to determine DOM properties as a function of molecular size (Wu et al. 2007). However, as discussed briefly above, the ability of the organics to be detected depends on their structure, e.g. conjugated structures are required for absorbance measurements. Therefore the disadvantage is that not all organics may be detected and accounted for by the technique chosen. However if the same detection technique is used consistently, this disadvantage is not severe and is outweighed by the ability to compare between samples and spectra collected at different times. Even more sophisticated techniques, including NMR spectroscopy, have been coupled to HPSEC (Conte et al. 2007; Woods et al. 2010; Nebbioso and Piccolo 2013). These techniques allow for the characterisation of DOM properties as a function of molecular size.

In order to account for all of the organics in a sample, carbon recovery is often calculated. This is done by measuring and comparing the amount of carbon in the original sample and in each of the fractions. However, commonly some of the carbon injected into the HPSEC is undetected or "lost" in the column. This can occur due to the irreversible sorption of organics to the column or by the small and continual elution of organics that can be hard to distinguish from the baseline (Alberts et al. 2002). There is also the possibility of flocculation, degradation and consumption of the organics in the column, which can introduce errors in the interpretation of spectral characteristics (Alberts et al. 2002; Allpike et al. 2005). Spectral features of the organics may naturally vary with molecular weight, particularly larger organics which are often more conjugated and therefore can have greater specific absorbances; this is a particular issue when UV detection is used (Alberts et al. 2002; Allpike et al. 2005).

Quantitative analysis of HPSEC profiles is possible by applying the approach known as peak fitting (Chow et al. 2008; Lewis et al. 2011; Lewis et al. 2012). Software is available that uses auto-fitting algorithms to optimise the fit to the spectra (Chow et al. 2008). However, the same procedure can be undertaken manually, for example using Microsoft Excel. The individual peaks that are fitted to the profile can have a Lorentzian (Lewis et al. 2011; Lewis et al. 2012) or Gaussian shape, depending on what best fits the chromatogram.

The individual peaks can be interpreted as different fractions of organics with different AMW that have been separated and eluted from the column at different times. This manual approach enables the fit of a fixed set of individual peaks to chromatograms, for a series of samples. By keeping the peak positions and widths constant, the peak heights can be changed to reflect the HPSEC results of another organic sample. In this way, the amount of organics represented by the individual peaks can be compared between samples. An example of this is shown in Figure 4 from Lewis et al. (2011), where the influent and effluent organics from two different wastewater treatment simulations were characterised by HPSEC coupled to fluorescence (at 260 nm). Direct comparisons between the different AMW organics can be undertaken by the comparison of the fitted peak heights for each of the samples.
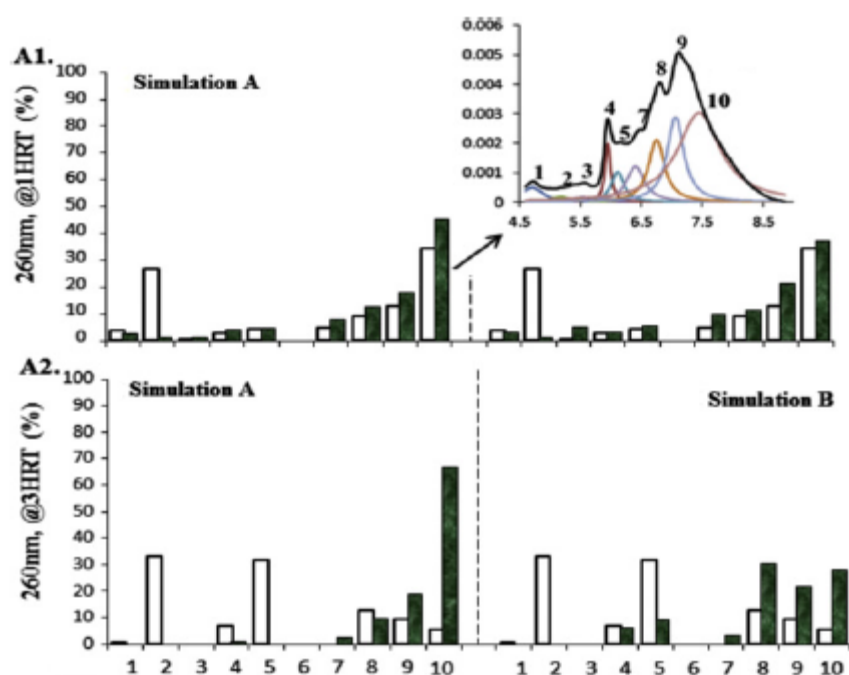
Figure 4 Peak fitting results from the HPSEC-fluorescence analysis of two wastewater treatment simulations (A and B) for the removal of organics. Influent (unfilled bars) and effluent (filled bars) waters from the two wastewater simulations studied by Lewis et al. (2011). One HPSEC profile is shown in the top-right corner, and the individual peaks fitted to this through peak fitting are numbered within this. The numbered fitted peaks correspond to the numbers shown in the column graphs.

Studies of OM using HPSEC have reported that separation of different sized constituents is key to understanding the origins and differences between organic materials (Woods et al. 2010). Of the different techniques that can be used to study the separated fractions, the combination of HPSEC with NMR spectroscopy has been particularly insightful. This

combination is thought to be a significant development for elucidating the structure of DOM from different environments (Woods et al. 2010).

In a study by Woods et al. (2010), it was found that there were clear differences in the composition of different size-distinguished fractions of DOM samples using solution NMR. The largest fraction contained more carbohydrate and aromatic-like structures, the midsized fractions included carboxyl-rich alicyclic molecules, and the smallest included material derived from linear terpenoids. In a separate study by Conte et al. (2007), differences in the structural groups of different sized molecular fractions of humic acid were determined using solid-state ^{13}C NMR spectroscopy. The larger molecular sized fractions were richer in alkyl structural groups, and the amount of oxidized Cs increased as the molecular size of the fractions decreased.

As an analytical tool, HPSEC has become particularly popular for the study of OM in water research (Allpike et al. 2005; Sarathy and Mohseni 2007; Chow et al. 2008). In practice, HPSEC has become an informative approach to separating and studying organics through different stages of water treatment (Chow et al. 2008). The advantage is that this technique, when coupled to a detection technique including UV and DOC, can provide readily available information to optimise water quality targets during the treatment process (Chow et al. 2008). For example, during drinking water treatment, the removal of OM can be determined by comparing the MW profiles before and after treatment. This approach to studying organics during treatment has been used by many (Gjessing et al. 1998; Chow et al. 1999; Bolto et al. 2001; Chow et al. 2002; van Leeuwen et al. 2002; Drikas et al. 2003; Chow et al. 2008). Therefore, with this rapid and reproducible separation technique, fractions can be not only detected but studied to as much detail as desired. This means that HPSEC is versatile and can be applied fit-for-purpose as a rapid separation and detection technique, as well as for more advanced studies of OM structure and role in the environment.

ORGANIC MATTER CHARACTERISATION - A NEW APPROACH

The analytical techniques available today are continually being developed and in turn are encouraging new ways in which we think of OM. In particular, the combined use of multiple analytical approaches has made it possible for more informative assessments on the nature of OM (Simpson et al. 2011). This way of thinking has stemmed from the

premise that any single analytical approach cannot provide complete characterisation (Gaffney et al. 1996; Allpike et al. 2010a; Greenwood et al. 2012). A multi-technique approach is now being used to address a range of different environmental and industry-related problems, including for the assessment of wastewater quality (Lewis et al. 2011; van Leeuwen et al. 2012) and especially for the detection of persistent organic pollutants (POPs) that can influence the health of receiving environments when released in effluents (Greenwood et al. 2012; Ratola et al. 2012).

A key advantage of using multiple characterisation techniques to study an OM sample is that the possibility of missing chemical or structural information is reduced. In order to make an informed decision of which techniques are best to apply, it is critical to have an understanding of the analytical strengths of each technique, so that they are utilised and the weaknesses of each are overcome, as well as an understanding of the chemical and structural features that are of interest to detect. By knowing these details, techniques that provide the most informative assessment of an organic sample will be applied.

Although there are clear benefits to a multi-technique approach to characterise OM, this approach brings with it the disadvantage of added complexity. This complexity comes from the increased amount of information that must be appropriately handled and interpreted. Suitable statistical techniques such as multivariate statistics provide the tools that are needed to address this issue of complexity.

Multivariate statistics provides a way to deal with a set of measurements (variables) for a number of individuals (or subjects) (Everitt 1975; Goodacre et al. 1996). Each variable is thought of as having a different dimension in space (Goodacre et al. 1996). Therefore for n variables, each individual may reside at a unique position in a space of n dimensions. Multivariate statistics provides simplification of these dimensions (which can become large and complex) via mathematical transformations that concentrate key information into two or three dimensions that can be displayed graphically with the least amount of information lost (Goodacre et al. 1996; Simeonov et al. 2003).

Multivariate statistics has been applied to the study of organic samples for decades, due to its ability to handle and study large data sets with multiple variables. For example, when applied to spectroscopy, the variables are generally properties such as wavelength or absorbance (Goodacre et al. 1996; Guo et al. 2013). There are a variety of multivariate

methods available and choosing an appropriate method to apply depends on the information-type to be used and the aim of the analysis. The most recognised and longest established technique of multivariate statistics that has been applied to the study of organics is principle component analysis (PCA) (Clarke and Warwick 2001; Jolliffe 2002).

Principle component analysis is an ordination method. An ordination method produces a map in two or three dimensions, where the placement of samples in the map is a reflection of the similarity between the data (Clarke and Warwick 2001). More precisely, the distances between the samples in the ordination is an attempt to reflect the corresponding dissimilarities in the data (Clarke and Warwick 2001). The purpose of PCA is to simplify the dimensionality of a data set that has a large number of variables, whilst preserving the variation present (Jolliffe 2002). The PCA method converts a set of observations of variables that are possibly correlated into a set of values that are linearly uncorrelated and referred to as principle components (Jolliffe 2002). There is either the same number or less principle components generated as the number of variables used. The results of PCA are displayed on an ordination plot where there is one principle component axis per subject (Clarke and Warwick 2001). For example, for a two-dimensional ordination plot, the two different axes represent two different principle components. Generation of principle component axes involves only linear transformations. The first principle component accounts for the greatest amount of variance in the results projected along the axis (Clarke and Warwick 2001). The next principle component fitted to the same data must be perpendicular, and therefore uncorrelated, to the previous principle component axis and accounts for the greatest amount of variance for the remaining variables (Clarke and Warwick 2001). As long as the data set is normally distributed, the principle component axes will be independent of each other. Although this method is conceptually simple, a linear representation has little flexibility and this can cause problems. There are numerous recent examples of its use in OM studies (Thomsen et al. 2002; Choi et al. 2013; Fernandez-Getino et al. 2013; Guo et al. 2013; Hu et al. 2013). In each of these studies, PCA was used to detect and describe differences between OM samples based on characteristics determined by different analytical techniques.

The more recently developed ordination method of non-parametric multidimensional scaling (nMDS) has the ability to accurately represent complex relationships between subjects in low-dimensions making it one of the best ordination methods available (Everitt 1978; Clarke and Warwick 2001; Clarke and Gorley 2006). The nMDS method

was developed by Shepard and Kruskal in the early 1960's for use in psychology (Shepard 1962a, b; Kruskal 1964; Clarke and Warwick 2001). In an nMDS plot, samples that are similar are positioned close together in the plot, and if they are dissimilar, they are positioned further apart.

To generate an nMDS plot, a resemblance matrix must be constructed from the data (Clarke and Warwick 2001; Clarke and Gorley 2006). A resemblance matrix is constructed by ranking the data for each sample based on similarities or dissimilarities between variables. There are different ways in which ranking can be determined, and this depends on the type of data being used. For example, for data sets where each variable is completely independent of other variables measured, such as for biological assemblage data, the resemblance matrix should be calculated using Bray-Curtis similarity (Clarke and Gorley 2006). When considering different characterisation techniques, this would be the case for py-GCMS-type data, as the occurrence of molecular ions in a chromatogram is independent of the occurrence of others. For data where variables are dependent on other variables measured or where variables do not have comparable ranges, Euclidean or Manhattan Distance is more appropriate (Clarke and Gorley 2006). This approach would be used to prepare NMR and HPSEC-type data, as their trace has points that are dependent on the previous point. All nMDS plots generated are done so with the lowest possible stress. The associated stress reflects the degree of confidence in the two dimensional representation of the multivariate data (Clarke and Gorley 2006). The degree of stress is represented by a number, which if below 0.2 (as a general rule) is thought to preserve the rank order of the data. The nMDS method can be preferred over the commonly used PCA method as outliers can be included in analyses and both types of data discussed (dependent and independent) can be analysed. With PCA, outliers are removed before analyses as they have a large effect on the ordination, and PCA can only represent data with completely independent variables.

Non-parametric multidimensional scaling ordination plots have been applied to the study of OM. As with the PCA method, these ordination plots have been applied to facilitate comparison between organic samples. In a study by Liu et al. (2012), the remediation of petroleum hydrocarbons in soil was studied to determine the effects of SOM and ageing of the hydrocarbons in the contaminated soil. Here, nMDS plots helped to demonstrate that the ageing of the hydrocarbons in the soil affected the bacterial communities present. A study by Pearson et al. (2007) demonstrated the ability of multivariate statistics to handle

large amounts of data. In this study, nMDS plots were used to explore relationships between lipid compounds collected from different lake surface sediments and the major trends in the source inputs. The nMDS approach allowed samples to be grouped according to their dominant lipid concentration, thus showing a distinction between compounds originating from bacterial, algal, macrophyte or terrestrial plant sources.

Non-parametric multidimensional scaling ordination plots can reflect similarities between organic samples determined by different analytical assessments. The major benefit of being able to generate an ordination plot based on results from each analytical technique is that it renders data from different techniques directly comparable. With the data in a similar format, multivariate statistics can be used to measure how closely related two different sets of multivariate data are. The Spearman Rank Correlation method is capable of doing this using the same information as used to generate the ordination plots (Clarke and Warwick 2001; Clarke and Gorley 2006). This method is a mantel-type test that measures the relationship between two sets of multivariate data (resemblance matrices) and displays this as a value: ρ (rho). More specifically, it has the capability to provide a measure of the amount of complementary information and distinctive information provided by each technique for a set of organic samples. This multivariate statistical approach is thought to be one of the first capable of assessing the amount of information overlap provided by different characterisation techniques.

OBJECTIVES OF THIS RESEARCH

The objectives of the research discussed in this thesis were to: (i) develop a multivariate statistical approach that facilitates the comparison of data on OM generated using multiple analytical techniques; and (ii) apply this approach to several different situations associated with OM characterisation during WW processing. The initial development phase of the approach involved data collected on a range of plant materials from two analytical techniques, solid-state ^{13}C NMR spectroscopy and flash py-GCMS. The different data were qualitatively compared with the aid of nMDS plots and quantitatively compared using the Spearman Rank Correlation method. The demonstrated approach was then applied to three separate case studies, the first of sediment organics in a pulp and paper mill receiving environment. The overall approach was expanded to include a third analytical technique, HPSEC, and applied to a suite of pulp and paper mill water and WW

dissolved organics, and finally applied to a suite of partially and fully treated sewage effluent organics from the United Kingdom.

REFERENCES

Ahangar AG (2010) Sorption of PAHs in the soil environment with emphasis on the role of soil organic matter: a review. World Appl Sci J 11:759-765

Akagi J, Zsolnay A, Bastida F (2007) Quantity and spectroscopic properties of soil dissolved organic matter (DOM) as a function of soil sample treatments: Air-drying and pre-incubation. Chemosphere 69:1040-1046

Alberts JJ, Takacs M, Egeberg PK (2002) Total luminescence spectral characteristics of natural organic matter (NOM) size fractions as defined by ultrafiltration and high performance size exclusion chromatography (HPSEC). Org Geochem 33:817-828

Alexander M (2000) Aging, bioavailability, and overestimation of risk from environmental pollutants. Environ Sci Technol 34:4259-4265

Allpike B, Busetti F, Heitz A, Berwick L, Warton B, Garbin S, Alessandrino M, Eckstaedt SVv, White D, Grice K, Greenwood P, Kagi R (2010a) Advanced Characterisation of Natural Organic Matter (NOM) in Australian Water Supplies. 80, Water Quality Research Australia Adelaide

Allpike B, Heitz A, Berwick L, Joll CA (2010b) Thermal and Chemical Degradation Techniques Applied to NOM Characterisation. Characterisation of Aquatic Natural Organic Matter using Direct Hydrous Pyrolysis., Water Quality Research Australia Limited

Allpike BP, Heitz A, Joll CA, Kagi RI, Abbt-Braun G, Frimmel FH, Brinkmann T, Her N, Amy G (2005) Size Exclusion Chromatography to Characterise DOC Removal in Drinking Water Treatment. Environ Sci Technol 30:2334-2342

Almendros G, Guadalix ME, GonzalezVila FJ, Martin F (1996) Preservation of aliphatic macromolecules in soil humins. Org Geochem 24:651-659

Alvarez AM, Carral P, Hernandez Z, Almendros G (2013) Assessment of Soil Organic Matter Molecular Characteristics Related to Hydrophysical Properties in Semiarid Soils (Central Spain). Arid Land Res Manag 27:303-326

Arnold WR, Cotsifas JS, Ogle RS, DePalma SGS, Smith DS (2010) A comparison of the copper sensitivity of six invertebrate species in ambient salt water of varying dissolved organic matter concentrations. Environmental Toxicology and Chemistry 29:311-319

Baldock JA, Currie GJ, Oades JM (1991) Organic Matter as Seen by Solid State ^{13}C NMR and Pyrolysis Tandem Mass Spectrometry. In: Wilson WS (ed) Advances in Soil Organic Matter Research: The Impact on Agriculture and the Environment, Royal Society of Chemistry, Cambridge, pp 45-60

Barton DHR, Schnitzer M (1963) A new experimental approach to humic acid problem. Nature 198:217-&

Beckett R, Jue Z, Giddings JC (1987) Determination of molecular-weight distributions of fluvic and humic acids using flow field-flow fractionation. Environ Sci Technol 21:289-295

Berden M, Berggren D (1990) Gel-filtration chromatography of humic substances in soil solutions using HPLC-determination of the molecular-weight distribution. J Soil Sci 41:61-72

Berwick L, Greenwood PF, Smernik RJ (2010a) The use of MSSV pyrolysis to assist the molecular characterisation of aquatic natural organic matter. Water Res 44:3039-3054

Berwick LJ, Greenwood PF, Meredith W, Snape CE, Talbot HM (2010b) Comparison of microscale sealed vessel pyrolysis (MSSVpy) and hydropyrolysis (Hypy) for the characterisation of extant and sedimentary organic matter. J Anal Appl Pyrol 87:108-116

Bolan NS, Adriano DC, Kunhikrishnan A, James T, McDowell R, Senesi N (2011) Dissolved organic matter: biochemistry, dynamics, and environmental significance in soils. In: Sparks DL (ed) Advances in Agronomy, Vol 110, Advances in Agronomy, Elsevier Academic Press Inc, San Diego, pp 1-75

Bolto B, Dixon D, Eldridge R, King S (2001) Cationic polymer and clay or metal oxide combinations for natural organic matter removal. Water Res 35:2669-2676

Bonin JL, Simpson MJ (2007) Variation in phenanthrene sorption coefficients with soil organic matter fractionation: The result of structure or conformation? Environ Sci Technol 41:153-159

Brady NC, Weil RR (2008) The Nature and Properties of Soils. Upper Saddle River, N.J. : Pearson Prentice Hall

Brauniger T, Jansen M (2013) Solid-state NMR Spectroscopy of Quadrupolar Nuclei in Inorganic Chemistry. Z Anorg Allg Chem 639:857-879

Brown SP (2012) Applications of high-resolution H-1 solid-state NMR. Solid State Nucl Mag 41:1-27

Cao XY, Yang J, Mao JD (2013) Characterization of kerogen using solid-state nuclear magnetic resonance spectroscopy: A review. Int J Coal Geol 108:83-90

Carpenter LJ, Archer SD, Beale R (2012) Ocean-atmosphere trace gas exchange. Chem Soc Rev 41:6473-6506

Celerier J, Rodier C, Favetta P, Lemee L, Ambles A (2009) Depth-related variations in organic matter at the molecular level in a loamy soil: reference data for a long-term experiment devoted to the carbon sequestration research field. Euro J Soil Sci 60:33-43

Chefetz B, Chen Y, Hadar Y, Hatcher PG (1998a) Characterization of Dissolved Organic Matter Extracted from Composted Municipal Solid Waste. Soil Sci Soc Am J 62:326-332

Chefetz B, Hader Y, Chen Y (1998b) Dissolved Organic Carbon Fractions Formed during Composting of Municipal Solid Waste: Properties and Significance. Acta hydroch hydrob 26:172-179

Chefetz B, Xing B (2009) Relative Role of Aliphatic and Aromatic Moieties as Sorption Domains for Organic Compounds: A Review. Environ Sci Technol 43:1680-1688

Chen Y, Senesi N, Schnitzer M (1978) Chemical and physical characteristics of humic and fluvic acids extracted from soils of mediterranean region. Geoderma 20:87-104

Choi BG, Cho J, Song KG, Maeng SK (2013) Correlation between effluent organic matter characteristics and membrane fouling in a membrane bioreactor using advanced organic matter characterization tools. Desalination 309:74-83

Chow C, Cook D, Drikas M (2002) Evaluation of magnetic ion exchange resin (MIEX (R)) and alum treatment for formation of disinfection by-products and bacterial regrowth. In: Wilderer P, 2nd World Water Congress: Water and Health-Microbiology, Monitoring and Disinfection. Water Science and Technology: Water Supply. IWA Publishing, London, pp. 267-274

Chow CWK, van Leeuwen JA, Drikas M, Fabris R, Spark KM, Page DW (1999) The impact of the character of natural organic matter in conventional treatment with alum. Water Sci Technol 40:97-104

Chow CWK, Fabris R, van Leeuwen J, Wang D, Drikas M (2008) Assessing Natural Organic Matter Treatability Using High Performance Size Exclusion Chromatography. Environ Sci Technol 42:6683-6689

Chow CWK, van Leeuwen JA, Fabris R, Drikas M (2009) Optimised coagulation using aluminium sulfate for the removal of dissolved organic carbon. Desalination 245:120-134

Clark LL, Ingall ED, Benner R (1999) Marine organic phosphorus cycling: Novel insights from nuclear magnetic resonance. American Journal of Science 299:724-737

Clarke KR, Warwick RM (2001) Change in Marine Communities: An Approach to Statistical Analysis and Interpretation. PRIMER-E Ltd, Plymouth Marine Laboratory, UK

Clarke KR, Gorley RN (2006) PRIMER v6: User Manual/Tutorial. PRIMER-E, Plymouth

Comte S, Guibaud G, Baudu M (2007) Effect of extraction method on EPS from activated sludge: An HPSEC investigation. J Hazard Mater 140:129-137

Conte P, Spaccini R, Smejkalova D, Nebbioso A, Piccolo A (2007) Spectroscopic and conformational properties of size-fractions acid separated from a lignite humic. Chemosphere 69:1032-1039

de Blas E, Almendros G, Sanz J (2013) Molecular characterization of lipid fractions from extremely water-repellent pine and eucalyptus forest soils. Geoderma 206:75-84

Derenne S, Largeau C (2001) A review of some important families of refractory macromolecules: Composition, origin, and fate in soils and sediments. Soil Science 166:833-847

Deshmukh AP, Chefetz B, Hatcher PG (2001) Characterization of organic matter in pristine and contaminated coastal marine sediments using solid-state 13C NMR, pyrolytic and thermochemolytic methods: a case study in the San Diego harbor area. Chemosphere 45:1007-1022

Dilly O, Bloem J, Vos A, Munch JC (2004) Bacterial diversity in agricultural soils during litter decomposition. Appl Environ Microb 70:468-474

Dinar E, Taraniuk I, Graber ER, Katsman S, Moise T, Anttila T, Mentel TF, Rudich Y (2006) Cloud Condensation Nuclei properties of model and atmospheric HULIS. Atmos Chem Phys 6:2465-2481

Dinar E, Taraniuk I, Graber ER, Anttila T, Mentel TF, Rudich Y (2007) Hygroscopic growth of atmospheric and model humic-like substances. J Geophys Res-Atmos 112:D05211

Drikas M (2003) Natural organic matter in drinking water: problems and solutions, Water Quality Research Australia Limited.

Drikas M, Chow CWK, Cook D (2003) The impact of recalcitrant organic character on disinfection stability, trihalomethane formation and bacterial regrowth: An evaluation of magnetic ion exchange resin (MIEX (R)) and alum coagulation. J Water Supply Res T 52:475-487

Duarte RMBO, Duarte AC (2011) A critical review of advanced analytical techniques for water-soluble organic matter from atmospheric aerosols. Trends Anal Chem 30:1659-1671

Ebersviller S, Lichtveld K, Sexton KG, Zavala J, Lin YH, Jaspers I, Jeffries HE (2012) Gaseous VOCs rapidly modify particulate matter and its biological effects - Part 2: Complex urban VOCs and model PM. Atmos Chem Phys 12:12293-12312

Enowashu E, Poll C, Lamersdorf N, Kandeler E (2009) Microbial biomass and enzyme activities under reduced nitrogen deposition in a spruce forest soil. Appl Soil Ecol 43:11-21

Esteves VI, Otero M, Duarte AC (2009) Comparative characterisation of humic substances from the open ocean, estuarine water and fresh water. Org Geochem 40:942-950

Everitt (1978) Graphical Techniques for Multivariate Data. Heineman, London

Everitt BS (1975) Multivariate Analysis: the Need for Data, and other Problems. Brit J Psychiat 126:237-240

Feitosa RS, Sodre FF, Maldaner AO (2013) Drugs of abuse in waters and wastewaters: occurrence, analytical determination and forensic applications. Quim Nova 36:291-305

Feng XJ, Simpson AJ, Simpson MJ (2006) Investigating the role of mineral-bound humic acid in phenanthrene sorption. Environ Sci Technol 40:3260-3266

Fernandez-Getino AP, Hernandez Z, Buena AP, Almendros G (2013) Exploratory analysis of the structural variability of forest soil humic acids based on multivariate processing of infrared spectral data. Eur J Soil Sci 64:66-79

Fomba KW, Galvosas P, Roland U, Karger J, Kopinke FD (2009) New Option for Characterizing the Mobility of Organic Compounds in Humic Acids. Environ Sci Technol 43:8264-8269

Fors EO, Rissler J, Massling A, Svenningsson B, Andreae MO, Dusek U, Frank GP, Hoffer A, Bilde M, Kiss G, Janitsek S, Henning S, Facchini MC, Decesari S, Swietlicki E (2010) Hygroscopic properties of Amazonian biomass burning and European background HULIS and investigation of their effects on surface tension with two models linking H-TDMA to CCNC data. Atmos Chem Phys 10:5625-5639

Gaffney JS, Marley NA, Clark SB (1996) Humic and Fulvic Acids and Organic Colloidal Materials in the Environment. In: Gaffney JS, Marley NA , Clark SB (ed) Humic and Fulvic Acids. Isolation, Structure and Environmental Role, ACS Symposium Series, American Chemical Society, Washington, pp 2-16

Galletti GC, Bocchini P (1995) Pyrolysis/gas chromatography/mass spectrometry of lignocellulose. Rapid Commun Mass Sp 9:815-826

Ghiorse WC, Wilson JT (1988) Microbial Ecology of the terrestrial subsurface. Adv Appl Microbiol 33:107-172

Gigliotti G, Kaiser K, Guggenberger G, Haumaier L (2002) Differences in the chemical composition of dissolved organic matter from waste material of different sources. Biol Fert Soils 36:321-329

Gjessing ET, Alberts JJ, Bruchet A, Egeberg PK, Lydersen E, McGown LB, Mobed JJ, Munster U, Pempkowiak J, Perdue M, Ratnawerra H, Rybacki D, Takacs M, Abbt-Braun G (1998) Multi-method characterisation of natural organic matter isolated from water: Characterisation of reverse osmosis isolates from water of two semi-identical dystrophic lakes basins in Norway. Water Res 32:3108-3124

Golding CJ, Smernik RJ, Birch GF (2004) Characterisation of sedimentary organic matter from three south-eastern Australian estuaries using solid-state ^{13}C-NMR techniques. Mar Freshwater Res 55:285-293

Goodacre R, Neal MJ, Kell DB (1996) Quantitative Analysis of Multivariate Data Using Artificial Neural Networks: A Tutorial Review and Applications to the Deconvolution of Pyrolysis Mass Spectra. Zentralb Bakteriol 284:516-539

Grandy A, Sinsabaugh R, Neff J, Stursova M, Zak D (2008) Nitrogen deposition effects on soil organic matter chemistry are linked to variation in enzymes, ecosystems and size fractions. Biogeochemistry 91:37-49

Greenwood PF, van Heemst JDH, Guthrie EA, Hatcher PG (2002) Laser micropyrolysis GC-MS of lignin. J Anal Appl Pyrol 62:365-373

Greenwood PF, Leenheer JA, McIntyre C, Berwick L, Franzmann PD (2006) Bacterial biomarkers thermally released from dissolved organic matter. Org Geochem 37:597-609

Greenwood PF, Berwick LJ, Croué JP (2012) Molecular characterisation of the dissolved organic matter of wastewater effluents by MSSV pyrolysis GC–MS and search for source markers. Chemosphere 87:504-512

Gu BH, Schmitt J, Chen ZH, Liang LY, McCarthy JF (1994) Adsorption and desorption of natural organic-matter on iron-oxide - mechanisms and models. Environ Sci Technol 28:38-46

Guo XJ, Yuan DH, Jiang JY, Zhang H, Deng Y (2013) Detection of dissolved organic matter in saline-alkali soils using synchronous fluorescence spectroscopy and principal component analysis. Spectrochim Acta A 104:280-286

Hallquist M et al. (2009) The formation, properties and impact of secondary organic aerosol: current and emerging issues. Atmos Chem Phys 9:5155-5236

Hanninen K (2010) Historical and current progress in understanding the origin and structure of humic substances. Chem Ecol 26:1-11

Hatcher PG, Clifford DJ (1994) Flash pyrolysis and in-situ methylation of humic acids from soil. Org Geochem 21:1081-1092

Hatcher PG, Dria KJ, Kim S, Frazier SW (2001) Modern analytical studies of humic substances. Soil Sci 166:770-794

Hayes MHB, Clapp CE (2001) Humic substances: Considerations of compositions, aspects of structure, and environmental influences. Soil Sci 166:723-737

Hedges JI, Oades JM (1997) Comparative organic geochemistries of soils and marine sediments. Org Geochem 27:319-361

Hedges JI, Keil RG (1999) Organic geochemical perspectives on estuarine processes: sorption reactions and consequences. Marine Chemistry 65:55-65

Hedges JI, Eglinton G, Hatcher PG, Kirchman DL, Arnosti C, Derenne S, Evershed RP, Kögel-Knabner I, de Leeuw JW, Littke R, Michaelis W, Rullkötter J (2000) The molecularly-uncharacterized component of nonliving organic matter in natural environments. Org Geochem 31:945-958

Henrichs SM (1992) Early diagenesis of organic matter in marine sediments: progress and perplexity. Mar Chem 39:119-149

Hepplewhite C, Newcombe G (2001) Characterisation of NOM ultrafiltration fractions isolated using three techniques. In: Swift RS , Spark KM (ed) Understanding and Managing Organic Matter in Soils, Sediments and Waters, International Humic Substances Society, pp 97-106

Hong P, Koza S, Bouvier ESP (2012) A review size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. J Liq Chromatogr R T 35:2923-2950

Houghton RA (2003) The Contemporary Carbon Cycle. In: Holland HD , Turekian KK (ed) Treatise on Geochemistry, Peramon, Oxford, pp 473-513

Hu LM, Shi XF, Guo ZG, Wang HJ, Yang ZS (2013) Sources, dispersal and preservation of sedimentary organic matter in the Yellow Sea: The importance of depositional hydrodynamic forcing. Mar Geol 335:52-63

Ito T, Shichi W, Ichida M, Gotoh H, Kamada H, Ando H (2011) Effect of electron spin-spin exchange interaction on spin precession in coupled quantum well. Journal of Applied Physics 110:033109

Iwai H, Fukushima M, Yamamoto M, Komai T, Kawabe Y (2013) Characterization of seawater extractable organic matter from bark compost by TMAH-py-GC/MS. J Anal Appl Pyrol 99:9-15

Jimenez JL et al. (2009) Evolution of Organic Aerosols in the Atmosphere. Science 326:1525-1529

Joll CA, Huynh T, Heitz A (2003) Off-line tetramethylammonium hydroxide thermochemolysis of model compound aliphatic and aromatic carboxylic acids:: Decarboxylation of some ortho- and/or para-substituted aromatic carboxylic acids. J Anal Appl Pyrol 70:151-167

Jolliffe IT (2002) Principle Component Analysis. Springer-Verlag New York, Inc, New York

Jurado E, Dachs J, Duarte CM, Simo R (2008) Atmospheric deposition of organic and black carbon to the global oceans. Atmos Environ 42:7931-7939

Kalbitz K, Solinger S, Park JH, Michalzik B, Matzner E (2000) Controls on the dynamics of dissolved organic matter in soils: A review. Soil Sci 165:277-304

Karathanasis AD (1999) Subsurface migration of copper and zinc mediated by soil colloids. Soil Sci Soc Am J 63:830-838

Kharayat Y (2012) Distillery wastewater: bioremediation approaches. J Integr Environ Sci 9:69-91

Kogel-Knabner I, Knicker H (2001) Characterisation of native and non-native organic matter in soils be means of ^{13}C CPMAS NMR spectroscopy. In: Swift RS , Spark KM (ed) Understanding and managing organic matter in soils, sediments and waters, International Humic Substances Society, pp 141-148

Kögel-Knabner I (1997) 13C and 15N NMR spectroscopy as a tool in soil organic matter studies. Geoderma 80:243-270

Kopinke F-D, Georgi A, Mackenzie K (2001) Sorption of Pyrene to Dissolved Humic Substances and Related Model Polymers. 1. Structure-Property Correlation. Environ Sci Technol 35:2536-2542

Koprivnjak J-F, Pfromm PH, Ingall E, Vetter TA, Schmitt-Kopplin P, Hertkorn N, Frommberger M, Knicker H, Perdue EM (2009) Chemical and spectroscopic characterisation of marine dissolved organic matter isolated using coupled reverse osmosis-electrodialysis. Geochim Cosmochim Ac 73:4215-4231

Kordel W, Dassenakis M, Lintelmann J, Padberg S (1997) The importance of natural organic material for environmental processes in waters and soils. Pure Appl Chem 69:1571-1600

Krol S, Zabiegala B, Namiesnik J (2010) Monitoring VOCs in atmospheric air II. Sample collection and preparation. Trac-Trend Anal Chem 29:1101-1112

Krull E, Haynes D, Lamontagne S, Gell P, McKirdy D, Hancock G, McGowan J, Smernik R (2009) Changes in the chemistry of sedimentary organic matter within the Coorong over space and time. Biogeochemistry 92:9-25

Kruskal JB (1964) Multidimensional-scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29:1-27

Kunhi Mouvenchery Y, Kucerik J, Diehl D, Schaumann GE (2012) Cation-mediated cross-linking in natural organic matter: a review. Reviews in Environmental Science and Bio-Technology 11:41-54

Lange FT, Scheurer M, Brauch HJ (2012) Artificial sweeteners-a recently recognized class of emerging environmental contaminants: a review. Analytical and Bioanalytical Chemistry 403:2503-2518

Leinweber P, Kruse J, Baum C, Arcand M, Knight JD, Farrell R, Eckhardt KU, Kiersch K, Jandl G (2013) Advances in Understanding Organic Nitrogen Chemistry in Soils Using State-of-the-art Analytical Techniques. In: Sparks DL (ed) Advances in Agronomy, Vol 119, Advances in Agronomy, Elsevier Academic Press Inc, San Diego, pp 83-151

Lewis R, van Leeuwen JA, Smernik RJ, Chow CWK, Everson A, Nothrop SC, Beecham S (2011) Changes in the organic character of post-coagulated Pinus radiata sulfite pulp mill wastewater under aerated stabilization basin treatment A laboratory scale study. Chem Eng J 175:160-168

Lewis R, van Leeuwen JA, Chow CWK, Everson A, Lewis DM (2012) Assessment of coagulated and non-coagulated ASB performance used to treat Pinus radiata sulfite pulp and paper mill effluent by resin fractionation and HPSEC techniques. Chem Eng J 213:109-117

Li XG, Zhang T, Sun SW, Lan HQ, Yu T (2012) Lignin in marine environment and its analysis-A review. Journal of Ocean University of China 11:501-506

Lille U, Heinmaa I, Pehk T (2003) Molecular model of Estonian kukersite kerogen evaluated by C-13 MAS NMR spectra. Fuel 82:799-804

Lindqvist B, Storgards T (1955) Molecular-sieving properties of starch. Nature 175:511-512

Liu PWG, Wang SY, Huang SG, Wang MZ (2012) Effects of soil organic matter and ageing on remediation of diesel-contaminated soil. Environ Technol 33:2661-2672

Liu Z-h, Ogejo JA, Pruden A, Knowlton KF (2011) Occurrence, fate and removal of synthetic oral contraceptives (SOCs) in the natural environment: A review. Sci Total Environ 409:5149-5161

Loh PS, Chen CTA, Anshari GZ, Wang JT, Lou JY, Wang SL (2012) A comprehensive survey of lignin geochemistry in the sedimentary organic matter along the Kapuas River (West Kalimantan, Indonesia). J Asian Earth Sci 43:118-129

Lueking AD, Huang WL, Soderstrom-Schwarz S, Kim MS, Weber WJ (2000) Relationship of soil organic matter characteristics to organic contaminant sequestration and bioavailability. J Environ Qual 29:317-323

Lynch JM (1991) Sources and Fate of Soil Organic Matter. In: Wilson WS (ed) Advances in Soil Organic Matter Research: The Impact on Agriculture and the Environment, Royal Society of Chemistry (Great Britain), Royal Society of Chemistry, Cambridge, pp 231-237

Mackenzie RC, Mitchell BD (1972) Differential Thermal Analysis. Soil. Academic Press, New York

Mao J, Chen N, Cao X (2011a) Characterization of humic substances by advanced solid state NMR spectroscopy: Demonstration of a systematic approach. Org Geochem 42:891-902

Mao J, Tremblay L, Gagné J-P (2011b) Structural changes of humic acids from sinking organic matter and surface sediments investigated by advanced solid-state NMR: Insights into sources, preservation and molecularly uncharacterized components. Geochim Cosmochim Acta 75:7864-7880

Mao J, Kong X, Schmidt-Rohr K, Pignatello JJ, Perdue EM (2012) Advanced Solid-State NMR Characterization of Marine Dissolved Organic Matter Isolated Using the Coupled Reverse Osmosis/Electrodialysis Method. Environ Sci Technol 46:5806-5814

Mao JD, Hu WG, Schmidt-Rohr K, Davies G, Ghabbour EA, Xing BS (2000) Quantitative characterization of humic substances by solid-state carbon-13 nuclear magnetic resonance. Soil Sci Soc Am J 64:873-884

Mao JD, Xing BS, Schmidt-Rohr K (2001) New structural information on a humic acid from two-dimensional H-1-C-13 correlation solid-state nuclear magnetic resonance. Environ Sci Technol 35:1928-1934

Mao JD, Hundal LS, Thompson ML, Schmidt-Rohr KS (2002) Correlation of poly(methylene)-rich amorphous aliphatic domains in humic substances with sorption of a nonpolar organic contaminant, phenanthrene. Environ Sci Technol 36:929-936

Mao JD, Tremblay L, Gagné JP, Kohl S, Rice J, Schmidt-Rohr K (2007) Humic acids from particulate organic matter in the Saguenay Fjord and the St. Lawrence Estuary investigated by advanced solid-state NMR. Geochim Cosmochim Ac 71:5483-5499

Marquet R, Mietton-Peuchot M, Wheatley AD (1999) Characterisation of trickling filter effluent by particle size distribution and high performance size exclusion chromatography. Water Res 33:1415-1424

Matilainen A, Gjessing ET, Lahtinen T, Hed L, Bhatnagar A, Sillanpää M (2011) An overview of the methods used in the characterisation of natural organic matter (NOM) in relation to drinking water treatment. Chemosphere 83:1431-1442

McDonald S, Bishop AG, Prenzler PD, Robards K (2004) Analytical chemistry of freshwater humic substances. Anal Chim Acta 527:105-124

McKinney DE, Carson DM, Clifford DJ, Minard RD, Hatcher PG (1995) Off-line thermochemolysis versus flash pyrolysis for the in-situ methylation of lignin - is pyrolysis necessary? J Anal Appl Pyrol 34:41-46

Mulder MM, Vanderhage ERE, Boon JJ (1992) Analytical in source pyrolytic methylation electron-impact mass-spectrometry of phenolic-acids in biological matrices. Phytochem Analysis 3:165-172

Nebbioso A, Piccolo A (2013) Molecular characterization of dissolved organic matter (DOM): a critical review. Anal Bioanal Chem 405:109-124

Northcott GL, Jones KC (2000) Experimental approaches and analytical techniques for determining organic compound bound residues in soil and sediment. Environmental Pollution 108:19-43

Orlov DS, Sadovnikova LK (2005) Soil organic matter and protective functions of humic substances in the bioshere - A critical review and prospects. In: Perminova IV, Hatfield K , Hertkorn N (ed) Use of Humic Substances to Remediate Polluted Environments: From Theory to Practice, NATO Science Series IV Earth and Environmental Sciences, Springer, Dordrecht, pp 37-52

Otto A, Simpson MJ (2007) Analysis of soil organic matter biomarkers by sequential chemical degradation and gas chromatography - mass spectrometry. Journal of Separation Science 30:272-282

Page D (2006) Characterization of organic matter from air, water, soils, and waste materials by analytical pyrolysis. In: Nollet LMN (ed) Chromatographic Analysis of the Environment CRC Press, Boca Raton, Florida, pp 287–310

Parsi Z, Hartog N, Górecki T, Poerschmann J (2007) Analytical pyrolysis as a tool for the characterization of natural organic matter—A comparison of different approaches. J Anal Appl Pyrol 79:9-15

Pearson EJ, Farrimond P, Juggins S (2007) Lipid geochemistry of lake sediments from semi-arid Spain: Relationships with source inputs and environmental factors. Org Geochem 38:1169-1195

Pignatello JJ (2012) Dynamic interactions of natural organic matter and organic compounds. J Soils Sediments 12:1241-1256

Plante AF, Fernandez JM, Leifeld J (2009) Application of thermal analysis techniques in soil science. Geoderma 153:1-10

Potschka M (1993) Mechanism of size-exclusion chromatography. 1. role of convection and obstructed diffusion in size exclusion chromatography. J Chromatogr 648:41-69

Ratola N, Cincinelli A, Alves A, Katsoyiannis A (2012) Occurrence of organic microcontaminants in the wastewater treatment process. A mini review. J Hazard Mater 239:1-18

Rocha JD, Brown SD, Love GD, Snape CE (1997a) Hydropyrolysis: A versatile technique for solid fuel liquefaction, sulphur speciation and biomarker release. Journal of Analytical and Applied Pyrolysis 40-1:91-103

Rocha JD, Brown SD, Love GD, Snape CE (1997b) Hydropyrolysis: a versatile technique for solid fuel liquefaction, sulphur speciation and biomarker release. J Anal Appl Pyrol 40–41:91-103

Rodriguez-Zaraqoza S, Gonzalez-Ruiz T, Gonzalez-Lozano E, Lozada-Rojas A, Mayzlish-Gati E, Steinberger Y (2008) Vertical distribution of microbial communities under the canopy of two legume bushes in the Tehuacan Desert, Mexico. Eur J Soil Biol 44:373-380

Ruiz SH, Wickramasekara S, Abrell L, Gao XD, Chefetz B, Chorover J (2013) Complexation of trace organic contaminants with fractionated dissolved organic matter: Implications for mass spectrometric quantification. Chemosphere 91:344-350

Rumpel C, Kogel-Knabner I (2011) Deep soil organic matter-a key but poorly understood component of terrestrial C cycle. Plant Soil 338:143-158

Sanderman J, Baldock JA, Amundson R (2008) Dissolved organic carbon chemistry and dynamics in contrasting forest and grassland soils. Biogeochemistry 89:181-198

Sanderman J, Amundson R (2009) A comparative study of dissolved organic carbon transport and stabilization in California forest and grassland soils. Biogeochemistry 92:41-59

Sanderman J, Lohse KA, Baldock JA, Amundson R (2009) Linking soils and streams: Sources and chemistry of dissolved organic matter in a small coastal watershed. Water Resour Res 45

Santos IR, Eyre BD, Huettel M (2012) The driving forces of porewater and groundwater flow in permeable coastal sediments: A review. Estuar Coast Shelf S 98:1-15

Sarathy SR, Mohseni M (2007) The impact of UV/H_2O_2 advanced oxidation of molecular size distribution of chromophoric natural organic matter. Environ Sci Technol 41:8315-8320

Schaumann GE (2006) Soil organic matter beyond molecular structure Part I: Macromolecular and supramolecular characteristics. J Plant Nutr Soil Sc 169:145-156

Schimmelmann A, Sessions AL, Mastalerz M (2006) Hydrogen isotopic (D/H) composition of organic matter during diagenesis and thermal maturation Annual Review of Earth and Planetary Sciences, Annual Review of Earth and Planetary Sciences, Annual Reviews, Palo Alto, pp 501-533

Schmidt MWI, Knicker H, Hatcher PG, KogelKnabner I (1997) Improvement of C-13 and N-15 CPMAS NMR spectra of bulk soils, particle size fractions and organic material by treatment with 10% hydrofluoric acid. European Journal of Soil Science 48:319-328

Schnitzer M, Schulten HR (1992) The analysis of soil organic matter by pyrolysis-field ionization mass spectrometry. Soil Sci Soc Am J 56:1811-1817

Schulten HR, Gleixner G (1999) Analytical pyrolysis of humic substances and dissolved organic matter in aquatic systems: structure and origin. Water Res 33:2489-2498

Shadkami F, Helleur R (2010) Recent applications in analytical thermochemolysis. J Anal Appl Pyrol 89:2-16

Shepard RN (1962a) The analysis of proximities - multidimensional-scaling with an unknown distance function 1. Psychometrika 27:125-140

Shepard RN (1962b) The analysis of proximities - multidimensional-scaling with an unknown distance function 2. Psychometrika 27:219-246

Silverio FO, Barbosa LCA, Pilo-Veloso D (2008) Pyrolysis as an analytical technique. Quim Nova 31:1543-1552

Simeonov V, Stratis JA, Samara C, Zachariadis G, Voutsa D, Anthemidis A, Sofoniou M, Kouimtzis T (2003) Assessment of the surface water quality in Northern Greece. Water Res. 37:4119-4124

Simpson AJ, Simpson MJ (2009) Nuclear Magnetic Resonance Analysis of Natural Organic Matter Biophysico-Chemical Processes Involving Natural Nonliving Organic Matter in Environmental Systems, John Wiley & Sons, Inc., pp 589-650

Simpson AJ, McNally DJ, Simpson MJ (2011) NMR spectroscopy in environmental research: From molecular interactions to global processes. Prog Nucl Mag Res Sp 58:97-175

Simpson AJ, Simpson MJ, Soong R (2012) Nuclear Magnetic Resonance Spectroscopy and Its Key Role in Environmental Research. Environ Sci Technol 46:11488-11496

Simpson M, Simpson A (2012) The Chemical Ecology of Soil Organic Matter Molecular Constituents. J Chem Ecol 38:768-784

Smernik RJ, Oades JM (1999) Effects of added paramagnetic ions on the 13C CP/MAS NMR spectrum of a de-ashed soil. Geoderma 89:219-248

Smernik RJ, Oades JM (2000) Effect of paramagnetic cations on solid state 13C nuclear magnetic resonance spectra of natural organic materials. Commun Soil Sci Plan 31:3011-3026

Smernik RJ, Oades JM (2001) Solid-state 13C-NMR dipolar dephasing experiments for quantifying protonated and non-protonated carbon in soil organic matter and model systems. Eur J Soil Sci 52:103-120

Smernik RJ, Oades JM (2002) Paramagnetic effects on solid state carbon-13 nuclear magnetic resonance spectra of soil organic matter. J Environ Qual 31:414-420

Smernik RJ, Oades JM (2003) Spin accounting and RESTORE - two new methods to improve quantitation in solid-state ^{13}C NMR analysis of soil organic matter. Eur J Soil Sci 54:103-116

Smernik RJ, Oliver IW, Merrington G (2003) Characterization of sewage sludge organic matter using solid-state carbon-13 nuclear magnetic resonance spectroscopy. Journal of Environmental Quality 32:1516-1522

Smernik RJ, Olk DC, Mahieu N (2004) Quantitative solid-state 13C NMR spectroscopy of organic matter fractions in lowland rice soils. Eur J Soil Sci 55:367-379

Smernik RJ (2005) Solid-state 13C NMR spectroscopic studies of soil organic matter at two magnetic field strengths. Geoderma 125:249-271

Smernik RJ, Schwark L, Schmidtc MWI (2006) Assessing the quantitative reliability of solid-state C-13 NMR spectra of kerogens across a gradient of thermal maturity. Solid State Nuc Mag 29:312-321

Smith MTE, Smernik RJ, Merrington G, Tibbett M (2008) Changes in sewage sludge carbon forms along a treatment stream. Chemosphere 72:981-985

Stevenson FJ (1994) Extraction, Fractionation, and General Chemical Composition of Soil Organic Matter Humus Chemistry: Genesis, Composition, Reactions., John Wiley & Sons, Inc, New York, pp 24-33

Sutton R, Sposito G (2005) Molecular structure in soil humic substances: The new view. Environ Sci Technol 39:9009-9015

Swift MJ (1979) Decomposition in terrestrial ecosystems / M.J. Swift, O.W. Heal & J.M. Anderson. Studies in ecology (Oxford, England); v. 5. Blackwell, Oxford

Tas AC, van der Greef J (1994) Mass spectrometric profiling and pattern recognition. Mass Spectrom Rev 13:155-181

Templier J, Derenne S, Croue J-P, Largeau C (2005) Comparative study of two fractions of riverine dissolved organic matter using various analytical pyrolytic methods and a ^{13}C CP/MAS NMR approach. Org Geochem 36:1418-1442

Thevenon F, Pote J (2012) Water Pollution History of Switzerland Recorded by Sediments of the Large and Deep Perialpine Lakes Lucerne and Geneva. Water Air Soil Poll 223:6157-6169

Thomsen M, Lassen P, Dobel S, Hansen PE, Carlsen L, Mogensen BB (2002) Characterisation of humic materials of different origin: A multivariate approach for quantifying the latent properties of dissolved organic matter. Chemosphere 49:1327-1337

Thurman EM (1985) Organic Geochemistry of Natural Waters. Kluwer Academic Publishers, The Netherlands

Uyguner-Demirel CS, Birben C, Bekbolet M (2013) Key role of common anions on the photocatalytic degradation profiles of the molecular size fractions of humic acids. Catal Today 209:122-126

van Leeuwen J, Chow C, Fabris R, Withers N, Page D, Drikas M, Iwa Programme C, Iwa Programme C, Iwa Programme C (2002) Application of a fractionation technique for better understanding of the removal of natural organic matter by alum coagulation, 3rd World Water Congress: Drinking Water Treatment. Water Science and Technology: Water Supply. I W a Publishing, London, pp. 427-433

van Leeuwen J, Lewis R, Goonan P, Struve N, Everson A, Nothrop S, Smernik R, Chow CK, Fabris R, Rupasinghe M (2012) Changes in character of organics in the receiving environment of effluent from a sulphite pulp mill. Environ Sci Pollut Res 19:2151-2158

Volkamer R, Jimenez JL, San Martini F, Dzepina K, Zhang Q, Salcedo D, Molina LT, Worsnop DR, Molina MJ (2006) Secondary organic aerosol formation from anthropogenic air pollution: Rapid and higher than expected. Geophys Res Lett 33

Vuorio E, Vahala R, Rintala J, Laukkanen R (1998) The evaluation of drinking water treatment performed with HPSEC. Environ Int 24:617-623

Wakeham SG, Lee C (1993) Production, transport, and alteration of particulate organic matter in the marine water column. Org Geochem: principles and applications:145-169

Wakeham SG, Lee C, Hedges JI, Hernes PJ, Peterson ML (1997) Molecular indicators of diagenetic status in marine organic matter. Geochim Cosmochim Ac 61:5363-5369

Wei ZB, Gao XX, Zhang DJ, Da J (2005) Assessment of thermal evolution of kerogen geopolymers with their structural parameters measured by solid-state C-13 NMR spectroscopy. Energ Fuel 19:240-250

Werner-Zwanziger U, Lis G, Mastalerz M, Schimmelmann A (2005) Thermal maturity of type II kerogen from the New Albany Shale assessed by C-13 CP/MAS NMR. Solid State Nuc Mag 27:140-148

Wershaw RL (1993) Model for Humus in Soils and Sediments. Environ Sci Technol 27:814-816

Wong S, Hanna JV, King S, Carroll TJ, Eldridge RJ, Dixon DR, Bolto BA, Hesse S, Abbt-Braun G, Frimmel FH (2002) Fractionation of Natural Organic Matter in Drinking Water and Characterization by 13C Cross-Polarization Magic-Angle Spinning NMR Spectroscopy and Size Exclusion Chromatography. Environ Sci Technol 36:3497-3503

Woods GC, Simpson MJ, Kelleher BP, McCaul M, Kingery WL, Simpson AJ (2010) Online high-performance size exclusion chromatography - Nuclear magnetic resonance for the characterization of dissolved organic matter. Environ Sci Technol 44:624-630

Wu FC, Evans RD, Dillon PJ, Cai YR (2007) Rapid quantification of humic and fulvic acids by HPLC in natural waters. Appl Geochem 22:1598-1605

Xue S, Zhao QL, Wei LL, Song YT, Tie M (2013) Fluorescence spectroscopic characterization of dissolved organic matter fractions in soils in soil aquifer treatment. Environ Monit Assess 185:4591-4603

Yakimenko OS (2001) Fulvic acids and the fulvic acid fraction of soil: Nature, properties, and isolation methods (A review). Eurasian Soil Sci 34:1289-1299

Yao AJ, Qiu RL, Qing CL, Mu SS, Reardon EJ (2011) Effects of humus on the environmental activity of mineral-bound Hg: influence on Hg plant uptake. Journal of Soils and Sediments 11:959-967

Zang X, Hatcher PG (2002) A Py-GC-MS and NMR spectroscopy study of organic nitrogen in Mangrove Lake sediments. Org Geochem 33:201-211

Zehr JP (2011) Nitrogen fixation by marine cyanobacteria. Trends Microbiol 19:162-173

Zehr JP, Kudela RM (2011) Nitrogen Cycle of the Open Ocean: From Genes to Ecosystems. In: Carlson CA , Giovannoni SJ (ed) Annual Review of Marine Science, Vol 3, Annual Review of Marine Science, Annual Reviews, Palo Alto, pp 197-225

Zsolnay A (2003) Dissolved organic matter: artefacts, definitions, and functions. Geoderma 113:187-209

Zujovic Z, Srejic R, Vucelic D, Vitorovic D, Jovancicevic B (1995) Structural-analysis of aleksinac oil-shale kerogen by high-resolution solid-state C-13 NMR-spectroscopy. Fuel 74:1903-1909

Zuloaga O, Navarro P, Bizkarguenaga E, Iparraguirre A, Vallejo A, Olivares M, Prieto A (2012) Overview of extraction, clean-up and detection techniques for the determination of organic pollutants in sewage sludge: A review. Anal Chim Acta 736:7-29

CHAPTER TWO

THE ORGANIC CHEMISTRY OF ORGANIC RESIDUES:
COMPARISON OF NMR AND PYROLYSIS DATA USING
MULTIVARIATE STATISTICAL APPROACHES

✳

The work in this chapter has been published in *Current Organic Chemistry*.

STATEMENT OF AUTHORSHIP

Title of Paper	The organic chemistry of plant residues: comparison of NMR and pyrolysis data using multivariate statistical approaches.
Publication Status	⊙ Published, ○ Accepted for Publication, ○ Submitted for Publication, ○ Publication style
Publication Details	Current Organic Chemistry

AUTHOR CONTRIBUTIONS

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	Emma L Plant		
Contribution to the Paper	Experimental development, performed analysis on all samples, data analysis and critical interpretation, wrote manuscript.		
Signature		Date	8/10/13

Name of Co-Author	Ronald J Smernik		
Contribution to the Paper	Supervised development of work, NMR data analysis and interpretation, reviewed manuscript.		
Signature		Date	10/10/13

Name of Co-Author	Paul Greenwood		
Contribution to the Paper	Supervised development of work, pyrolysis data analysis and interpretation, reviewed manuscript.		
Signature		Date	8/10/13

Name of Co-Author	Lynne M Macdonald		
Contribution to the Paper	Supervised development of work, multivariate data analysis and interpretation, reviewed manuscript.		
Signature		Date	8/10/13

Name of Co-Author	John van Leeuwen		
Contribution to the Paper	Supervised development of work, data analysis and interpretation, reviewed manuscript.		
Signature		Date	9/10/13

ABSTRACT

To effectively characterise and distinguish between different organic matter (OM) samples, multiple chemical characterisation techniques are often employed. Due to the structural complexity of OM and the unique information provided by different characterisation techniques, it is often difficult to compare and combine data obtained from different analytical methods. In this study, we show how non-parametric multivariate statistical approaches can be used to compare the relative pattern of similarity/dissimilarity between organic samples characterised by two common solid-state analytical techniques: [13]C nuclear magnetic resonance (NMR) spectroscopy and flash pyrolysis-gas chromatography mass spectrometry (py-GCMS). These analytical methods were used to characterise a suite of plant residues including the leaf, flower, bark and wood of several species. Using non-parametric multivariate statistical approaches we identified similarities between the plant residue data using ordination plots, which enabled us to identify where NMR and py-GCMS distinguished between residues differently. A mantel-type test called RELATE showed that there was significant ($P<0.05$) similarity between the NMR and py-GCMS data in terms of their ability to differentiate between plant residues of different type; 61% of the sample discrimination was common to both profiling techniques, while 39% of discrimination was method specific. Further multivariate comparisons indicated that NMR was more sensitive to detecting differences in the organic composition of the plant residues.

INTRODUCTION

Organic matter (OM) consists of structurally complex compounds of plant, microbial or animal origin at various stages of decomposition (Lynch 1991; Wershaw 1993; Wong et al. 2002; Allpike et al. 2010; Matilainen et al. 2011), and is an important constituent of both soils (Wershaw 1993; Simpson and Simpson 2012) and natural waters (Allpike et al. 2010; Mao et al. 2011). Soil organic matter (SOM) is a key influence on the chemical and physical properties of soils, including the stability of soil aggregates, water holding capacity, metal binding capacity, and buffering capacity (Wershaw 1993), all of which influence soil fertility (Simpson and Simpson 2012). Aqueous natural organic matter (NOM) is a potential cause of many water quality problems (Allpike et al. 2010). Water utilities strive to remove NOM from source waters for domestic use, as the carbon in NOM is a substrate for bacterial growth and the presence of NOM in pipes can lead to fouling (Drikas 2003;

Allpike et al. 2010). Removal of NOM is costly, particularly for drinking water purposes, and chemical treatments can lead to the formation of harmful disinfection by-products (Allpike et al. 2010). Research interest in understanding how management of soil and water resources influences the nature of OM continues to grow as analytical characterisation methods improve.

The complexity of OM brings analytical challenges, and no single approach can provide complete characterisation (Gaffney et al. 1996; Allpike et al. 2010; Greenwood et al. 2012). As a result, it can be beneficial to use several analytical techniques to minimise the possibility of missing information or particular chemical differences that are better detected by one method over another (Baldock et al. 1991). Despite the significance of OM and the variety of characterisation techniques used for OM analysis, ranging from quick measurements to more detailed spectroscopic analyses (Matilainen et al. 2011), there is still a large portion that is "molecularly uncharacterised" (Hedges et al. 2000; Simpson and Simpson 2012). Further development of characterisation methodology will lead to an improved understanding of the structure and chemistry of OM. Potentially this could lead to improved knowledge about SOM stability and recalcitrance (Simpson and Simpson 2012), and new approaches in management of drinking water catchments, water treatment processes and potable water distribution (Allpike et al. 2010).

Two of the most commonly used analytical techniques for the characterisation of OM are solid-state ^{13}C nuclear magnetic resonance (NMR) spectroscopy and flash pyrolysis-gas chromatography mass spectrometry (py-GCMS) (Baldock et al. 1991; Wong et al. 2002; Berwick et al. 2010; Mao et al. 2011; Greenwood et al. 2012). Solid-state ^{13}C NMR spectroscopy has been used to study a range of organic materials (Kögel-Knabner 1997; Matilainen et al. 2011). The distribution of ^{13}C nuclei in chemical classes (alkyl, O-alkyl, aromatic and carbonyl) is determined based on differences in chemical shift in a standard NMR experiment (Smernik and Oades 2001). A major advantage of NMR is that all ^{13}C nuclei potentially produce signals of equal intensity regardless of their chemical environment (Smernik and Oades 2000c). However, some organic structures (e.g. charcoal, high molecular mobility lipids) are under-represented when the cross polarization (CP) technique is used (Smernik and Oades 2000b) and the presence of paramagnetic species can also affect CP detection (Smernik and Oades 1999, 2000a). Nonetheless, spin counting experiments have demonstrated that the majority of organic carbon in isolated OM samples is detected in ^{13}C CP NMR spectra, and comparisons

between CP spectra and the more quantitatively reliable direct polarisation (DP) spectra show that discrepancies can usually be attributed to the presence of charcoal and high mobility lipids for OM from a diversity of sources including soils (Smernik and Oades 2000c; Smernik et al. 2004), sediments (Golding et al. 2004; Dickens et al. 2006) and sewage sludges (Smernik et al. 2003; Smith et al. 2008). Furthermore, for plant residues, which contain little or none of these components, CP detection is near-quantitative (Johnson et al. 2005; Preston et al. 2011). The main limitation of NMR characterisation is its limited ability to discriminate structures within the broad definitions of carbon (C) types (alkyl, O-alkyl, aromatic and carboxyl). Thus NMR characterisation of OM provides a relatively quantitative but "broad-brush" assessment of OM composition. These features, particularly the potential for unbiased results, have led to solid-state ^{13}C NMR spectroscopy becoming a powerful research tool for OM characterisation (Smernik and Oades 2000b, 2001; Smernik 2005; Matilainen et al. 2011).

In comparison to solid-state ^{13}C NMR spectroscopy, py-GCMS gives a more detailed molecular level analysis of organic samples. It involves the thermal degradation of OM into simpler units for identification (Gaffney et al. 1996) on rapidly heating the OM samples to temperatures >500°C for a short time (seconds) (Berwick et al. 2007). Pyrolysis produces volatile fragments that are characteristic of the organic structures originally present. The products are then able to be GC separated based on their thermal volatility and structural polarity, and identified by MS analysis. In a typical py-GCMS analysis of OM, tens to hundreds of individual pyrolysis products are detected, many of which can be fully identified through comparison of their mass spectra to large MS databases of authentic compounds. The main disadvantage of py-GCMS analysis is that only a small fraction of the original OM gets converted into small, volatile compounds amenable to GCMS detection. Furthermore, this fraction of OM that is detected may vary between samples and between components within samples. Thus py-GCMS provides a detailed assessment of OM composition, but one that is likely to be non-quantitative or biased.

As the information these characterisation techniques provide is complex and detailed, there is a need for tools that can facilitate interpretation, integration or comparison of large data sets. Multivariate statistics are useful in condensing large and highly complex data sets with minimal information loss, and in summarising the overall differences between samples or groups of samples (Goodacre et al. 1996; Simeonov et al. 2003). Multivariate statistics have been used in soil and water research for decades and advances

in statistical software interfaces are facilitating greater opportunities to apply them more widely. Of the multivariate methods available, principle component analysis (PCA) has become quite popular for studying large data sets with multiple variables. However, there are a number of multivariate methods available and choosing among these depends on the type of information obtained and the purpose of the analysis.

With this study we aim to demonstrate how formal multivariate statistical analyses, such as ordination plots, can be a useful approach to i) assess the relative similarity/dissimilarity among a range of leached plant samples based on NMR and py-GCMS profiling; and ii) to compare the ability of the two techniques to discriminate between the samples and identify which analytical technique is more sensitive to the chemical differences within the sample set.

MATERIALS AND METHODS

SAMPLES

In this study, we analysed a suite of leached plant residues, produced by prolonged submersion in Mill-Q^{TM} water. These materials are the subject of previous analytical studies (Miles 2005; Langsa 2008) and are listed in Table 1. The plant samples comprised wood, bark, leaf, flower or nut elements of three different tree types (*Eucalyptus wandoo*, *Corymbia calophylla* and *Pinus radiata*) found on the shores of Mundaring Weir, a major water reservoir and source water supply for the Western Australian capital, Perth. Sample preparations included cutting the wood samples into thin slices and removal of the woody stem from the Eucalyptus leaves. For each plant sample, 10 g was separately leached in 1.5 L of Milli-Q^{TM} water for 140 days. The residue material was dried first at room temperature and then by heating in an oven at 50°C for 3 days.

Table 1 Milli-Q^{TM} water leached plant residue samples studied.

Eucalyptus wandoo	1. Fresh leaf	*Pinus radiata*	7. Bark
	2. Dead leaf		8. Fresh leaf
	3. Fresh flower		9. Dead leaf
	4. Dead flower		10. Wood
	5. Bark		11. Nut
Corymbia calophylla	6. Bark		

SOLID-STATE ^{13}C CP NMR SPECTROSCOPY

Solid-state ^{13}C CP NMR spectra were acquired with magic angle spinning (MAS) at a ^{13}C frequency of 100.6 MHz on a Varian Unity INOVA 400 spectrometer. Samples were packed in a 7 mm diameter cylindrical zirconia rotor with Kel-F end-caps, and spun at 6500±100 Hz in a Doty Scientific supersonic MAS probe. Spectra were acquired using a ramped-amplitude cross polarization (CP-ramp) pulse sequence, in which the ^{1}H spin lock power was varied linearly during the contact time. A 1 ms contact time was used and 2000 transients were collected for each spectrum. Recycle delays (2 s) were chosen to be › 5 × T_1H, as determined in preliminary inversion-recovery experiments. Free induction decays were acquired with a sweep width of 50 kHz; 1216 data points were obtained over an acquisition time of 12 ms. All spectra were zero-filled to 8192 data points and processed with a 100 Hz Lorentzian line broadening and a 0.01 s Gaussian broadening. Chemical shifts were externally referenced to the methyl resonance of hexamethylbenzene at 17.36 ppm.

For multivariate analysis, data was normalised and baseline corrected to ensure all data points were non-negative and further smoothed by taking the average of each five consecutive points throughout the spectra. The NMR trace in the spectral region of 0-250 ppm was used for multivariate analyses.

FLASH PY-GCMS

Flash pyrolysis of small sample quantities (0.5 – 0.8 mg) was conducted at 650 ^{o}C for 10 s using a Chemical Data System pyrolysis unit. The 160 Pyroprobe inserted into a dedicated pyrolysis chamber mounted directly onto the vaporisation injector of a Hewlett Packard

(HP) 5890 Series II GC which was maintained at 250 $^{\circ}$C. The GC was coupled to a 5971 mass selective detector (MSD) with a 60 m, 0.25 mm i.d., 0.25 µm phase ZB-5MS capillary column (Phenomenex). Helium carrier gas (9 psi) and split injections of between 20 and 50 mL min^{-1} were used and the oven was programmed to increase from 40 $^{\circ}$C (2 min hold) at 4 $^{\circ}$C min^{-1} to a final temperature of 310 $^{\circ}$C (15 min). Full scan m/z 50–550 mass spectra, with an electron energy of 70 eV were acquired.

Products were tentatively identified by mass spectral correlation to the NIST05 mass spectral library; see Supplementary Material for list of products (Table S1). Only high quality peak matching (typically ›70%) between measured and library mass spectra are reported.

The integrated peak areas (from the total ion chromatogram (TIC)) of all identified products (Table S1) were used for multivariate analysis. Products not detected in any particular sample were given a zero value in the statistical treatments. For each sample, peak areas were log transformed and normalised.

MULTIVARIATE STATISTICAL ANALYSIS

The purpose of the multivariate statistical analysis was to assess the relative power of the different techniques to discriminate between samples in terms of their organic chemical composition. Non-parametric multidimensional scaling (nMDS) was used to analyse the NMR and py-GCMS data and was carried out in Primer v6 (Clarke and Gorley 2006). An nMDS ordination plot aims to represent the relative relationship between samples based on a rank order of the corresponding dissimilarities between samples. Unlike PCA, it does not assume data linearity and is not constrained to a Euclidean distance basis. There are fewer assumptions/limitations and greater flexibility in selecting underlying similarity/dissimilarity matrices that best suit the data type. The nMDS approach has been recommended as one of the best ordination methods available (Everitt 1978). Its benefits include an ability to represent more complex relationships accurately in low-dimensions (Clarke and Warwick 2001). Furthermore, nMDS provides greater flexibility to avoid distance distortion and 'horse-shoe' artifacts in highly multidimensional data types.

For the NMR data, the resemblance matrix was based on Manhattan Distance because with spectral type data, each data point is part of a continuous trace rather than being

independent. For the py-GCMS data, the resemblance matrix was based on Bray-Curtis dissimilarity, since the input data consisted of a set of individual peak intensities and thus the data is a series of discrete values where adjacent points are completely independent. The nMDS ordination plots display the relationships between samples based on NMR or py-GCMS chemical profiles: the closer the samples are in the ordination plot the more similar their chemical compositions as identified by that analytical technique. The 'stress' level associated with the nMDS plots indicate the degree of confidence in the 2-D representation of the multivariate data: a stress of <0.2 is considered robust whereas a stress of >0.2 indicates some caution should be used in interpreting the 2-D plot due to a reduced ability to display relationships in the complex data. Comparison of the relationship between the NMR and py-GCMS ordinations was carried out using the Spearman rank correlation method (termed RELATE in Primer), reported as ρ (rho). It provides a non-parametric measure of statistical dependence between the similarity matrices derived from the NMR and GCMS data.

RESULTS AND DISCUSSION

SOLID-STATE ^{13}C CP NMR SPECTROSCOPY

Differences among NMR spectra can be identified directly; however, spectral information becomes too complex to allow direct quantitative assessment of the degree of similarity or difference among samples. This can be achieved by generating a resemblance matrix and displaying these resemblances by way of an nMDS ordination plot (Figure 1a). An nMDS plot is a 2-D representation of the full multivariate information in the resemblance matrix, and provides an effective way of representing relative overall similarity across samples. Samples grouped closer together in the nMDS plot are more similar than those spaced further apart. The generation of such a plot makes identification of similarities between samples easier to interpret, and is especially beneficial to those not familiar with certain spectral data as the 2-D nMDS plot is not technique specific.
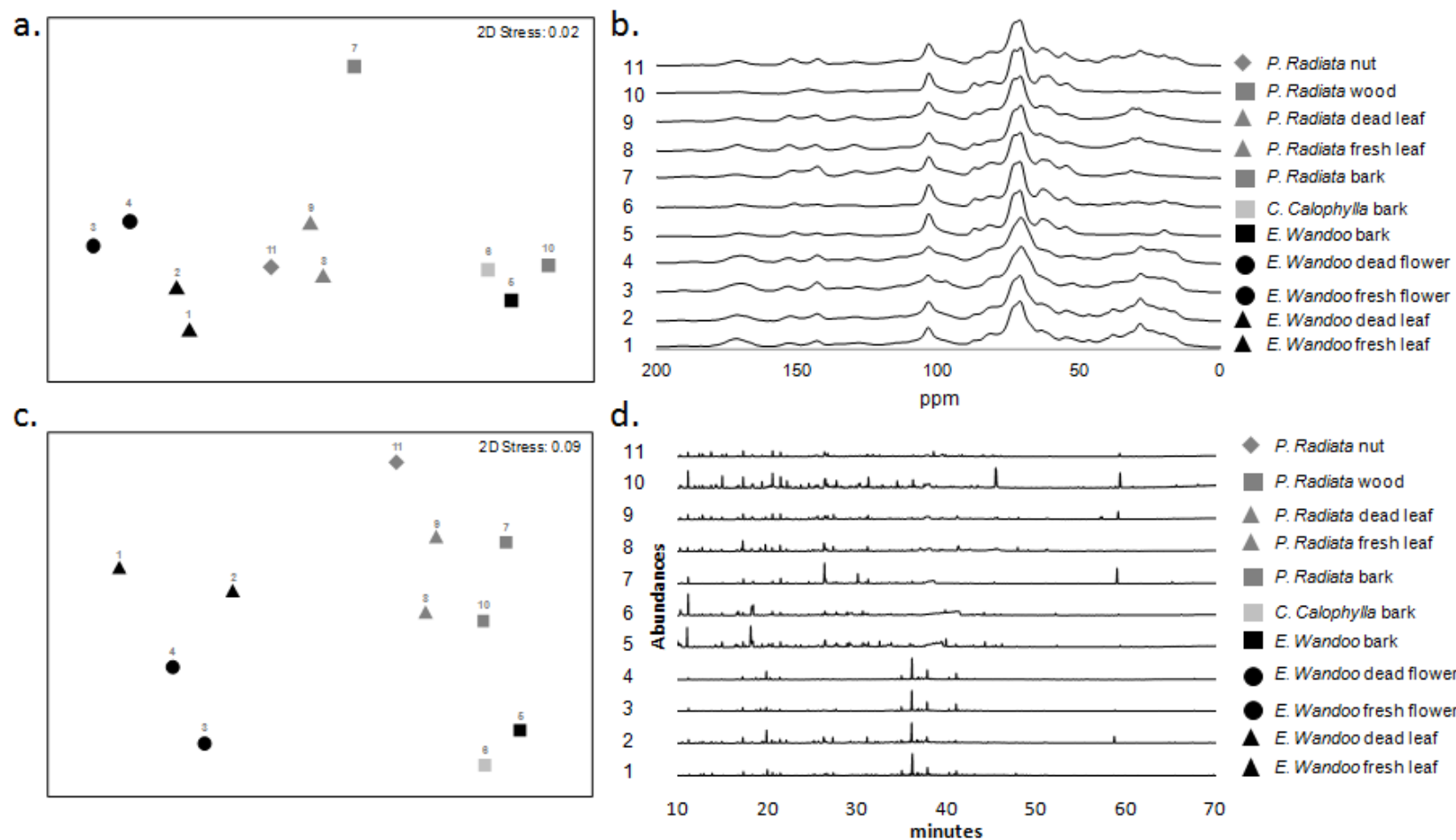
Figure 1 (a) Non-parametric multidimensional scaling (nMDS) ordination plot, based on the Manhattan Distance resemblance matrix, of the NMR spectra (b) demonstrating the relative similarity/dissimilarity between 11 plant residue samples; (c) nMDS ordination plot, based on Bray-Curtis Dissimilarity resemblance matrix, of py-GCMS spectra (d) of the same 11 plant residue samples.

The nMDS plot (Figure 1a) shows that NMR spectra of fresh and dead *E. wandoo* leaves (samples 1-2) are very similar, as are fresh and dead *E. wandoo* flowers (samples 3-4); these two pairs of samples are quite similar to each other as well (i.e. samples 1-4 plot relatively close together). This is consistent with the appearance of the NMR spectra of samples 1-4, which are similar across the entire spectrum (Figure 1b). In contrast, the nMDS plot (Figure 1a) indicates that the NMR profile of *E. wandoo* bark (sample 5) is quite different from those of the other *E. wandoo* samples (1-4). A source of this difference is clearly evident from the NMR spectrum of sample 5, which lacks the variety of peaks in the alkyl regions (0-45 ppm) of samples 1-4, apart from the peak at ~20 ppm (Figure 1b). Most of the woody residues, specifically *E. wandoo* bark (5), *C. calophylla* bark (6) and *P. radiata* wood (10), are grouped together in the nMDS plot, but away from *P. radiata* bark (7), which appears as the most dissimilar of all the plant residues studied (Figure 1a). Compared to the grouped woody residues, the NMR spectrum of *P. radiata* bark shows different alkyl (0-45 ppm) and more aromatic (110-160 ppm) signals (Figure 1b). The *P. radiata* leaf and nut samples (8, 9 and 11) appear close to those of the *E. wandoo* samples (1 and 2) in the nMDS plot (Figure 1a). The NMR spectra of these samples (Figure 1b) contain similar peaks across the whole spectral region, with a complex of alkyl peaks (0-40 ppm), and aromatic peaks of moderate intensity (~145 and ~152 ppm).

In general, the NMR spectra of the plant residues (Figure 1b) are consistent with previously published ^{13}C NMR spectra (Thomsen et al. 2002; De Marco et al. 2012). They are all similar in that the strongest peaks are in the O-alkyl region (45-110 ppm). These can be attributed mainly to carbohydrate and especially cellulose (De Marco et al. 2012). In particular, they all contain a complex set of signals between 70 and 80 ppm, which are due to C2-C6 carbons in carbohydrate and a peak at ~105 ppm, which is due to the anomeric carbon or C1 carbon of the glucose unit in cellulose (De Marco et al. 2012). Differences between the spectra are clearest in the alkyl (0-45 ppm) and aromatic (110-160 ppm) regions and also in the relative size of the carbonyl peak at ~170 ppm.

It should be noted that the rate of magic angle spinning (6.5 MHz) was insufficient to completely overcome the chemical shift anisotropy of the aromatic and carbonyl carbon nuclei and hence these carbons are affected by spinning sideband (SSB) artifacts. This was most evident in the appearance of a small (~10% of the intensity of the central band) downfield SSB for the carbonyl C at ~235 ppm (outside the chemical shift range presented

in Figure 1b; see expanded NMR spectra in Supplementary Material, Figure S1). The corresponding upfield carbonyl SSB underlies the di-O-alkyl peak at ~105 ppm. Due to the low aromatic content of the samples, aromatic SSBs are barely visible.

FLASH PY-GCMS

The peak areas of identified products were used for multivariate analysis. The discrete nature of the pyrolysis data (as opposed to the continuous NMR spectral data) meant the resemblance matrix was generated differently; hence the plot generated from the pyrolysis data (Figure 1c) is based on Bray-Curtis dissimilarity.

The *E. wandoo* samples (1-4) are grouped together in the nMDS plots derived from the py-GCMS data (Figure 1c); this is consistent with their grouping in the NMR based nMDS plot (Figure 1a). Both data sets also show *E. wandoo* bark (5) to be different from the other *E. wandoo* samples (1-4) (Figure 1a and 1c). Obvious product differences include an absence in sample 5 of the abundant peaks between 35 and 42 minutes present in samples 1-4 (Figure 1d). The woody-type samples (5, 6, and 10) are grouped in both the NMR and py-GCMS nMDS plots, but their discrimination from the other samples is not quite as clear with py-GCMS (Figure 1c). *P. radiata* bark (7) again appears to be the most distinct of the bark samples, but again py-GCMS does not distinguish it from the other samples as clearly as NMR does. However, py-GCMS does group *P. radiata* leaf and nut samples (8, 9, and 11) more closely to *P. radiata* bark (7) and wood (10) samples than NMR.

STATISTICAL COMPARISON OF NMR AND PY-GCMS RESULTS

Inspection of the NMR and py-GCMS data (Figures 1b and 1d respectively) provides for a qualitative assessment of the organic composition of the plant residues. We can begin to quantitatively assess the relative similarity/dissimilarity in overall sample compositions by using multivariate statistical approaches. This begins with the generation of nMDS plots (Figure 1a and 1c), which express similarities among the resemblance matrices of the NMR and py-GCMS data of the plant residues. However this does not provide an indication or comparison on how powerful the discrimination is between the organic samples using these techniques. We can further quantitatively test the relationship between the two resemblance matrices using a comparative mantel-type test called RELATE. This provides a measure of the strength of the relationship between the two resemblance matrices,

expressed in the parameter ρ (rho), which ranges from 0 (no relationship) to 1 (strong relationship). A RELATE analysis was chosen for this study as the resemblance matrices for comparison were based on both distance (NMR) and dissimilarity (py-GCMS) measures.

RELATE analysis revealed a significant ($P<0.05$) relationship between the two matrices (rho= 0.61) indicating that there was a degree of rank sample order between the NMR and py-GCMS data. From this we can conclude that 61% of the discrimination between the organic samples was common to both analytical techniques. This suggests that while the information gained had some overlap, each method provided some level of method-specific discrimination. Considering a wider range of analytical methods, this approach could be very useful in helping to identify which methods provide the most complementary and independent data for different organic sample types. The most appropriate choice of analytical techniques would be those of low statistical dependence that provide complimentary and independent data, as demonstrated by the NMR and py-GCMS techniques in this study.

Both of these aspects – the degree of similarity between analytical techniques and the degree of difference – have important implications regarding the use of multiple techniques for OM characterisation. On the one hand, a high degree of similarity between measurements is reassuring, given that both techniques are widely used on their own in studies of OM. It would be disconcerting if these techniques gave completely unrelated assessments of similarity for a set of simple materials such as this. Insomuch as the techniques gave similar assessments, the two techniques can be considered interchangeable, and for practical purposes one would gain the majority of information by implementing just one of these analyses. On the other hand, the fact that some extra information can be gained by carrying out both analyses is also potentially useful. In truth, this dichotomy is innately understood by most practitioners: NMR and py-GCMS have well-known strengths and weaknesses in terms of discrimination ability. However, the power of the approach taken here is that it allows a rigorous, objective and quantitative assessment of these aspects of redundancy and complimentarity. So although many studies of organics have used multivariate statistics to help manage multiple variables and analyse large data sets (Thomsen et al. 2002; Simeonov et al. 2003; Wang et al. 2004; Kujawinski et al. 2009; Lin et al. 2011), these studies do not provide a quantitative

assessment of the extent that the information provided by different characterisation techniques differs.

Another way to express the relationship between the two techniques is to plot the individual pair wise resemblances from the resemblance matrices for the two techniques against each other (Figure 2). This allows for an assessment of the linear trend of the resemblance information, as well as an assessment of which method provided a greater discrimination of samples. For example, the *E. wandoo* and *C. calophylla* bark samples exhibit the greatest slope (2.26 and 2.20 respectively) and strongest linear trends (r^2 = 0.87 and 0.9 respectively) of the residues (Figure 2). Moderate trends (r^2 = 0.4-0.6) are apparent for *E. wandoo* fresh and dead leaf, and fresh and dead flower, while the weakest trends (r^2 < 0.4) are seen for *P. radiata* fresh and dead leaf, bark, wood and nut. The slope of the lines for these moderate and weak trends is variable.

The characterisation technique able to provide greatest sample distinction could also be identified. For strong linear trends (e.g., *E. wandoo* bark, *C. calophylla* bark), NMR provided the greatest range of normalised distance resemblance measurements for the plant samples, showing spreads of 0.81 and 0.74, respectively (Figure 2). The pyrolysis resemblance measurements for these samples did not provide as great a range, showing spreads of 0.32 and 0.33, respectively (Figure 2). This suggests that solid-state ^{13}C NMR spectroscopy was more sensitive than py-GCMS at detecting differences in the organic chemistry of most of the samples analysed. For the majority of samples, NMR was more sensitive to organic differences between plant residues than py-GCMS. The only instances where py-GCMS was more sensitive were for the samples with the smaller slopes and moderate to weak linear trends (e.g., *P. radiata* fresh and dead leaf, bark and nut; bottom row of Figure 2).
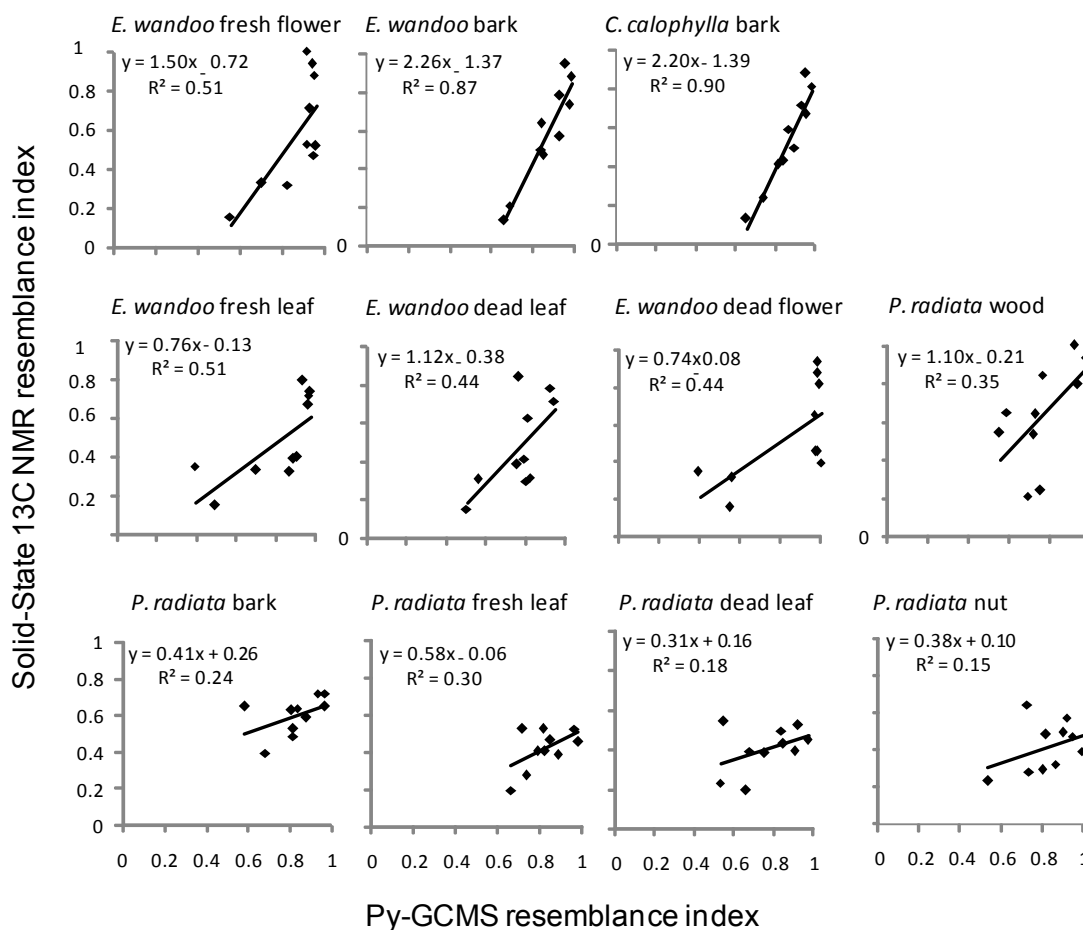
Figure 2 Regression plots between the NMR resemblance index and the py-GCMS resemblance index for the pair-wise comparison of individual plant residue samples against other plant residues in the test set. Resemblance indices represent the normalised Manhattan Distance and normalised Bray-Curtis Dissimilarity matrices for NMR and py-GCMS respectively. The samples have been grouped based on the slope of the linear line, greatest ($\geq$1.5) at the top and weakest (<0.6) at the bottom. The top row shows regression plots where NMR analysis dominated distinction between samples (greater spread), and the bottom row shows where py-GCMS dominated.

It was expected that NMR and py-GCMS would distinguish between the plant residues to different degrees, even though the nMDS plots of the resemblance data indicated similar sample correlations by both NMR and py-GCMS. As discussed earlier, NMR and py-GCMS provide a different evaluation on the chemistry of OM. NMR analysis provides a relatively unbiased distribution of ^{13}C nuclei into broad chemical classes, and in comparison, py-GCMS provides a more detailed molecular level characterisation of organic fragments. During pyrolysis, thermal rearrangements may have led to the loss of

chemical information that was unique to a sample (Saiz-Jimenez 1994), possibly resulting in the lower distinction between samples when characterised by this technique. Further assessment of the data interpretation would be best addressed with replicate samples which will take into account any reproducibility issues.

CONCLUSION

This study demonstrates the capacity of multivariate statistics to combine and compare characterisation data from two very different analytical techniques. This statistical approach compared the molecular profiles of leached plant samples by solid-state ^{13}C NMR spectroscopy and flash py-GCMS. The approach was used to determine the relative similarities/differences between samples as determined by each analytical technique, and quantitatively assess how well these analytical techniques discriminate between the samples. The respective nMDS plots showed solid-state ^{13}C NMR spectroscopy and flash py-GCMS data provide a similar discrimination of the plant samples. In quantitative terms, 61% of the discrimination between the samples was identified by both profiling techniques, emphasising the complementary nature of the two characterisation methods. This meant that 39% of the discrimination was method specific, and each technique also provided unique results to the analysis. Further inspection of resemblance matrices for each individual plant residue showed that although there was a positive correlation between the characterisation of the residues using these two techniques, the strength of this correlation varied substantially, with r^2 between 0.15 (weakest trend) and 0.9 (strongest). Of the two characterisation methods assessed, solid-state ^{13}C NMR spectroscopy generally provided the greater discrimination of the sample set.

Statistical consideration of the analytical data from ^{13}C NMR and py-GCMS would be applicable to the characterisation of other organic materials. Robust statistical treatment can be used to measure and compare the sensitivity with which different analytical methods are able to resolve differences between chemically similar samples. This encourages an innovative extension to the way OM is characterised and will improve the ability to molecularly define complex OM.

ACKNOWLEDGEMENTS

REFERENCES

Allpike B, Busetti F, Heitz A, Berwick L, Warton B, Garbin S, Alessandrino M, Eckstaedt SVv, White D, Grice K, Greenwood P, Kagi R (2010) Advanced Characterisation of Natural Organic Matter (NOM) in Australian Water Supplies. 80, Water Quality Research Australia Adelaide

Baldock JA, Currie GJ, Oades JM (1991) Organic Matter as Seen by Solid State ^{13}C NMR and Pyrolysis Tandem Mass Spectrometry. In: Wilson WS (ed) Advances in Soil Organic Matter Research: The Impact on Agriculture and the Environment, Royal Society of Chemistry, Cambridge, pp 45-60

Berwick L, Greenwood P, Kagi R, Croué J-P (2007) Thermal release of nitrogen organics from natural organic matter using micro scale sealed vessel pyrolysis. Org. Geochem. 38:1073-1090

Berwick L, Greenwood PF, Smernik RJ (2010) The use of MSSV pyrolysis to assist the molecular characterisation of aquatic natural organic matter. Water Res 44:3039-3054

Clarke KR, Warwick RM (2001) Change in Marine Communities: An Approach to Statistical Analysis and Interpretation. PRIMER-E Ltd, Plymouth Marine Laboratory, UK

Clarke KR, Gorley RN (2006) PRIMER v6: User Manual/Tutorial. PRIMER-E, Plymouth

De Marco A, Spaccini R, Vittozzi P, Esposito F, Berg B, Virzo De Santo A (2012) Decomposition of black locust and black pine leaf litter in two coeval forest stands on Mount Vesuvius and dynamics of organic components assessed through proximate analysis and NMR spectroscopy. Soil Biol. Biochem. 51:1-15

Dickens AF, Baldock JA, Smernik RJ, Wakeham SG, Arnarson TS, Gélinas Y, Hedges JI (2006) Solid-state 13C NMR analysis of size and density fractions of marine sediments: Insight into organic carbon sources and preservation mechanisms. Geochimica et Cosmochimica Acta 70:666-686

Drikas M (2003) Natural organic matter in drinking water: problems and solutions, Water Quality Research Australia Limited.

Everitt BS (1978) Graphical Techniques for Multivariate Data. Heinman, London

Gaffney JS, Marley NA, Clark SB (1996) Humic and Fulvic Acids and Organic Colloidal Materials in the Environment. In: Gaffney JS, Marley NA , Clark SB (ed) Humic and Fulvic Acids. Isolation, Structure and Environmental Role, ACS Symposium Series, American Chemical Society, Washington, pp 2-16

Golding CJ, Smernik RJ, Birch GF (2004) Characterisation of sedimentary organic matter from three south-eastern Australian estuaries using solid-state ^{13}C-NMR techniques. Mar Freshwater Res 55:285-293

Goodacre R, Neal MJ, Kell DB (1996) Quantitative Analysis of Multivariate Data Using Artificial Neural Networks: A Tutorial Review and Applications to the Deconvolution of Pyrolysis Mass Spectra. Zentralb Bakteriol 284:516-539

Greenwood PF, Berwick LJ, Croué JP (2012) Molecular characterisation of the dissolved organic matter of wastewater effluents by MSSV pyrolysis GC–MS and search for source markers. Chemosphere 87:504-512

Hedges JI, Eglinton G, Hatcher PG, Kirchman DL, Arnosti C, Derenne S, Evershed RP, Kögel-Knabner I, de Leeuw JW, Littke R, Michaelis W, Rullkötter J (2000) The molecularly-uncharacterized component of nonliving organic matter in natural environments. Org Geochem 31:945-958

Johnson CE, Smernik RJ, Siccama TG, Kiemle DK, Zhihong X, Vogt DJ (2005) Using 13C nuclear magnetic resonance spectroscopy for the study of northern hardwood tissues. Canadian Journal of Forest Research 35:1821-1831

Kögel-Knabner I (1997) 13C and 15N NMR spectroscopy as a tool in soil organic matter studies. Geoderma 80:243-270

Kujawinski EB, Longnecker K, Blough NV, Vecchio RD, Finlay L, Kitner JB, Giovannoni SJ (2009) Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution mass spectrometry. Geochim. Cosmochim. Act. 73:4384-4399

Langsa M (2008) Molecular and isotopic characterisation of leachates in water. Masters Thesis Thesis, Curtin University, Perth

Lin DS, Greenwood PF, George S, Somerfield PJ, Tibbett M (2011) The development of soil organic matter in restored biodiverse Jarrah forests of South-Western Australia as determined by ASE and GCMS. Environ. Sci. Pollut. Res. 18:1070-1078

Lynch JM (1991) Sources and Fate of Soil Organic Matter. In: Wilson WS (ed) Advances in Soil Organic Matter Research: The Impact on Agriculture and the Environment, Royal Society of Chemistry (Great Britain), Royal Society of Chemistry, Cambridge, pp 231-237

Mao J, Tremblay L, Gagné J-P (2011) Structural changes of humic acids from sinking organic matter and surface sediments investigated by advanced solid-state NMR: Insights into sources, preservation and molecularly uncharacterized components. Geochim Cosmochim Acta 75:7864-7880

Matilainen A, Gjessing ET, Lahtinen T, Hed L, Bhatnagar A, Sillanpää M (2011) An overview of the methods used in the characterisation of natural organic matter (NOM) in relation to drinking water treatment. Chemosphere 83:1431-1442

Miles J (2005) Characterisation studies to investigate the origin of aquatic biopolymers. Honours Thesis Thesis, Curtin University, Perth

Preston CM, Smernik RJ, Powers RF, McColl JG, McBeath TM (2011) The decomposition of windrowed, chipped logging slash and tree seedling response: A plant growth and nuclear magnetic resonance spectroscopy study. Organic Geochemistry 42:936-946

Saiz-Jimenez C (1994) Analytical Pyrolysis of Humic Substances: Pitfalls, Limitations, and Possible Solutions. Environ. Sci. Technol. 28:1773-1780

Simeonov V, Stratis JA, Samara C, Zachariadis G, Voutsa D, Anthemidis A, Sofoniou M, Kouimtzis T (2003) Assessment of the surface water quality in Northern Greece. Water Res. 37:4119-4124

Simpson M, Simpson A (2012) The Chemical Ecology of Soil Organic Matter Molecular Constituents. J Chem Ecol 38:768-784

Smernik RJ, Oades JM (1999) Effects of added paramagnetic ions on the 13C CP/MAS NMR spectrum of a de-ashed soil. Geoderma 89:219-248

Smernik RJ, Oades JM (2000a) Effect of paramagnetic cations on solid state 13C nuclear magnetic resonance spectra of natural organic materials. Commun Soil Sci Plan 31:3011-3026

Smernik RJ, Oades JM (2000b) The use of spin counting for determining quantitation in solid state 13C NMR spectra of natural organic matter 2. HF-treated soil fractions. Geoderma 96:159-171

Smernik RJ, Oades JM (2000c) The use of spin counting for determining quantitation in solid state 13C NMR spectra of natural organic matter: 1. Model systems and the effects of paramagnetic impurities. Geoderma 96:101-129

Smernik RJ, Oades JM (2001) Solid-state 13C-NMR dipolar dephasing experiments for quantifying protonated and non-protonated carbon in soil organic matter and model systems. Eur J Soil Sci 52:103-120

Smernik RJ, Oliver IW, Merrington G (2003) Characterization of sewage sludge organic matter using solid-state carbon-13 nuclear magnetic resonance spectroscopy. J Environ Qual 32:1516-1522

Smernik RJ, Olk DC, Mahieu N (2004) Quantitative solid-state 13C NMR spectroscopy of organic matter fractions in lowland rice soils. Eur J Soil Sci 55:367-379

Smernik RJ (2005) Solid-state 13C NMR spectroscopic studies of soil organic matter at two magnetic field strengths. Geoderma 125:249-271

Smith MTE, Smernik RJ, Merrington G, Tibbett M (2008) Changes in sewage sludge carbon forms along a treatment stream. Chemosphere 72:981-985

Thomsen M, Lassen P, Dobel S, Hansen PE, Carlsen L, Mogensen BB (2002) Characterisation of humic materials of different origin: A multivariate approach for quantifying the latent properties of dissolved organic matter. Chemosphere 49:1327-1337

Wang WJ, Baldock JA, Dalal RC, Moody PW (2004) Decomposition dynamics of plant materials in relation to nitrogen availability and biochemistry determined by NMR and wet-chemical analysis. Soil Biol. Biochem. 36:2045-2058

Wershaw RL (1993) Model for Humus in Soils and Sediments. Environ Sci Technol 27:814-816

Wong S, Hanna JV, King S, Carroll TJ, Eldridge RJ, Dixon DR, Bolto BA, Hesse S, Abbt-Braun G, Frimmel FH (2002) Fractionation of Natural Organic Matter in Drinking Water and Characterization by 13C Cross-Polarization Magic-Angle Spinning NMR Spectroscopy and Size Exclusion Chromatography. Environ Sci Technol 36:3497-3503

SUPPLEMENTARY MATERIAL

Table **S1** Tentatively identified organic products from the flash pyrolysis analysis of the plant residues. The mass spectral match (%), based on correlation with the NIST05 mass spectral library, is shown in brackets.

Compound Number	Retention time (min)	Major Ions	Compound name (major ions)
1	8.786	92	toluene (90%)
2	9.823	84	3-amino-s-triazole (7%)
3	9.853	96	furfural (86%)
4	10.284	222	hexamethylcyclotrisiloxane (83%)
5	11.083	96	furfural (80%)
6	11.117	96	1,4-dimethylpyrazole (50%)
7	11.202	96	furfural (80%), 3,5-dimethyl-1H-pyrazole (80%)
8	11.255	96	1,4-dimethylpyrazole (64%)
9	12.405	106	ethylbenzene (64%)
10	12.768	106	p-xylene (97%)
11	12.805	106	(o or p-) xylene (87%)
12	12.93	106	1,3-dimethylbenzene (97%)
13	13.71	104	1,3,5,7-cyclooctatetraene (93%)
14	13.77	140	styrene (90%)
15	14.971	98	1,2-cyclopentanedione (72%)
16	15.466	136	1R-alpha-pinene (93%)
17	16.688	114	butanedioic acid, cycloc hydrazide (37%)
18	16.832	169	cyclobutanecarboxylic acid, morpholide (39%)
19	17.134	94	phenol (86%)
20	17.303	94	phenol (91%)
21	18.292	112	1,2-cyclohexanedione (38%)
22	18.301	114	dihydro-2,4(1H,3H)-pyrimidinedione (47%)
23	18.468	114	dihydro-2,4(1H,3H)-pyrimidinedione (38%)
24	18.479	171	2-butyl-2-ethyl-3-methyloxazolidine (40%)
25	18.795	136	alpha-phellandrene (91%)
26	19.245	112	2-methylcyclohexanone (9%)
27	19.247	136	1,7,7-trimethylbicyclo[2.2.1]hepta-2-ene (94%)
28	19.27	136	(+)-4-carene (98%)
29	19.442	112	3-methyl-1,2-cyclopentanedione (97%)
30	19.546	134	1-methyl-2-(1-methylethyl)benzene (95%)
31	19.833	136	D-limonene (96%)
32	19.928	154	eucalyptol (98%)
33	20.582	108	2-methylphenol (64%)
34	20.631	116	Indene (55%)
35	20.635	116	1-propynylbenzene (91%)
36	21.046	136	1-methyl-4-(1-methylethyl)-1,4-cyclohexadiene (91%)
37	21.438	108	4-methylphenol (95%)
38	22.18	124	2-methoxyphenol (90%)
39	22.257	124	mequinol (91%)
40	24.581	144	2,3-dihydro-3,5-dihydroxy-6-methyl-4H-pyran-4-

			one (80%)
41	24.687	122	2,4-dimethylphenol (94%)
42	25.295	122	4-ethylphenol (64%)
43	25.787	129	N-(3-methylbutyl)acetaminde (43%)
44	26.358	110	1,2-benzenediol (91%)
45	26.499	110	2-amino-3-hydroxypyridine (30%)
46	26.519	138	4-methoxy-1,3-benzenediamine (49%)
47	26.544	138	2-methoxy-3-methylphenol (47%)
48	26.557	138	2-methoxy-4-methylphenol (92%)
49	26.753	128	naphthalene (95%)
50	26.77	128	Azulene (81%)
51	27.357	120	2,3-dihydrobenzofuran (86%)
52	27.713	126	5-(hydroxymethyl)-2-furancarboxaldehyde (90%)
53	27.758	126	4,5-dihydro-3-methyl-1-propyl-1H-pyrazole (72%)
54	27.779	126	4-mercaptophenol (72%)
55	28.952	156	4-oxononanal (38%)
56	28.957	86	1,2,3-trimethyldiaziridine (43%)
57	30.164	124	4-methyl-1,2-benzenediol (94%)
58	30.721	114	ethyl cyclopropanecarboxylate (38%)
59	30.729	86, 282	3-aminopyrrolidine (38%), heptafluorobutyric (32%)
60	31.109	142	2-methylnaphthalene (96%)
61	31.218	150	2-methoxy-4-vinylphenol (91%)
62	31.72	142	1-methylnaphthalene (96%)
63	32.219	136	2,5,6-trimethyl-1,3,6-heptatriene (90%)
64	32.594	154	2,6-dimethoxyphenol (94%)
65	33.889	87, 116	thiocyanic acid (35%), 2-ethoxy-2-methylbutane (35%)
66	34.536	152	vanillin (90%)
67	34.635	204	tricyclo[6.3.0.0(2,4)]undec-8-ene (cas#1000152-25-6) (97%)
68	35.05	204	1H-cycloprop[e]azulene,... (cas#000489-40-7) (99%)
69	35.339	204	patchoulene (92%)
70	35.978	168	4-hydroxy-3-methoxy-benzoic acid (59%)
71	36.053	204	1H-cycloprop[a]naphthalene (cas#017334-55-3) (97%)
72	36.221	204	1H-cycloprop[e]azulene,... (cas#000489-39-4) (99%)
73	36.226	204	1H-cycloprop[e]azulene (cas#072747-25-2) (99%)
74	36.266	164	2-methoxy-4-(1-propenyl)phenol (96%)
75	36.785	204	1H-cycloprop[e]azulene,... (cas#021747-46-6) (93%)
76	36.791	204	1S,2S,5R-1,4,4-trimethyltricyclo[6.3.1.0(2,5)]dodec-8(9)-ene (92%)
77	36.802	204	azulene,... (cas#003691-11-0) (92%)
78	36.965	204	1H-cycloprop[e]azulene,... (cas#025246-27-9) (99%)
79	37.43	204	naphthalene... (cas#005951-61-6) (93%)
80	37.937	204	1H-cycloprop[e]azulene,... (cas#021747-46-6) (99%)
81	38.625	166	durohydroquinone (59%)
82	38.95	180	2,3,5,6-tetrafluoroanisole (64%)
83	40.382	222	epiglobulol (96%)

84	41.04	202	2(1H)-naphthalene… (cas#055220-87-6) (53%)
85	41.14	194	2,6-dimethoxy-4-(2-propenyl)phenol (83%)
86	41.185	222	globulol (97%)
87	41.445	186	2-heptyl-1,3-dioxane (43%)
88	43.237	204, 222	1H-cycloprop[e]azulene,… (cas#021747-46-6) (83%), 2-naphthalenemethanol,… (cas#001209-71-8) (81%)
89	43.314	204	naphthalene… (cas#005951-61-6) (96%)
90	44.292	194	2,6-dimethoxy-4-(2-propenyl)phenol (87%)
91	45.438	178	4-hydroxy-2-methoxycinnamaldehyde (98%)
92	46.275	210	1-(2,4,6-trihydroxyphenyl)-2-pentanone (50%)
93	47.988	138, 252	2,6,6-trimethylbicyclo[3.1.1]heptane (64%), 11-tetradecyn-1-ol (64%)
94	48.084	142	(E)-6-nonen-1-ol (58%)
95	52.319	208	3,5-dimethoxy-4-hydroxycinnamaldehyde (81%)
96	59.025	286	1,1'-sulfonylbis[4-chloro-benzene,.. (cas#000080-07-9) (99%)
97	65.422	277, 278	triphenylphosphine imide (93%), triphenylphosphine oxide (91%)
98	67.913	296	heneicosane (94%)
99	71.863	254	octadecene (93%)

6. *C. calophylla* bark

5. *E. wandoo* bark

11. *P. radiata* nut

4. *E. wandoo* dead flower

10. *P. radiata* wood

3. *E. wandoo* fresh flower

9. *P. radiata* dead leaf

2. *E. wandoo* dead leaf

8. *P. radiata* fresh leaf

1. *E. wandoo* fresh flower

7. *P. radiata* bark

250 200 150 100 50 0 ppm

250 200 150 100 50 0 ppm
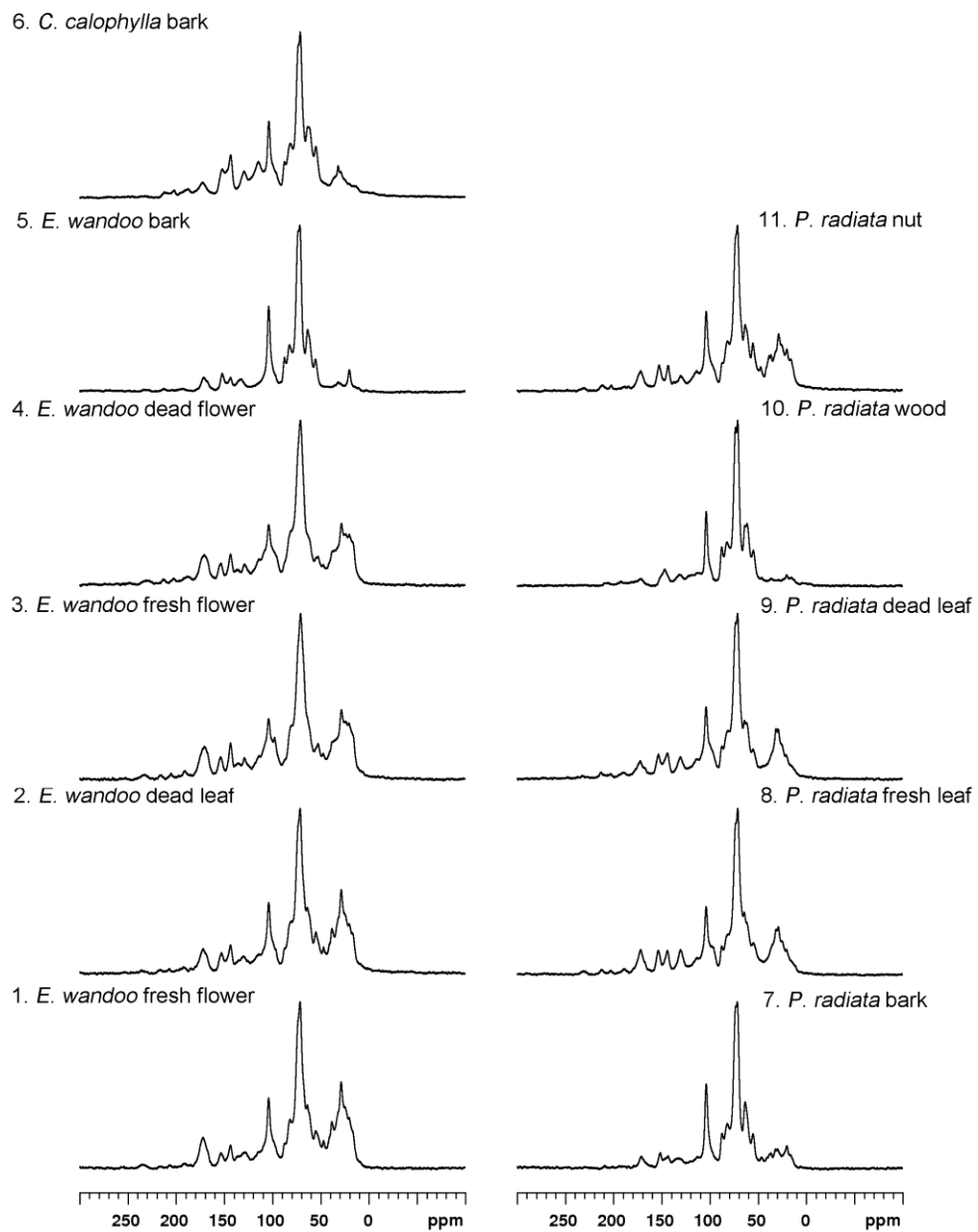
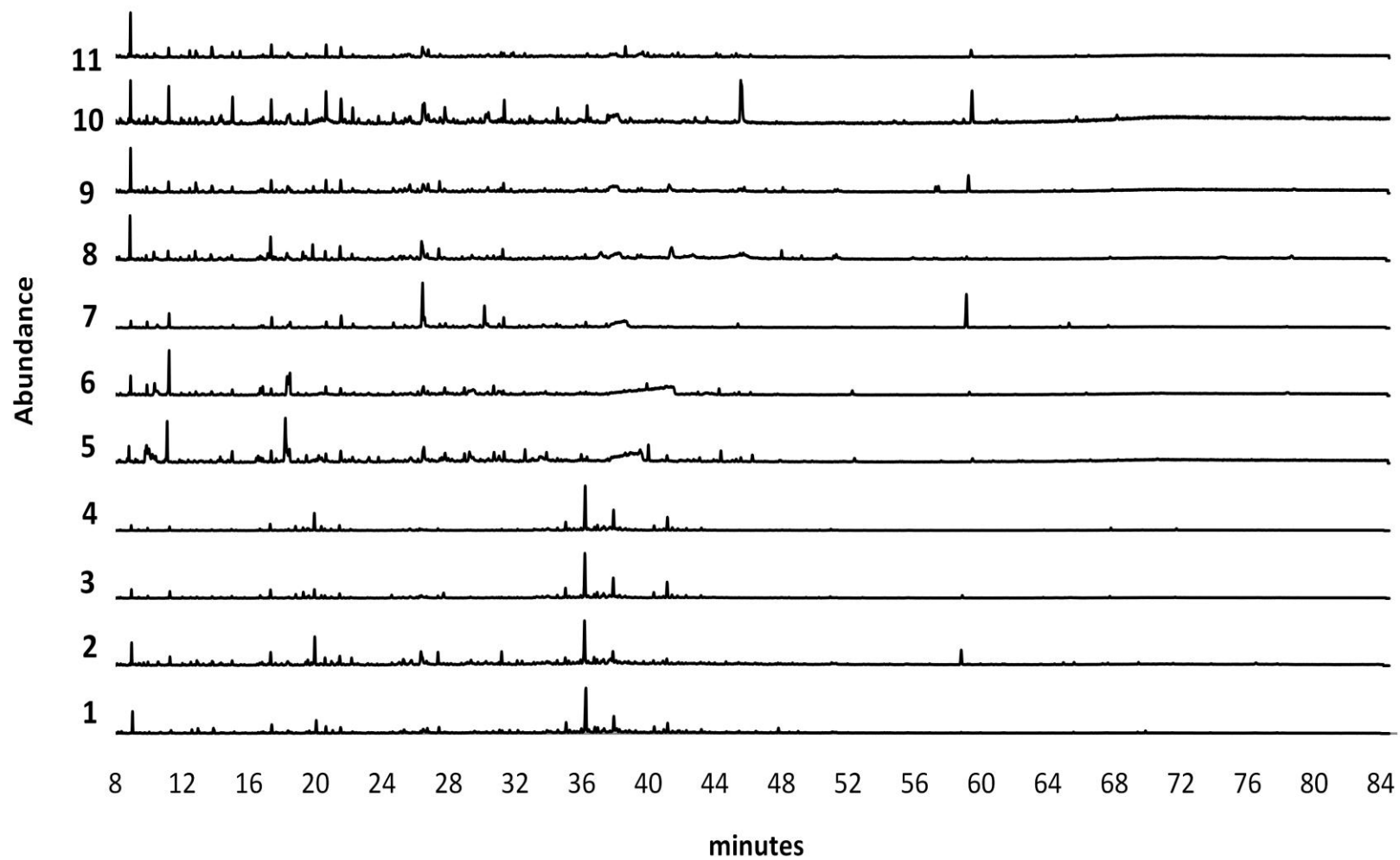Figure S1 Full solid-state ^{13}C NMR spectra of the plant residues.

Figure S2 Expansion of the flash py-GCMS results.

CHAPTER THREE

CAN ORGANIC RESIDUES FROM A PULP AND PAPER MILL BE IDENTIFIED IN THE SEDIMENTS OF THE RECEIVING ENVIRONMENT USING NMR AND PYROLYSIS TECHNIQUES?

*

The work in this chapter has been submitted to *Environmental Science and Pollution Research.*

STATEMENT OF AUTHORSHIP

Title of Paper	Can organic residues from a pulp and paper mill be identified in the sediments of a receiving environment using NMR and pyrolysis techniques?
Publication Status	O Published, O Accepted for Publication, ⊙ Submitted for Publication, O Publication style
Publication Details	Environmental Science and Pollution Research

AUTHOR CONTRIBUTIONS

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	Emma L Plant
Contribution to the Paper	Experimental development, performed analysis on all samples, data analysis and critical interpretation, wrote manuscript.
Signature	Date 8/10/13

Name of Co-Author	Ronald J Smernik
Contribution to the Paper	Supervised development of work, NMR data analysis and interpretation, reviewed manuscript.
Signature	Date 10/10/13

Name of Co-Author	Paul Greenwood
Contribution to the Paper	Supervised development of work, pyrolysis data analysis and interpretation, reviewed manuscript.
Signature	Date 8/10/13

Name of Co-Author	Lynne M Macdonald
Contribution to the Paper	Supervised development of work, multivariate data analysis and interpretation, reviewed manuscript.
Signature	Date 8/10/13

Name of Co-Author	John van Leeuwen
Contribution to the Paper	Supervised development of work, data analysis and interpretation, reviewed manuscript.
Signature	Date 9/10/13

ABSTRACT

The foreign materials or exaggerated loadings that can be found in wastewater (WW) can have long lasting ecosystem impacts when discharged. Lake Bonney SE is a well-mixed, shallow coastal lake in South Australia that has received WW from pulp and paper mills for many years. In recent years, enhanced WW treatment has greatly reduced organic loading to the lake. In this study, sedimentary organic matter (SeOM) sampled from near the WW discharge point was isolated and characterised using two sophisticated analytical techniques, for the purpose of determining whether persistent organic features of mill origin could be identified. A small amount of aromatic material that may have been from lignin-rich effluent was detected by solid-state ^{13}C nuclear magnetic resonance (NMR) spectroscopy, although these organics were not concentrated in the deeper sediments, which might have been anticipated if derived from the WW discharge. Methoxy-phenolic compounds, which may also result from the degradation of mill-derived lignin, were detected by flash pyrolysis-gas chromatography mass spectrometry (py-GCMS). Again, however, the concentrations of these markers were not obviously higher at depth. This suggests that either SeOM from the mill discharge has been distributed widely across the lake, transported through the sediment profile beyond the sampling depth, or that these organic features are typical of natural levels of lignin input.

INTRODUCTION

Pulp and paper manufacturing involves processes that separate wood into cellulose and non-cellulose components. These processes can be water intensive, with some commentators speculating the paper industry will become one of the largest manufacturing users of water during the twenty-first century (Kallas et al. 1994; Thompson et al. 2001). However, there is a trend towards improving water efficiency by recycling water within the plants, which also allows for the recovery of fibres that would have otherwise been lost in the wastewater (WW).

Pulp WW contains a range of compounds extracted from the wood, some of which have been transformed by the processes of pulping and bleaching. A key component of this material is dissolved lignin, which is responsible for discolouration of the pulp WW (Thompson et al. 2001). Modern pulp and paper mills incorporate primary and secondary effluent treatment technologies to remove bleaches, dissolved woody substances and other

wastes that are potentially harmful to the environment (Hall et al. 2009b; Lewis et al. 2011; Thompson et al. 2001). In most cases, modern WW treatment technologies meet or surpass regulated water quality goals (Hall et al. 2009b).

Concern still remains about the legacy of pulp and paper effluent in receiving environments (Hall et al. 2009b). The literature is rich with research on the effects on organisms exposed to effluent and field-based assessments (Hall et al. 2009a). Much of this focus has been on the impacts on fish populations, which can include endocrine disruption and the impairment of development, reproduction and survival (Costigan et al. 2012; Pollock et al. 2010). However, despite the many dedicated studies and reports, there is still only limited understanding of the impacts of pulp and paper mill effluents in receiving environments, especially in terms of longevity, due to the inconsistent use of approaches and monitoring tools (Hall et al. 2009b). It is only recently that questions about the longevity of pulp and paper mill effluent discharge effects on receiving waters have been seriously considered (Hall et al. 2009a; Hall et al. 2009b).

The issue of long-term effects, if any, of pulp and paper mill effluent discharge into receiving waters has attracted recent interest in the south east of South Australia. Here, the local pulp and paper industry has been discharging WW to Lake Bonney SE since 1939 (Glover 1975; Struve 2007). By 1970, the mills were discharging up to 18 ML/day of filtered WW with total suspended solids (TSS) as high as 36 t/day (Glover 1975). Struve (2007) reported that following the installation of primary WW treatment (in 1973) and secondary WW treatment (in 1992), the lake received an average effluent flow rate of 31.5 ML/day with TSS measuring on average 22 mg/L.

Solid-state ^{13}C nuclear magnetic resonance (NMR) spectroscopy and pyrolysis-gas chromatography mass spectrometry (py-GCMS) are two techniques that have been widely used for organic matter (OM) characterisation. NMR spectroscopy can provide quantitative estimates of broad classes of carbon functional groups, such as carbonyl, aromatic, O-alkyl and alkyl (Simpson et al. 2011). On the other hand, py-GCMS provides semi-quantitative speciation level information often including information on the functional groups (Gaffney et al. 1996).

The complexity of the data provided by these analytical techniques does require specialised interpretation, more so when trying to combine information from both

techniques. Multivariate statistical methods can help in the interpretation of complex data through formal quantification of the overall similarity/difference between samples and identification of the main chemical drivers contributing to the differences. This can help those who are unfamiliar with data outputs from specific techniques to identify relationships or features of interest. We recently demonstrated how multivariate methods can be used to quantify similarity within a set of plant-derived organic materials, as well as to quantitatively assess the complementary nature of results from different characterisation techniques (Plant et al. in press).

The aim of the study reported in this paper was to characterise the SeOM of Lake Bonney, with a particular interest in identifying any chemical markers indicating a legacy effect of lignin rich WW of pulp and paper mill origin. To do this, the characterisation techniques of solid-state ^{13}C NMR spectroscopy and flash py-GCMS were used. Multivariate methods were used to assess similarities in the chemistry of the SeOM as determined by each of the techniques. An assessment of the complementary nature of these analytical techniques was also undertaken by multivariate methods.

MATERIALS AND METHODS

DESCRIPTION AND HISTORY OF THE SAMPLE SITE

Lake Bonney SE is a large, shallow lake situated in the south-east of South Australia, about 10 km south of Millicent (Figure 1). The lake is ~23 km long, up to 4.5 km wide, and has a maximum depth of 3.6 m. It is well mixed due to frequent coastal winds. Prior to the discharge of pulp and paper mill effluent into Lake Bonney SE from the late 1930s/early 1940s it was a popular community recreational location (EPA 2004). Since European settlement, in the early 1800's, the hydrology and condition of the lake had been substantially altered due to drainage schemes for flood control and impacts from land use (EPA 2004; Struve 2007).
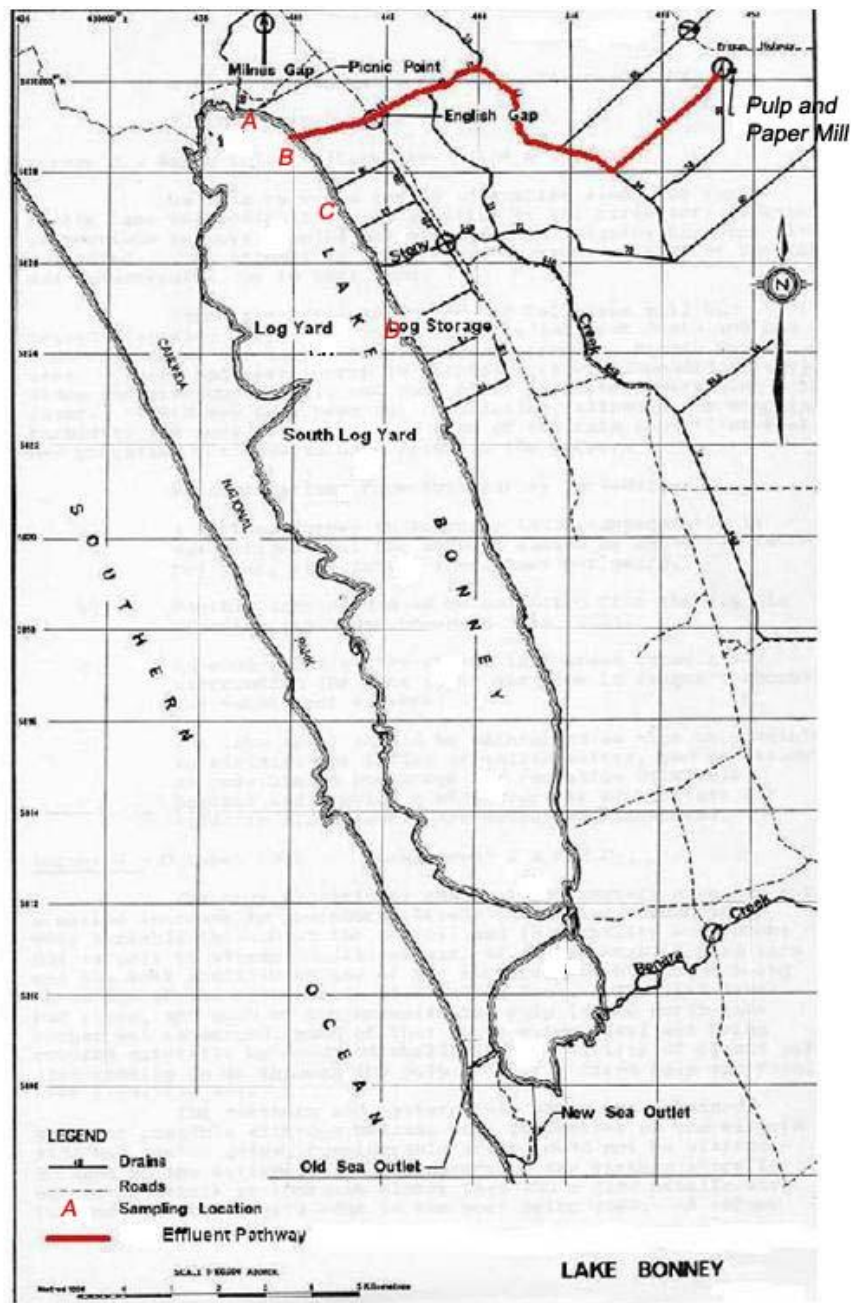
Figure 1 Map of Lake Bonney SE showing sediment sampling sites and the path of the main drain used for the discharge of effluent from the pulp and paper mill. Adapted from Struve (2007).

The first pulp and paper mill in the Millicent area began to legally discharge WW to Lake Bonney in 1939 (Glover 1975; Struve 2007). At this time, the WW that was discharged was colourless after filtration, although it contained a considerable amount of dissolved waste, fine pulp and other debris (Glover 1975; Struve 2007). As early as 1945 there were localised impacts at the north-eastern shore of the lake, near the WW discharge point (Struve 2007). In 1958, a second pulp and paper mill was established and permitted to discharge

its untreated WW into drains that lead to Lake Bonney (Glover 1975). By 1970, both mills were discharging a similar quantity of WW (18 ML/day) with TSS measuring up to 18 t/day (Glover 1975).

In order to limit the TSS load of the WW discharged, both mills installed clarifiers in 1973 (Glover 1975). This led to a marked reduction in the deposition of mill-derived organic fibre material on the north-eastern shore of Lake Bonney. Further environmental initiatives to improve WW quality were carried out at the mills, including changing the bleaching agent from chlorine to peroxide in 1991 (EPA 2004). An additional clarifier was installed in 1992 to reduce the discharge of suspended solids. In 1993, three aerated stabilisation basins (ASBs) were installed after the clarification process to reduce the biological oxygen demand (BOD) of the effluent. At the time of this study, the ASB system had a retention time of ~30 days, consisting of three treatment ponds: a continuously-stirred tank reactor, followed by a plug flow reactor and a stabilisation lagoon (Slade 2006; van Leeuwen et al. 2012). After the ASB system, the treated WW is discharged to Lake Bonney.

The impacts of this historic and continuing WW discharge into Lake Bonney have been the focus of many studies (EPA 2004; Glover 1975; Haynes et al. 2007; Struve 2007; van Leeuwen et al. 1993; van Leeuwen et al. 2012). These include investigations into the nature of the organics at the mill including the high WW colour, and treatment practices targeting their removal (Struve 2007; van Leeuwen et al. 1993; van Leeuwen et al. 2012). The condition and health of the lake, including both the water and sediment ecologies has been an issue of on-going concern (EPA 2004; Haynes et al. 2007; Struve 2007). However, there have been few investigations on the long term effects and the residence time of the effluent OM discharged to Lake Bonney, particularly the relatively high concentration outflows prior to the implementation of contemporary treatment strategies. There has been some speculation that the organics discharged from the mills to Lake Bonney over time may have been incorporated and accumulated in the sediments, as has been reported elsewhere (Judd et al. 1998; Struve 2007), where they may be protected from mineralisation and persist for some time in these environments. To address this knowledge gap, here we investigate this issue using advanced analytical techniques which provide sensitive characterisation of SeOM at the molecular level.

SEDIMENT SAMPLES: COLLECTION AND PREPARATION

Sediment profiles were collected in ~1 m of water from four locations on Lake Bonney on 31 August 2010. Sediments were collected using a hand-held corer, consisting of a 50 mm diameter PVC pipe with a screw-top stopper on one end and a rubber stopper on the other end. Sediment samples were all collected approximately 100 m from the lake's eastern shoreline; four sites (A-D) represent an approximate north-south transect taken 500 m, 1.5 km, 3 km and 6 km south of the northern extremity of the lake (Figure 1). The discharge drain of the mill is located nearest site B (Figure 1). Sediments were cored to a depth of up to 12 cm; deeper sampling was prevented by the presence of stones. Sediment cores were divided into 2-4 layers based on visually distinguishable coloured layers (each ~3 cm deep). For each layer, duplicate sediment samples (3 g each) were treated with approximately 30 mL of 1M hydrochloric acid and shaken end-over-end for 30 minutes. Following centrifugation, the supernatant was removed and the duplicate sediment samples for each layer were combined and rinsed twice with deionised water. The residue was demineralised with hydrofluoric acid, using the method of (Skjemstad et al. 1994). The remaining organic material was separated from small stones and sand that resisted HF-treatment by suspending it in a small volume of deionised water and decanting. The particulate OM was isolated by centrifugation and freeze-dried for analysis.

SOLID-STATE ^{13}C CP NMR SPECTROSCOPY

Solid-state ^{13}C cross polarization (CP) NMR spectra were acquired with magic angle spinning (MAS) at a ^{13}C frequency of 100.6 MHz on a Varian Unity INOVA 400 spectrometer. Samples were packed in a 7 mm diameter cylindrical zirconia rotor with Kel-F end-caps, and spun at 6500±100 Hz in a Doty Scientific supersonic MAS probe. Spectra were acquired using a ramped-amplitude cross polarization (CP-ramp) pulse sequence, in which the ^{1}H spin lock power was varied linearly during the contact time. A 1 ms contact time was used and 2000 transients were collected for each spectrum. Recycle delays (2 s) were chosen to be > 5 × T_1H, as determined in preliminary inversion-recovery experiments. Free induction decays were acquired with a sweep width of 50 kHz; 1216 data points were obtained over an acquisition time of 12 ms. All spectra were zero-filled to 8192 data points and processed with a 100 Hz Lorentzian line broadening and a 0.01 s Gaussian broadening. Chemical shifts were externally referenced to the methyl resonance of hexamethylbenzene at 17.36 ppm.

For multivariate analysis, data was normalised and baseline corrected to ensure all data points were not negative and further smoothed by taking the average of each five consecutive points throughout the spectra. The NMR trace in the spectral region of 0-250 ppm was used for multivariate analyses.

FLASH PY-GCMS

Flash pyrolysis of small sub-samples of isolated sediment organics (0.5 – 0.8 mg) was conducted at 650 ^{o}C for 10 s using a Chemical Data Systems 160 Pyroprobe inserted into a dedicated pyrolysis chamber which was maintained at 250 ^{o}C. The pyrolysis chamber was mounted directly onto the vaporisation injector of a Hewlett Packard (HP) 6890 Series II GC, coupled to a 5973b mass selective detector used for pyrolysate detection. A 60 m, 0.25 mm i.d., 0.25 μm phase DB-5MS capillary column (J&W) was used with helium carrier gas (9 psi). Injection splits of between 20 and 50 mL min^{-1} were used. The GC oven was temperature programmed from 40 ^{o}C (2 min hold) at 4 ^{o}C min^{-1} to 310 ^{o}C (15 min) and the transfer line was kept at 310 ^{o}C. Full scan m/z 50–550 mass spectra were acquired with an electron energy of 70 eV. Products were identified by mass spectral correlation to the NIST05 mass spectral library; see Supplementary Material for list of products (Table S1). Assignments were only made for high quality peak matching (typically ›70%) between measured and library mass spectra.

The integrated peak areas (from the total ion chromatogram (TIC)) of all identified products (Online Resource 1) were used for multivariate analysis. Products not detected in any particular sample were given a zero value in the statistical treatments. For each sample, peak areas were log transformed and normalised.

MULTIVARIATE STATISTICAL COMPARISON

Multivariate statistical analysis was undertaken to assess the relative power of the two analytical techniques to discriminate between samples based on their organic chemical composition, as detailed by Plant et al. (in press). Non-parametric multidimensional scaling (nMDS) was used to analyse the NMR and py-GCMS data and was carried out in Primer v6 (Clarke et al. 2006). The nMDS approach provides an ordination plot that represents the relative relationship between samples based on the rank order of the

corresponding dissimilarities between samples. The nMDS analysis was conducted on resemblance matrices suitable for the individual data types. The resemblance matrix for continuous trace spectral-NMR data was based on Manhattan Distance, while the resemblance matrix for the py-GCMS data, consisting of a set of individual peak intensities, was based on Bray-Curtis dissimilarity. The nMDS ordination plots display the relationships between samples based on their NMR or py-GCMS chemical profiles: the closer the samples are in the ordination plot, the more similar their chemical compositions as identified by that analytical technique. The 'stress' level associated with the plots indicates the degree of confidence in the multivariate data being expressed in the 2-D ordination: a stress of <0.2 is considered robust whereas a stress of >0.2 indicates that caution should be used in interpreting the 2-D plot, as there is a reduced ability to display relationships in the complex data.

Comparison of the relationship between the NMR and py-GCMS ordinations was carried out using the Spearman Rank Correlation method (termed RELATE in Primer), reported as ρ (rho). It provides a non-parametric measure of statistical dependence between the similarity matrices derived from the NMR and py-GCMS data.

RESULTS AND DISCUSSION

SOLID-STATE ^{13}C CP NMR SPECTROSCOPY

Solid-state ^{13}C cross polarisation (CP) NMR spectra of isolated organics from the lake's sediments from locations A to D (Figure 1) are shown in Figure 2. They contain numerous common features. All contain clear peaks due to long-chain alkyl carbon (~30 ppm), O-alkyl carbon (~80 ppm) and carboxyl carbon (~180 ppm). Most also contain smaller peaks due to methoxyl carbon (~56 ppm), di-O-alkyl carbon (~105 ppm), aryl carbon (~130 ppm) and O-aryl carbon (~150 ppm). For sediments from sites A, B and C, there is a general increase in alkyl carbon (~30 ppm), and a decrease in O-alkyl carbon (~80 ppm) with depth. There is also a slight decrease in carboxyl carbon (~180 ppm) with depth at all sites. The greatest variation in composition with depth is at Site D. Here, the alkyl carbon peak at ~30 ppm of the middle layer is weaker than for the upper or lower layers.
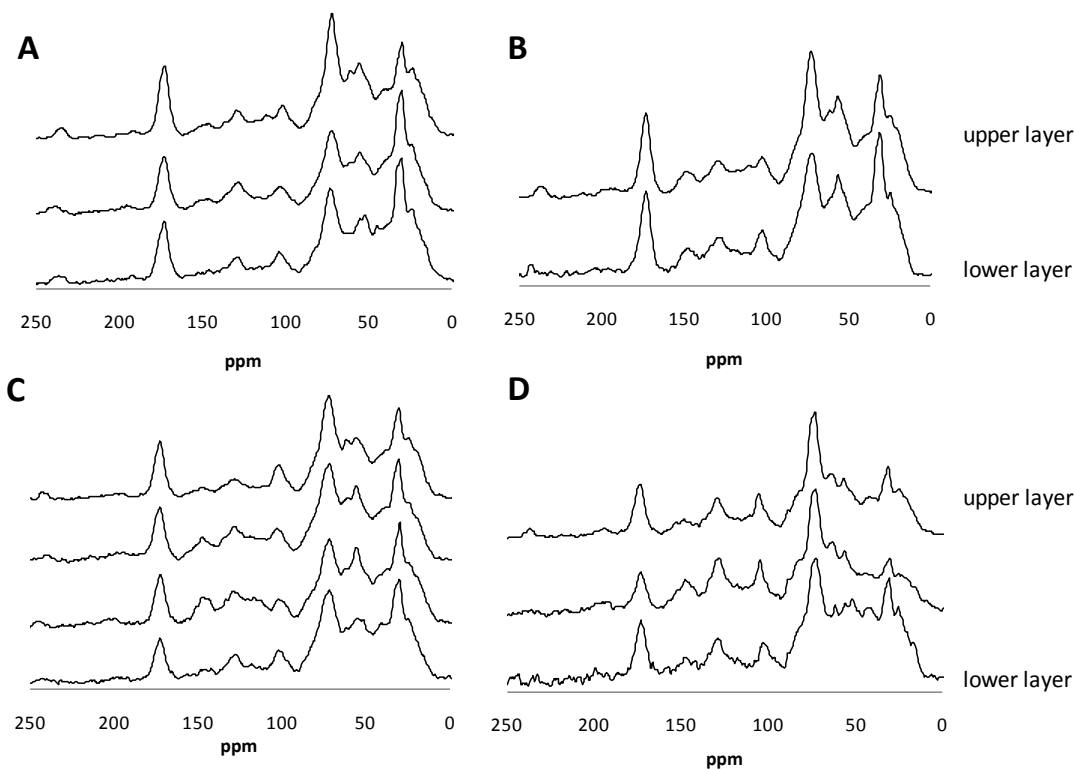
Figure 2 Solid-state ^{13}C CP NMR spectra of each of the sediment layers isolated from the four sites (A-D).

Before primary treatment of pulp mill WW, large amounts of suspended particulate OM (as previously detailed) were discharged that subsequently settled around the lake (Kinhill Engineers Pty Ltd 1990). The colour of the lake's waters had increased to very high levels (to >800 Hazen units) after effluent discharge from chemical pulping since the 1960s (Glover 1975; Kinhill Engineers Pty Ltd 1990). Total organic carbon in the lake waters reached up to several hundred milligrams per litre (Kinhill Engineers Pty Ltd 1990), which demonstrated a historic time when there were very high particulate and dissolved organic loading discharged into the lake's waters. In 1993, the lake's aquatic organic loading was of high colour and had a strong aromatic signature, but by 2007, the colour was much lower and aromatic signatures (determined by ^{13}C NMR) were predominantly absent in lake waters (van Leeuwen et al. 2012). With the history of the lake having waters of high colour and high aromaticity, one might expect to see an aromatic-rich layer of organics in the sediments of Lake Bonney. Although aromatic carbons were detected in this study, including the O-aryl (145 ppm) and methoxy (56 ppm) carbons which together are indicative of lignin-derived methoxy-phenol products (Deshmukh et al. 2001), no strongly aromatic layer can be seen in the NMR spectra shown in Figure 2.

Organic sediments from the same region as Lake Bonney, but removed from the WW source, were analysed by NMR in a study by Krull et al. (2009), who reported a similarly low amount of aromatic carbon to that detected in the Lake Bonney sediments here. These results indicate that the sediments within our study do not contain more aromatic carbon than adjacent sediments unaffected by pulp and paper mill effluent. The lack of aromatic-rich organic indicators within the sediment layers of the pulp and paper mill receiving environment suggest that the mill organics were diluted by the organic loading from other sources such as the inflow from creeks or the addition of autochthonous organics to the lake, or decomposed. The lake's waters experienced high levels of algae and between 2003 and 2007; chlorophyll concentrations at this time were in the 35-75 µg/L range (P Goonan, personal communication). This is likely to have arisen from the lake being high in nutrients but with much lower colour levels than reported in the 1970s to early 1990s. The resultant autochthonous organics may then have substantially diluted those derived from the mill's processes within the sediments. However, organics derived from algae would be generally more degradable by microbial activity than lignin derived macro-molecules of polyaromatic character.

Sediment sampling for this study was conducted in 2010, three years after it had been found that the colour and aromatic nature of organics present in the lake's waters had reduced dramatically from historically high levels as reported in van Leeuwen et al. (2012). This demonstrates there had been significant removal process(es) from the water column.

Louchouarn et al. (1997, 1999) investigated sediments of the Lower St. Lawrence Estuary and the Saguenay Fjord in order to evaluate the fate of solid organic wastes from pulp and paper mills that discharged into this system. They reported (in 1997) elevated levels of lignin and lignin-signature compounds indicative of gymnosperm wood in deeper sediment horizons (to about 60 cm) of the upper Saguenay basin compared with background estuary levels, and these were directly attributed to intense pulp and paper mill activity. An increase of lignin and lignin-derived compounds such as vanillyl phenols (by CuO oxidation) with sediment depth was most rapid at depths greater than 30 cm. With increased control of solids discharged from mills in the late 1980s and early 1990s, lignin organic carbon decreased in the upper basin surface sediments (Louchouarn et al. 1997).

The mills discharging WW to Lake Bonney had adopted primary treatment for reduction of solid organic wastes (fibre material removal) from 1973 and in our study; samples were collected nearly 40 years later. In contrast, the findings of Louchouarn et al. (1997) were from samples collected in 1994 only several years after the reported increase in control measures for solid waste discharge by pulp mill industries. The highest concentrations of lignin and lignin-derived compounds reported by Louchouarn et al. (1997) occurred in sediments deeper than the maximum depth we were able to sample (12 cm) and so it is possible that deeper sediments of Lake Bonney may be richer in lignin-derived aromaticity.

The ^{13}C NMR spectra (Figure 2) have greater similarity to spectra of lake water organics collected in 2007 rather than in 1993 (van Leeuwen et al. 2012), with notable diminishing of the aromatic signal since 1993. Coinciding with the prominent aromatic signatures in mill discharge waters in 2007 and in lake waters in 1993, were high colour levels. Hence there appears to have been no build up and persistence of aromatic compounds that would also contribute to high colour, within the sediments of Lake Bonney (to 12 cm). Lignin, although often considered a recalcitrant biopolymer, can undergo degradation processes under either aerobic or anaerobic conditions (Louchouarn et al. 1997) and can be evidenced by increases in relative abundances of the acid to aldehyde forms of the constituent cinnamyl, syringyl and vanillyl phenols (Loh et al. 2012). The microbiologically mediated transformations of chlorinated phenolics representative of pulp mill sources, under aerobic (O-methylation) and anaerobic (de-O-methylation) conditions and their degradation have been reported by Neilson et al. (1983, 1987) and Remberger et al. (1986). Nonetheless, there is potential for long term persistence of lignin derivatives within sediments. Lignin phenols, identified as tracers for land-ocean transfer of terrestrial OM, have been shown to display radiocarbon ages of approximately 300-1200 years in surface sediments in carbon isotope studies by Feng et al. (2013).

FLASH PY-GCMS

The TICs from py-GCMS analysis of the isolated organic sediments are shown in Figure 3. A list of the pyrolysis products is provided in the Supplementary Material. The TICs of all samples are generally similar with depth, with products displaying a broad molecular weight range. In particular, the site A pyrolysis data appears similar with depth. However, the products are of lowest concentration in the deepest layer, which also showed its major product at a retention time of ~8 min which was not present in the sediments above

(Figure 3). Pyrolysis-GCMS distributions for sites B and C are similar to that of A, with only subtle differences seen with depth (Figure 3). The chromatograms for site D are quite weak compared to those for the other sites, and the separate depths showed the largest differences (Figure 4). Across the sampling transect, the upper layer organics appear very similar from site A to D. For the deeper layers, there appears to be similarity among sites A to C, with site D showing more notable differences.

Significant differences in the organic speciation of the sediments are not clearly evident from the chromatograms shown in Figure 3. To scrutinise the product distributions with sediment depth and along the transect in more detail, the products were classified into broad classes and the total relative abundances of these are shown in Figure 4. Some trends with depth were observed: nitrogen (N) products decreased from the upper to the next lowest layer at sites A, C and D; and aliphatics increased with depth at sites A and C. The relative abundance of aromatics and phenols was relatively constant with depth at each location; again, there was no indication of an aromatic-rich layer. The site B organics, being closest to the source of mill discharge, do not substantially vary with depth, which could be expected due to the well-mixed nature of the lake.
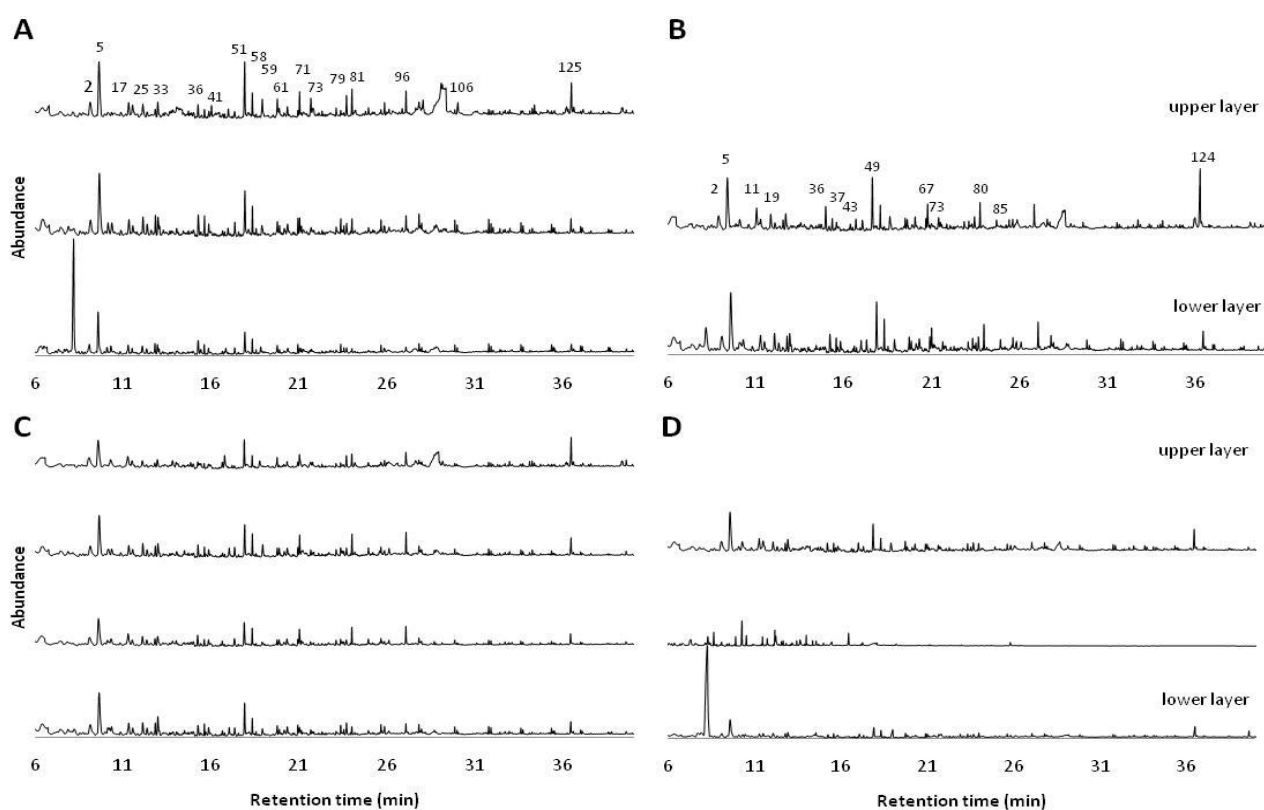
Figure 3 Normalised gas chromatograms from the pyrolysis of sediment organics isolated from the four sites (A-D) on Lake Bonney. Peak assignments correspond to products listed in the Supplementary Material.

Methoxy or hydroxyl-phenols are typical lignin pyrolysis products which, if detected in high abundance, could be attributed to organic wastes from the paper mill (Brudin et al. 2010; Deshmukh et al. 2001; Greenwood et al. 2002). Methoxy-phenol products were present in the pyrograms from the majority of the sediments but showed no obvious concentration increase in sediments at depths which might have corresponded to periods prior to the implementation of more advanced treatment strategies at the mill. Overall, therefore, the py-GCMS analysis provided no evidence of an identifiable contribution of mill organics to the sediments in this area.

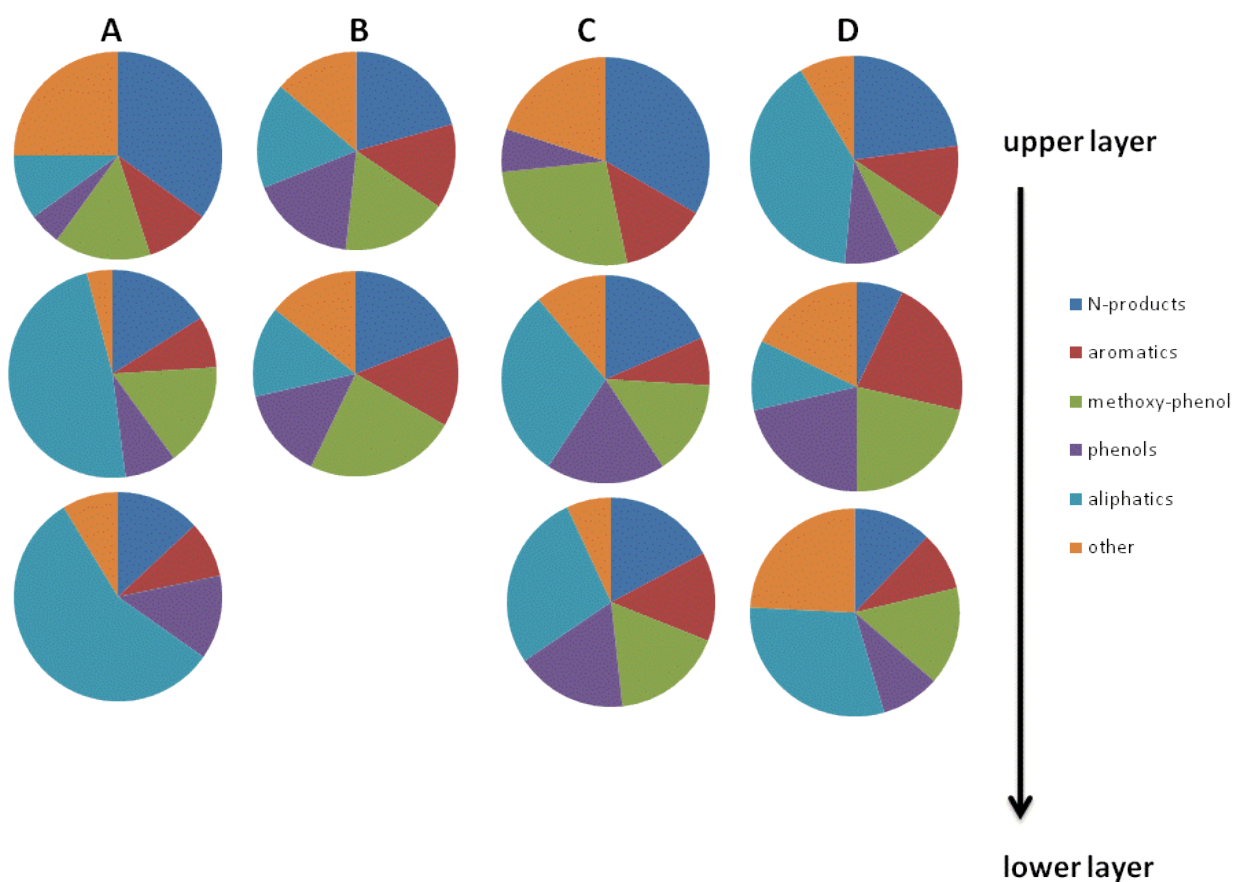Figure 4 Functional group distributions derived from pyrolysis-gas chromatography mass spectroscopy for each of the sediment organics isolated from the four sites on Lake Bonney (A to D).
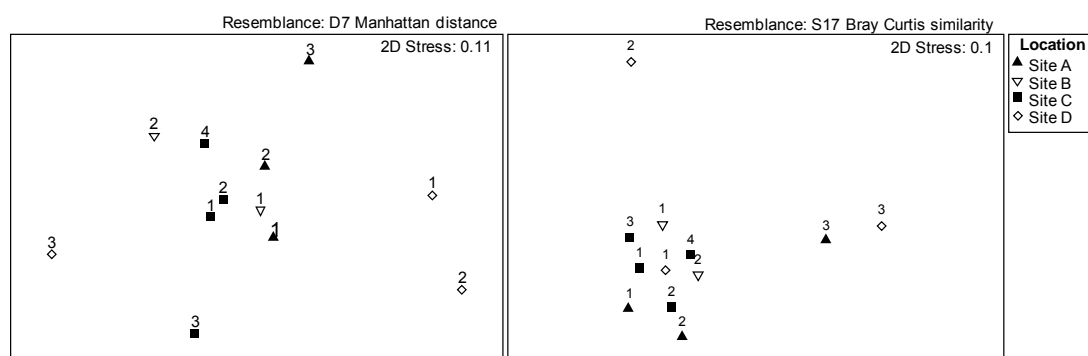
Figure 5 Non-parametric multidimensional scaling ordination plots representing the similarities between the sediment organics based on the NMR (left) and py-GCMS (right) characterisation. The closer the points in the plot, the more chemically similar the organics as determined by that technique. Numbers 1 to 4 are labelling the depth of the organic layers isolated from the four sites (A to D), 1 being the top-most layer and 4 being the deepest.

In order to quantitatively assess the relative similarity in the overall organic profiles measured by NMR and py-GCMS, a multivariate statistical comparison was undertaken similar to that described by Plant et al. (in press). This involved the generation of nMDS plots for each characterisation technique to facilitate visual comparison of the similarities among the samples as expressed in the respective data sets. The nMDS plots show similarities between the top two layers (points labelled 1 and 2) at all locations, except for site D by NMR (Figure 5). A greater dissimilarity of site D is evident in the NMR data (Figure 2). In this case, NMR may be more sensitive to the variation in alkyl and O-alkyl carbons in these layers. Both techniques also distinguished layer 3 from site A and layers 2 and 3 from site D as dissimilar to all other samples. The chemical markers that distinguish these three sediment organics from the remainder include layer 2 from site D having a more defined aromatic region and less alkyl carbon by NMR, and a smaller molecular weight range by py-GCMS; layer 3 from site A lacks O-aryl carbon at 150 ppm in the NMR spectrum; and layer 3 from sites A and D are both dominated by a major product at ~8 min in their py-GCMS chromatograms.

A quantitative comparison of the resemblance matrices was carried out using the Spearman Rank Correlation method (Plant et al. in press). This provides a measure, rho (ρ), of the strength of the relationship between the two resemblance matrices ($\rho=0$ indicating no relationship and $\rho=1$ indicating the strongest correlation). Here, moderate

correlation between NMR and py-GCMS resemblance matrices was indicated by the significant (P < 0.05) relationship with ρ = 0.42. This can be interpreted as 42% of the discrimination between the organic sediment samples being common to both NMR and py-GCMS data sets. In other words, 42% of information provided by these two techniques is redundant. This level of redundancy in the information provided by these two techniques provides confidence in the conclusions drawn about the chemical nature of the organic samples. This demonstrates broad (42%) consistency in the sample discrimination provided by both techniques, but also shows that 58% of the discrimination is unique to each technique, highlighting the benefit of using both analytical methods together.

CONCLUSIONS

NMR and py-GCMS analysis of the organic sediments of a lake (Lake Bonney SE) that received WW (primary and secondary treated) from a pulp and paper mill provided no clear evidence (e.g. high aromatic content, methoxy-phenol speciation) typical of wood/lignin sourced waste. A weak aromatic signal was detected by NMR and methoxy-phenols were detected by py-GCMS at only moderate abundances. There was no indication of an aromatic-rich layer in the sediments, which might have reflected a period (to at least the early 1990's) when the lake water was highly coloured and rich in dissolved aromatic compounds. The lake is generally well mixed and has a water quality that is similar along its length. It appears that the highly coloured aromatic compounds derived from pulping processes and released into lake waters in the past have now been largely degraded or transported to deeper sediments; their persistence in lake sediments to 12 cm was not evident in this study.

ACKNOWLEDGEMENTS

REFERENCES

Brudin S, Schoenmakers P (2010) Analytical methodology for sulfonated lignins. J Sep Sci 33:439-452

Clarke KR, Gorley RN (2006) PRIMER v6: User Manual/Tutorial. PRIMER-E, Plymouth

Costigan SL, Werner J, Ouellet JD, Hill LG, Law RD (2012) Expression profiling and gene ontology analysis in fathead minnow (Pimephales promelas) liver following exposure to pulp and paper mill effluents. Aquat Toxicol 122:44-55

Deshmukh AP, Chefetz B, Hatcher PG (2001) Characterization of organic matter in pristine and contaminated coastal marine sediments using solid-state 13C NMR, pyrolytic and thermochemolytic methods: a case study in the San Diego harbor area. Chemosphere 45:1007-1022

EPA (2004) Lake Bonney South East, South Australia Past, Present and Possible Future. In: Environmental Protection Authority SA (Hrsg.), Adelaide, pp. 1-16

Feng XJ, Benitez-Nelson BC, Montlucon DB, Prahl FG, McNichol AP, Xu L, Repeta DJ, Eglinton TI (2013) C-14 and C-13 characteristics of higher plant biomarkers in Washington margin surface sediments. Geochim Cosmochim Acta 105:14-30

Gaffney JS, Marley NA, Clark SB (1996) Humic and Fulvic Acids and Organic Colloidal Materials in the Environment. In: Gaffney JS, Marley NA , Clark SB (Editors) Humic and Fulvic Acids. Isolation, Structure and Environmental Role, ACS Symposium Series. American Chemical Society, Washington, pp 2-16

Glover B (1975) Lake Bonney (S.E.) Ecological Study, Engineering and Water Supply Department, Adelaide

Greenwood PF, van Heemst JDH, Guthrie EA, Hatcher PG (2002) Laser micropyrolysis GC-MS of lignin. J Anal Appl Pyrol 62:365-373

Hall TJ, Fisher RP, Rodgers JH, Jr., Minshall GW, Landis WG, Kovacs TG, Firth BK, Dube MG, Deardorff TL, Borton DL (2009a) A Long-Term, Multitrophic Level Study to Assess Pulp and Paper Mill Effluent Effects on Aquatic Communities in Four US Receiving Waters: Background and Status. Integr Environ Assess Manag 5:189-198

Hall TJ, Landis WG (2009b) Introduction to Special Series: Measurement and Analysis of the Potential Long Term Impact of Pulp and Paper Mill Effluent on Receiving Waters. Integr Environ Assess Manag 5:186-188

Haynes D, Gell P, Tibby J, Hancock G, Goonan P (2007) Against the tide: the freshening of naturally saline coastal lakes, southeastern South Australia. Hydrobiologia 591:165-183

Judd MC, Stuthridge TR, Price RW (1998) Pulp mill sourced organic compounds from new zealand sediments part 3: Mechanical pulp mills and remote sites. Chemosphere 36:2311-2320

Kallas J, Munter R (1994) Posttreatment of pulp and paper-industry wastewaters using oxidation and adsorption processes Water Sci Technol 29:259-272

Kinhill Engineers Pty Ltd (1990) APCEL Pulp Mill Redevelopment Project, Draft Environmental Impact Statement. ISBN 0949397415 Concord Printing Pty Ltd Adelaide

Krull E, Haynes D, Lamontagne S, Gell P, McKirdy D, Hancock G, McGowan J, Smernik R (2009) Changes in the chemistry of sedimentary organic matter within the Coorong over space and time. Biogeochemistry 92:9-25

Lewis R, van Leeuwen JA, Smernik RJ, Chow CWK, Everson A, Nothrop SC, Beecham S (2011) Changes in the organic character of post-coagulated Pinus radiata sulfite pulp mill wastewater under aerated stabilization basin treatment A laboratory scale study. Chem Eng J 175:160-168

Loh PS, Chen CTA, Anshari GZ, Wang JT, Lou JY, Wang SL (2012) A comprehensive survey of lignin geochemistry in the sedimentary organic matter along the Kapuas River (West Kalimantan, Indonesia). J Asian Earth Sci 43:118-129

Louchouarn P, Lucotte M, Canuel R, Gagne JP, Richard LF (1997) Sources and early diagenesis of lignin and bulk organic matter in the sediments of the lower St. Lawrence Estuary and the Saguenay Fjord. Mar Chem 58:3-26

Louchouarn P, Lucotte M, Farella N (1999) Historical and geographical variations of sources and transport of terrigenous organic matter within a large-scale coastal environment. Org Geochem 30:675-699

Neilson AH, Allard AS, Hynning PA, Remberger M, Landner L (1983) Bacterial methylation of chlorinated phenols and guaiacols - formation of veratroles from guaiacols and high-molecular-weight chlorinated lignin. Appl Environ Microb 45:774-783

Neilson AH, Allard AS, Lindgren C, Remberger M (1987) Transformations of chloroguaiacols, chloroveratroles, and chlorocatechols by stable consortia of anaerobic-bacteria. Appl Environ Microb 53:2511-2519

Pollock MS, Dube MG, Schryer R (2010) Investigating the link between pulp mill effluent and endochrine disruption: attempts to explain the presence of intersex fish in the Wabigoon River, Ontario, Canada. Environ Toxicol Chem 29:952-965

Remberger M, Allard AS, Neilson AH (1986) Biotransformations of chloroguaiacols, chlorocatechols, and chloroveratroles in sediments. Appl Environ Microb 51:552-558

Simpson AJ, McNally DJ, Simpson MJ (2011) NMR spectroscopy in environmental research: From molecular interactions to global processes. Prog Nucl Mag Res Sp 58:97-175

Skjemstad JO, Clarke P, Taylor JA, Oades JM, Newman RH (1994) The removal of magnetic-materials from surface soils - a solid-state C-13 CP/MAS NMR-study. Aust J Soil Res 32:1215-1229

Slade AH (2006) A review of Kimberly Clark Australia Millicent's wastewater treatment system, Scion

Struve N (2007) Relationship between UV-visible Properties of Organic Matter with Treatment of Water/Wastewaters. Diploma Thesis, University of South Australia, Adelaide

Thompson G, Swain J, Kay M, Forster CF (2001) The treatment of pulp and paper mill effluent: a review. Bioresource Technol 77:275-286

van Leeuwen J, Nicholson B, Hayes K (1993) Distribution of chlorophenolic compounds, from a pulp mill, in Lake Bonney, South Australia. Mar Freshwater Res 44:825-834

van Leeuwen J, Lewis R, Goonan P, Struve N, Everson A, Nothrop S, Smernik R, Chow CK, Fabris R, Rupasinghe M (2012) Changes in character of organics in the receiving environment of effluent from a sulphite pulp mill. Environ Sci Pollut Res 19:2151-2158

Supplementary Material

Table 1 Tentatively identified organic products from the flash pyrolysis analysis of the sedimentary organics. The mass spectral match (%), based on correlation with the NIST05 mass spectral library, is shown in brackets.

Compound Number	Retention Time (min)	Major Ions	Pyrolysate name (% match)
1	8.117	100	2-methyl-2-propanoic acid (90%)
2	8.868	67	pyrrole (55%)
3	8.989	67, 79	pyrrole (46%), pyrimidine (45%)
4	9.021	79	pyrimidine (55%)
5	9.37	92	toluene (95%)
6	10.018	112	trans-1-butyl-2-methylcyclopropane (95%)
7	10.191	55	propanenitrile (59%)
8	10.239	170	4-ethyl-5-methylnonane (50%)
9	10.268	99, 129	2-methyl-3-(1-methylethyl)aziridine (25%), 2-methoxymethyl-2-methylpyrrolidine (27%)
10	10.291	114	3-ethylhexane (58%)
11	11.034	95	2-ethyl-1H-pyrrole (72%)
12	11.124	96	furfural (90%)
13	11.204	95	3-ethyl-1H-pyrrole (72%)
14	11.222	81, 93	3-methyl-1H-pyrrole (70%), 2-ethyl-1H-pyrrole (72%)
15	11.227	81	2-methyl-1H-pyrrole (87%)
16	11.25	81, 95	2-methyl-1H-pyrrole (87%), 1-ethyl-1H-pyrrole (72%)
17	11.253	81	3-methyl-1H-pyrrole (64%)
18	11.482	81	2-methyl-1H-pyrrole (90%)
19	11.838	92, 93	spiro[3,3]hepta-1,5-diene (46%), 3-methyl-pyridine (64%)
20	11.916	146	1,2-ethanediol (64%)
21	12.027	93	3-methylpyridine (60%)
22	12.028	106	ethylbenzene (55%)
23	12.033	92, 106	spiro[3.3]hepta-1,5-diene (42%), ethylbenzene (42%)
24	12.055	92	spiro[3.3]hepta-1,5-diene (42%)
25	12.061	93	2-methylpyridine (70%)
26	12.064	93, 106	3-methylpyrrole (53%), ethylbenzene (50%)
27	12.692	104	styrene (97%)
28	12.746	98, 144	1-heptene (73%), 1-nonanol (80%)
29	12.75	98, 112	isopropylcyclobutane (64%), 3-octene (60%)
30	12.76	98	1-heptene (92%)
31	12.767	98, 112	isopropylcyclobutane (70%), 1,2-diethylcyclobutane

			(64%)
32	12.768	98	isopropylcyclobutane (64%)
33	12.885	104	styrene (97%)
34	13.757	98	2-hydroxy-2-cyclopenten-1-one (87%)
35	14.734	110	5-methyl-2-furancarboxaldehyde (95%)
36	14.965	94	phenol (96%)
37	15.338	140	1-decene (89%)
38	15.796	120	1,2,4-trimethylbenzene (46%)
39	15.801	142	decane (64%)
40	15.895	114	2-methyliminoperhydro-1,3-oxazine (64%)
41	15.966	112	3,5-dimethyl-1-hexene (43%)
42	16.581	112	2-hydroxy-3-methyl-2-cyclopenten-1-one (55%)
43	16.685	128	2-methoxy-5-methylthiophene (25%)
44	16.716	128	2-octanone (32%)
45	16.978	128	2-methyl-5-(methylthio)furan (32%)
46	16.994	128	2-aminomethyl-5-methylamino-1,3,4-oxadiazole (38%)
47	17.051	108	2-methylphenol (95%)
48	17.27	108	3-methylphenol (94%)
49	17.62	108	4-methylphenol (97%)
50	17.833	108	3-methylphenol (96%)
51	17.858	112	3,5-dimethyl-1-hexene (43%)
52	17.858	108	2-methylphenol (97%)
53	17.871	108	methylphenol (97%)
54	18.072	124	2-methoxyphenol (92%)
55	18.289	154	1-undecene (55%)
56	18.292	126	1-methyl-2-pentyl-cyclopropane (64%)
57	18.292	126, 154	1-methyl-2-pentylcyclopropane (93%), 5-undecene (89%)
58	18.294	124	2-methoxyphenol (78%)
59	18.616	126	levoglucosenone (64%)
60	19.491	117	benzyl nitrile (97%)
61	19.715	117	4-methylbenzonitrile (72%)
62	19.761	122	2,4-dimethylphenol (96%)
63	20.054	122	2-ethylphenol (64%)
64	20.204	122	4-ethylphenol (94%)
65	20.264	122	3-ethylphenol (76%)
66	20.667	168	3-dodecane (94%)
67	20.765	138	2-methoxy-4-methylphenol (95%)
68	20.881	168	1-dodecene (95%)
69	20.883	168	2-dodecene (96%)
70	20.928	138	2-methoxy-4-methylphenol (97%)
71	20.985	138	2-methoxy-3-methylphenol (78%)

72	21.022	110	1,2-benzenediol (78%)
73	21.39	144	1,4:3,6-dianhydro-alpha-d-glucopyranose (90%)
74	21.547	120	2,3-dihydrobenzofuran (80%)
75	22.848	152	4-ethyl-2-methoxyphenol (87%)
76	22.897	206	1-bromo-2,2-dimethyl-3-carboxylic acid (38%)
77	22.997	152	4-ethyl-2-methoxyphenol (91%)
78	23.108	182	1-tridecene (95%)
79	23.432	117	indole (94%)
80	23.746	150	2-methoxy-4-vinylphenol (90%)
81	23.967	150	2,4,6-trimethyl-1,3-phenylenediamine (78%)
82	23.967	150	2-methoxy-4-vinylphenol (91%)
83	23.97	150	1-(3-methoxyphenyl)ethanone (78%)
84	24.502	134	4-(2-propenyl)phenol (95%)
85	24.688	164	eugenol (96%)
86	25.402	168	1-dodecene (96%)
87	25.527	196, 168	3-tetradecene (89%), nonyl-cyclopropane (89%)
88	25.604	131	3-methyl-1H-indole (97%)
89	25.615	196, 200	3-tetradecene (93%), 1-tridecanol (91%)
90	25.617	196	4-tetradecene (93%)
91	25.617	168	cyclododecane (94%)
92	25.621	196	cyclotetradecane (94%)
93	25.622	196	1-tetradecene (96%)
94	25.823	131	3-methyl-1H-indole (95%)
95	26	164	2-methoxy-4-(1-propenyl)phenol (98%)
96	26.982	164	2-methoxy-4-(1-propenyl)phenol (97%)
97	27.777	182, 210	1-tridecene (93%), 1-pentadecene (91%)
98	27.778	182	1-tridecene (97%)
99	27.778	168	cyclododecane (93%)
100	27.784	280	9-Eicosene (83%)
101	28.019	142, 210	2-methyl-2-butenoic acid (72%), 3-methyl-2-butenoic acid (50%)
102	29.818	196	2-tetradecene (96%)
103	29.82	196, 256	2-tetradecene (98%), 1-heptadecanol (91%)
104	29.822	196, 280, 238	2-tetradecene (93%), 3-Eicosene (93%), 1-heptadecene (93%)
105	29.826	224	cyclohexadecane (94%)
106	30.011	170	cyclopentanecarboxylic acid (59%)
107	31.751	238	1-heptadecene (98%)
108	31.752	238, 270	1-heptadecene (94%), 1-octadecanol (91%)
109	31.753	238, 266	1-heptadecene (94%), 1-nonadecene (94%)
110	31.758	266	1-nonadecene (94%)
111	31.878	240	heptadecene (95%)

112	31.879	226, 240	7-methylpentadecene (86%), heptadecene (95%)
113	31.88	226, 240	hexadecane (89%), heptadecane (89%)
114	32.731	228	tetradecanoic acid (99%)
115	33.588	252, 266, 238	E-15-heptadecenal (91%), 1-nonadecene (91%), E-14-hexadecenal (91%)
116	33.589	238, 266	1-heptadecene (95%), 1-nonadecene (94%)
117	33.592	252, 266	5-octadecene (98%), 1-nonadecene (94%)
118	33.592	266	1-nonadecene (94%)
119	34.137	152	6,6-dimethylbicyclo[3.1.1]heptane-2-carboxaldehyde (42%)
120	34.232	228	2-pentadecanol (55%)
121	35.33	210, 266, 386	cyclopentadecane (93%), 1-nonadecene (94%), 1-heptadecene (95%)
122	35.335	266	nonadecene (99%)
123	36.352	214, 228, 256	tridecanoic acid (91%), tetradecanoic acid (93%), n-hexadecanoic acid (87%)
124	36.415	228, 256	tetradecanoic acid (94%), n-hexadecanoic acid (95%)
125	36.442	256	n-hexadecanoic acid (95%)
126	36.443	214, 256	tridecanoic acid (94%), n-hexadecanoic acid (95%)
127	39.621	284, 372	octadecanoic acid (93%), 2-(2-hydroxyethoxy)ethyl ester octadecanoic acid (72%)

CHAPTER FOUR

CHANGES IN THE NATURE OF THE DISSOLVED ORGANICS DURING PULP AND PAPER MILL WASTEWATER TREATMENT: A MULTIVARIATE STATISTICAL STUDY COMBINING DATA FROM THREE ANALYTICAL TECHNIQUES

*

STATEMENT OF AUTHORSHIP

Title of Paper	Changes in the nature of dissolved organics during pulp and paper mill wastewater treatment: a multivariate statistical study combining data from three analytical methods
Publication Status	⊙ Published, ○ Accepted for Publication, ○ Submitted for Publication, ○ Publication style
Publication Details	Environmental Science and Pollution Research

AUTHOR CONTRIBUTIONS

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	Emma L Plant	
Contribution to the Paper	Experimental development, performed analysis on all samples, data analysis and critical interpretation, wrote manuscript.	
Signature		Date 8/10/13

Name of Co-Author	Ronald J Smernik	
Contribution to the Paper	Supervised development of work, NMR data analysis and interpretation, reviewed manuscript.	
Signature		Date 10/10/13

Name of Co-Author	Paul Greenwood	
Contribution to the Paper	Supervised development of work, pyrolysis data analysis and interpretation, reviewed manuscript.	
Signature		Date 8/10/13

Name of Co-Author	Lynne M Macdonald	
Contribution to the Paper	Supervised development of work, multivariate data analysis and interpretation, reviewed manuscript.	
Signature		Date 8/10/13

Name of Co-Author	John van Leeuwen	
Contribution to the Paper	Supervised development of work, data analysis and interpretation, reviewed manuscript.	
Signature		Date 9/10/13

CHAPTER FIVE

CHARACTERISATION OF ORGANIC MATTER IN SEWAGE EFFLUENT: COMBINING AND COMPARING NMR, PYROLYSIS, AND HPSEC ANALYTICAL APPROACHES

*

The work in this chapter has prepared as a paper for submission to *Environmental Monitoring and Assessment.*

STATEMENT OF AUTHORSHIP

Title of Paper	Characterisation of organic matter in sewage effluent: combining and comparing NMR, pyrolysis, and HPSEC analytical approaches.
Publication Status	O Published, O Accepted for Publication, O Submitted for Publication, ⊙ Publication style
Publication Details	prepared for submission to Environmental Monitoring and Assessment. This publication has not been submitted.

AUTHOR CONTRIBUTIONS

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	Emma L Plant	
Contribution to the Paper	Experimental development, performed analysis on all samples, data analysis and critical interpretation, wrote manuscript.	
Signature		Date 8/10/13

Name of Co-Author	Ronald J Smernik	
Contribution to the Paper	Supervised development of work, NMR data analysis and interpretation, reviewed manuscript.	
Signature		Date 10/10/13

Name of Co-Author	Paul Greenwood	
Contribution to the Paper	Supervised development of work, pyrolysis data analysis and interpretation, reviewed manuscript.	
Signature		Date 8/10/13

Name of Co-Author	Lynne M Macdonald	
Contribution to the Paper	Supervised development of work, multivariate data analysis and interpretation, reviewed manuscript.	
Signature		Date 8/10/13

ABSTRACT

Organic matter (OM) makes up a large component of sewage wastewater (WW) and it is important that the high organic loads are reduced before the water re-enters the environment. However, due to the variability and complexity of sewage OM, it is common for residual organics to remain in the final effluents. Despite the well established analytical techniques used to study OM, they have not been adequately applied to the study of residual OM in sewage effluent. Therefore, this study provides a detailed characterisation of two partly treated and five fully treated sewage effluent residual organics from separate treatment plants using solid-state ^{13}C nuclear magnetic resonance (NMR) spectroscopy, flash pyrolysis-gas chromatography mass spectrometry (py-GCMS), and high performance size exclusion chromatography (HPSEC). Each technique showed little difference in the nature of the partly treated and residual organics. They were each dominated by O-alkyl carbon as determined by NMR; dominated by aromatics by py-GCMS; and were dominated by 2000 Da organics by HPSEC. In particular, pyrolysis-GCMS was able to provide an informative molecular analysis identifying anthropogenic organics in each of the samples. These anthropogenic organics are of interest for environmental reasons and would have left the sewage treatment system unnoticed if only conventional detection techniques were used.

INTRODUCTION

Domestic wastewater (WW) has characteristically high organic matter (OM) levels and therefore an important part of domestic WW treatment is the removal of OM through physical, chemical and biological treatment. Domestic WW contains a complex mixture of high molecular weight (MW) OM derived from organic human waste, partially degraded food organics and OM originally present in drinking water. It also contains an array of small organic compounds, including compounds referred to as "emerging" contaminants, which includes personal care products, laundry detergent components, and pharmaceuticals (Langdon et al. 2011; Zuloaga et al. 2012), synthetic organic molecules including surfactants and pesticides (Shon et al. 2006b), and illicit drugs (Ratola et al. 2012; Feitosa et al. 2013). The presence of synthetic organic molecules in WW, in particular, has complicated removal of organics as many of these compounds are resistant to, or are very slow to, biologically degrade (Shon et al. 2006b). Despite the extensive amount of research that has gone into optimising the domestic WW treatment process

(Shon et al. 2006b; Laughlin and Abella 2007; Pearson et al. 2007; Gardner et al. 2012), conventional WW treatment does not remove all OM (Shon et al. 2006b). This is in part due to its highly complex and heterogeneous nature (Simpson and Simpson 2012; Nebbioso and Piccolo 2013). The residual OM that remains in the final effluent is released with the water and has the ability to influence the health of the receiving environment and persist for long periods of time (Hall et al. 2009; Hall and Landis 2009; Chiang et al. 2010). However, treatment that completely removes all carbon (C) and nutrients may have an unhealthy influence on the receiving waters as these nutrients may be essential for the health of the aquatic environment (Shon et al. 2006b).

The residual OM in sewage effluents is typically measured in terms of total or dissolved organic carbon (TOC or DOC), and chemical or biological oxygen demand (COD or BOD) (Marquet et al. 1999; Shon et al. 2006b; Laughlin and Abella 2007; Henderson et al. 2011; Gardner et al. 2012). These techniques provide a relatively quick measure of the organics present in a sample, and their application to studying organics in water is well established. Generally these measurements are made to ensure that the sewage WW treatment processes are efficient and that the resulting effluent meets quality standards. More sophisticated analytical techniques, including spectroscopic and chromatographic techniques, are also used to study OM in water (Shon et al. 2006a; Matilainen et al. 2011). Spectroscopic and chromatographic techniques provide more detailed chemical or structural-type information compared to TOC, DOC, COD or BOD.

Three sophisticated analytical techniques that are at the forefront of OM characterisation are (i) solid-state ^{13}C nuclear magnetic resonance (NMR) spectroscopy; (ii) pyrolysis coupled to gas chromatography mass spectrometry (py-GCMS); and (iii) high performance size exclusion chromatography (HPSEC). Solid-state ^{13}C NMR spectroscopy provides information on OM composition at the broad functional group level (Matilainen et al. 2011; Simpson et al. 2011). Pyrolysis-GCMS also provides structural information, but at a more detailed molecular level, for example to detect molecular biomarkers and pharmaceuticals in WW discharge (Greenwood et al. 2006; Greenwood et al. 2012). Separation of dissolved organic matter (DOM) and subsequent determination of MW can be achieved with HPSEC analysis, which has become a popular means of studying DOM in water research (Chow et al. 2008; Lewis et al. 2012). Despite the application of these spectroscopic and chromatographic analytical techniques to the study of OM in general,

they have not often been adequately applied to the study of effluent OM in particular (Shon et al. 2006a; Xue et al. 2011).

One of the few studies that has incorporated these analytical techniques to characterise sewage effluent OM is a study by Navalon et al. (2011). In this study, NMR techniques were used to determine the structural nature of the recalcitrant organics that are not removed by biological treatment. In a different study by Greenwood et al. (2012), pyrolysis techniques were able to detect the degradation product, iminostilbene, of the antidepressant and antiepileptic clinical drug Carbamazepine. Common industrial chemicals used in the manufacture of cosmetics, detergents and industrial solvent stabilisers were also detected in these effluents. These studies show how diverse effluent OM can be and emphasises the importance of applying such sophisticated tools to the study of effluent OM.

Nuclear magnetic resonance and py-GCMS have, however, been more widely applied to the study of organic sludges produced during WW treatment processes (Marquet et al. 1999; Pelekani et al. 1999; Smernik et al. 2003a, b; Smith et al. 2008). Treated sewage sludge can be recycled and applied to land as an environmentally sustainable way of disposing of these organics (Smith et al. 2008; Caricasole et al. 2011; Liu et al. 2012), and due to this recycling of organics, sludges have been studied using solid-state ^{13}C NMR spectroscopy (Smernik et al. 2003b, a; Smith et al. 2008; Caricasole et al. 2011) and less so by py-GCMS (Pognani et al. 2011). The structure of three different sludges from one sewage treatment stream was shown to vary by solid-state ^{13}C NMR spectroscopy (Smith et al. 2008). The first sludge sample collected from an earlier stage of OM removal was the more distinct and the other two sludges from the later stages of OM removal were more similar, both with their C distribution dominated by O-alkyl C compared to the dominance of alkyl C in the first sludge sample. Pyrolysis-GCMS analysis of sewage sludge organics by Pognani et al. (2011) was able to identify more structural features within the aromatic and alkyl classes compared to NMR analysis, including proteins and polysaccharides (Pognani et al. 2011).

This paper details the characterisation of OM contained in the final effluent of five sewage WW treatment plants in the United Kingdom along with two partially treated effluents, using the three analytical techniques mentioned above (solid-state ^{13}C NMR spectroscopy, py-GCMS and HPSEC).

METHODS AND MATERIALS

SAMPLE PREPARATION

The final effluent from five sewage treatment plants and partially treated WW (one post ozone treatment, the other post granular activated carbon (GAC) treatment) from two of these plants were provided as freeze-dried material for this analysis. The samples included in this study are subject to commercial in confidence, therefore details of the individual sample sites and sewage WW treatments have been left out intentionally. High molecular weight (>1000 Da) OM was concentrated by ultrafiltration (UF). Approximately 2 g of freeze-dried sample was dissolved in 500 mL of MilliQ^{TM} water and filtered first to 0.45 μm then to 0.2 μm. These filtered samples were then placed in a 600 mL UF cell with a 1000 Da molecular weight cut-off. The UF cell was pressurised to approximately 0.4 MPa using nitrogen. Each sample was filtered to 50% cell capacity (approximately 300 mL) then diluted to 600 mL with MilliQ^{TM} water. This procedure was carried out seven times, and was followed by measuring electrical conductivity (EC) of the filtrate. The EC of the final filtrate was < 40 μScm^{-1}. At this point the retentate was concentrated to 100 mL and separated into two parts: 90 mL for freeze-drying and 10 mL for wet chemical analyses.

Table 1. Sewage wastewater samples that were analysed.

Sample #	Sample Site	Sewage Process/Stage Type
1	A	Post ozone treatment
2	B	Post granular activated carbon treatment
3	A	Final effluent
4	B	Final effluent
5	C	Final effluent
6	D	Final effluent
7	E	Final effluent

BASIC CHEMISTRY

Electrical conductivity was determined using an ATI Orion model 170 (TÜV Product Services, Germany) EC meter. Three replicate measurements were collected for each water sample before desalting and of the retentate after desalting. Dissolved organic carbon measurements were made on 0.45 μm filtered samples before desalting and on a diluted

aliquot of retentate (1 mL of retentate diluted to 10 mL with MillQ[TM] water) using a Formacs[HT] series combustion TOC/TN (total nitrogen) analyser (Skalar Analytical Breda, Netherlands).

SOLID-STATE ^{13}C CP NMR SPECTROSCOPY

Solid-state ^{13}C cross polarization (CP) NMR spectra were acquired with magic angle spinning (MAS) at a ^{13}C frequency of 100.6 MHz on a Varian Unity INOVA 400 spectrometer. Samples were packed in a 7 mm diameter cylindrical zirconia rotor with Kel-F end-caps and ceramic spacers, and spun at 6500±100 Hz in a Doty Scientific supersonic MAS probe. Spectra were acquired using a ramped-amplitude cross polarization (CP-ramp) pulse sequence, in which the ^{1}H spin lock power was varied linearly during the contact time. A 1 ms contact time was used and 10000 transients were collected with a recycle delay of 2 s. Free induction decays (FIDs) were acquired with a sweep width of 50 kHz; 8000 data points were obtained over an acquisition time of 160 ms. An FID of an empty rotor background spectrum was subtracted and all spectra were zero-filled to 8192 data points and processed with a 100 Hz Lorentzian line broadening and a 0.01 s Gaussian broadening. Chemical shifts were externally referenced to the methyl resonance of hexamethylbenzene at 17.36 ppm.

For each sample, the NMR spectra were integrated to determine the relative proportion of carbon signal in the following regions: carbonyl (190-160 ppm), aromatic (160-110 ppm), O-alkyl (110-45 ppm) and alkyl (45-0 ppm).

For multivariate analysis, data was normalised and baseline corrected to ensure all data points were non-negative and further smoothed by taking the average of each five consecutive points throughout the spectra. The NMR trace in the spectral region of 0-250 ppm was used for multivariate analyses.

FLASH PY-GCMS

Flash pyrolysis of small sample quantities (0.5 – 0.8 mg) of freeze-dried material was conducted at 650 $^{\circ}$C for 10 s using a Chemical Data Systems 160 Pyroprobe inserted into a dedicated pyrolysis chamber maintained at 250 $^{\circ}$C. The pyrolysis chamber was mounted directly onto the vaporisation injector of a Hewlett Packard (HP) 5890 Series II GC,

coupled to a 5971 mass selective detector (MSD) used for pyrolysate detection. A 60 m, 0.25 mm i.d., 0.25 μm phase ZB-5MS capillary column (Phenomenex) was used with helium carrier gas (9 psi). Injection splits of between 20 and 50 mL min^{-1} were used. The GC oven temperature program was 40 ^{o}C (2 min hold) heated at 4 ^{o}C min^{-1} to 310 ^{o}C (15 min) and the transfer line was kept at 310 ^{o}C. Full scan m/z 50–550 mass spectra, with an electron energy of 70 eV were acquired.

Products were identified by mass spectral correlation to the NIST05 mass spectral library. Assignments were only made where there was high quality peak matching (typically ›70%) between measured and library mass spectra; nevertheless, these may still be regarded as tentative in the absence of correlation to authentic standards. Part of the variance evident in the mass spectral correlation can be attributed to variations in the m/z regions of the acquired (m/z 50–550) and library mass spectra.

The identified products were categorised into broad chemical classes, such as aromatic, phenolic, aliphatic, N compound etc, as an alternate way of representing the changes in the chemical nature of the organics.

The integrated peak areas (from the total ion chromatogram (TIC)) of all identified products (Supplementary Material) were used for multivariate analysis. Products not detected in any particular sample were given a zero value in the statistical treatments. For each sample, peak areas were log transformed and normalised.

HPSEC

Analysis of WW by HPSEC was carried out on diluted aliquots of retentate (1 mL retentate diluted to 10 mL with MillQ^{TM} water). A Waters Acquity H-Class system was used with a photodiode array detector (Waters Corporation, USA) operating between 200 and 460 nm, and extracting at 260 nm. A 0.1 M NaCl solution containing phosphate buffer (0.02 M) was passed at 1 mL min^{-1} through a Shodex KW802.5 column packed with silica (Showa Denko, Japan). An effective separation range of approximately 50 to 50,000 Da was provided by this column. To determine the apparent molecular weight (AMW) of the eluted organics, calibration was achieved with polystyrene sulphonate molecular weight standards of 35, 18, 8 and 4.6 kDa (Polysciences Inc., USA).

The HPSEC chromatograms for each of the organic samples were analysed by peak fitting as described previously (Lewis et al. 2011) using Microsoft Excel. The data for each chromatogram was fitted as the sum of five Gaussian functions, Equations 1 and 2, each of which has three defined peak parameters: position (p), height (h) and width (w).

Equation 1: $$I_x = he-\frac{(x-p)^2}{2w^2}$$

Equation 2: $$I_{fit} = \sum_{i=1}^{5} I_x$$

Initial values of p, h and w were estimated for each peak from the position, height and width of the visible peaks in the chromatogram. In Microsoft Excel, the Solver Routine was used to minimise the sum of the squares of residuals between the fitted intensities (I_{fit}) and the actual intensities (I) by varying values of p, h and w for sample 7. Values of p and w were held constant for the remaining samples, i.e. only h was allowed to vary between samples.

For multivariate analysis, the chromatogram trace itself was used, with AMWs log transformed, and the absorbance results (at 260 nm) baseline corrected and normalised.

RESULTS AND DISCUSSION

Initial attempts to characterise the whole freeze-dried WW samples by NMR were unsuccessful because only a small fraction of the samples were OM, with the majority being inorganic salts. Furthermore, these whole freeze-dried WW samples contained substantial quantities of carbonate, resulting in carbonate peaks dominating the ^{13}C NMR spectra. Therefore it was necessary to desalt the samples. Ultrafiltration was used to remove the carbonate and the majority of salt, and to concentrate the organic material. A 1000 Da molecular weight cut off membrane was used. This procedure was very successful in removing salt, with only 0.7-1.4% of the mass retained in the retentate (Table 2). It was also reasonably effective in retaining OM, with around half (39-57%) of the DOC recovered in the retentate (Table 2). However, this does mean that around one half of the DOC was not recovered, the majority of which is likely to have been low MW material (<1000 Da) that passed through the membrane. Overall, the concentration of OM in the desalted samples was around 25 times greater than in the whole freeze-dried WW samples.

Table 2 Water quality parameters before and after the process of desalting, and the resulting DOC and mass recoveries for each of the WW samples. Measurements are recorded for the concentrations or volumes specified.

Sample	Site	Original EC (μS.cm^{-1}) (2 g in 500 mL)	Desalted EC (μS.cm^{-1}) (100 mL retentate)	DOC (mg/L) (2 g in 500 mL)	DOC (mg/L) (100 mL retentate)	DOC recovery (%)	Mass recovery (%)
1	A	5350	166	28.7	56.3	39	1.2
2	B	5030	138	11.3	23.3	41	0.7
3	A	2580	166	20.5	58.2	57	1.0
4	B	5180	153	20.5	39.8	39	0.8
5	C	4840	150	31.0	41.9	27	0.9
6	D	4800	156	32.1	70.8	44	1.4
7	E	4980	161	25.7	70.9	55	1.4

The NMR spectra for the desalted WW organics are shown in Figure 1. Each of the samples is dominated by O-alkyl peaks at ~75 and ~105 ppm, and contains minor carbonyl (~180 ppm) and alkyl C (10-40 ppm) signals (Figure 1). Each WW sample contains at least one alkyl peak at ~25 ppm; however, samples 1, 2 and 7 have more distinct alkyl peaks, including a second peak at ~20 ppm. The final effluents of sites A and B appear to have more alkyl C signal (3 and 4) compared to their partially treated organics (1 and 2). Carbonate (~168 ppm) was not detected in the ^{13}C NMR spectra of any of the desalted WW samples.

By integrating the NMR spectra across broad chemical shift regions, the distribution of basic C types can be estimated. The integration results for the alkyl (0-45 ppm), O-alkyl (45-110 ppm), aromatic (110-160 ppm) and carbonyl (160-190 ppm) C regions for each of the samples are shown in Table 3. Here, the dominance of the O-alkyl C region is confirmed with at least half of the signal for a sample appearing in this region. The integration results confirm the greater proportion of alkyl signal in the fully treated samples of 3 and 4, compared to the partially treated effluents (1 and 2). Samples 2, 3 and 7 have the greatest proportion of carbonyl signal.

As discussed above, there are no previously reported solid-state ^{13}C NMR spectra of sewage effluent OM. However, comparisons can be made to sewage sludge organics and the dominance of alkyl and O-alkyl Cs and the limited amount of aromatic C is consistent with results previously reported for solid-state ^{13}C NMR spectra of these (Caricasole et al. 2011; Jindo et al. 2012).
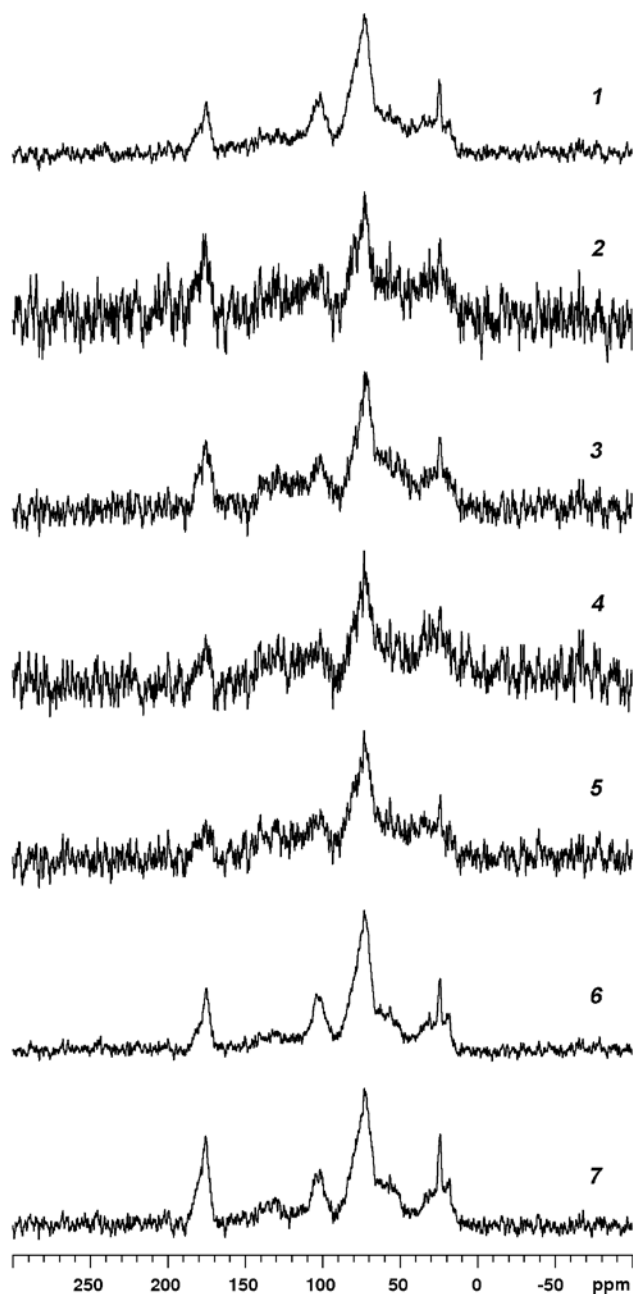

Figure 1 Solid-state ^{13}C CP NMR spectra of the desalted sewage WW samples.

Table 3 Results from integration of solid-state ^{13}C CP NMR spectra of the desalted sewage WW samples across four broad chemical shift regions.

Samples	Carbonyl (190-160 ppm)	Aromatic (160-110 ppm)	O-alkyl (110-45 ppm)	Alkyl (45-0 ppm)
			%	
1	6.8	10.8	64.3	18.0
2	10.9	15.8	51.6	21.7
3	10.4	16.2	57.6	15.9
4	5.5	19.9	52.8	21.8
5	4.8	16.9	59.7	18.6
6	9.2	9.4	64.3	17.2
7	12.6	12.9	59.5	15.0

The py-GCMS TICs for the desalted WW samples are shown in Figure 2. For each sample, a broad range of products were detected, particularly at retention times below 28 minutes (Figure 2). Many of the products detected are common to most or all of the samples, but the intensity of the peaks vary. Some differentiation among the samples is clearly evident for products with retention times from 28 minutes onwards. For example, the peak at ~30 minutes is most pronounced in samples 3, 4 and 5; samples 2, 3, and 7 show a large peak at ~34.5 minutes; and samples 3 and 5 have two distinct peaks at ~35.5 and ~36 minutes (Figure 2). Generally, assignment of these products is achieved by correlation of mass spectral patterns to library data bases (Supplementary Material). An alternative way to identify differences in the chemical nature of these products is to collate the common functional groups identified in each of the products. From the relative distribution of these groups in each of the samples (Figure 3), the nature of the organics can be compared. The common functional groups identified were aromatic, nitrogen, ketone, phenol, alkene, acid and other (those organics that did not fall into the other groups listed). Aromatic products, excluding those with a phenol group, dominate the pyrolysis products for all samples (Figure 3). This contrasts with the dominance of O-alkyl C indicated by NMR (Figure 1). However, this is not unexpected as pyrolysis identifies fragments of the organic material, not the entire sample as NMR does. Aromatic fragments of DOM may be more easily cleaved from the original organic material for analysis, whereas the entire sample (including these fragments as seen by pyrolysis) is seen by NMR.

The pyrolysis data showed several modest differences between partly and fully treated samples from the same site. Compared to the post-ozone sample (sample 1), the final effluent (sample 3) at site A had a decrease in the relative proportion of the aromatic group and a corresponding increase in the phenolic group (Figure 3). Both contribute to the ^{13}C NMR aromatic signal which consequently showed little change. There was also a small decrease in alkene pyrolysates, which was consistent with the decrease in alkyl carbons seen by NMR (Table 2), and a small increase in the proportions of acidic and nitrogen-containing (N) pyrolysates with increasing treatment (Figure 3). The pyrolysis results showed more significant differences between the treatments of site B. Again the proportion of phenolic pyrolysates increased with treatment (Figure 3); which did mirror the increasing aromatic signal measured by NMR (Table 2). There was an absence of "other" (i.e., fluorine and isocyanate, see Supplementary Material) and decrease of N-containing pyrolysates (Figure 3). Sources of Fluorine in WW include lubricants, adhesives and cosmetics. Fluorinated compounds are preferentially adsorbed to sewage sludges during WW treatment (Schroder 2003) which may explain the absence of fluorinated compounds in the final effluent sample (3). Isocyanate may be a secondary product of a rearrangement or some other reaction during pyrolysis. Secondary reactions including rearrangements can be favoured at elevated temperatures and can dissipate the structural detail of the parent sample (Leinweber et al. 2013).

The pyrolysis group distribution of the final effluent organics of samples 3 – 7 were all generally similar, particularly samples 3-5 (Figure 3). NMR analysis similarly showed these samples to be molecularly quite alike (Figure 1).

Metabolites of pharmaceuticals were detected in all samples (Supplementary Material). These products included indolizine (23.2 minutes) and indole-type fragments at 23.3, 23.5, 25.4, and 27.6 minutes. It should be noted that the compounds themselves have molecular weights lower than the molecular weight cut-off used to isolate the samples. Therefore the presence of these fragments indicates that they are attached to, or associated with, larger organics. Pharmaceuticals and personal care products contribute to a diverse group of unregulated pollutants, often referred to as emerging contaminants, that can be introduced to receiving environments through WW treatment plants (Ratola et al. 2012; Zuloaga et al. 2012). Many emerging contaminants are known to associate strongly with DOM (Ratola et al. 2012) making them difficult to remove by traditional treatment

practices (Ratola et al. 2012). Their general recalcitrance and sequestration by DOM was further demonstrated by their survival of the desalting pre-analysis step used here.
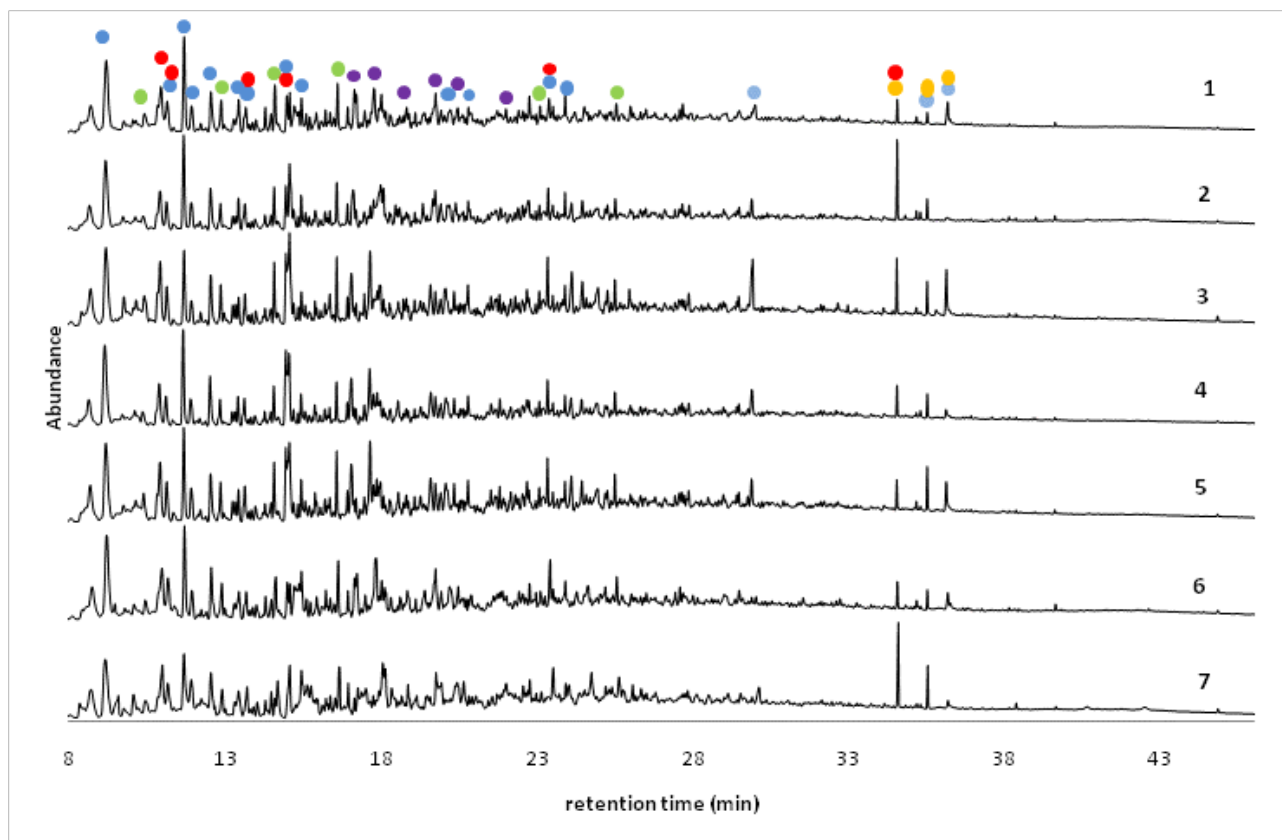

Figure 2 Total ion chromatograms of each of the desalted sewage WW samples. The coloured peak assignments represent the common functional groups: blue – aromatic; red – nitrogen; green – ketone; purple – phenol; light blue – alkene; and orange – acid.
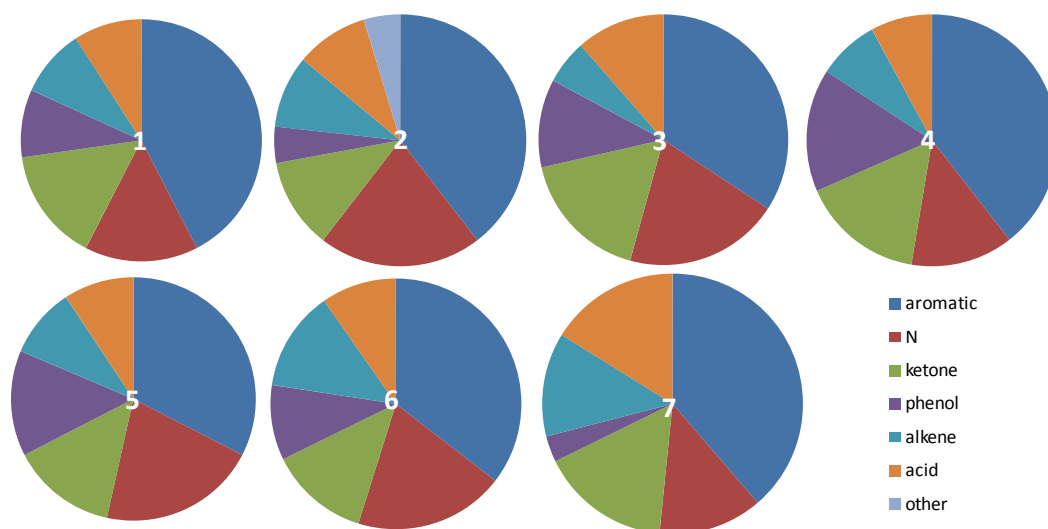
Figure 3 Functional group distributions for the desalted sewage WW samples as determined by flash pyrolysis-GCMS. Colours correspond to the labelling of fragments in Figure 2.

The HPSEC chromatograms show that there is little variation in the AMW of OM in the sewage WW samples (Figure 4). They all contain a very small peak at ~80000 Da, a dominant peak at ~2000 Da and a minor peak at ~400 Da (Figure 4). It should be kept in mind that these samples were desalted by UF using a 1000 Da membrane and this resulted in the loss of ~50% of DOC; presumably the DOM components lost were smaller than this 1000 Da cut-off. The presence of the small peak at ~400 Da is thus somewhat surprising and may indicate some degradation of DOM during HPSEC analysis or it may again be due to the presence of "host-guest" associations between large and small molecules. In any case, all HPSEC chromatograms are very similar to that of untreated reservoir surface water as presented in Chow et al. (2008). The HPSEC profile of this surface water DOM contains all three peaks seen here for the sewage WW samples and in similar relative proportions. It is interesting that the HPSEC profiles of partially and fully treated sewage OM, in this case, are so similar to that of drinking water source DOM. As discussed above, it has been suggested that drinking water OM makes up a large amount of sewage effluent OM (Shon et al. 2006b), and the results reported here are consistent with this.

Quantification of the HPSEC chromatograms was carried out by deconvolution of each HPSEC trace using a previously described protocol (Lewis et al. 2011). Each chromatogram was fitted as the sum of five peaks with fixed AMW and peak width. The AMW of each the fitted peaks (in $\log_{10}$Da) was 5.50 (peak 1), 6.00 (peak 2), 6.50 (peak 3),

7.55 (peak 4) and 11.00 (peak 5). Some small differences in AMW distributions are apparent from these fits (Figure 5). There is little difference in the relative proportion of organics represented by peaks 1 and 2, and most difference was observed for the higher AMW organics represented by peaks 3, 4 and 5 (Figure 5). Peak 4 represents the peak at ~2000 Da that dominated each of the spectra (Figure 4 and 5). The overall lack of variation in HPSEC chromatograms is consistent with the NMR and py-GCMS results.
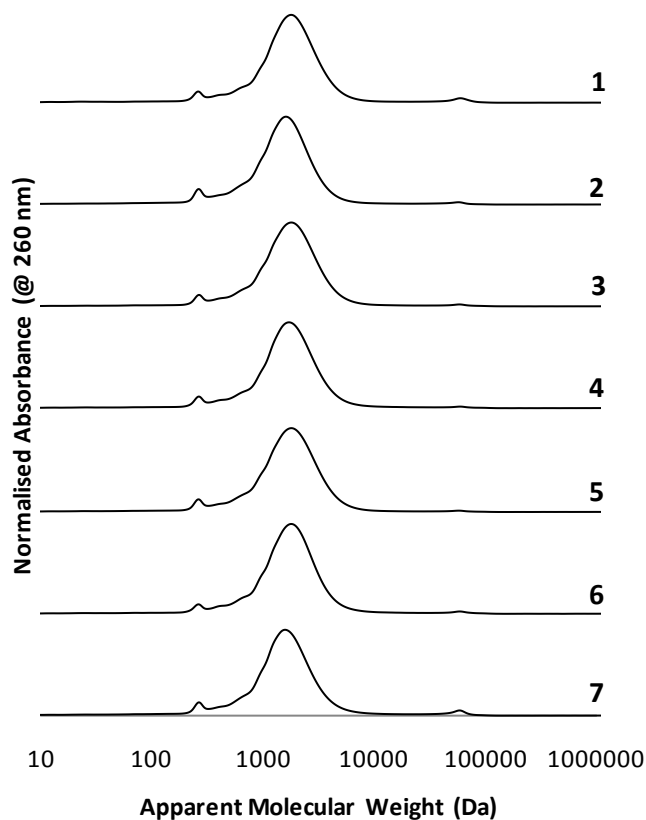

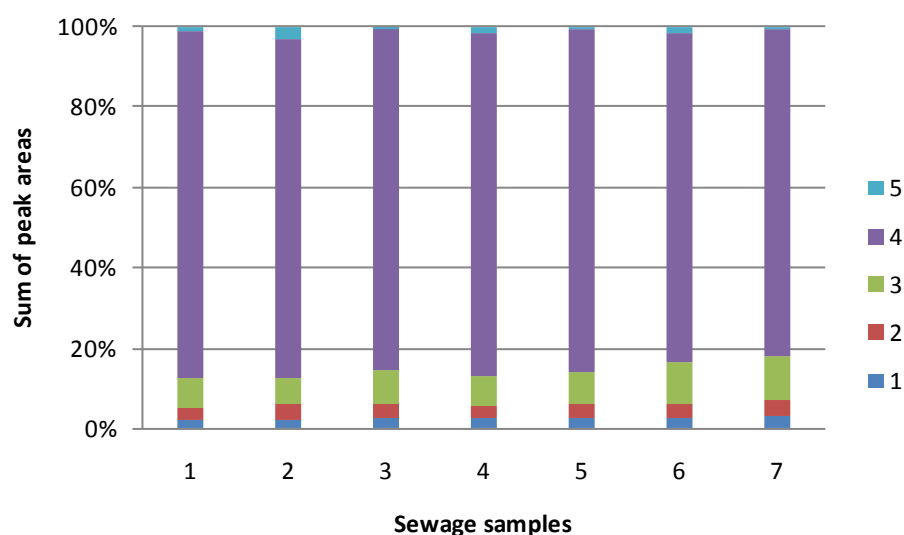Figure 4 The HPSEC chromatograms of the desalted sewage WW samples.

Figure 5 Relative proportions of the five Gaussian peaks fitted to each of the desalted sewage WW HPSEC profiles (1-7). The AMW of the fitted Gaussian peaks are 5.50 (1), 6.00 (2), 6.50 (3), 7.55 (4), and 11.00 (5) $\log_{10}$Da.

Characterisation of the partially and fully treated sewage effluent organics provided an more detailed information than conventional detection techniques commonly applied to study effluent OM. The degree of similarity between the effluent samples was further assessed using multivariate statistical approaches. This analysis, however, was beyond the scope of this study but is included in Supplementary Material II.

Despite OM being a variable and highly complex mixture, the five different sewage treatment plants produced residual organics that were very similar. An understanding of the residual organics that re-enter the environment after sewage treatment is important, as they have the potential to influence the health of the ecosystem in a positive or negative way. Pyrolysis-GCMS was able to provide a more detailed molecular-level analysis on the complexity of the residual organics, identifying anthropogenic residual organics in all samples. It is only with the use of sophisticated analytical techniques, like those used in this study, that structural and chemical information on the nature of residual organics can be determined. Without these more sophisticated techniques, there is the potential for residual organics of concern, including anthropogenic organics, to pass through the treatment unnoticed. This study was carried out to demonstrate the analytical benefit of applying sophisticated analytical techniques to the study of residual effluent organics.

CONCLUSION

Solid-state ^{13}C NMR spectroscopy, py-GCMS, and HPSEC provided valuable and complementary molecular information about the organic concentrates isolated from partially and fully treated sewage effluent of several WWTPs. The respective analytical data showed the organic composition of all the WW samples were generally similar. Pyrolysis-GCMS provided the most detailed information, including the consistent presence of emerging pollutants in all WW effluents. It should be noted that only 27-57% of DOC was obtained by UF, therefore differences in the nature of the OM may be present in the lower MW fraction. However, the fact that there was a difference in the quantity of OM isolated indicates that there was variation in the nature of the OM in each of the samples.

ACKNOWLEDGEMENTS

REFERENCES

Caricasole P, Provenzano MR, Hatcher PG, Senesi N (2011) Evolution of organic matter during composting of different organic wastes assessed by CPMAS C-13 NMR spectroscopy. Waste Manage 31:411-415

Chiang G, Munkittrick KR, Orrego R, Barra R (2010) Monitoring of the Environmental Effects of Pulp Mill Discharges in Chilean Rivers: Lessons Learned and Challenges. Water Quality Research Journal of Canada 45:111-122

Chow CWK, Fabris R, van Leeuwen J, Wang D, Drikas M (2008) Assessing Natural Organic Matter Treatability Using High Performance Size Exclusion Chromatography. Environ Sci Technol 42:6683-6689

Feitosa RS, Sodre FF, Maldaner AO (2013) Drugs of abuse in waters and wastewaters: occurrence, analytical determination and forensic applications. Quim Nova 36:291-305

Gardner M, Comber S, Scrimshaw MD, Cartmell E, Lester J, Ellor B (2012) The significance of hazardous chemicals in wastewater treatment works effluents. Science of the Total Environment 437:363-372

Greenwood PF, Leenheer JA, McIntyre C, Berwick L, Franzmann PD (2006) Bacterial biomarkers thermally released from dissolved organic matter. Org Geochem 37:597-609

Greenwood PF, Berwick LJ, Croué JP (2012) Molecular characterisation of the dissolved organic matter of wastewater effluents by MSSV pyrolysis GC–MS and search for source markers. Chemosphere 87:504-512

Hall TJ, Fisher RP, Rodgers JH, Jr., Minshall GW, Landis WG, Kovacs TG, Firth BK, Dube MG, Deardorff TL, Borton DL (2009) A Long-Term, Multitrophic Level Study to Assess Pulp and Paper Mill Effluent Effects on Aquatic Communities in Four US Receiving Waters: Background and Status. Integr Environ Assess Manag 5:189-198

Hall TJ, Landis WG (2009) Introduction to Special Series: Measurement and Analysis of the Potential Long Term Impact of Pulp and Paper Mill Effluent on Receiving Waters. Integr Environ Assess Manag 5:186-188

Henderson RK, Subhi N, Antony A, Khan SJ, Murphy KR, Leslie GL, Chen V, Stuetz RM, Le-Clech P (2011) Evaluation of effluent organic matter fouling in ultrafiltration treatment using advanced organic characterisation techniques. J Membrane Sci 382:50-59

Jindo K, Martim SA, Navarro EC, Perez-Alfocea F, Hernandez T, Garcia C, Aguiar NO, Canellas LP (2012) Root growth promotion by humic acids from composted and non-composted urban organic wastes. Plant Soil 353:209-220

Langdon KA, Warne MS, Smernik RJ, Shareef A, Kookana RS (2011) Selected personal care products and endocrine disruptors in biosolids: An Australia-wide survey. Sci Total Environ 409:1075-1081

Laughlin DC, Abella SR (2007) Abiotic and biotic factors explain independent gradients of plant community composition in ponderosa pine forests. Ecol Model 205:231-240

Leinweber P, Kruse J, Baum C, Arcand M, Knight JD, Farrell R, Eckhardt KU, Kiersch K, Jandl G (2013) Advances in Understanding Organic Nitrogen Chemistry in Soils Using State-of-the-art Analytical Techniques. In: Sparks DL (ed) Advances in Agronomy, Vol 119, Advances in Agronomy, Elsevier Academic Press Inc, San Diego, pp 83-151

Lewis R, van Leeuwen JA, Smernik RJ, Chow CWK, Everson A, Nothrop SC, Beecham S (2011) Changes in the organic character of post-coagulated Pinus radiata sulfite pulp mill wastewater under aerated stabilization basin treatment A laboratory scale study. Chem Eng J 175:160-168

Lewis R, van Leeuwen JA, Chow CWK, Everson A, Lewis DM (2012) Assessment of coagulated and non-coagulated ASB performance used to treat Pinus radiata sulfite pulp and paper mill effluent by resin fractionation and HPSEC techniques. Chem Eng J 213:109-117

Liu PWG, Wang SY, Huang SG, Wang MZ (2012) Effects of soil organic matter and ageing on remediation of diesel-contaminated soil. Environ Technol 33:2661-2672

Marquet R, Mietton-Peuchot M, Wheatley AD (1999) Characterisation of trickling filter effluent by particle size distribution and high performance size exclusion chromatography. Water Res 33:1415-1424

Matilainen A, Gjessing ET, Lahtinen T, Hed L, Bhatnagar A, Sillanpää M (2011) An overview of the methods used in the characterisation of natural organic matter (NOM) in relation to drinking water treatment. Chemosphere 83:1431-1442

Navalon S, Alvaro M, Garcia H (2011) Analysis of organic compounds in an urban wastewater treatment plant effluent. Environmental Technology 32:295-306

Nebbioso A, Piccolo A (2013) Molecular characterization of dissolved organic matter (DOM): a critical review. Anal Bioanal Chem 405:109-124

Pearson EJ, Farrimond P, Juggins S (2007) Lipid geochemistry of lake sediments from semi-arid Spain: Relationships with source inputs and environmental factors. Org Geochem 38:1169-1195

Pelekani C, Newcombe G, Snoeyink VL, Hepplewhite C, Assemi S, Beckett R (1999) Characterization of natural organic matter using high performance size exclusion chromatography. Environ Sci Technol 33:2807-2813

Pognani M, Barrena R, Font X, Adani F, Scaglia B, Sanchez A (2011) Evolution of organic matter in a full-scale composting plant for the treatment of sewage sludge and biowaste by respiration techniques and pyrolysis-GC/MS. Bioresource Technol 102:4536-4543

Ratola N, Cincinelli A, Alves A, Katsoyiannis A (2012) Occurrence of organic microcontaminants in the wastewater treatment process. A mini review. J Hazard Mater 239:1-18

Schroder HF (2003) Determination of fluorinated surfactants and their metabolites in sewage sludge samples by liquid chromatography with mass spectrometry and tandem mass spectrometry after pressurised liquid extraction and separation on fluorine-modified reversed-phase sorbents. J Chromatogr A 1020:131-151

Shon H-K, Kim S-H, Erdei L, Vigneswaran S (2006a) Analytical methods of size distribution for organic matter in water and wastewater. Korean J Chem Eng 23:581-591

Shon HK, Vigneswaran S, Snyder SA (2006b) Effluent organic matter (EfOM) in wastewater: Constituents, effects, and treatment. Crit Rev Env Sci Tec 36:327-374

Simpson AJ, McNally DJ, Simpson MJ (2011) NMR spectroscopy in environmental research: From molecular interactions to global processes. Prog Nucl Mag Res Sp 58:97-175

Simpson M, Simpson A (2012) The Chemical Ecology of Soil Organic Matter Molecular Constituents. J Chem Ecol 38:768-784

Smernik RJ, Oliver IW, Merrington G (2003a) Characterization of sewage sludge organic matter using solid-state carbon-13 nuclear magnetic resonance spectroscopy. J Environ Qual 32:1516-1522

Smernik RJ, Oliver IW, Merrington G (2003b) Advanced solid-state carbon-13 nuclear magnetic resonance spectroscopic studies of sewage sludge organic matter: Detection of organic "domains". J Environ Qual 32:1523-1533

Smith MTE, Smernik RJ, Merrington G, Tibbett M (2008) Changes in sewage sludge carbon forms along a treatment stream. Chemosphere 72:981-985

Xue S, Zhao QL, Ma XP, Li FY, Wang J, Wei LL (2011) Comparison of dissolved organic matter fractions in a secondary effluent and a natural water. Environ Monit Assess 180:371-383

Zuloaga O, Navarro P, Bizkarguenaga E, Iparraguirre A, Vallejo A, Olivares M, Prieto A (2012) Overview of extraction, clean-up and detection techniques for the determination of organic pollutants in sewage sludge: A review. Anal Chim Acta 736:7-29

SUPPLEMENTARY MATERIAL 1

Table 1 Tentatively identified compounds from pyrolysis of the various UK sewage WW samples. * indicates pharmaceutical fragments identified.

compound number	ret time (min)	M+	pyrolysate
1	9.106	92	toluene (91%)
2	10.764	82	2-cyclopenten-1-one (80%)
3	10.82	81	3-methyl-1H-pyrrole (86%)
4	10.827	81	2-methyl-1H-pyrrole (86%)
5	11.05	108	2-methylbicyclo[3.2.0]hept-2-ene (86%)
6	11.068	81	2-methyl-1H-pyrrole (86%)
7	11.128	81	3-methyl-1H-pyrrole (90%)
8	11.62	106	ethylbenzene (93%)
9	11.863	106	p-xylene (97%)
10	11.874	106	o-xylene (97%)
11	12.485	104	styrene (96%)
12	12.805	96	2-methyl-2-cyclopenten-1-one (87%)
13	13.382	120	(1-methylethyl)benzene (86%)
14	13.397	95	2,5-dimethyl-1H-pyrrole (68%)
15	13.588	95	2,3-dimethyl-1H-pyrrole (68%)
16	13.601	95	2,5-dimethyl-1H-pyrrole (76%)
17	14.246	120	propylbenzene (83%)
18	14.438	120	1-ethyl-3-methylbenzene (64%)
19	14.451	120	1,3,5-trimethylbenzene (60%)
20	14.53	96	3-methyl-2-cyclopenten-1-one (72%)
21	14.78	94	phenol (91%)
22	14.893	93	aniline (91%)
23	15.027	118	alpha-methylstyrene (93%)
24	15.034	94	phenol (76%)
25	15.166	103	benzonitrile (46%)
26	15.167	140	1-decene (87%)
27	15.405	120	1,2,3-trimethylbenzene (78%)
28	15.406	120	1-ethyl-4-methylbenzene (83%)
29	15.406	120	1,2,4-trimethylbenzene (50%)
30	16.323	118	1-propenylbenzene (95%)
31	16.325	118	7-methylbicyclo[4.2.0]octa-1,3,5-triene (96%)
32	16.33	118	1-ethenyl-2-methylbenzene (93%)
33	16.52	110	2,3-dimethyl-2-cyclopenten-1-one (90%)
34	16.885	116	indene (86%)
35	16.899	108	2-methylphenol (70%)
36	17.105	132	5,7-dimethylenebicyclo[2.2.2]oct-2-ene (50%)
37	17.184	108	2-methylphenol (96%)
38	17.452	108	3-methylphenol (91%)
39	17.48	108	4-methylphenol (95%)
40	17.903	154	1-hepta-2-methyl-cyclopropane (70%)
41	17.904	112	pentylcyclopropane (49%)
42	17.91	124	2-methoxyphenol (45%)
43	18.031	124	2,3,4-trimethyl-2-cyclopenten-1-one (81%)
44	18.121	108	4-methylphenol (94%)
45	19.562	122	2,4-dimethylphenol (95%)
46	19.708	130	2-methylindene (97%)

47	19.716	130	(1-methyl-2-cyclopropen-1-yl)benzene (95%)
48	19.883	130	1-methyl-1H-indene (94%)
49	20.051	122	2-ethylphenol (76%)
50	20.186	130	1,4-dihydronaphthalene (91%)
51	20.321	122	2,5-dimethylphenol (60%)
52	20.321	122	2,3-dimethylphenol (55%)
53	20.758	128	naphthalene (94%)
54	21.789	136	3,4,5-trimethylphenol (87%)
55	22.504	144	1,1-dimethyl-1H-indene (70%)
56	22.525	144	1,3-dimethyl-1H-indene (93%)
57	22.732	144	1,2-dihydro-3-methylnaphthalene (83%)
58	22.738	144	2,3-dimethyl-1H-indene (92%)
59	22.935	182	1-tridecene (95%)
60	22.937	144	1,1-dimethyl-1H-indene (60%)
61	23.02	132	2,3-dihydro-1H-inden-1-one (96%)
62	23.28	117	indolizine (91%)*
63	23.319	117	indole (93%)*
64	23.48	142	1-methylnaphthalene (90%)
65	23.505	117	indole (68%)*
66	23.866	142	1-methylnaphthalene (90%)
67	23.869	142	2-methylnaphthalene (95%)
68	24.085	125	3-methoxy-2(1H)-pyridone (46%)
69	24.282	139	4-ethyl-1-azabicyclo[2.2.2]octane (59%)
70	24.423	137	N,N-di-2-propenyl-2-propen-1-amine (53%)
71	25.238	196	(E)-3-tetradecene (95%)
72	25.24	196	1-tetradecene (91%)
73	25.357	154	biphenyl (90%)
74	25.438	131	6-methyl-1H-indole (93%)*
75	25.46	131	3-methyl-1H-indole (93%)*
76	25.481	131	4-methyl-1H-indole (94%)*
77	25.77	156	1-ethylnaphthalene (94%)
78	25.786	156	2-ethylnaphthalene (90%)
79	26.286	156	2,7-dimethylnaphthalene (97%)
80	26.749	156	2,3-dimethylnaphthalene (98%)
81	26.766	156	1,3-dimethylnaphthalene (98%)
82	27.406	182	(Z)-6-tridecene (96%)
83	27.408	210	1-pentadecene (92%)
84	27.421	182	1-tridecene (96%)
85	27.633	168	4-methyl-1,1'-biphenyl (95%)
86	27.646	145	2,6-dimethyl-1H-indole (64%)*
87	27.662	145	2,5-dimethyl-1H-indole (86%)*
88	27.868	147	2,6-dimethylphenol isocyanate (64%)
89	29.748	152	pulegone (46%)
90	29.78	110	ethylidene-cyclohexane (43%)
91	29.869	166	3-hexylcyclohexene (50%)
92	29.872	166	fluorine (70%)
93	29.916	152	5-methyl-2-(1-methylethenyl)cyclohexanone (45%)
94	34.535	376	phthalic acid, isobutyl undecyl ester (83%)
95	35.149	237	hexadecanenitrile (70%)
96	35.166	167	undecanenitrile (56%)
97	35.171	209	tetradecanenitrile (90%)
98	35.509	270	hexadecanoic acid (97%)
99	35.513	270	pentadecanoic acid (98%)
100	36.084	214	tridecanoic acid (94%)
101	36.122	256	n-hexadecanoic acid (93%)

102	38.368	296	(Z)-9-octadecanoic acid (99%)
103	38.997	312	hexadecanoic acid (99%)
104	39.595	255	hexadecanamide (93%)
105	39.605	227	tetradecanamide (90%)
106	44.832	278	1,2-benzenedicarboxylic acid (70%)

Supplementary Information II. A multivariate statistical analysis of the analytical data to further investigate similarities between the effluent organics and to evaluate the occurrence of complementary information from the combination of different analytical approaches.

The pattern of similarity among samples detected by each analytical technique was investigated through the use of multivariate statistics. Multivariate statistics was used to aid in the interpretation of differences between the fully and partially treated effluent organic samples due to the high degree of similarity in results as determined by each of the three analytical techniques used.

For a qualitative comparison between the separate characterisation techniques, a common format is required. An effective way of achieving this, which has been demonstrated previously by Plant et al. (in press), is by the conversion of their respective data sets into nMDS ordination plots (Figure 6). For the NMR and HPSEC data, the resemblance matrices were based on Manhattan Distance, as the spectral data type includes continuous data points rather than independent points. For the py-GCMS data, the resemblance matrix was based on Bray-Curtis dissimilarity, since the input data consisted of a set of individual peak intensities and thus the data is a series of discrete values where adjacent points are completely independent. For the HPSEC resemblance matrix, the data were first transformed to the fourth root, to avoid data clumping, which was observed with the original data.

The nMDS plots in Figure 6 show similarities within the sample set based on the chemical and structural information expressed by each of the analytical techniques, and allow for direct comparisons. In an nMDS plot, points close in space represent samples that are chemically or structurally similar as determined by that analytical technique, and those spaced further apart are less similar.
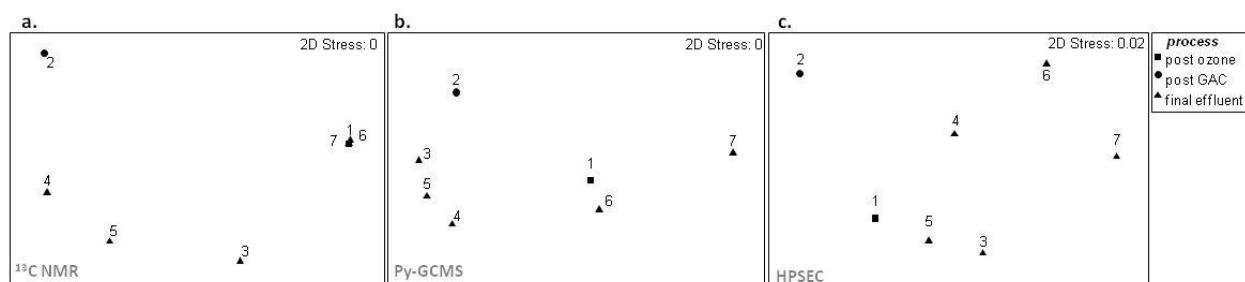
Figure 6 Non-metric multidimensional scaling ordination plots generated from the spectral data from NMR, py-GCMS, and HPSEC analyses; nMDS plots for NMR and HPSEC were based on Manhattan Distance resemblance matrices, whereas the nMDS plot for py-GCMS was based on the Bray-Curtis Dissimilarity matrix.

As discussed in the main text, each of the individual techniques indicated that the OM in these samples varied little. Due to the high degree of similarity between the samples, it was thought that the nMDS plots would aid in the differentiation between the effluent organic samples. However, the variation shown in the nMDS plots may actually reflect a degree of random variation or "noise" as the primary source of variation. A degree of similarity is apparent between the NMR and py-GCMS nMDS plots, with samples 1, 6 and 7 grouping together and samples 2-5 grouping together (Figure 6). As these two plots reflect similar sample distinctions, it suggests that the variation is due to structural features. However, when referring back to the original spectra, similarities between these particular groupings, especially for NMR, appear to reflect noise; 2-5 are noisier than 1, 6 and 7. The pattern of variation between samples evident in the HPSEC nMDS plot appears quite different to that observed for NMR and py-GCMS. This could be a result of the technique highlighting different aspects of difference among the samples, or that the variation is primarily random. Therefore, due to the variation in each of these nMDS plots possibly reflecting noise, multivariate analysis was found to be not appropriate for this data set.

One of the problems associated with the use of nMDS plots, is that the axes lack a scale. The variation, therefore, between samples is reflected by their closeness in space, rather than as a measure. Therefore, to put sample variation in an nMDS into context, reference organics could be added to the analysis. In this case, reference organics may show the effluent organics as being more similar or highlight the variation between the samples more effectively. Further to this, the most efficient test of random variation within the data sets would be to provide replicate samples for analysis. By increasing the number of

replicates, random variation would be reduced and structural variation would be better reflected. No replicate samples were available, so this approach could not be tested.

However, out of curiosity, a quantitative comparison of the information provided by the three different analytical techniques was carried out to investigate the amount of complementary information provided in the analysis. The Spearman Rank Correlation method was used to do this, as has been previously reported (Plant et al. in press). This method provides a measure, rho (ρ), of the strength of the relationship between two resemblance matrices ($\rho=0$: no relationship; and $\rho=1$: the strongest correlation). This analysis confirmed there to be a significant (P = 0.05) correlation between the NMR and py-GCMS resemblance matrices with $\rho=0.44$. This means that 44% of sample discrimination was common to NMR and py-GCMS, and 56% of discrimination was technique specific information. The correlation found between the NMR and py-GCMS data sets provides support for the argument that the variation between the effluent organic samples is due to structural variation and not random variation. In addition, multivariate analysis confirmed there was no correlation between either of these techniques and HPSEC. The lack of correlation supports the argument that the variation seen for the HPSEC data is primarily random variation.

CHAPTER SIX

SYNOPSIS

*

It has long been recognised that organic matter (OM) is one of the most complex natural mixtures on Earth. This material is known to play an important role in the natural world, and is ubiquitous in all natural environments. However, despite the rich literature on the role and function of OM in many environments, key aspects of its composition and structure remain unclear. It is now appreciated that OM is not a single, defined material, but rather exists as a complex mixture, whose structure varies according to source, the environment in which it was produced or transported to, and its stage of degradation.

There are many different analytical techniques that are used to study OM. Amongst these, there are several sophisticated techniques that can provide valuable chemical or structural information on OM. These sophisticated analytical techniques are well understood and have been widely applied to the study of OM. A common approach to OM characterisation now involves the combined use of these sophisticated techniques to provide for a more informative analysis. However, it remains unclear when they are used together, how much of the information provided by the different techniques are the same and how much information is unique.

The investigations outlined in this thesis describe the development of a protocol to qualitatively and quantitatively assess the information provided by a multi-technique approach to OM characterisation. The sophisticated analytical techniques chosen for the multi-technique approach are at the forefront of OM characterisation and include (i) solid-state ^{13}C nuclear magnetic resonance (NMR) spectroscopy, (ii) flash pyrolysis-gas chromatography mass spectrometry (py-GCMS), and (iii) high performance size exclusion chromatography (HPSEC). This approach to assess the extent of technique complimentarity is based on multivariate statistics and involves the use of ordination plots to assess the information qualitatively, and the Spearman Rank Correlation method to assess the information quantitatively. The developed methodology was applied to a range of different OM samples associated with wastewater (WW) processing and the findings of these studies are outlined below.

CHAPTER TWO - A METHODOLOGY FOR COMBINING AND COMPARING INFORMATION FROM NMR AND PYROLYSIS IS DEVELOPED AND DEMONSTRATED

This chapter describes the development and initial testing of the approach used to combine and compare information from different characterisation techniques using multivariate statistics. This work has been published (2013) in a special edition of *Current Organic Chemistry* titled "Recent advances in environmental organic and bio-organic chemistry".

In this study, a suite of plant residues were characterised using solid-state ^{13}C NMR spectroscopy and flash py-GCMS. The plant residues chosen for this study had been studied previously. They were chosen on the following bases: (i) they are all carbon-rich; (ii) ample material was available for all analyses; (iii) there was a reasonable degree of variation in composition among the samples; and (iv) the results of both NMR and pyrolysis could be readily anticipated and checked against analyses of similar materials published in the literature.

The first requirement for comparing information provided by the two techniques was transformation into a common format. This was achieved by constructing a non-parametric multidimensional scaling (nMDS) ordination plot for each analytical approach. The nMDS plots reflect the multivariate data for each technique in such a way that similarities and differences between the organic residues were more easily interpreted. However, care was taken when using the ordination plots as they are just a representation of the true multivariate information used to create them. Therefore the ordination plots were used as a guide only to assess the original spectral results for each technique. By using these ordination plots, similar sample comparisons could be made based on either NMR or py-GCMS data.

It should be noted here that to construct the nMDS plots, Manhattan Distance and Bray-Curtis Dissimilarity measurements were used for the NMR and py-GCMS resemblance matrices, respectively. In these representations, a smaller separation between points signifies less dissimilarity between samples and a larger separation signifies greater dissimilarity. For a distance measure, such as Manhattan Distance, there is no upper band to the level of resemblance, so the data is scaled from zero (no distance) to infinity (greater

distance). For a dissimilarity measure, such as Bray-Cutis dissimilarity, the level of resemblance is scaled from zero (no dissimilarity) to 100 (maximum dissimilarity).

What was novel about this study was the demonstration of a quantitative assessment of how the chemical information from the different analytical approaches complemented each other. This was achieved using the Spearman Rank Correlation method, which, in effect, assessed the power of the different analytical approaches to differentiate between the plant residues based on the chemical information they provided. For this particular set of samples, 61% of the discrimination between samples was common to both analytical methods. Therefore more than half of the chemical information provided by one of the techniques was also provided by the other. As well as the large amount of complementary information provided by the two techniques, this identifies some (39%) as being method-specific. This is a particularly important finding in terms of a multi-technique approach to OM characterisation. The degree of similarity measured between NMR and py-GCMS data is reassuring especially as both techniques are used on their own to characterise organics. It would also suggest that the use of these two techniques is to some extent interchangeable, as the majority of information could be provided by implementing just one of these techniques. However, it also shows that there is a substantial amount of technique-specific information, indicating there is benefit in using these techniques together.

To further assess the relationship between the multivariate information from the different techniques, the resemblance values were plotted against each other. This allowed for an assessment of which technique provided the greatest amount of discrimination between the organic residues. The technique that provided the greatest discrimination in this study was NMR, with the greatest spread of resemblance measures being 0.81 compared to 0.33 for py-GCMS. Therefore even though both techniques drew similar conclusions when discriminating between the organics, overall NMR was determined as being more sensitive to the chemical differences.

With the ability to study technique complementarity, informed decisions as to what techniques are best to apply to an organic sample are possible, saving time, money and resources. However, these findings relate specifically to this particular set of plant residues and may not be reflective of all OM samples. The amount of technique

complementarity may be different for other sets of organics, and so this approach was subsequently applied to other suites of organic materials.

CHAPTER THREE – APPLICATION OF THE DEVELOPED METHODOLOGY TO THE STUDY OF SEDIMENT ORGANICS.

The work in this study applied the methodology demonstrated in Chapter Two to qualitatively and quantitatively assess the same multi-technique approach to the study of sediment organics. The work in Chapter Three has been submitted (July 2013) to *Environmental Science and Pollution Research.*

The basis of the study in Chapter Three was to characterise the organics in the sediments of a pulp and paper mill receiving environment (Lake Bonney) in order to identify any organics of previous mill discharge. Lake Bonney was sampled around the vicinity of the discharge point of the mill. Sample cores were collected from four sites and the sediment profiles, which were distinguished based on colour, were separated and the organics from each were isolated. Isolation involved using the established methodologies of treating the matrix with hydrochloric acid, to remove the carbonate (shells etc) present, and then hydrofluoric acid, to demineralise the sample. This allowed for the separation of the particulate organics from the sediment matrix for analysis. The sediment organics were characterised using solid-state ^{13}C NMR spectroscopy and flash py-GCMS.

Both NMR and py-GCMS detected methoxy-phenol compounds, which are lignin-derived materials, in the sediments, which may be indicative of pulp and paper mill organics. However there was no obvious concentration increase of these aromatic organics with depth, which would suggest heavy organic loads from WW discharged prior to the use of more advanced treatment strategies. Neither were the higher concentrations aromatic organics near the discharge point. Therefore the contribution of mill organics to the receiving environment was determined to be small when all sources of organic inputs to the lake were considered.

The two analytical approaches were qualitatively compared by generating nMDS plots as demonstrated in Chapter Two. Again, ordination plots were used to demonstrate similarities and differences between the organic samples as determined by the different analytical methods. From the two plots, the different analytical methods were seen to

differentiate between the sediment organics to a similar degree. Both techniques distinguished between the top-most layers of each site similarly, with more variation seen between the deeper organics from the different sites.

A quantitative assessment of this information, as demonstrated in Chapter Two, showed that 42% of sample discrimination between the sediment organics was common to NMR and py-GCMS. This means that the two techniques provided unique information to the analysis (i.e. 58%). As with the previous study, this work highlighted the advantage of using solid-state ^{13}C NMR spectroscopy and flash py-GCMS together to characterise this suite of sediment organic samples.

Interestingly, the amount of complementary information provided by the two analytical techniques in this study was different than for the plant residues in the previous study (Chapter Two). This was in line with expectations. Organic matter is known to be highly complex, and therefore we would expect a degree of variation when the same technique is applied to different OM samples. Therefore the application of this multivariate statistics approach to further characterise OM and assess technique-complementarity was insightful and successfully applied to this set of unknown organics.

CHAPTER FOUR – EXPANSION OF THE METHODOLOGY TO INCLUDE HPSEC ANALYSIS FOR APPLICATION TO PULP AND PAPER MILL WASTEWATER ORGANICS.

In this chapter, the methodology developed in Chapter Two was expanded to include a third analytical approach, HPSEC, to the study of a different set of organic samples. The three analytical approaches in this study provided a qualitative and quantitative assessment of the organics before being assessed by the multivariate approach. The work in Chapter Four has been published (2013) in *Environmental Science and Pollution Research.*

The three analytical techniques used in Chapter Four were applied to dissolved organics collected at different points through a pulp and paper mill and its secondary WW treatment process to determine how the nature of the organics changed during treatment. Water samples were collected at six stages, following the water progress through the mill to the discharge point for release. Samples were collected from a tissue machine, a clarifier, each of three aerated stabilisation basins (ASBs), and at the discharge point to the

receiving environment, which is Lake Bonney, whose sediments were studied in the previous chapter. The dissolved organics from these samples were isolated using the standard practice of filtering the samples to 0.45 µm.

The multi-technique characterisation approach was expanded to include HPSEC for these unknown organics. High performance size exclusion chromatography is a wet chemical technique and was included in this study to provide a different perspective on the dissolved organics. This technique has also become a popular method for studying dissolved organics due to its ability to provide in-time and reproducible results. A limitation to the HPSEC approach used in this study is a bias towards UV detectable organics. There are other detection methods possible with HPSEC, including DOC and SUVA, which would provide for further opportunities for OM characterisation.

Characterisation by each of the analytical techniques was extended in this study to provide qualitative and quantitative information on the dissolved organics. For NMR, integration across broad chemical regions provided relative proportions of broad carbon (C) groups; for py-GCMS, the relative distribution of common functional groups was determined; and for HPSEC, a deconvolution procedure was used to discriminate overlapping peaks in the HPSEC profiles.

Each technique identified the tissue machine organics as being quite distinct, the clarifier and ASB 1 organics as being similar; and the ASB 2, ASB 3 and outlet organics as being similar also. However, some information was clearly technique-specific, including the detection of N and P-containing organics in the ASBs by py-GCMS. The residual organics at the outlet that leave for Lake Bonney contained carbohydrate and aromatic carbon, which are indicative of pulp and paper mill organics (lignin). This included methoxy-phenol species that were isolated in the sediments of Lake Bonney, as discussed in Chapter Three.

Again, data from the analytical techniques was first compared qualitatively by constructing nMDS plots for each technique. The nMDS plot for the HPSEC data was prepared, in the same way as the NMR data, based on Manhattan Distance as the raw data (the chromatogram) consisted of a smooth trace in which adjacent points are highly correlated. In addition, an nMDS plot was also constructed from the basic chemical properties (pH, EC and DOC) measured for the dissolved organic samples. This was to

follow changes in these aspects of water quality through the mill and secondary WW treatment. The resemblance matrix for the basic chemical properties was constructed the same way as for the py-GCMS data, using Bray-Curtis Dissimilarity, due to the independent nature of the data.

Three of the four nMDS plots indicated that the largest change in the nature of the organics occurred between the tissue machine and the clarifier. The exception was the nMDS plot for py-GCMS, which showed the largest change between the ASB 1 and ASB 2 organics. This change was due to the reoccurrence of smaller molecular fragments, which were present in the tissue machine, in ASB 2. The ability of py-GCMS to notice this difference probably reflects the higher sensitivity of this technique to specific molecular level features. The nMDS plots of the analytical data reflected the dissimilarity of the tissue machine organics and the two major groupings of the remaining organics as discussed above.

Using the approach described in Chapter Two, the three analytical techniques were quantitatively compared using the Spearman Rank Correlation method. It was determined that NMR and HPSEC were very closely correlated, with 86% of sample discrimination common to both techniques. This suggests that the size of the organic compounds was closely related to the distribution of broad functional groups present. With such a high level of technique complementarity measured in this study, it can be speculated that these two techniques could be used interchangeably for the study of similar dissolved organics in terms of discriminating between the samples. The lack of a relationship between these methods and py-GCMS shows that the latter provided only new information to the analysis of these dissolved organics. The unique information provided by py-GCMS is thought to be due to its ability to detect a wider range of products, which provides a greater degree of sensitivity to the analysis.

It is worth noting that this was the first study that did not show a relationship between data from solid-state ^{13}C NMR spectroscopy and flash py-GCMS. In other words, this was the first study that NMR and py-GCMS information did not complement each other. Obviously this would be due to the type of organics studied and may even be a result of the limitations of the techniques. In this case as py-GCMS provided unique information only, suggesting that the level of molecular detail it provided was far beyond the structural detail provided by NMR.

CHAPTER FIVE – APPLICATION OF THE METHODOLOGY TO SEWAGE EFFLUENT ORGANICS FROM THE UNITED KINGDOM.

The work included in this chapter details the application of the expanded methodology, including three analytical techniques, to the characterisation of a different set of WW organics. The work detailed in Chapter Five has been prepared for submission to *Environmental Monitoring and Assessment*.

The purpose of this work was to characterise a suite of partially and fully treated sewage effluent dissolved organics from the United Kingdom. Once the sewage organics were concentrated by the process of desalting, they were characterised by the three analytical techniques. As in Chapter Four, qualitative and quantitative information was provided by each analytical approach. The composition of the samples appeared to be very similar by all techniques. Each sample by NMR was found to be dominated by O-alkyl C; with pyrolysis, the fragments were dominated by aromatic material; and the apparent molecular weights (AMWs) were mainly clustered around 2000 Da, as determined by HPSEC. An interesting result of characterisation by py-GCMS was the detection of pharmaceutical metabolites in all of the organics.

The data from the three characterisation techniques were qualitatively compared with the aid of nMDS plots. Despite the large degree of similarity between the data, each of the plots showed variation between the samples. There appeared to be some similarity in the variation between the organics as depicted by the NMR and py-GCMS plots, with two major groupings of samples. However when referring back to the original spectra, variation was likely reflecting noise. On the other hand, HPSEC groupings were quite different to the other two plots. Therefore it was speculated that due to the high degree of similarity seen in the data, the variation in the nMDS plots may be a reflection of the random variation or "noise" rather than actual variation in composition. This shows that nMDS plots display all forms of variation between samples including random variation if it dominates the distinction between organic samples. As random variation was thought to be reflected in the HPSEC nMDS plot, the use of multivariate statistics in this case was deemed inappropriate for the sample set. The best way to test this is to analyse replicate samples; unfortunately there was an insufficient amount of material to do this for NMR analysis.

Out of interest and for completeness, the Spearman Rank Correlation method was also carried out on the multivariate data. Quantitatively, NMR and py-GCMS were the only techniques to be correlated, with 44% of sample discrimination common to both analytical methods. This amount of technique complementarity was very similar to that measured for the sediment organics in Chapter Three. However as both techniques provided a considerable amount (56%) of unique information, more-so than complementary information, it would be of more benefit to use these techniques together in this case. On the other hand, HPSEC was the only technique to not be correlated. This result was not surprising as the variation depicted in the qualitative analysis (nMDS plot) was different to the other techniques.

In this assessment of the effluent analytical data by multivariate statistical methods, ordination plots were likely pushed to their limits in distinguishing between the organic samples. This result makes for an interesting contrast to the results in the previous chapters. An aspect of nMDS plots that was not previously an issue is the lack of a scale on the axes. It wasn't until this particular study, involving very similar organic samples, that variation between the organics could not be effectively reflected in the nMDS results; there was little confidence in the differentiation displayed in the plots due to random variation dominating. An approach to address this limitation of nMDS would be to include some reference organics in the plots to help quantify the differences between the effluent organics. In an effort to do this, all of the organic sample sets studied in Chapters Two to Five were prepared into one nMDS plot for each technique (Figure 1-3).

With the addition of reference organics in the NMR nMDS plot (Figure 1), the similarities between the effluent organics appear to be reflected in a similar manner to the original nMDS in Chapter Five. Variation in the composition of the effluent organics is obvious in Figure 1 and appears expanded compared to the nMDS in Chapter Five; however, the variation here is greater than the variation between the organics in the other sample sets. This is a particularly unusual finding, as the NMR spectra of the effluent organics were more noticeably similar in comparison to the spectral results of the other organic sets, particularly those organics from the pulp and paper mill (Chapter Four).

For the py-GCMS results, the similarities between the effluent organics (Figure 2) are reflected differently to the original nMDS in Chapter Five. However, what is important

here is that the organics are more compressed; appear closer in space. This is even more so for the HPSEC data (Figure 3), where the effluent organics are compressed compared to the original nMDS and the reference organics sit further away in space. Therefore with the addition of the other organic sets studied as reference organics, the structural similarities between the effluent organics become more quantifiable. For the py-GCMS results (Figure 2), there is a degree of variation between the effluent organics which can be thought of as better reflecting structural information. For the HPSEC results (Figure 3), the structural variation is slight with all sample points overlapping. The lack of variation now shown in the nMDS plots is possibly more representative of the structural variation between the spectra, as there is a gauge of similarity provided by the reference organics.
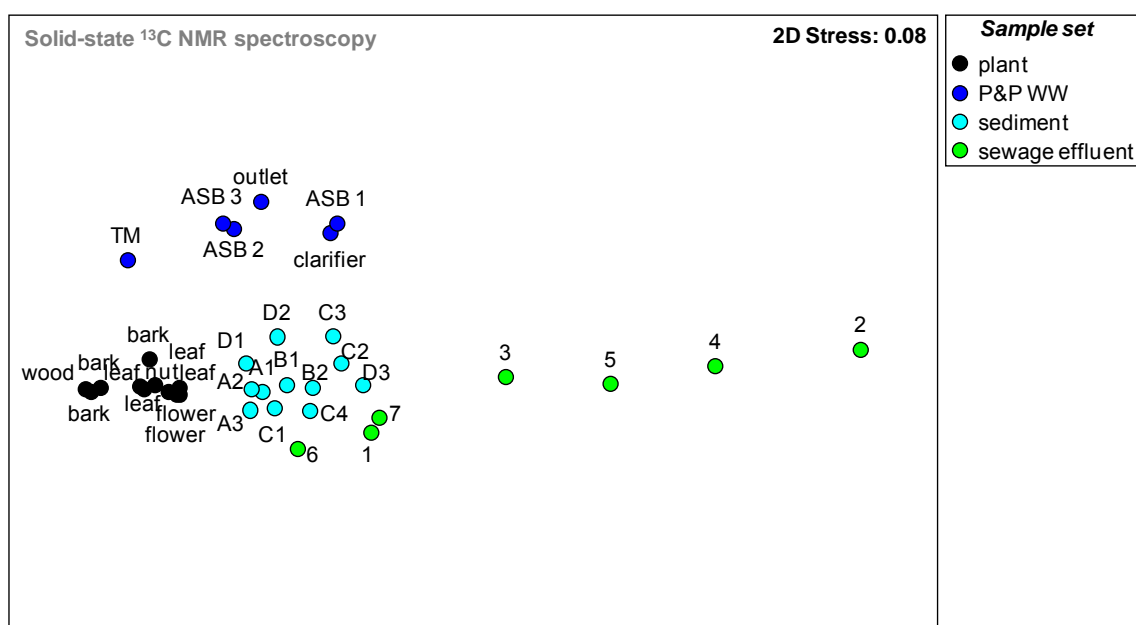

Figure 1 An nMDS plot displaying the differentiation between the four different organic sample sets as depicted by solid-state ^{13}C NMR spectroscopic characterisation. Sample labels have been altered for the plant and sediment samples only. Plant organics have been labelled as residue type, and sediment organics have been labelled as site (A-D) and layer from top to bottom (1-4). TM is short for tissue machine.
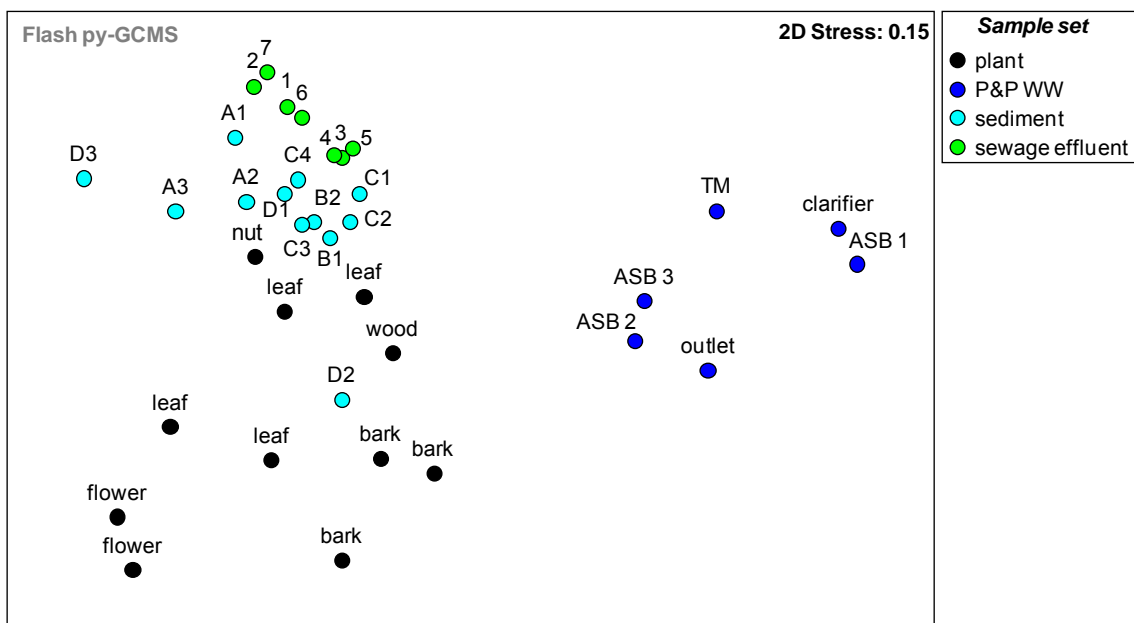
Figure 2 An nMDS plot displaying the differentiation between the four different organic sample sets as depicted by flash py-GCMS analysis. Sample labels have been altered for the plant and sediment samples only. Plant organics have been labelled as residue type, and sediment organics have been labelled as site (A-D) and layer from top to bottom (1-4). TM is short for tissue machine.
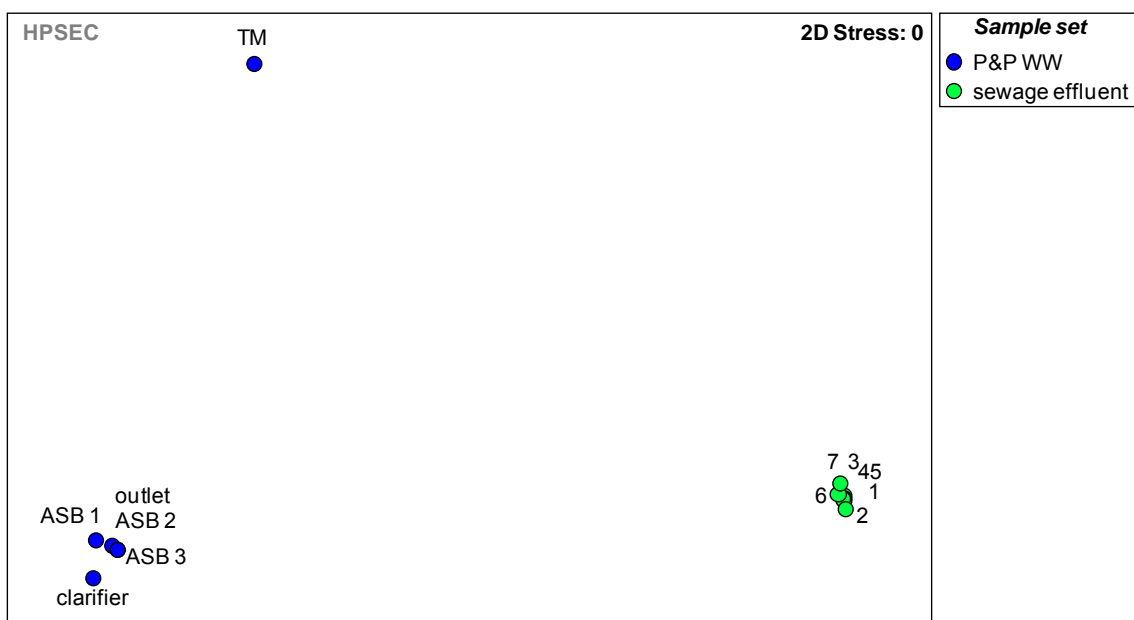

Figure 3 An nMDS plot displaying the differentiation between the two different organic sample sets as depicted by HPSEC analysis. The data was log transformed to prevent clumping. TM is short for tissue machine.

CONCLUSIONS AND FUTURE DIRECTIONS

As discussed above, solid-state ^{13}C NMR spectroscopy and flash py-GCMS have been widely applied to the study of OM and are each well understood. The addition of the increasingly popular wet chemical technique of HPSEC was included where appropriate to provide a different perspective on the nature of the organic samples compared to the two solid-state techniques. The only aspect of this study that was allowed to vary was the type of organics that were characterised. A range of different organic samples were chosen to provide a strong test of the multivariate statistical approach developed in Chapter Two. Wastewater organics were chosen due to their associated high organic loads, which proved helpful for isolating sufficient sample quantities for analysis by each of the analytical methods. Characterisation of WW organics by each of the analytical techniques in this study are common, and an understanding of how the data from these techniques compare during the treatment of WW organics were thought to be particularly beneficial for the understanding and optimisation of WW treatment.

Multivariate methods were able to combine the different types of data from the analytical techniques into a format that was directly comparable. This involved the generation of a resemblance matrix for each technique used. The multivariate information in each of these resemblance matrices was then displayed by way of nMDS ordination plots. The nMDS plots for each of the techniques were easily interpretable and helped in the identification of sample similarities. More importantly, they allowed for the direct comparison of data from different analytical techniques so that sample discrimination could be compared.

By preparing each of the sample sets into the same nMDS plot (Figures 1-3) each set become reference organics for the others. For the techniques of NMR and py-GCMS, the different organic sets are shown as being more similar compared to their original nMDS plots. However similarities within organic sets still appear, for example for the pulp and paper mill WW organics the tissue machine organics appear most dissimilar, the clarifier and ASB 1 are similar, and ASB 2, ASB 3 and the outlet are similar also for each of the techniques. In the HPSEC plot, the tissue machine (TM) organics appear considerably dissimilar to the other WW organics, to the same degree as to the sewage effluent organics (Figure 3).

The most novel aspect of the approach was that the multivariate information for each technique was quantitatively compared by using the Spearman Rank Correlation method. This method provided a measure of the amount of complementary information provided by the analytical techniques. For each of the different sets of organics characterised, the techniques provided a different amount of complementary information (Table 1). In Chapters Two and Three, where only NMR and py-GCMS techniques were used to characterise the organic samples, the results were correlated and hence a degree of complementary information was determined. In the last two studies, Chapters Four and Five, the data from two of the three techniques were correlated, and the third was seen to provide only new information to the analysis. For the organics studied in Chapters Two (plant residues), Three (sediment) and Five (sewage WW) the broad structural groups determined by NMR and the more detailed molecular fragments of py-GCMS were able to differentiate between the organic samples to a similar degree (Table 1). In Chapter Four (pulp and paper mill WW), the broad structural groups determined by NMR and the molecular weight distributions determined by HPSEC differentiated between the organic samples to a similar degree (Table 1). This diversity in complementary information was expected due to the different organic samples studied. If diversity was not seen, the developed methodology would need to be questioned as the variable nature of the OM would not have been reflected. Therefore, as an initial attempt to combine and quantitatively compare data from different analytical techniques in a meaningful way, this multivariate approach seemed reasonable and was insightful.

Table 1 Summarising the amount of complementary information provided by the analytical methods measured by the Spearman Rank Correlation method for the four different sets of OM. The light grey coloured squares indicate the techniques that a significant correlation was measured; the blank squares show that no correlation was measured; and the dark grey show the technique that was not applied.

Type of OM	Solid-state ^{13}C NMR spectroscopy	Flash py-GCMS	HPSEC
Plant residues	61%		
Sediment organics	42%		
Pulp and paper mill WW DOM	86%		86%
Sewage WW DOM	44%		

From applying the developed approach to different sets of organic samples, it was determined that the most beneficial combination of analytical techniques was when a degree of complementary information as well as some new technique-specific information was provided. This was the case for each set of organic samples studied. Therefore, solid-state ^{13}C NMR spectroscopy, flash py-GCMS, and HPSEC were shown to be a good combination for a multi-technique approach to OM characterisation in each case. If each of the techniques provided technique-specific information only, there would be the benefit of more structural information on the organics, however, there may be a degree of uncertainty in the results if no information drew similar conclusions.

The work in this thesis has set the foundations of an approach to assess technique complementarity. This research has gone so far as to identify when and which of the three analytical techniques provide complementary information and then provide a measure of how much discrimination was common. The future directions of this research would be to: (i) determine what information, from the correlated techniques, is responsible for providing the same sample discriminations, (ii) study a wider range of organic samples with the three analytical techniques in order to assess when they are best applied together and whether predictions can be made for their combined use, and finally (iii) apply this methodology to other characterisation techniques when used to study OM.

It is unclear at this stage if multivariate statistics would be able to determine the information responsible for correlations between techniques. An ideal approach would be to characterise a set of OM that is well understood, potentially Suwannee River organics, or even simple organic structures with predictable results. Technique complementarity could be assessed with the characterisation of simple structures, and complexity of the organics could be increased with subsequent tests.

For a greater understanding of the three analytical techniques used in this thesis, a wider range of organic samples would need to be studied. By increasing their organic repertoire, predictions of when these techniques may be correlated could be made. A major benefit of making informed predictions means that these techniques can be more suitably applied to the characterisation of OM. For example in the case of the pulp and paper mill organics studied in this work (Chapter Four), the structure of the organics from this mill may be better studied in the future by the use of NMR and py-GCMS only, as HPSEC was shown

to provide 86% of the same discrimination information as NMR. This would save time and money by avoiding the application of HPSEC.

The methodology developed and demonstrated in this thesis has the potential to be applied to other characterisation techniques that are used to study OM. As the structure and role of OM has been a topic rigorously studied for decades, an assessment of the techniques commonly applied to study it would be beneficial. This would allow for the understanding of the strengths and limitations of different characterisation techniques when applied to the study of OM structure and role in the environment.

Although it is common knowledge that a multi-technique approach is more informative when characterising OM, understanding what techniques are better applied together for different organic sets has the potential to allow for more specific and informative analyses. By improving the understanding of how and when to apply different analytical techniques to the characterisation of OM, all areas of OM research discussed in the Introduction have the potential to be improved. This includes optimising the removal of OM through WW treatment processes, and improving the quality of drinking water, as well as understanding the role and function of OM in soils, sediments, natural waters and the atmosphere.