

**CHARACTERISING WATER TREATMENT WORKS
PERFORMANCE USING FLUORESCENCE
SPECTROSCOPY**

By

MAGDALENA ZOFIA BIEROZA

**A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY**

**School of Civil Engineering
The University of Birmingham
November 2009**

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ABSTRACT

Organic matter (OM) in drinking water treatment is a common impediment responsible for increased coagulant and disinfectant dosages, formation of carcinogenic disinfection-by-products (DBPs), and microbial re-growth in distribution system. The inherent heterogeneity of OM implies the utilization of advanced analytical techniques for its characterisation and assessment of removal efficiency. Here, the application of simple fluorescence excitation-emission (EEM) spectroscopy to OM characterisation in drinking water treatment was presented. Monthly raw and clarified water samples were obtained for 16 UK surface water treatment works. Fluorescence EEM spectroscopy was used for the assessment of total organic carbon (TOC) removal and OM characterisation. Fluorescence peak C intensity was found to be a sensitive and reliable measure of OM content and hence an indicator of DBPs presence. Fluorescence peak C emission wavelength and peak T intensity (reflecting the degree of hydrophobicity and the microbial fraction respectively) were found to characterise the OM; the impact of both on TOC removal efficiency was apparent. OM fluorescence properties were shown to predict TOC removal, and identify spatial and temporal variations. The simplicity, sensitivity, speed of analysis and low cost, combined with potential for incorporation into on-line monitoring systems, mean that fluorescence spectroscopy offers distinct advantages over other THM precursors characterisation techniques.

*Moim Rodzicom,
Wiesławie i Leonowi*

*To my parents,
Wiesława and Leon*

ACKNOWLEDGMENTS

I would like to thank my supervisors, Dr John Bridgeman and Prof. Andy Baker, for all their support, endless enthusiasm for the project, and kindness.

I would like also to thank my friends, Lina Lindfors for proofreading of this thesis and Shahad Aljanabi for her help with running of samples.

I would like also to thank my beloved friend, Prof. Juliusz Lukasiewicz, for his optimism, faith, and friendship.

Finally I would like to thank my family, my parents and my sister Ania, for their endless support, whole hearted encouragement and always being there for me.

TABLE OF CONTENTS

1. INTRODUCTION	1
1.1. BACKGROUND	1
1.2. SCOPE OF STUDY	2
1.3. PUBLICATIONS	3
1.4. THESIS STRUCTURE	4
2. LITERATURE REVIEW	6
2.1. ORGANIC MATTER IN FRESHWATER ECOSYSTEMS.....	6
2.2. ORGANIC MATTER CHARACTERISATION	10
2.2.1. <i>Isolation and fractionation of organic matter</i>	10
2.2.2. <i>Organic matter characterisation with spectroscopic techniques</i>	14
2.2.3. <i>Molecular and environmental effects on organic matter fluorescence</i>	20
2.3. ORGANIC MATTER IN DRINKING WATER TREATMENT	28
2.3.1. <i>Formation of disinfection by-products</i>	28
2.3.2. <i>Organic matter removal in drinking water treatment</i>	39
2.4. APPLICATION OF FLUORESCENCE SPECTROSCOPY IN ORGANIC MATTER CHARACTERISATION IN DRINKING WATER TREATMENT	45
2.5. <i>Characterising water treatment works performance using fluorescence spectroscopy</i>	55
2.6. CHAPTER 2 FIGURES	61
3. MATERIALS AND METHODS	67
3.1. SOURCE WATERS	67
3.2. LABORATORY METHODS	73
3.3. DATA PRE-PROCESSING	77
3.4. CHAPTER 3 FIGURES	79
3.5. CHAPTER 3 TABLES.....	89
4. FLUORESCENCE PROPERTIES OF RAW AND PARTIALLY-TREATED WATER AND ORGANIC MATTER REMOVAL EFFICIENCY	92
4.1. FLUORESCENCE CHARACTERISATION OF RAW AND CLARIFIED WATER.....	92
4.2. ORGANIC MATTER REMOVAL EFFICIENCY.....	98
4.3. CHAPTER CONCLUSIONS	107

4.4.	CHAPTER 4 FIGURES	109
4.5.	CHAPTER 4 TABLES	122
5.	ADVANCED DATA MINING TECHNIQUES IN PATTERN RECOGNITION AND CALIBRATION OF FLUORESCENCE DATA	128
5.1.	APPLICATION OF DATA MINING TECHNIQUES FOR ADVANCED ANALYSIS OF FLUORESCENCE DATA	128
5.1.1.	<i>Pattern recognition techniques</i>	131
5.1.2.	<i>Calibration techniques</i>	138
5.2.	PATTERN RECOGNITION OF DRINKING WATER FLUORESCENCE DATA..	142
5.2.1.	<i>Principal components analysis</i>	142
5.2.2.	<i>Parallel factor analysis</i>	144
5.2.3.	<i>Self-organizing map analysis</i>	150
5.2.4.	<i>Comparison of pattern recognition methods</i>	155
5.3.	CALIBRATION OF DRINKING WATER FLUORESCENCE DATA	159
5.3.1.	<i>Comparison of calibration methods</i>	159
5.4.	CHAPTER CONCLUSIONS	162
5.5.	CHAPTER 5 FIGURES	164
5.6.	CHAPTER 5 TABLES.....	184
6.	FLUORESCENCE IN CHARACTERISATION OF WATER TREATMENT OPTIMISATION PROCESSES.....	187
6.1.	OPTIMISATION OF DRINKING WATER TREATMENT	187
6.2.	CAMPION HILLS TOC REMOVAL OPTIMISATION INVESTIGATION	189
6.3.	CHAPTER CONCLUSIONS	196
6.4.	CHAPTER 6 FIGURES	198
6.5.	CHAPTER 6 TABLES.....	205
7.	VARIATION IN FLUORESCENCE PROPERTIES THROUGH WATER TREATMENT WORKS	207
7.1.	INTRODUCTION	207
7.2.	EFFECTS OF CHLORINE ADDITION ON FLUORESCENCE	208
7.3.	RESULTS	210
7.3.1.	<i>Visual inspection of EEMs</i>	210

7.3.2.	<i>Fluorescence regional integration</i>	212
7.3.3.	<i>Self-organizing map</i>	215
7.4.	CHAPTER CONCLUSIONS	216
7.5.	CHAPTER 7 FIGURES	218
7.6.	CHAPTER 7 TABLES	225
8.	DISCUSSION OF THE RESULTS	226
8.1.	RELATING FLUORESCENCE OF RAW AND CLARIFIED WATER TO ORGANIC MATTER REMOVAL EFFICIENCY.....	226
8.2.	RELATING ORGANIC MATTER FLUORESCENCE TO DISINFECTION-BY PRODUCTS FORMATION.....	231
8.3.	COMPARISON BETWEEN FLUORESCENCE EEM SPECTROSCOPY AND STANDARD ORGANIC MATTER CHARACTERISATION TECHNIQUES	238
8.4.	COMPARISON BETWEEN DIFFERENT DATA MINING TECHNIQUES FOR FLUORESCENCE DATA ANALYSIS	240
8.5.	CHAPTER 8 FIGURES	243
8.6.	CHAPTER 8 TABLES	244
9.	CONCLUSIONS	248
9.1.	SUMMARY OF THE FINDINGS	248
9.2.	IMMEDIATE APPLICATIONS FOR WATER COMPANIES.....	250
9.3.	FUTURE RESEARCH POTENTIAL.....	252
	LIST OF REFERENCES	253

LIST OF FIGURES

Figure 2.1:	The river continuum concept (after Junk and Wantzen, 2003)	61
Figure 2.2:	Theoretical humic (a) and fulvic acid (b) (after Hudson et al., 2007)	62
Figure 2.3:	Jablonski energy diagram (after Hudson et al., 2007 and Lakowicz, 1999)	63
Figure 2.4:	Fluorescence excitation-emission matrix with typical fluorescence features (peak A, C, T and Rayleigh-Tyndall effect) (after Coble, 1996)	64
Figure 2.5:	The effect of aluminium on fluorescence of humic acid (9 mg/l; Reynolds and Ahmad, 2005)	64
Figure 2.6:	Four most common THMs: chloroform (a), dichlorobromomethane (b), dibromochloromethane (c), bromoform (d)	65
Figure 2.7:	Removal of DOC from a sample of Middle River Reservoir (South Australia) water under jar test conditions at pH 5.5 (after van Leeuwen et al., 2005)	65
Figure 2.8:	The two patterns of relationship between TOC removal and coagulant dosage (after Randtke, 1988)	66
Figure 2.9:	Schematic diagram of the drinking water treatment stages	66
Figure 3.1:	Location of the WTWs with the selected land use types: urban fabric (red), forests (green), and wetlands (purple)	79
Figure 3.2:	Severn Trent Water WTWs network (Severn Trent Water 2008)	80
Figure 3.3:	Total monthly rainfall at Sutton Bonington station (blue columns), average monthly rainfall (grey columns), and mean air temperature (red line) in the study period (July 06 – February 08) (Met Office data)	81
Figure 3.4:	Total monthly rainfall at Sutton Bonington station (blue columns), and at Ross-on-Wye (red columns) in the study period (July 06 – February 08) (Met Office data)	82
Figure 3.5:	Raman peak intensity (au) during the study period	83
Figure 3.6:	Instrumental correction factors for emission (a) and excitation (b) wavelengths	84
Figure 3.7:	Excitation-emission matrices (EEMs) with the Rayleigh-Tyndall scatter for raw and clarified water for Bamford and Little Eaton WTWs. Bamford raw (a) and clarified water (b), Little Eaton raw (c) and clarified water (d)	85-86

Figure 3.8: Changes in fluorescence intensity of peak C (a) and peak T (b) due to sample storage for Campion Hills and Draycote WTWs. Error bars (dashed line Draycote, solid line Campion Hills)	87
Figure 3.9: Raw water EEM for Bamford WTW after scatter removal (compare with Figure 3.5 a)	88
Figure 4.1: Fluorescence EEMs characterising OM changes during the treatment process. Bamford raw water (a), Bamford clarified water (b), Little Eaton raw water (c), and Little Eaton clarified water (d)	109- 110
Figure 4.2: Fluorescence properties of raw and clarified water for all WTWs. Peak C intensity vs peak T2 intensity for raw (a) and clarified water (b), peak C emission wavelength vs T2/C intensity ratio for raw (c) and clarified water (d). Fluorescence intensity in arbitrary units (au). Linear regression with 95 % prediction lines	111- 113
Figure 4.3: Mean peak C emission wavelength vs mean T2/C intensity ratio for raw (a) and clarified water (b). For explanation of the labels see Figure 4.2	114
Figure 4.4: TOC concentrations vs peak C intensity for raw (a) and clarified water (b). For explanation of the labels see Figure 4.2. Linear regression with 95 % prediction lines	115
Figure 4.5: TOC removal vs peak C intensity (FLU) removal and absorbance measured at 254 nm (UV-Vis) removal between raw and clarified water. TOC removal vs FLU removal for sites (a) and months (b). TOC removal vs UV-Vis removal for sites (c) and months (d). For explanation of the labels see Figure 4.2. Linear regression with 95 % prediction lines a, b, c, d	116- 118
Figure 4.6: Fluorescence OM properties vs TOC removal. Raw water peak C emission for sites (a) and months (b), clarified water peak C intensity for sites (c) and months (d), clarified water peak T2 intensity for sites (e) and months (f). For explanation of the labels see Figure 4.2. Linear regression with 95 % prediction lines	118- 121
Figure 5.1: Three layer feed-forward neural network consisting of input, hidden and output layer (a). Triangles indicate passive neurons, circles – active neurons. Active neuron with connection weights (arrows) that sums up the inputs (Σ) and transforms the the sum with a sigmoid function (ς) (b)	164
Figure 5.2: Self-organising map. Black dots indicate neurons	165
Figure 5.3: Total variance (%) and sum of variance (%) explained by the first ten PCA components	165
Figure 5.4: Loading plots for the first four principal components. PC1 (a), PC2 (b), PC3 (c), PC4 (d)	166- 167

Figure 5.5: Score plots for the first two principal components from PCA analysis of raw (a) and clarified water (b) fluorescence excitation-emission spectra. For explanation of the labels see Figure 4.2	168
Figure 5.6: Score plots for the second and third principal components from PCA analysis of raw (a) and clarified water (b) fluorescence excitation-emission spectra. For explanation of the labels see Figure 4.2	169
Figure 5.7: Three-component PARAFAC model excitation and emission loadings of raw, clarified and all data. Component 1: emission a), excitation b); Component 2: emission c), excitation d); Component 3: emission e), excitation f)	170- 172
Figure 5.8: Four-component PARAFAC model excitation and emission loadings of raw, clarified and all data. Component 1: emission a), excitation b); Component 2: emission c), excitation d); Component 3: emission e), excitation f); Component 4: emission g), excitation h)	173- 175
Figure 5.9: PARAFAC scores of raw and clarified water. a) peak C fluorescence (first component, X axis) and peak T2 fluorescence (third component, Y axis) – raw water, b)) peak C fluorescence (first component, X axis) and peak T2 fluorescence (third component, Y axis), c) peak C fluorescence (first component, X axis) and peak A fluorescence (second component, Y axis) – clarified water. For explanation of the labels see Figure 4.2	176- 177
Figure 5.10: PARAFAC scores of raw and clarified water for two contrasting WTWs Bamford (a) raw, b) clarified) and Church Wilne (c) raw, d) clarified)	178- 179
Figure 5.11: Visualisation of the SOM map for fluorescence data: U matrix (a), sample distribution and clusters (b), hit histogram (c). Notation used, e.g. S 1 and S 1c denotes raw and clarified water of site 1 – Bamford (sites numbers can be found in table 3.1)	179
Figure 5.12: Reference vector plots for selected SOM neurons with the highest number of hits. Neuron 1 (a), neuron 109 (b), neuron 9 (c) and neuron 120 (d)	180
Figure 5.13: Component planes for the excitation-emission wavelengths with fixed excitation at 280 nm. Emission wavelength of 300 (a), 350 (b), 400 (c), 450 (d), and 500 (e) nm	180
Figure 5.14: Hit histograms for raw (red) and clarified (green) water. Sites: 1 – Bamford, 2 – Campion Hills, 3 – Church Wilne, 4 – Cropston, 5 – Draycote, 6 – Frankley, 7 – Little Eaton, 8 – Melbourne, 9 – Mitcheldean, 10 – Mythe, 11 – Ogston, 12 – Shelton, 13 – Strensham, 14 – Tittesworth, 15 – Trimpley, 16 – Whitacre	181- 183
Figure 6.1: A relationship between coagulation pH and TOC removal efficiency (May 2007 – February 2008). For explanation of the labels see Figure 4.2	198

Figure 6.2: Champion Hills raw water spectral properties, peak C emission wavelength and peak T intensity vs TOC removal for all data August 06 – February 08 (black dots) and coagulation experiment data (red dots)	198
Figure 6.3: Raw, clarified, and final water EEMs for two coagulation optimisation episodes	199-201
Figure 6.4: TOC removal predicted (raw-clarified water fluorescence model) vs actual TOC removal. CH data August 06 – February 08 (red dots), coagulation experiment data (blue dots)	202
Figure 6.5: Organic matter removal vs pH during the coagulation experiment	202
Figure 6.6: TOC removal actual (black squares) and predicted from the fluorescence data of raw water (red triangles) vs coagulation pH during Champion Hills coagulation experiment	203
Figure 6.7: TOC removal actual (black squares) and predicted from the fluorescence data of raw and clarified water (red triangles) vs coagulation pH during Champion Hills coagulation experiment	203
Figure 6.8: Relative TOC removal vs coagulation pH through Champion Hills WTW. A blue line (July 07), a red dashed-line demonstrates hypothesized TOC removal for pH value of 5.5	204
Figure 7.1: Relative change in peak A intensity (%) through different WTWs	218
Figure 7.2: Relative change in peak C intensity (%) through different WTWs	218
Figure 7.3: Relative change in peak T intensity (%) through different WTWs	219
Figure 7.4: Excitation-emission matrices of raw (a), clarified (b), filtered (c), post-GAC water (d), pre-contact tank (e), post contact tank pre reducing agent addition (f), post contact tank post de-chlorination (g), and final water (h). Strensham WTW, 05.08.07	220-223
Figure 7.5: Location of EEM peaks and five EEM regions (after Chen et al., 2003)	224
Figure 7.6: SOM analysis of Strensham fluorescence data through WTW. The U matrix (a) and samples distribution (b): raw (F1), clarified (F2), filtered (F3), post-GAC water (F4), pre-contact tank (F5), post contact tank pre reducing agent addition (F6), post contact tank post reducing agent addition (F7), and final water (F8). Strensham WTW, 05.08.07	224
Figure 8.1: The relationship between clarified water peak T2 intensity and formation potential of dichlorobromomethane	243

LIST OF TABLES

Table 3.1: Summary of sampling sites: * - direct abstraction from river to WTW; Typical catchment land use, selected, types, of the largest percentage in total catchment area : A – non-irrigated arable land; P – pastures; C – other cultivated areas; U – urban fabric; I – industrial, transport or commercial units; G – green urban areas; F – forests; O – other areas	89
Table 3.2: Organic matter characteristics: mean TOC: > 6.0 mg/l indicates high TOC, mean TOC < 3.0 mg/l indicates low TOC; mean SUVA > 3.1 mg/l.m indicates hydrophobic OM (HPO), mean SUVA < 3.0 mg/l.m indicates hydrophilic OM (HPI)	90
Table 3.3: WTWs' performance parameters during the investigation period	91
Table 4.1: Kruskal-Wallis ANOVA test results. Mean ranks and test statistics (df – degrees of freedom, Asymp. Sig. – the asymptotic significance)	122
Table 4.2: Organic matter removal (%) calculated as a decrease in fluorescence intensity, UV-Vis absorbance and TOC concentrations for 16 WTWs. The highest values in red, the lowest values in blue	123
Table 4.3: Organic matter removal (%) calculated as a decrease in fluorescence intensity, UV-Vis absorbance and TOC concentrations for particular months and raw water peak C emission (nm). The highest values in red, the lowest values in blue. “–” indicates no measurements	124
Table 4.4: TOC removal prediction from raw and clarified water fluorescence and absorbance properties. Significant correlations with coefficients > 0.36 highlighted in red	125
Table 4.5: TOC removal prediction from different raw and clarified water fluorescence properties - sites. Significant correlations with coefficients > 0.36 highlighted in red	126
Table 4.6: TOC removal prediction from different raw and clarified water fluorescence properties - months. Significant correlations with coefficients > 0.36 highlighted in red	127
Table 5.1: Examples of application of multivariate techniques to EEMs analysis (CAL – calibration, CLAS - classification)	184
Table 5.2: Details of the fluorescence dataset used in the calibration analysis	185
Table 5.3: Variance explained (VARE (%)), residual sum of squares (SSQ) and CORCONDIA (COR (%)) result of three PARAFAC models for first seven PARAFAC components (C)	185

Table 5.4: Statistical details of PCA and PARAFAC models. VARE – variance explained (%), SSQ – residual sum of squares, COR – CORCONDIA (%)	186
Table 5.5: Characteristics of three PCA and PARAFAC components (C) identified for drinking water fluorescence dataset. Excitation and emission wavelengths maxima and identified fluorophores. Wavelength in brackets denotes secondary maximum	186
Table 5.6: Prediction accuracy and prediction error for selected decomposition and calibration methods: MLR – multiple linear regression, SR – stepwise regression, PLS – partial least squares analysis, BPNN – artificial neural network with back-propagation	186
Table 6.1: Required TOC removals by enhanced coagulation (%) for plants using conventional treatment (Step 1 of US EPA guidelines) (US EPA, 1999)	205
Table 6.2: TOC (mg/l) and SUVA (mg/l.m) of raw (R), clarified (C), and final water (F). Organic matter removal prediction (%) from fluorescence analysis (peak C int, raw water model Flu1, raw and clarified water model Flu2), TOC removal and UV-Vis absorbance reduction	205- 206
Table 6.3: Raw and clarified water properties for Campion Hills and all 16 water treatment sites (August 06 – February 08)	206
Table 6.4: Fluorescence peak emission (nm), excitation wavelength (nm) and intensity (au) for raw, clarified, and final water for baseline (18.11.08) and optimised (12.02.09) coagulation conditions	206
Table 7.1: Fluorescence Regional Integration parameters	225
Table 7.2: Relative distribution of normalized fluorescence volumes through Strensham WTW (05.08.07)	225
Table 8.1: OM properties of raw and clarified water and its removal (TOC removal %). HPO – hydrophobic OM, HPI - hydrophilic OM, Int. – intermediate OM character, Micr – microbially-derived OM, Stab. – stable OM properties, Var. – variable OM properties	244
Table 8.2: Mean disinfection by-products formation potentials ($\mu\text{g/l}$) in raw water	245
Table 8.3: Disinfection by-products ($\mu\text{g/l}$) formation potential prediction from raw and clarified water fluorescence and absorbance properties. Significant correlations with coefficients > 0.36 highlighted in red	246- 247

LIST OF ABBREVIATIONS

^{13}C NMR	^{13}C -nuclear magnetic resonance
Δ_{Cem}	change in mean peak C emission wavelength (nm)
$\Delta_{\text{T2int/Cint}}$	change in mean ratio of peak T2 intensity to peak C intensity
ANNs	artificial neural networks
BDOC	biodegradable dissolved organic carbon
BOD	biological oxygen demand
BPNN	back-propagation neural network
CDOM	chromophoric dissolved organic matter
CHA	charged hydrophilics
CORCONDIA	core consistency analysis
DAF	dissolved air flotation
DBPs	disinfection by-products
DCAA	dichloroacetic acid
DCAN	dichloroacetonitrile
DOC	dissolved organic carbon (mg/l)
DOM	dissolved organic matter
EEM	fluorescence excitation-emission matrix
FRI	fluorescence regional integration
GAC	granular activated carbon
HAAFP	formation potential of haloacetic acids
HAAs	haloacetic acids
HANs	haloacetonitriles
HBC	hopper bottomed clarifiers

HPI	hydrophilic organic matter
HPLC	high-performance liquid chromatography
HPO	hydrophobic organic matter
HPSEC	high-performance size exclusion chromatography
IHSS	International Humic Substances Society
MCR	multivariate curve resolution
MLR	multiple linear regression
NaN	Not A Number
NDMA	N-nitrosodimethylamine
NEU	neutral hydrophilics
NOM	natural organic matter
OM	organic matter
OMrem_{FLU}	organic matter removal derived from fluorescence measurements (%)
OMrem_{TOC}	organic matter removal derived from TOC measurements (%)
PARAFAC	parallel factor analysis
PC	principal component
PCA	principal components analysis
peak A	humic-like fluorescence (with absorption in UV range)
peak C	humic-like fluorescence (with absorption in visible range)
peak T	tryptophan-like fluorescence
PLS	partial least squares
RMSEP	root mean squared error of prediction
SHA	slightly hydrophobic acids
SOM	self-organizing map

SPLITT	split-flow thin-cell separation technique
SR	stepwise regression
SRFA	Suwannee River fulvic acid
SRHA	Suwannee River humic acid
SUVA	specific ultraviolet absorbance (mg/l.m)
T2_{int}/C_{int}	the ratio of peak T2 intensity to peak C intensity
THAAFP	total HAAFP (µg/l)
THMFP	trihalomethanes formation potential (µg/l)
THMs	trihalomethanes
TOC	total organic carbon (mg/l)
TOX	total organic halide (µg/l)
TTHM	total trihalomethanes
TTHMFP	total THMFP (µg/l)
US EPA	US Environmental Protection Agency
UV	ultraviolet
UV₂₅₄	ultraviolet absorbance measured at 254 nm
UV-Vis	ultraviolet-visible absorbance spectroscopy
VHA	very hydrophobic acids
WTWs	water treatment works

1. INTRODUCTION

1.1. BACKGROUND

Organic matter (OM) is a complex heterogeneous mixture of chemical compounds ubiquitous in all natural and anthropogenically-transformed environments, from marine deepwater, through estuarial to freshwater ecosystems. OM plays an important role in aquatic biochemical processes, controlling the availability of nutrients for living organisms and carbon transfer between different ecosystems. In drinking water treatment OM impairs the effectiveness of several treatment processes. During the water treatment process, raw water OM should be effectively removed prior to disinfection due to its propensity for forming toxic and carcinogenic disinfection by-products (DBPs) as a result of its chemical reaction with chlorine.

An increased focus has been placed on DBPs compliance in the UK following the tightening of the regulatory standard for trihalomethanes (THMs) from 100 µg/l measured on a three monthly rolling mean to an absolute 100 µg/l standard. It is considered likely that this standard will be further tightened to bring the UK more into line with other European nations and the US. Analytical techniques for the identification of THMs can take in excess of three days. In the face of tightening standards, there exists the need for a surrogate THM parameter which can be measured accurately and quickly at the water treatment works (WTWs) and which will give a satisfactory indication of the THMs concentration leaving the water treatment works. Total organic carbon (TOC) is commonly used as a measure of WTWs performance and, in particular, of THMs precursors (Carlson and Hardy, 1998; Golfinopoulos and Arhinditsis, 2002a,b; Hua, 2000). Turnaround periods for TOC samples can be of

a similar timescale to those for THMs. Thus, there is a need in drinking water treatment for a simple and rapid OM removal characterisation and THMs assessment technique.

The inherent complexity of organic matter structure and function determines the substantial number of different physical and chemical methods used in organic matter characterisation. Recent advances in spectrofluorometric techniques have concentrated on the development of more accurate, portable and faster instruments, enhanced optical analysis efficiency and have stimulated academic and industrial interest in utilization of intrinsic spectral properties of organic matter in characterisation of its composition and role in a variety of ecosystems (Mopper and Schultz, 1993; Coble, 1996; McKnight et al., 2001; Stedmon et al., 2003; Boehme et al., 2004; Hudson et al., 2007).

The fluorescence properties of OM can be useful in determining the structure and function of particular OM compounds. These properties control the chemical reactivity (i.e. propensity to complex specific organic and inorganic substances), and mobility of OM in aquatic environments. Furthermore, fluorescence spectroscopy can determine physico-chemical properties (molecular weight and degree of aromaticity) influencing OM reactivity and thus be indicative of the OM propensity for THMs formation in drinking water treatment.

1.2. SCOPE OF STUDY

In the work presented here, fluorescence spectroscopy was used for the quantitative and qualitative characterisation of OM during conventional water treatment. As fluorescence measurements are rapid and non-invasive with the possibility for incorporation into on-line monitoring system, fluorescence spectroscopy can provide an accurate assessment of organic matter removal efficiency during treatment processes and facilitate on-line prediction of DBPs formation potential. Moreover, organic matter characterisation can provide an insight into the

dependence of organic matter removal efficiency on the character of organic matter, based on the presence, relative importance and spectral properties of particular fluorophores in raw water. The purpose of this thesis is to assess the use and application of fluorescence spectroscopy at WTWs, to provide a rapid indication of OM / TOC removal and hence an assessment of THMs formation, as previous research in this field is limited (Goslan, 2003; Goslan et al., 2004).

1.3. PUBLICATIONS

This is the first work summarising the application of fluorescence spectroscopy to characterise organic matter properties and removal in the field of drinking water treatment. An important part of the work was application of advanced data mining techniques for robust fluorescence analysis. The novelty of the work resulted in publication of two papers in peer-review journals (Bierozza et al., 2009a, b) and presentation of two papers (Bierozza et al., 2009c, d) and two posters on national and international conferences (*Developments in Water Treatment and Supply*, Bath, July 07 and AGU Fall Meeting, San Francisco, December 07).

The work was also presented on:

- British Hydrological Society National Meeting, Surface Water Quality: Modelling, Monitoring and Management, May 08 (*Freshwater organic matter characterisation using fluorescence spectroscopy – implications for water quality studies and drinking water treatment*).
- AGU Chapman Conference, Organic Matter Fluorescence, Birmingham, October 08 (*Self-organizing Map Analysis of EEM Fluorescence Spectra of Raw and Partially-Treated Water as a Basis for the Determination of Organic Matter Removal Efficiency at Water Treatment Works*).

- British Hydrological Society, Peter Wolf Young Hydrologists Symposium, Liverpool, June 09 (*Assessing freshwater organic matter properties and removal at water treatment works with fluorescence spectroscopy and self-organizing maps*).
- Annual conference of the Operations Research Society, Warwick, September 09 (*Optimising water treatment works performance using data mining and artificial neural networks*).

1.4. THESIS STRUCTURE

The thesis comprises nine chapters. Following the Introduction (Chapter 1), in chapter 2, a literature review was presented that summarises the importance of OM characterisation in aquatic environments and problems related to OM removal in drinking water treatment. In particular, the following topics were discussed based on literature findings: structure and functions of OM in various environments (Section 2.1), evaluation of common OM characterisation techniques with special attention given to fluorescence spectroscopy (Section 2.2), relation between fluorescence and OM properties (Section 2.2.3), occurrence and assessment of OM and DBPs in drinking water treatment (Section 2.3). Finally, previous applications of fluorescence spectroscopy to water studies were presented (Section 2.4) and thesis aims and objectives were formulated (Section 2.5).

Chapter 3 presents the characterisation of sampling sites (Section 3.1) and discusses the implementation of all analytical techniques used in this study (Section 3.2). In addition, special attention was given to the pre-processing of fluorescence data (Section 3.3).

Results presentation was divided into four main chapters. In chapter 4, fluorescence data was analysed with simple statistical methods to derive characterisation of raw and partially-treated water samples and OM removal efficiency for the 16 WTWs analysed in this study.

The application of advanced data mining techniques for robust pattern recognition and calibration analysis of fluorescence data was shown in chapter 5. Chapter 6 shows the application of fluorescence spectroscopy for OM removal process optimisation, whereas chapter 7 demonstrates the ability of fluorescence spectroscopy to characterise unit process performance at water treatment works. The results obtained in this study are summarised and discussed in chapter 8, with emphasis put on the ability of fluorescence spectroscopy to characterise OM at WTWs (Section 8.1), the relation between fluorescence OM properties and DBPs formation (Section 8.2), comparison between standard OM characterisation techniques and fluorescence spectroscopy (Section 8.3), and efficacy of different fluorescence data analysis methods (Section 8.4). Finally, in chapter 9 the main thesis findings and results were summarised.

In addition, CD-rom containing series of appendices was prepared to present the original data (Appendix A), Matlab® and Java scripts used for fluorescence data pre-processing (Appendix B), and copies of all published papers and conference presentations (Appendix C).

2. LITERATURE REVIEW

2.1. ORGANIC MATTER IN FRESHWATER ECOSYSTEMS

Natural organic matter (NOM) comprises a complex mixture of heterogeneous chemical compounds in dissolved, colloidal, and particulate states, naturally occurring in many aquatic environments, i.e. soil water, groundwater, freshwater, estuarine and marine. The heterogeneous structure of NOM results from the multiple allochthonous and autochthonous sources contributing to the overall organic matter (OM) pool in the environment. NOM can originate from the decomposition of plants and dead particulate OM in the catchment (allochthonous sources), and can be produced *in situ* by living aquatic organisms (bacteria, phytoplankton). These two main pools of NOM have been found to have distinctively different structural and functional properties.

Allochthonous NOM comprises refractory, humified, high-molecular weight material and is a highly aromatic fraction, whereas autochthonous, microbially-derived NOM is more labile, low-molecular weight, and less aromatic (McKnight et al., 2001; Komada et al., 2002; Waiser and Robarts, 2004). The major constituents of allochthonous NOM are humic substances, that can be further subdivided into three groups of substances depending on the solubility at different pH (fulvic, humic, and hydrophilic acids) (Thurman, 1985). Humic substances in riverine ecosystems account for 40 to 70 % of the total dissolved organic matter (DOM) content measured as dissolved organic carbon (DOC) and the fulvic acid is their major constituent (Thurman, 1985; Mobed et al., 1996; McKnight et al., 2001). Aromatic moieties of humic substances contain several chemical groups (i.e. carboxylic and phenolic) and comprise compounds of different molecular sizes (100 – 100 000 Da) (Wu et al., 2003a; Kelton et al., 2007). Compared to humic acids, fulvic acids have lower molecular weights

(500 – 2000 Da) and lower content of highly conjugated aromatic structures (Senesi et al., 1991; Elkins and Nelson, 2001). Blaser et al. (1999) found that with the increasing residence time in the water, humic acids demonstrate reduction in the highly conjugated structures as a result of precipitation or decomposition to simple phenolics.

Autochthonous NOM comprises mainly amino acids that, along with carbohydrates, carboxylic acids, hydrocarbons, sterols, alcohols, ketones, ethers, pigments, and anthropogenic organic pollutants constitute the remaining NOM pool (Thurman, 1985; Thacker et al., 2005).

The amount and structure of OM present in the environment depends on the efficiency of biogeochemical processes in the catchment (primary production, biodegradation, mineralization, physicochemical interactions, photodegradation), which are controlled by the hydrological and climatic regimes, geology, land cover, and anthropogenic impacts. Thus, the NOM content measured as total organic carbon (TOC) or dissolved organic carbon (DOC) undergoes complex short- and long-term variations. Several studies have investigated the impact of environmental variability on NOM concentration and composition (Battin, 1998; Komada et al., 2002; Waiser and Robarts, 2004; Kelton et al., 2007; Fellman et al., 2008; Hernes et al., 2008).

Terrestrial NOM derived from soil and released from the biomass is transported with lotic ecosystems (streams and rivers), modified and retained within lentic ecosystems and soil, and finally merged with the NOM pool in the marine ecosystems. Therefore the location in the catchment, along the river continuum from source to sea, controls both the amount and character of the NOM (Figure 2.1). The relative significance of allochthonous NOM sources generally decreases along the length of lotic ecosystems and with increasing size of the lentic ecosystems (lakes, wetlands) (Thomas, 1997). Junk and Wantzen (2003) determined three

main riverine zones with different OM composition patterns (Figure 2.1). In the upland catchments the inputs of terrestrially-derived OM from the surrounding catchment are relatively higher compared to the lowland catchments, where the importance of channel and floodplain produced autochthonous OM is distinctively greater.

Moreover, a number of studies showed that DOM quantity measured as DOC along the lotic systems gradually decreases (Katsuyama and Ohte, 2002; Maurice et al., 2002; Wu et al., 2007; Baker et al., 2008; Battin et al., 2008), however the impact of flow regimes, land use, and anthropogenic pollution can obscure this relationship. For example, during periods of prolonged rainfall, wet soil conditions, and higher river flows, more allochthonous, aromatic NOM is transported in streams as the discharge pattern shifts from groundwater to soil pore water source (Maurice et al., 2002; Baker et al., 2008). Maurice et al. (2002) found that during low flow conditions, not only did the degree of aromaticity and molecular weight of NOM decrease, but DOC concentrations were also lower.

The highest DOC concentrations are predominant in urbanised catchments with evidence of pollution and in natural DOC-rich riparian wetlands (Jaffe et al., 2004; Waiser and Robarts, 2004; Baker et al., 2008).

Wilson and Xenopoulos (2009) investigated the effects of agricultural land use on NOM concentration and composition. They correlated increasing proportions of continuous cropland with decreasing structural complexity (lower molecular weights and lower aromaticity) of NOM and higher contribution of autochthonous fractions as nitrogen enrichment in soil promotes microbial activity.

NOM partitioning between allochthonous and autochthonous fractions in lentic ecosystems is highly dependent upon the size of the system, as hydrological factors and physicochemical properties of the catchment have greater influence on smaller lakes and wetlands. Within

lakes, DOM is an important link in the microbial loop, providing the source of energy for aquatic bacteria, limiting ultraviolet (UV) light penetration and oxygen concentrations (Thomas, 1997; Elkins and Nelson, 2001; Gondar et al., 2008).

OM plays important environmental roles in many ecological processes including light absorption and photochemical reactivity, proton binding, binding of heavy metals, binding of organic contaminants, adsorption at surfaces and aggregation, and finally represents an important link in global carbon and nitrogen biogeochemical cycles (Elkins and Nelson, 2001; White et al., 2003; Thacker et al., 2005). Complexation of metal ions to DOM controls the bioavailability of many toxic metals in the environment (Elkins and Nelson, 2001). NOM in lakes and wetlands attenuates both photosynthetically active and UV light, reducing the depth of autotrophic production and protecting aquatic organisms from harmful UV radiation (Gao and Zepp, 1998; Waiser and Robarts, 2004). Photochemical reactions involving DOM control the environmental fate of several chemical species (i.e. carbon, nitrogen and redox sensitive metals), including production of highly reactive transients (i.e. oxidants) (Amon and Benner, 1996; White et al., 2003). In particular, photo-Fenton reactions in iron-rich coastal waters were found to play an important role in DOM photodegradation (Gao and Zepp, 1998; White et al., 2003).

Recently, much scientific debate has focused on explanations for the long-term increase of dissolved organic carbon concentrations in aquatic ecosystems observed in many places in Europe and North America (Freeman et al., 2004; Evans et al., 2006; Monteith et al., 2007). Evans et al. (2006) linked increasing OM content in the UK's freshwater by an average of 91% since 1988 mainly with the depletion of sulphur deposition and reduced acidity. With the uncertain future rates and trends in environmental change induced by climate change, much research concentrates on the better understanding of the role of NOM in carbon storage and

transformation. Evans et al. (2006) suggested that changes related to the climate warming can increase the release of DOC from upland wetlands, and variation in rainfall and flow patterns can significantly affect the available OM pool. Anticipated future increases in air temperature and prolonged droughts can also change the predominant DOM character in freshwater to lower molecular weight, more aliphatic and hydrophilic fractions (Maurice et al., 2002). Increased levels of atmospheric carbon dioxide were hypothesized to increase the net primary productivity in wetlands, and thus to increase the leachate DOC concentrations in streams (Freeman et al., 2004).

2.2. ORGANIC MATTER CHARACTERISATION

2.2.1. Isolation and fractionation of organic matter

To understand the OM importance in the environment a facilitation of the adequate laboratory techniques of OM characterisation is required. However, the inherent complexity of OM impedes the application of simple characterisation tools, and necessitates the application of numerous analytical techniques for comprehensive OM characterisation.

Heterogeneous OM character complicates the identification and quantification of its structure and functionality. Differences in elemental composition, charge, and secondary and tertiary structure of various OM compounds influence their functionality, reactivity in the environment and treatability in drinking water and wastewater. Therefore, detailed information on the role of NOM in the environment is necessary and can be obtained from laboratory experiments with the fractions isolated from the overall NOM pool.

Several physico-chemical properties of NOM can be used to categorize and describe NOM in terms of its role in the environment and reactivity, e.g. molecular weight, degree of aromaticity, proton capacity, charge density (Gjessing et al., 1999; Thacker et al., 2005;

Henderson et al., 2008a). In practice, OM characterisation properties obtained for a given sample are compared with the reference material provided by the International Humic Substances Society (IHSS): Suwannee River fulvic acid (SRFA) and Suwannee River humic acid (SRHA) (Figure 2.2). In this way, the results of independent experiments on humic substances properties can be compared. The SRFA standard was for example used to evaluate the reproducibility of functional assays for DOM characterisation by Thacker et al. (2005).

Different isolation and fractionation techniques have been used to characterise NOM structure, functionality and reactivity in the environment, i.e. ion exchange resins, reverse osmosis, rotary evaporation, tangential flow ultrafiltration (Krasner et al., 1996; Gjessing et al., 1999; Patel-Sorrentino, 2002; Wu et al., 2003a; Waiser and Robarts, 2004). Basic NOM characterisation includes elemental characterisation of OM to quantify the DOC or TOC, nitrogen, phosphorus, stable and radio isotopes components, and molecular discrimination into several functional compounds, i.e. amino acids, proteins, carbohydrates or lipids (Yamashita and Tanoue, 2003). Elemental analysis of the humic substances revealed a typical weight composition comprising 52-56 % carbon, 35-40 % oxygen, 5 % hydrogen, 1-2 % nitrogen, and 1-2 % sulphur (Reckhow et al., 1990). The differences between humic and fulvic acids included higher content of nitrogen in the humic and higher content of oxygen in the fulvic fraction.

Resin fractionation has become the most common method to characterise OM. It involves the separation of OM into its humic/non-humic and hydrophobic/hydrophilic fractions (Leenheer 1981, Chow et al. 2004, Soh et al. 2008). In particular, the XAD-8 resin adsorption method was adopted in the isolation procedure of IHSS humic substances (Aiken et al, 1985). In this method, OM is fractionated into hydrophobic and hydrophilic isolates based on the preferential adsorption of the hydrophobic fraction on the XAD-8 resin. Both fulvic and

humic acids are eluted from the XAD-8 resin by alkaline extraction with aqueous NaOH (Malcolm and MacCarthy, 1992; Aiken et al., 1985). Subsequent low pH precipitation and desalting steps involving cation exchange, enable extraction of humic and fulvic acid respectively. Aromatic content determined with ^{13}C -nuclear magnetic resonance (^{13}C NMR) data confirmed higher partition of aromatic carbon in humic acids (30-35 %) compared to fulvic acids (14-19 %) (Reckhow et al., 1990).

Furthermore, the molecular size distribution of OM can be determined by high-performance size exclusion chromatography (HPSEC) (Huber and Frimmel, 1996; Her et al., 2002; Matilainen et al., 2002; Nagao et al., 2003). Mass spectrometry and chromatography (i.e. high-performance liquid chromatography (HPLC)) have been used in identification of individual OM compounds (Wu et al., 2003b; Yamashita and Tanoue, 2003; Kelton et al., 2007).

Several studies report comprehensive comparative NOM characterisation studies with the use of several laboratory techniques available (Abbt-Braun and Frimmel, 1999; Gjessing et al., 1999; Thacker et al., 2005; Gondar et al., 2008; Henderson et al., 2008a). Thacker et al. (2005) proposed a series of 11 standardised, reproducible assays characterising environmental functions of DOM in freshwaters: photochemical activity, role in the transport and bioavailability of toxic metals and hydrophobic organic contaminants (i.e. copper, benzo(a)pyrene), aggregation and sorption processes (i.e. adsorption to alumina). Gondar et al. (2008) used earlier work to discriminate between allochthonous (upper catchment stream water) and autochthonous DOM (lake water). It was found that the latter is less light-absorbing, less fluorescent, more hydrophilic, and has fewer proton-dissociating groups (Gondar et al., 2008). Abbt-Braun and Frimmel (1999) showed that several OM characterising parameters can exhibit significant changes with respect to the isolation technique used, i.e.

elemental analysis, spectral properties, proton capacity, complexation capacity, adsorption behaviour and disinfection by-product formation potential.

The applicability of the isolation techniques in the assessment of OM can be limited due to time-consuming preparation and laboratory analysis, distinctive alterations to water samples (i.e. pH, concentration) and question of the representativeness of the raw water properties by the obtained fractions (Gjessing et al., 1999; Hautala et al., 2000; Kalbitz et al., 2000; Marhaba et al., 2000; Kitis et al. 2001; Nagao et al., 2003; Baker and Spencer, 2004; Rosario-Ortiz et al. 2007). Analytical methods based on isotopic or molecular DOM characterisation require large sample volumes and complex preparation procedures that may lead to sample contamination and significant changes to the original organic material (Coble, 1996). Thus, the prepared isolates might not fully reflect the actual structure, functionality and reactivity of DOM present in the aquatic environment (Baker and Spencer, 2004). Blaser et al. (1999) compared the influence of different isolation techniques on the removal of particular compounds from the DOM pool. They found that reverse osmosis removed a small portion of fluorescent organic material compared to the evaporation technique. Gjessing et al. (1999) concluded that the OM loss during isolation with evaporation or reverse osmosis techniques can be significant (10-30 %). More recently, in the study of Song et al. (2009) the efficacy of the reverse osmosis and subsequent fractionation with resin adsorption chromatography was evaluated. Reverse osmosis efficacy was found to depend on both the OM characteristics (more efficient for aromatic OM) and pH (reduced OM recoveries under acidic conditions). Likewise, the efficacy of resin adsorption fractionation and the overall distribution of different OM fractions (the relative contribution of hydrophobic and hydrophilic OM) were influenced by the column operational parameters.

2.2.2. *Organic matter characterisation with spectroscopic techniques*

Optical methods can provide an alternative to isolation techniques in OM characterisation (Kalbitz et al., 2000; Kitis et al., 2001; Baker, 2002; Her et al., 2003; Jaffe et al., 2008). With the recent advances in spectroscopic techniques, ultraviolet-visible absorbance spectroscopy (UV-Vis) and fluorescence spectroscopy have become common OM characterisation tools, providing rapid, non-invasive and sensitive analysis of bulk OM properties.

Optical methods of OM characterisation are based on the measurement of several parameters describing absorption and emission of energy by OM molecules. When an OM molecule is exposed to an external source of light, a photon is absorbed by the molecule and the electron configuration changes. An electron from the ground state is promoted to an upper excited singlet state. The reverse process, in which the photon is emitted during transition of an electron from an excited energy level to the lower level i.e. ground state, is called luminescence (Lakowicz, 1999). The two types of luminescence are observed. The direct electron transition from the singlet state to lower energy level is called fluorescence, whereas the phosphorescence process involves an additional transition to a triplet state with electron spin change (Figure 2.3) (Lakowicz, 1999). Although fluorescence is the reverse of absorption, absorption (excitation) occurs at a shorter wavelength than the corresponding fluorescence emission due to energy loss in non-radiative decay (the Stoke's shift) (Senesi, 1990; Lakowicz, 1999; Hudson et al., 2007). Both absorption and fluorescence wavelengths are specific to a particular molecule. The inherent optical properties of bulk OM samples result from the superposition of the optical properties of individual compounds and the intermolecular interactions. Absorption exponentially decreases with increasing wavelength producing a characteristic long wavelength absorption tail (> 350 nm) (Twardowski et al., 2004; Helms et al., 2008). The emission maxima are observed to shift continuously to the red

with increasing excitation wavelength. Both absorption and emission demonstrate featureless spectra with no distinct bands. The origins of these intrinsic optical properties of humic substances have been investigated by several authors (del Vecchio and Blough, 2004; Boyle et al., 2009). The results indicate that the absorption and emission spectra result from the superposition of the particular spectra of independent chromophores and the electronic interactions between chromophores. Intermolecular donor-acceptor interactions i.e. between hydroxyl-aromatic donors and quinoid acceptors, were hypothesized to play an important role in the formation of the specific spectral properties of humic substances (del Vecchio and Blough, 2004).

The fraction of DOM that absorbs ultraviolet and visible light is often referred to as chromophoric or coloured DOM (CDOM), whereas the DOM fraction exhibiting fluorescence in both ultraviolet and visible range is described as fluorescent DOM (FDOM) (Helms et al., 2008). CDOM comprises 10-90 % of the total DOM pool and therefore constitutes a significant DOM fraction in aquatic ecosystems controlling the photochemical reactions of the surface water and nutrient and light availability for aquatic organisms (Thurman, 1985; Twardowski et al., 2004).

To extract information about DOM properties from absorption and fluorescence spectra, several spectral parameters can be defined. In the absorption spectra analysis, a common approach is to record the absorption values at particular wavelengths. Absorption at 254 nm (or sometimes at 272 nm) has been widely used as an indicator of aromaticity and humification (Weishaar et al., 2003; Helms et al., 2008). Likewise, absorption at 254 nm normalized to DOC and referred to as a specific UV absorbance (SUVA) was shown to correlate strongly with the percentage of DOM aromaticity determined by ^{13}C -NMR of humic isolates ($R^2 = 0.97$) (Weishaar et al., 2003). By fitting the absorption values to the exponential

function, a spectral slope parameter can be obtained (the exponent) describing the relative steepness of the spectra (Twardowski et al., 2004). The spectral slope is often used for characterisation of CDOM composition and molecular weight. Helms et al. (2008) used the absorption spectral slope ratio as an indicator of NOM from dissimilar water sources along the estuarine transect. The ratio of absorption spectral slopes of 275 – 295 nm and 350 – 400 nm, was correlated with molecular weight obtained from size-fractionation analyses.

Over the decades of research on the application of fluorescence spectroscopy to environmental studies, the technique of measuring the sample fluorescence has shifted from simple measurements of selected excitation or emission wavelengths, synchronous fluorescence scans to simultaneous collection of fluorescence data over the whole range of different excitation and emission wavelengths (Senesi et al., 1989; Miano and Senesi, 1992; Blaser et al., 1999; Kalbitz et al., 2000; Hudson et al., 2007; Kelton et al., 2007; Henderson et al., 2009). The latter technique is well known as excitation-emission fluorescence spectroscopy and is considered as the best available fluorescence technique (Hudson et al., 2007).

Fluorescence spectroscopy is becoming an increasingly popular method in NOM studies and can facilitate rapid, accurate and on-line OM prediction. NOM in surface and ground waters exhibits distinctive fluorescence properties as a result of absorption of high-energy photons by an OM molecule and re-emission of lower-energy photons at longer wavelengths. This inherent spectral property can be utilized in fingerprinting of NOM with fluorescence spectroscopy. The main applications of fluorescence spectroscopy include the characterisation of NOM composition and sources and determination of the general water quality. The method is non-invasive, rapid and accurate with potential for online monitoring (Baker, 2001; Her et al., 2003). No sample pre-treatment is required and lack of any significant perturbation in

NOM composition and structure are the main advantages of fluorescence over different isolation procedures. The degree to which NOM samples can be altered with the isolation techniques was investigated by Lombardi and Jardim (1999). They found that the C18 Sep-Pak extraction procedure resulted in a loss of 65 % of the fluorescence intensity scanned at 350 nm excitation wavelength and 34 % respectively at 450 nm excitation wavelength.

A fluorescence excitation-emission matrix (EEM) contains a substantial amount of information on OM composition and structure. To retrieve the most important fluorescence information, a number of methods can be utilized, from simple visual inspection and recording the peak maxima values to more advanced statistical methods. The simplest spectral parameter derived from an EEM for DOM characterisation is the wavelength-independent position of each of the fluorophores recorded as emission and excitation wavelengths of the fluorescence intensity maximum (Coble, 1996). The ratios of the fluorescence peaks were found to be useful in distinguishing between OM sources in rivers (Baker and Spencer, 2004), lakes (McKnight et al., 2001), marine ecosystems (Coble, 1996). In particular the fluorescence index has been used to indicate the degree of aromaticity (McKnight et al., 2001). It is defined as the ratio of fluorescence intensity at 450 nm emission wavelength to fluorescence intensity at 500 nm emission wavelength, both at 370 nm excitation wavelength. High fluorescence indices have been shown (1.7 – 2.0) to indicate microbially derived DOM (usually lower aromaticity), whereas lower indices (lower than 1.3) are typical of terrestrial sources (usually higher aromaticity).

Analysis of multi-dimensional fluorescence data is often coupled with the application of advanced statistical and computational techniques for exploration of the patterns within fluorescence spectra, for classification and calibration purposes. In particular, methods from multi-way analysis have become the most popular in modelling of fluorescence data, with

parallel factor analysis (PARAFAC) being the current state-of-the-art technique (Bro, 1997; Stedmon et al., 2003; Fellman et al., 2008). In the PARAFAC model fluorescence data are decomposed into a set of independent components reflecting the changes in composition and source material between water samples. Each component therefore represents a group of fluorophores of similar, specific fluorescence properties (Stedmon and Markager, 2005). Combined EEM-PARAFAC modelling is often used as a tool to trace DOM fractions in different environments and in DOM mixing studies (Fellman et al., 2008; Engelen et al., 2009).

Fluorescence excitation-emission measurements provide valuable information on the DOM composition that allows for ecological fingerprinting of aquatic samples. Each EEM scan demonstrates a specific combination of fluorescence intensities over a range of excitation and emission wavelengths, however in freshwater three main fluorescence regions are typically present: humic-, fulvic-, and protein-like fluorescence. Coble (1996) introduced an alternative classification of fluorescence regions, where humic-like fluorescence is subdivided into peak A and peak C, and protein-like fluorescence collectively denotes peak B (tyrosine-like) and peak T (tryptophan-like) (Figure 2.4). Peak A fluorescence maximum occurs at 237-260 nm excitation and 400-500 nm emission wavelength, whereas peak C comprises humic-like fluorescence absorbing in visible light range (300-370 nm excitation and 400-500 nm emission wavelength). Both protein-like regions present double peaks: peak B at 225-237 nm excitation and 309-321 nm emission; 275 nm excitation and 310 nm emission wavelength, peak T 275 nm excitation and 340 nm emission; 225-237 nm excitation and 340-381 nm emission wavelength (Hudson et al., 2007; Henderson et al., 2009). In highly coloured, peaty upland sources, peak C fluorescence can commonly be subdivided into two peaks: the first at 320-340 nm excitation and 410-430 nm emission wavelength and the second at 370-390 nm

excitation and 460-480 nm emission wavelength (Baker, 2001; Baker et al., 2008). In marine water an additional peak M (marine humic-like) is present at 290-310 nm excitation and 370-410 nm emission wavelength (Coble, 1996; Lombardi and Jardim, 1999). The blue shift towards lower emission wavelengths in the humic-like fluorescence (from peak M to peak C region) can be utilized in distinguishing marine from riverine water samples. Moreover, Komada et al. (2002) linked these optical changes with DOM ageing and humification.

Fluorescence spectroscopy can be successfully used to obtain information about biodegradability of DOM (Kalbitz et al., 2003; Wu et al., 2003a; Fellman et al., 2008), environmental roles of DOM, i.e. metal binding (Elkins and Nelson, 2001; Maurice et al., 2002; Baker et al., 2008), DOM sourcing and fingerprinting (Pullin and Cabaniss, 1997; del Castillo et al., 1999; Newson et al., 2001; Komada et al., 2002; Maurice et al. 2002; Stedmon et al., 2003; Baker and Spencer, 2004; Kelton et al., 2007). The inherent fluorescence properties of DOM fractions provide valuable information on the structure, conformation and heterogeneity of the DOM pool (Mobed et al., 1996).

In several studies, a combined application of several analytical techniques for comprehensive OM characterisation has been presented, i.e. absorption and spectrofluorometric techniques or HPSEC measurements with online fluorescence detection (del Castillo et al., 1999; Gjessing et al., 1999; Her et al., 2002; Nagao et al., 2003; Wu et al., 2003b; Fellman et al., 2008). Those comparative studies provide the opportunity to compare the efficacy of DOM characterisation techniques.

DOM characterisation via HPSEC coupled with fluorescence EEM spectroscopy provides presentation of fluorescence properties as a function of molecular weight (Her et al., 2003; Nagao et al., 2003; Wu et al., 2003a). However, research carried out by Wu et al. (2003a) suggested that HPSEC can underestimate the concentrations of humic- and protein-like

fluorescence and larger molecular weights due to their hydrophobic character and adsorption onto HPSEC column.

In summary, fluorescence spectroscopy is more sensitive compared to UV-Vis absorbance spectroscopy (10-1000 times) with the possibility of single-molecule detection (Kalbitz et al., 2000; Henderson et al., 2009). Fluorescence measurements are more robust than absorbance at low DOM concentrations (del Castillo et al., 1999). Furthermore, fluorescence spectroscopy provides discrimination between chromophoric OM absorbing at similar wavelengths (McKnight et al., 2001; Stedmon and Markager, 2005; Henderson et al., 2009).

2.2.3. Molecular and environmental effects on organic matter fluorescence

To understand the relationship between fluorophore composition and particular chemical, structural and functional properties of DOM, attempts have been made to correlate EEM fluorescence features with the results of several laboratory DOM analysis techniques, i.e. isolation techniques, HPSEC, ¹³C-NMR etc. (Stewart and Wetzel, 1980; Mobed et al., 1996; del Castillo et al., 1999; Kalbitz et al., 1999; Her et al., 2002; Nguyen et al., 2005; Baker et al., 2008; Hudson et al., 2008).

Peak C fluorescence intensity normalized to absorbance at 340 nm was found to be a good indicator of DOM molecular weight (Stewart and Wetzel, 1980; Wu et al., 2003a; Belzile and Guo, 2006). Wu et al. (2003a) investigated the optical properties of lake water DOM in relation to molecular size distribution determined with tangential flow ultrafiltration. They found that fluorescence/absorbance ratio increased with molecular weight indicating that fluorophores and chromophores are not evenly distributed over different molecular weights as previously reported by Stewart and Wetzel (1980).

Fluorescence emission wavelength was related to the degree of aromaticity and hydrophobicity (Kalbitz et al., 1999). The increase in the emission wavelength indicates an increase in the number of highly substituted aromatic nuclei and conjugated unsaturated systems (Senesi et al., 1991; Kalbitz et al., 1999; Baker et al., 2008). Lower emission wavelengths are indicative of simpler fluorophore structure and microbially-derived fulvic acids (Mobed et al., 1996; McKnight et al., 2001; Kelton et al., 2007). Additionally, higher emission wavelengths of both peak A and C were related by Senesi et al. (1991) to the greater proximity of fluorophores in higher molecular DOM that results in increased probability of internal quenching and lower fluorescence intensities. The lower peak C fluorescence intensities of higher molecular weight structure were explained by Stewart and Wetzel (1980) by the presence of a masking effect where hydrophobic structures are hidden within the macromolecular structure of hydrophilic groups and their fluorescence is quenched. Tryptophan-like fluorescence was correlated with both algal and microbially-derived organic material, thus was found to be a good surrogate of biological oxygen demand (BOD) (Determann et al., 1998; Nguyen et al., 2005; Elliot et al., 2006; Hudson et al., 2008; Henderson et al., 2008a,b).

Cumberland and Baker (2007) demonstrated that a relationship between fluorescence intensity and DOC (both gradient i.e. fluorescence intensity per gramme of carbon and the correlation coefficient) can be a useful discrimination tool of DOM sources in freshwater, i.e. river, wetland, spring, pond and sewage. The strongest correlation coefficients between fulvic-like or humic-like fluorescence and DOC were found for natural DOM source dominated samples, whereas the highest fluorescence intensities per gramme of carbon were observed for DOM-rich wetland samples. The sewage-impacted water samples showed generally poorer relationships between fluorescence and DOC.

More recently, Baker et al. (2008) compared fluorescence properties of lake and river samples in the UK with the functional properties of DOM described in more detail by Thacker et al. (2005) and including information on photochemical fading, buffering capacity, metal binding, hydrophilicity and adsorption to alumina. For the data normalized to DOC concentration, they found strong linear relationships between optical properties of DOM samples with benzo[a]pyrene binding, alumina adsorption, aromaticity and buffering capacity. In particular, both absorbance measured at 340 nm and peak C emission wavelength were good predictors of DOM aromaticity ($R^2 = 0.92$ and 0.93 respectively) and alumina adsorption ($R^2 = 0.90$ and 0.84 respectively).

Despite several decades of research, the chemical nature of fluorophores and their structure are not well described due to analytical difficulties (Wu et al., 2003a; Baker, 2005). Fluorescence properties of amino-, humic and fulvic acids have been investigated by several authors (Senesi et al., 1991; Coble, 1996; Mobed et al., 1996; Pullin and Cabaniss, 1997). Fluorescence EEM spectra reflect the contribution of individual fluorophores in the sample that results from chemical changes and mixing processes (del Casitllo et al., 1999). The difference in optical properties between particular DOM fractions have been elucidated with differences in chemical composition (presence of particular functional groups), structure (conformation and location within DOM pool), and degree of chemical conjugation (presence of highly conjugated systems vs. linear systems). Chemical changes can produce a hypsochromic shift in the fluorescence spectra, towards shorter wavelengths (blue shift) as a result of shortening of the conjugated systems, elimination of substituents that result in a more linear structure, and conformational transformations that increase the molecules rigidity (del Castillo et al., 1999). The bathochromic shift towards longer wavelengths (red

shift) can result from coiling of OM structure, increase in the extent of conjugated systems and therefore vibrational energy losses of the promoted electrons (del Castillo et al., 1999).

Observed differences in fluorescence spectra between humic and fulvic acids (red shift in both excitation and emission wavelengths of humic acids) reflect the predominance in humic acids of high molecular weight and highly conjugated fractions comprising electron-withdrawing, meta-directing functional groups, i.e. hydroxyl, distributed coumarins and xanthenes, quinines and Schiff-Base derivatives (Senesi, 1990; Senesi et al., 1991; Mobed et al., 1996; Blaser et al., 1999; Kalbitz et al., 1999). As a result of self-quenching present in highly substituted aromatic moieties of humic acids, the fluorescence intensities of fulvic acids derived from similar sources are distinctively higher (Mobed et al., 1996). In the structure of fulvic acids, the electron-donating functional groups (ortho- and para-directing) are more typical (i.e. carboxyl, hydroxycoumarin-like structures and variously substituted phenolics originating from lignin), resulting in shorter fluorescence maxima wavelengths (blue shift; Senesi, 1990). The difference in fluorescence of aromatic and aliphatic structures results from the difference in the emission energy. In aromatic structures with extended conjugated systems (π -electron systems) the energy difference between the ground and first singlet state is lower thus producing the red shift in emission wavelength (Senesi, 1990). Conversely, aliphatic, linear systems fluoresce at lower emission wavelengths.

Senesi and colleagues (1989, 1991) extensively investigated structural changes in molecule fluorescence including effects of the substitution with halogens and heavy metals. Carbonyl-containing substituents, hydroxide, alkoxide and amino groups were found to shift the emission to longer wavelengths (red shift). The effect of heavy atom, i.e. halogen, incorporation into fluorescent compounds varies depending on the location of the substitution. The internal heavy-atom effect reduces fluorescence intensity by the promotion of electron

spin conversions (intersystem crossing) leading to the delayed energy release of the phosphorescence (Senesi, 1990).

The location of amino acids within the DOM structure determines the availability of carbon and nitrogen to the aquatic organisms (Baker, 2005). Amino acids comprising protein-like fluorescence in DOM can be bound or absorbed to humic substances, present in living organisms or as free amino acids in a solution (Reynolds, 2003; Wu et al., 2003a; Baker, 2005; Elliot et al., 2006). Traditional techniques of amino acids determination, i.e. high performance liquid chromatography (HPLC), underestimate the amount of labile tryptophan fraction and the EEM technique alone provides scant structural information on the location of the observed tryptophan-like fluorescence in DOM pool (Reynolds, 2003). However, thermal quenching of fluorescence experiments with river, wastewater and IHSS standards samples revealed that in riverine samples tryptophan is more likely to be bound to the humic substances than in tryptophan standard, where free tryptophan molecules prevail (Baker, 2005). These observations were based on the assumption that the degree of thermal quenching of fluorescence depends on the exposure of tryptophan in DOM structure. Lead et al. (2006) used the split-flow thin-cell (SPLITT) separation technique to investigate molecular size distribution of fluorophores in freshwater samples. They found that tryptophan-like fluorescence comprises mainly larger, particulate DOM compounds, whereas both fulvic- and humic-like fluorescence cover a range of molecular sizes (dissolved, colloidal, particulate) which is in accordance with the findings of Wu et al. (2003a). Unlike tryptophan-like fluorescence, both peak A and C fluorophores were found to be highly dependent on the solution pH with fluorescence intensity increasing for higher pH values (Lead et al., 2006).

Yamashita and Tanoue (2003) examined the composition of amino acids from estuarine and marine environments with a combined EEM and HPLC approach. They correlated the actual

concentrations of tyrosine and tryptophan with tyrosine-like and tryptophan-like fluorescence respectively. Similar findings were reported by Reynolds (2003), who found that normalised fluorescence intensities in the tryptophan-like region correlated with the free tryptophan concentrations determined with HPLC ($R^2 = 0.99$). In the study of Yamashita and Tanoue (2003), a tryptophan-like peak was found only in estuarine waters, whereas tyrosine-like fluorescence was ubiquitous and observed also in marine samples. The authors concluded that tryptophan-like fluorescence is derived from the non-living high molecular mass DOM rather than excreted from living microorganisms (Yamashita and Tanoue, 2003).

More recently, Spencer et al. (in press) found that absorbance measurements at 355 nm can provide a good indication of lignin phenol degradation. Lignin phenol was found to be a biomarker for highly photochemically reactive terrigenous OM.

Fellman et al. (2008) used SUVA (absorbance measured at 254 nm) and fluorescence EEM data modelled with parallel factor analysis (PARAFAC) to investigate the chemical properties of DOM in soil water from different wetland types. From the PARAFAC modelling it was demonstrated that a significant amount of protein-like fluorescence is biodegradable ($R^2 = 0.82$).

Although the molecular structure is of the greatest importance in determining the location and efficiency of fluorescence maxima, the impact of environmental factors, e.g. temperature, metal ions, pH and light exposure, can significantly alter the molecular structure and therefore the fluorescence spectra. Environmental factors can change the rates of individual processes involving excited molecular states or affect the energy levels involved in fluorescence and other competitive processes (Senesi, 1990).

In general, increasing temperature increases the likelihood of non-radiative energy decay within the molecule which will result in a reduction of the observed fluorescence intensity

(Senesi, 1990; Baker, 2005). Baker (2005) investigated the effect of thermal quenching on fluorescence intensity over a range of temperatures (10-45° C) for freshwater and wastewater samples. It was found that tryptophan-like fluorescence is more sensitive to the temperature changes (in the range between 20 ± 4 and 35 ± 5 %) compared with the fulvic- and humic-like fluorescence regions (19 ± 4 to 26 ± 3 %) (Baker, 2005; Sereczynska-Sobecka et al., 2007).

Similar to thermal quenching, the fluorescence signal can be substantially diminished by the presence of metal ions, such as Fe, Cu, Zn, Pb that can form strong complexes with fulvic acids compounds over a range of pH values (Blaser et al., 1999; Reynolds and Ahmad, 1995; Kelton et al., 2007; Henderson et al., 2009) (Figure 2.5). Acidic functional groups, i.e. the phthalates and the salicylates, within DOM structure were found to be the main target for metal complexation (Elkins and Nelson, 2001). Unlike other metal ions, aluminium ions with NOM produce fluorescent complexes leading to more profound changes in fluorescence excitation-emission spectra (Blaser et al., 1999; Sharpless and McGown, 1999; Elkins and Nelson, 2001). The degree of fluorescence quenching depends on the type of fluorophore, metal ion concentration and speciation, and pH (Henderson et al., 2009). However, the published results are tentative and reflect metal ion-NOM reactions under laboratory conditions that can be substantially different from the naturally occurring interactions (Reynolds and Ahmad, 1995; Henderson et al., 2009).

An increase in pH results in increased fluorescence intensities in both fulvic- and humic-like fluorescence regions, with no significant changes in tryptophan-like fluorescence (Patel-Sorrentino et al., 2002; Lead et al., 2006). Those changes were found to be independent of molecular size. pH effects on fluorescence can result from the changes in the molecular orbits of excitable electrons, changes in molecular shape and charge density, and finally from the combined pH effects on the metal ions solubility (Patel-Sorrentino et al., 2002). The degree of

fluorescence changes induced by pH variation is higher for peak C and reflects higher deprotonation rates of functional groups comprising fulvic-like compared to humic-like fluorophore. However, contradictory results were presented by Patel-Sorrentino et al. (2002) who found that peak A is more sensitive to pH changes and the increase in fluorescence intensity with increasing pH is higher than that of peak C. Different pH generates conformational changes in the structure of fluorophores reflected by the changes in the exposure or isolation of functional groups (Mobed et al., 1996). As a result, changes in fluorescence emission wavelength can be observed, i.e. red shift observed for humic acid. Ghosh and Schnitzer (1980) observed structural changes in humic acids with increasing pH. At higher pH simpler, more linear structures are predominant producing an increase in fluorescence intensity compared to more conjugated structures at lower pH where a self-quenching effect is observed as a result of masking some fluorophores inside conjugated structures (Stewart and Wetzel, 1980; Senesi, 1990; Patel-Sorrentino et al., 2002). Similar to thermal changes, pH effects on fluorescence in range from 2 to 12 are reversible, indicating that major fluorophores structure is not altered by temperature and pH (Patel-Sorrentino et al., 2002; Henderson et al., 2009).

Irradiation of DOM samples by UV-Vis light results in a reduction in the peak fluorescence intensity and both excitation and emission wavelengths that can reflect the shift towards lower molecular weights and a simpler, less conjugated structure of fluorophores (Amon and Benner, 1996; Kelton et al., 2007). The light-induced changes can include loss of colour (photobleaching), reduction of aromatic constituents, and shift towards lower molecular weight fractions (Waiser and Robarts, 2004). Amon and Benner (1996) identified two main pathways of DOM photochemical transformations, direct and indirect changes. Direct changes were demonstrated by compounds directly excited by the light absorption leading to

photoisomerization, electron transfer, intramolecular decomposition and rearrangement (Zika, 1981). Indirect reactions result in formation of highly reactive species, i.e. free hydroxyl radicals (White et al., 2003). Those results suggest that light exposure plays an important role in DOM degradation and can significantly affect optical properties of aquatic samples (Wu et al., 2003a; Waiser and Robarts, 2004). Furthermore, Patel-Sorrentino et al. (2004) observed changes in fluorescence after irradiation that followed a three-step pattern: sharp decrease in intensity within first few minutes, followed by increase and final decrease. The authors elucidated those changes with the initial photodegradation of the weakest fluorescent sites (decrease in fluorescence intensity) followed by conformation transformation from spherocolloidal to linear structure and activation of previously inhibited fluorophores (i.e. an increase in fluorescence intensity) (Patel-Sorrentino et al., 2004). Prolonged irradiation results in further degradation of fluorophores (further decrease in fluorescence intensity). The sensitivity of fluorescence in the identification of DOM sources is highly dependent on the intensity and depth of light penetration in the euphotic zone.

For samples with high absorbance (i.e. high content of humic substances) a red shift of fluorescence emission spectra is observed (inner-filter effect). Mobed et al. (1996) suggests that absorbance correction is needed for highly concentrated humic substances samples in order to enable comparison between fluorescence spectra. However, for freshwater samples with TOC concentrations of well below 25 ppm the inner filter effect is negligible (Hudson et al., 2008).

2.3. ORGANIC MATTER IN DRINKING WATER TREATMENT

2.3.1. Formation of disinfection by-products

NOM is commonly present in freshwater ecosystems used by water treatment companies to provide clean and reliable sources of drinking water. However, during drinking water treatment, OM has to be effectively removed in the early stages of the treatment process as NOM interferes with most of the treatment processes. NOM can be responsible for unpleasant odour and taste in water, fouling of filtration membranes, formation of disinfection by-products (DBPs) upon OM reaction with disinfectant, increased disinfectant demands and microbial re-growth in the distribution system (Widrig et al., 1996; Gjessing et al., 1999; Fabris et al., 2008). Therefore, the presence of OM in drinking water treatment can increase the operational costs of water treatment works (WTWs), reduce the useful life of treatment assets, impede the delivery of desired levels of service and compliance with existing water quality regulations, and finally can produce a serious health-risk to their customers. Thus, an ongoing focus of water treatment research has been on improving the OM removal efficiency (i.e. enhanced coagulation, UV irradiation, membrane filtration) and the development of novel techniques of NOM characterisation.

In the studies of OM in drinking water treatment, special attention has been given to the formation of DBPs in drinking water as some of the chlorinated and brominated species have been reported to demonstrate adverse health effects (US EPA, 2006). DBPs are formed in drinking water when the residual OM (not removed in the initial treatment stages) chemically reacts with the disinfectant, usually chlorine (Figure 2.6). As a result, different functional groups within OM undergo complex chemical transformations and several, both volatile and stable chemical species are formed. The propensity for DBPs formation as a result of OM reaction with disinfectant was first reported by Rook (1974) and Bellar (1974). Since then, extensive research has been conducted to identify the factors influencing DBPs formation, OM reactivity with disinfectants, and the effects of different DBPs on human health.

Disinfection-by products comprise several chemical species, including trihalomethanes (THMs, i.e. chloroform CHCl_3 , dichlorobromomethane CHCl_2Br , dibromochloromethane CHBr_2Cl , bromoform CHBr_3), haloacetic acids (HAAs) and haloacetonitriles (HANs) (Rook, 1974; Krasner et al., 1989; Nikolaou et al., 1999). The recognized adverse health effects of THMs have been reported by many authors and resulted in tightening of regulatory standards for THMs to 100 $\mu\text{g/l}$ in the UK (absolute standard) and 80 $\mu\text{g/l}$ in the US (based on a running annual average) (Dunnick and Melnick, 1993; Moudgal et al., 2000; van Leeuwen et al., 2005; US EPA, 2006; Jackson et al., 2008). Chloroform was found to comprise the preponderance of THMs (up to 50%) (Reckhow and Singer, 1990). The current Disinfectants/DBPs rule regulates four of the THMs (chloroform, dichlorobromomethane, dibromochloromethane, bromoform) and five HAAs (monochloro- ClAA , monobromo- BrAA , dichloro- Cl_2AA , dibromo- Br_2AA , and trichloroacetic acid Cl_3AA , the US Environmental Protection Agency (US EPA) limit on selected HAAs is 60 $\mu\text{g/l}$) (US EPA, 2006). However, the number of the identified DBPs species exceeds 600 including compounds of various toxicity and reactivity (Bond et al., 2009).

DBPs formation is controlled by multiple factors including water quality parameters (e.g. type and concentration of organic precursor, background concentration of bromide, temperature) and treatment conditions (e.g. coagulant type and dose, coagulation pH, disinfectant type, disinfectant dosage and contact time, residual disinfectant concentration, ratio of disinfectant concentration to TOC, pH, point of treatment process at which chlorine is added) (Engerholm and Amy, 1983; Carlson and Hardy, 1998; Nikolaou et al., 1999; Golfopoulos and Arhonditsis, 2002a,b; Liang and Singer, 2003; Roccaro et al., 2008; Francis et al., 2009). Thus, traditional methods of DBPs formation control rely mainly on OM precursor removal prior to the disinfection stage (Hubel and Edzwald, 1987; Randtke, 1988;

Rizzo et al., 2004) and, the use of alternative disinfectants (i.e. monochloramine, Carlson and Hardy, 1998; Chen and Valentine, 2007). Chlorine is often added at the pre-filtration stage as an oxidant to aid precipitation of metal ions. Thus, the postponement of the chlorine addition point in the treatment processes sequence has also been proposed as one of the methods directly mitigating the harmful effects of DBPs as fewer DBPs are formed when the chlorine is added later (Singer, 1994; Lliang and Singer, 2003).

The DBPs formation mechanism involves a series of complex chemical reactions highly dependent on the OM structure (Rook, 1974; Kavanaugh et al., 1980; Engerholm and Amy, 1983; Carlson and Hardy, 1998; Li et al., 2000a). The relationship between DBPs formation and the reaction time is an exponential function with rapid initial DBPs formation followed by slower DBPs production as both the OM and disinfectant concentrations become depleted with time (Lliang and Singer, 2003). The observed dual reactivity of OM with chlorine was elucidated by Gang et al. (2002) with the presence of two OM functionalities of different chlorine demand. The high molecular weight, OM structures containing carbonyl, hydroxyl and amine functional groups were found to be the sites preferentially attacked by chlorine molecules (Reckhow et al., 1990; Senesi, 1990; Gallard and von Gunten, 2002). Therefore, this OM fraction demonstrates rapid rate of chlorine consumption, whereas less reactive OM structures (i.e. containing methyl group) are responsible for slow and prolonged chlorine demand (Gang et al., 2002). The fast reacting THMs precursors comprise resorcinol-like structures (meta-dihydroxybenzene structures) and β -diketones, whereas other phenolic structures contribute to slower reacting THMs precursors (Korshin et al., 1999; Gallard and von Gunten, 2002; Chang et al., 2006; Bond et al., 2009). During the primary chlorination reactions either substitution or oxidation can occur and the subsequent reaction pathway heavily depends on the type, amount, and reactivity of intermediate chlorination species

produced. However, Wu et al. (2003b) suggested that the direct chlorine incorporation into OM molecule structure is less common. Reckhow et al. (1990) suggested that the chlorination mechanism depends on the degree of protonation of the reacting species, as the increase in pH shifts the reaction pathway from substitution to oxidation. For alkaline pH conditions more THMs were found to be produced, whereas under acidic conditions both THMs and HAAs can be formed depending on the presence in the OM pool of the readily oxidizable functional groups (Liang and Singer, 2003). Several authors corroborated the preferential formation of brominated DBPs and higher total DBPs yields in waters containing higher levels of bromide (Chang et al., 2001; Clark et al., 2001; Ates et al., 2007b).

The different reactivity of particular OM fractions with disinfectant results from the inherent heterogeneity of OM. Thus, the DBPs formation upon reaction with a disinfectant depends on both physical (e.g. molecular weight, degree of hydrophobicity) and chemical (elemental composition, functional groups, conformation) properties of OM. A number of studies attempted to correlate various OM properties to DBPs formation. Reckhow et al. (1990) conducted a comprehensive study of the THMs formation for different humic substances, determining the properties mostly affecting the formation of DBPs. They found a significant relationship between THMs formation, chlorine consumption and physico-chemical properties of humic and fulvic acids. Aromatic content, molecular size, and the presence of carboxyl and phenolic functional groups (i.e. hydroxyl-, carboxy-, and methoxy-substituted aromatic structures) were found to have the greatest importance in determining OM reactivity with disinfectant. Aromatic moieties of the humic substances containing phenolic functional groups have been reported to be highly reactive with chlorine in producing DBPs (Reckhow et al. 1990; Korshin et al. 1999; Liang and Singer, 2003; Wu et al. 2003b; Soh et al. 2008). In experimental studies, strong linear relationships between chlorine consumption and aromatic

carbon content, phenolic group content and activated aromatic rings content were commonly observed. Furthermore, humic acids were shown to produce higher amounts of DBPs (measured as total organic halide (TOX)) and consume more chlorine than fulvic acids (Reckhow et al., 1990). Different fractions of the humic substances were found to form different types of DBPs. In particular, the hydrophobic fraction is a more significant HAA precursor, whereas the hydrophilic fraction containing more aliphatic moieties, plays a major role in THMs formation (Stevens et al., 1976; Reckhow et al., 1990; Gang et al., 2002; Lliang and Singer, 2003; Parsons et al., 2004). However, the total yield of THMs and HAAs is always higher for hydrophobic OM than for the corresponding hydrophilic fractions (Lliang and Singer, 2003). In the research of Bond et al. (2009) the reactivity of various OM compounds was characterized in terms of chlorine demand, halogen substitution efficiency and THMs formation potential. The activated aromatic compounds including β -dicarbonyl species and amino acids were shown to be the most important DBPs precursors (Gallard and von Gunten, 2002; Bond et al., 2009). The combination of halogenation parameters, chlorine demand and halogen substitution efficiency, was found to be a useful predictor of OM reactivity with chlorine (Bond et al., 2009). The activated aromatic compounds exhibiting the greatest propensity for THMs formation (i.e. ferulic acid, L-tryptophan, resorcinol) showed the highest levels of chlorine demand and substitution efficiency.

Similar to aromatic structure content, molecular weight was reported to be an important factor in DBPs formation (Gang et al., 2003). Gang et al. (2003) found that the total THM (TTHM) yield coefficients (the ratio of the concentration of TTHM formed to the concentration of chlorine consumed) decreased with increasing molecular weight. Thus, higher THMs formation rates are observed from smaller molecular weight OM, as a result of easier decomposition of the low molecular weight halogenated intermediates. Furthermore,

the molecular size distribution of OM determined by HPSEC analysis has been related to OM removal efficiency and DBPs formation potential (Her et al., 2002).

Overall, higher reactivity of humic acids with disinfectant in DBPs formation does not pose a problem to their removal during the conventional treatment by coagulation. The hydrophobic and high molecular weight fraction of humic OM is preferentially removed, significantly reducing the DBPs formation potential prior to disinfection with chlorine (Fearing et al., 2004; Kim and Yu, 2005; Roe et al., 2007). The opposite reactivity and removal efficiency demonstrates the hydrophilic OM fraction, which comprises lower molecular weight material of carbohydrates, proteins and amino acids. The hydrophilic fraction is in general recalcitrant to coagulation removal and hence if not effectively removed in other treatment process, significantly contributes to formation of DBPs (Liang and Singer, 2003; Kim and Yu 2005; Soh et al. 2008). The relative proportion of both OM fractions is an inherent property of raw drinking water, and results from a variety of biogeochemical processes in the catchment. The hydrophobic fraction comprises recalcitrant, allochthonous organic material originates from soils or decomposition of higher plants, whereas the majority of the hydrophilic fraction is derived from autochthonous primary production in aquatic ecosystem (algae, bacteria). Significant amounts of the hydrophilic fraction originate from the domestic wastes and agricultural production. The relative importance of different NOM fractions undergoes seasonal variations that can impact upon OM removal efficiency and concentrations of the formed DBPs, however no consistent patterns were discerned (Parsons et al., 2004; Ates et al., 2007a; Baytak et al., 2008).

Several attempts have been made to apply kinetic models of THM formation, calibrated and validated on site-specific basis. Several studies have employed empirical models derived from the basic stoichiometric relationships relating THM formation to TOC levels (Kavanaugh et

al., 1980) or linear and nonlinear regression analysis incorporating different set of THM formation predictors (Engerholm and Amy, 1983; Amy et al., 1987, 1998; Clark et al., 2001; Golfinopoulos and Arhonditsis, 2002a,b; Sohn et al., 2004; Uyak et al., 2007). The main advantage of such models is the determination of the maximum allowable TOC level that still complies with the THMs standard, in conjunction with the quantification of operational treatment parameters. However, the complexity of chlorine/NOM reactions and the equivocal character of the rate constants for those reactions necessitates site-specific recalibration.

The trihalomethanes formation potential (THMFP) is a common measure of OM propensity for THMs formation under laboratory conditions (HAAFP is the formation potential of HAAs). The THMFP determination includes the preparation of the samples with added chlorine and measurement of the THMs concentrations in samples with gas chromatography followed by electron capture detection or mass spectrometry detection (Hu et al., 1999; van Leeuwen et al., 2005; Ates et al., 2007b; Pérez Pavón et al., 2008). Water sample is buffered to pH 7 and spiked to approximately 5mg/l free Cl_2 prior to storage for 7 days. Then samples are rechecked on the third day and extra free Cl_2 is added if necessary. Portions of sample are then transferred to a septum vial to reach equilibrium with its headspace at 80 °C. Finally, a sample of vapour is injected by auto-sampler on to a gas chromatography capillary column fitted with electron capture detector, to determine quantitatively THMFP present in the sample. To achieve DBPs levels detectable by the analytical methods, a pre-concentration of the analytes is required. However, the method is labour-intensive and time-consuming and hence attempts have been made to facilitate other, more rapid techniques of THMs formation prediction.

WTW performance with respect to OM removal is usually monitored with the use of routinely measured parameters, such as TOC concentration and different UV-Vis absorbance.

Those analytical methods provide high instrumental sensitivity, minimal, non-destructive sample pre-treatment and water sample analysis under ambient conditions of pH, ionic strength and concentration.

The simplicity of TOC and UV-Vis absorbance methods results in their common utilization in drinking water and wastewater treatment for performance assessment and process monitoring (Korshin et al., 1997; Kitis et al., 2001; Matilainen et al., 2002; Weishaar et al., 2003; Fearing et al., 2004). In particular, TOC and UV absorbance are used for the quantification of the overall OM concentration and estimation of OM reactivity in DBPs formation (Gang et al., 2002). UV-Vis absorbance has been reported to be more sensitive and accurate surrogate parameter for predicting the OM reactivity with disinfectant than TOC (Reckhow et al., 1990; Najm et al., 1994; Marhaba et al., 2000; Kitis et al., 2001; Roccaro and Vagliasindi, 2009). A relationship between absorbance measured at 254 nm (or 272 nm) and DBPs formation potential is well established in the literature (Edzwald et al., 1985; Reckhow et al., 1990; Wu et al. 2000; Roccaro and Vagliasindi, 2009).

Differential UV spectroscopy has been utilized in the characterisation of the reaction between OM and disinfectant and correlated with TOX (Korshin et al., 1997; Li et al., 2000a; Wu et al., 2003b; Roccaro et al., 2008; Roccaro and Vagliasindi, 2009). The decrease in UV absorbance at 280 nm before and after chlorination was found to correlate with both THMs and HAAs formations (Kitis et al., 2000). Roccaro et al. (2008) used differential absorbance at 272 nm as a predictive tool for DBPs formation and chlorine consumption.

Weishaar et al. (2003) proposed specific ultraviolet absorbance (SUVA, the absorbance at 254 nm per unit DOC in mg/l) as an indicator of aromatic carbon content. SUVA can be useful as an indirect rapid measure of OM removal efficiency coagulation. Waters of high

SUVA value ($> 4 \text{ mg/l.m}$) were found to comprise highly reactive and mainly hydrophobic in nature OM fraction (Edzwald and Tobiason, 1999).

Some authors proposed simple mechanistic models for THMFP prediction based on the chlorine decay (Clark, 1998; Gang et al., 2002). However, the DBPs formation prediction from the chlorine demand is site-specific and only valid for the specific treatment conditions used in the model calibration.

More recently, Korshin et al. (2009) demonstrated the use of the absorbance slope index for OM reactivity characterisation. The absorbance slope index was defined as the ratio of the relative contribution of activated aromatic groups (absorbance wavelength $> 250 \text{ nm}$) to the content of both carboxylic and aromatic groups at lower absorbance wavelengths (200 – 230 nm). The index was shown to characterise OM reactivity in drinking water treatment and was correlated with both SUVA and THMFP on site-specific basis.

Although both TOC and UV-Vis absorbance provide invaluable information on OM content, character and reactivity with disinfectant, several disadvantages with those methods have been reported. The main limitation of absorbance spectroscopy in OM characterisation includes the interference from UV absorbing compounds commonly present in treated water. The presence of turbidity and inorganic substances e.g. nitrate nitrogen, can increase the observed absorbance and affect the accuracy of OM determination (Eaton, 1995; Wang and Hsieh; 2001). Because only a fraction of TOC and UV-Vis is reactive with the disinfectant, those parameters can not completely represent OM propensity for DBPs formation (Gang et al., 2002). Thus, no correlation was found between TOC removal by coagulation and THMFP reduction as the latter is highly dependent on OM character rather than the OM quantity (Gang et al., 2002). Moreover, the relationship between UV absorbance surrogates and DBPs formation is dependent on the site-specific OM properties and impedes the direct comparison

of the results obtained for various locations. The critique of this method is also found in literature (Weishaar et al. 2003, Ates et al. 2007b).

Edzwald et al. (1985) successfully tested the use of UV (254 nm) absorbance as a surrogate parameter for THMFP determination. However, contradictory results were obtained by several authors (Parsons et al., 2004; Ates et al., 2007b) who postulated that the aforementioned relationships might be water-specific and only valid for particular treatment works. They found that for low SUVA waters (< 3 mg/l.m in the study of Parsons et al. (2004) and < 2.3 mg/l.m in the study of Ates et al. (2007b)) there was no observable relationship between any of the UV-Vis-derived parameters (UV₂₅₄ nm, SUVA, differential absorbance at 254 nm) and THMFP. Similar results were found by Chen and Valentine (2007) indicating that the relationship between SUVA and N-nitrosodimethylamine formation potential is dependent on the specific OM composition. N-nitrosodimethylamine (NDMA) is disinfection by-product formed when OM is disinfected with monochloramine. Korshin et al. (1997) suggested that chlorination may result in destruction of electron-rich sites within NOM (activated aromatic sites, conjugated double bonds) absorbing UV light effectively at 254-280 nm wavelengths. Thus, the utilization of the absorbance spectroscopy *per se* can significantly underestimate the THMFP prediction as the recalcitrant OM fractions that do not absorb at 254 nm pose the greatest problems in DBPs removal.

While SUVA was found to be a good overall predictor of aromaticity, it does not provide indication of DOC reactivity with chlorine (Weishaar et al., 2003). Moreover, SUVA measurements are sensitive to sample pH, nitrate and iron concentration changes, which precludes wide use of this parameter in drinking water treatment and direct comparison of samples aromaticity (Weishaar et al. 2003, Ates et al. 2007b). Ates et al. (2007b) suggested

that SUVA as a bulk surrogate parameter of OM reactivity may not provide a good indication of OM moieties highly reactive with chlorine.

2.3.2. *Organic matter removal in drinking water treatment*

In conventional drinking water treatment, OM removal is achieved by coagulation, flocculation, clarification and filtration processes. The efficiency of OM removal by coagulation is significantly dependent on the characteristics of OM. The low molecular weight, hydrophilic OM is recalcitrant to removal by coagulation process OM fraction. Therefore additional treatment techniques have to be facilitated for raw waters containing highly hydrophilic OM, i.e. granular activated carbon (GAC), ion exchange, membranes, and advanced oxidation processes. Below, the most common OM removal techniques were evaluated with particular emphasis on the most commonly utilized coagulation process.

Coagulation with inorganic metal salts is a conventional treatment process removing turbidity, NOM and other soluble organic and inorganic pollutants (van Leeuwen et al., 2005). The overall coagulation process comprises dosing and rapid mixing of certain chemical additives (coagulants, usually trivalent metal salts), physical and chemical reactions of NOM with the coagulant, slow mixing to facilitate particle agglomeration and floc formation (flocculation), followed by downstream solid-liquid separation of OM and reaction precipitates. Depending on the structure and composition of the OM and coagulation conditions (coagulant dosage and pH, water temperature and alkalinity), the OM removal by coagulation involves three major mechanisms: charge neutralisation; precipitation and co-precipitation (Randtke, 1988; Edzwald and Tobiasson, 1999; Uyak and Toroz, 2006).

The colloidal fraction (particulate OM) carries a negative surface charge as a result of ionisation of functional groups (e.g. carboxylic) on the surface of the solid and the adsorption

of ions or other charged species such as polymers onto the colloid surface (Parsons and Jefferson, 2006). Therefore the removal of the colloidal OM by coagulation requires electrical destabilisation of the particles, which can be achieved by electrical double layer compression; adsorption and charge neutralisation; adsorption and bridging and enmeshment in a precipitate (Randtke, 1988; Gregory, 1997). The type of mechanism governing the colloidal OM removal depends on the coagulation pH. For metal salts, at lower pH (5-6) the charge is neutralised by positively charged hydrolysis products, whereas at higher pH (7-8) an immobilisation in a coagulant precipitate, adsorption and bridging are more probable mechanisms.

Particulate OM constitutes solely a few percent of the total organic carbon (Edzwald and Tobiason, 1999). Thus, the removal of soluble OM fraction controls the efficiency of the coagulation process and determines the DBPs formation potential, disinfectant demand and biofilm growth in distribution system. To remove DOM by coagulation a conversion of dissolved substance into a solid (aluminium or iron humates and fulvates) by causing its solubility product to be exceeded (precipitation) or adsorption of dissolved substance onto metal hydroxide surface (co-precipitation) is required. With the increasing pH, the organic humates and fulvates precipitation process is impeded by precipitation of hydroxides and co-precipitation becomes more effective.

The degree of removal achieved in the coagulation processes depends both on the characteristics of the OM (i.e. TOC, molecular weight, aromaticity) and operating conditions. Therefore, for the predominant type of OM, the coagulation condition for the particular treatment plant can be optimised and thus yield significant improvements in efficiency of TOC removal. Some enhancement in TOC removal can be achieved by increasing the applied coagulant dose, changing the coagulant and adjusting pH. Edzwald and Tobiason (1999) defined the term of optimised coagulation as coagulation conditions of dosage and pH that

achieve maximum particle and turbidity removals, maximum TOC and DBPs precursor removals, and minimum residual coagulant, sludge production and operating costs.

The US Environment Protection Agency (US EPA, 1999) prepared a guideline for determination of enhanced coagulation conditions for drinking water treatment works. In the document the required TOC removals for waters of different initial TOC concentrations and alkalinity were defined. The detailed procedures of the assessment of the current work performance and optimisation process were also presented. In practice, determination of optimised coagulation condition involves evaluation of both the optimum pH and coagulant dose using a series of bench-scale jar tests (van Leeuwen et al., 2005).

As reported by several studies, for a given coagulant type, coagulation pH is the most important parameter controlling the optimisation of coagulation in terms of OM removal (Volk et al. 2000; Uyak and Toroz, 2006). However, the optimum coagulation pH for OM removal is reported to be in range 5-6 for alum and 4-5 for ferric salts (Randtke, 1988; Sharp et al., 2006). pH controls both the amount (resulting in a certain coagulant demand), speciation and charge of OM, and the availability of the coagulant hydrolysis products for neutralisation and adsorption of OM. The interactions between OM and coagulant include two parallel, competitive mechanisms: between hydrogen ions (H^+) and cationic coagulant hydrolysis with products competing for anionic organic compounds, and between hydroxyl ions (OH^-) and organic anions competing for cationic metal hydrolysis products. Below and above the optimum coagulation pH ranges, the interaction between OM and coagulant is dominated by hydrogen-OM and hydroxyl-coagulant hydrolysis products reactions resulting in poorer TOC removal.

The coagulation pH affects the coagulant dosage and the relationship between dosage and OM concentration. At lower pH, a relatively lower coagulant dosage is required and there is

a stoichiometric relationship between the required coagulant dosage and the TOC of the water to be treated (Randtke, 1988). With increasing pH, the required coagulant dosage is higher as a result of redistribution of surface charges of both OM and coagulant species. Increased pH promotes deprotonation of OM reducing its negative charge and decreases the positive charge on metal coagulants (O'Melia et al., 1999; Yu et al., 2007).

Gregor et al. (1997) demonstrated the effect of adjusting pH (to 4-5) prior to coagulation on the OM removal. A lower pH enhances the formation of soluble NOM-coagulant complexes. For different coagulation pHs, it was shown that lowering to pH 4 before addition of coagulant improved OM removal efficiency and reduced required coagulant dosages. Lower coagulation pH can result in a higher residual concentration of coagulant, therefore a trade-off between enhanced OM removal and coagulant concentration should be achieved (van Leeuwen et al., 2005) (Figure 2.7).

Increasing the coagulant dosage generally improves removal of OM as more metal compounds are available for OM complexation. However, for OM demonstrating heterogeneous character, increased dosages of coagulant may lead to overdosing as colloidal humic acids and soluble fulvic acids produce different coagulant demands. Higher coagulant doses that are required to neutralise the high anionic charge of soluble fulvic acids, are likely to lead to overdosing of humic acid resulting in restabilisation of the colloids (Randtke, 1998; Gregor et al., 1997). Soh et al. (2008) tested the impact of increased dosages of alum on coagulation efficiency. It was found that increasing the alum dose improved OM removal. However, a further increase in coagulant dose did not result in a significant improvement of OM removal due to the presence of a fraction recalcitrant to removal by coagulation. Randtke (1988) discussed two patterns of relationship between OM removal and coagulant dosage (Figure 2.8). The most common for surface water treatment works with moderate to high

turbidity and alkalinity is a gradual increase in OM removal with increasing dosage, approaching a maximum asymptotically. This type of relationship is associated with high pH, relatively low concentrations of OM and heterogeneous OM compositions. In the second pattern, OM removal increases very sharply for the particular coagulant concentration and is indicative of precipitation-dominated coagulation mechanism. It is commonly observed for low pH, high TOC, and homogeneous OM with the predominance of humic substances.

The impact of the coagulant type on the degree of OM removal is less obvious and strongly depends on the particular raw water properties (OM properties, general water quality including presence of divalent cations, temperature) and coagulation conditions (presence of pH adjustment) (Randtke, 1988; Volk et al., 2000; Rizzo et al., 2004).

The efficiency of OM removal by coagulation is strictly related to the character and concentration of OM present in raw water (Kim and Yu, 2005; van Leeuwen et al., 2005; Fabris et al., 2008). Higher OM removal is observed for samples with moderate and high OM concentrations (Chow et al., 1999; Edzwald, and Tobiason, 1999; Volk et al., 2000; Rizzo et al., 2004). The higher the TOC concentrations, the greater contribution of humic substances, preferentially removed by coagulation due to their high molecular weights.

The differences in OM properties (molecular weight, aromaticity, hydrophobicity, presence of particular functional groups) determine the varying impact of coagulation processes on removal of different OM fractions. The higher removal efficiencies occur for the high molecular-weighted aromatic humic substances, of greater THMFP and greater affinity for hydrophobic compounds (Liang and Singer, 2003; Parsons et al., 2004; Kim and Yu, 2005; Fabris et al., 2008). In particular, the phenolic fraction in the hydrophobic OM was found to be removed preferentially by coagulation compared to the carboxylic moieties (Kim and Yu, 2005). Thus, the concentration of the hydrophilic fraction can provide an indication of the

potential coagulation efficiency and expected TOC removals (Sharp et al., 2006; Fabris et al. 2008). For dissimilar raw waters from selected Australian and Norwegian treatment works, Fabris et al. (2008) demonstrated that following coagulation, the recalcitrant OM character was similar.

In several studies, the effect of oxidation pre-treatment, i.e. pre-ozonation prior to coagulation on the efficiency of OM removal has been investigated (Widrig et al., 1996; Win et al., 2000; Gallard and von Gunten, 2002). The chemical transformations of OM on reaction with ozone were found to improve coagulation efficiency in the range of 10-15 % (Widrig et al., 1996). Win et al. (2000) found that compared to ozonation and short UV irradiation, the long UV irradiation with the presence of H₂O₂ significantly decreased OM amount, resulted in breakdown of the refractory OM fractions and improved its bioavailability, and finally enhanced the overall removal by coagulation. In the study of Gallard and von Gunten (2002) the pre-treatment with UV-Vis irradiation showed no effect on THMs formation albeit produced higher chlorine demand. The pre-ozonation led to lower THMs formation with stable chlorine demand and pre-chlorination reduced both the THMFP and chlorine demand.

Soh et al. (2008) analysed the alum removal of four OM isolates derived from rapid fractionation technique (Chow et al., 2004): very hydrophobic acids (VHA), slightly hydrophobic acids (SHA), charged hydrophilics (CHA), and neutral hydrophilics (NEU). They found that the NEU fraction comprising low molecular and hydrophilic OM was recalcitrant to removal by alum coagulation (for the operational alum dose of 100 mg/l the removals of four fractions were 72% VHA, 59% SHA, 38% CHA and 16% NEU respectively). It was also demonstrated that a poorly removed NEU fraction had the second greatest THMFP (after VHA) and significantly contributed to biodegradable dissolved organic carbon fraction (85%, BDOC), responsible for microbial growth and biofilm

formation in the distribution system (Bolto et al., 2002). Thus, the higher the proportion of VHA, SHA and CHA fractions in raw water, the greater the amount of TOC removed by coagulation, lower potential of THMs formation and better DBPs precursors control, lower disinfectant demand and lower residual disinfectant is needed to maintain water quality in the distribution system.

2.4. APPLICATION OF FLUORESCENCE SPECTROSCOPY IN ORGANIC MATTER CHARACTERISATION IN DRINKING WATER TREATMENT

Fluorescence spectroscopy appears to be a promising monitoring technique in drinking water treatment as it is rapid, reagentless and requires no sample preparation prior to measurements (Sharpless and McGown, 1999; Baker, 2001; Hudson et al., 2007; Henderson et al., 2009). However, to date the application of fluorescence spectroscopy in the water industry has been limited to water quality assessment, wastewater characterisation, and monitoring of biological activity in bioreactors (Reynolds and Ahmad 1997; Galapate et al., 1998; Ahmad and Reynolds, 1999; Hudson et al., 2007; Wolf et al., 2007).

Fluorescence EEM spectroscopy has been used for fingerprinting of wastewater sources due to distinctively increased levels of protein-like fluorescence (Baker and Spencer, 2004; Elliot et al., 2006; Hur et al., 2008; Nam and Amy, 2008). Henderson et al. (2009) evaluated the application of fluorescence spectroscopy as a recycled water monitoring tool for process performance assessment, cross-connection detection and overall water quality in indirect potable reuse systems. Fluorescence spectroscopy can detect small changes in DOM/TOC concentration and characteristics (Henderson et al., 2009). With the potential for discrimination of water sources (e.g. drinking and recycled water, riverine water and wastewater), fluorescence spectroscopy can be a useful monitoring tool in water quality

studies (Henderson et al., 2009). Likewise, in the study of Nam and Amy (2008) fluorescence properties of wastewater effluent were compared to NOM properties. The increased levels of protein-like fluorescence and humic-like fluorescence location at lower excitation and emission wavelengths were found to be the most important features discriminating between wastewater and freshwater OM.

The application of fluorescence spectroscopy in bioprocess monitoring has been investigated by several authors (Li and Humprey, 1992; Khoury et al., 1992; Wolf et al., 2007; Denkhaus et al., 2007; Wolf et al., 2007). Biofilms are populations of microorganisms that can accumulate at phase boundaries and biodegrade organic compounds from water i.e. biologically active carbon or in the distribution system (Hallam et al., 2001; Denkhaus et al., 2007; Velten et al., 2007; Simpson, 2008). Organisms in the biofilm were found to contain natural intracellular fluorophores (tryptophan, pyridoxine, NAD(P)H, riboflavin) that can provide indication on microbial activity, biomass development phase, and cell concentrations. For the exponential phase of cell growth, fluorescence intensity of tryptophan-like fluorescence was correlated with cell concentrations (Li and Humprey, 1991; Khoury et al., 1992). More recently, fluorescence spectroscopy, coupled with chemometric techniques (i.e. partial least squares (PLS), principal components analysis (PCA), artificial neural networks (ANNs)), has been utilized for online monitoring of an anaerobic digestion process (Skibsted et al., 2001; Morel et al., 2004; Wolf et al., 2007). Morel et al. (2004) found a good correlation between synchronous fluorescence spectra characterising the digestion process with chemical oxygen demand, volatile fatty acids content, and methane production.

In algal monitoring, portable spectrophotometers have been commonly used for in situ assessment of algae composition and quantity. Algae do not pose a problem to water treatment processes provided populations are relatively low. However if present in high

concentrations, algogenic OM is known to interfere with many of the drinking water treatment processes i.e. coagulation, filtration. Algogenic OM was found to contain a relatively high nitrogen content and low aromatic carbon and phenolic contents (Fabris et al., 2008; Henderson et al., 2008a, 2009). The fluorescence spectra of algogenic OM exhibit a predominance of tryptophan-like fluorescence of higher excitation wavelengths and intensities in the exponential phase compared to the stationary phase (Determan et al., 1998; Smith et al., 2004; Nguyen et al., 2005; Elliot et al., 2006; Henderson et al., 2008a). A strong affinity of autochthonous OM for hydrophilic material was also corroborated (Laabs et al., 2004; Henderson et al., 2008a).

Of special importance is the detection of cyanobacteria occurrence and dynamics in freshwater bodies used as drinking water supplies (Gregor et al., 2007; Henderson et al., 2008a,b). Cyanobacterial outbreaks can pose a potential threat, while many cyanobacteria are toxin-producing (most common *Planktothrix rubescens* and *Microcystis aeruginosa*) (Leboulanger et al, 2002). Standard methods of cyanobacteria detection, including monitoring of phytoplankton assemblages and direct chromatographic measurements of cyanotoxins are scarcely sensitive and selective, expensive and time-consuming (Gregor et al., 2007). Fluorescence spectroscopy utilizes the excitation and emission signatures of chlorophyll *a* and other chloroplastic pigments, responsible for light absorption and conversion in autotrophs. Different phytoplankton groups, e.g. green algae and cyanobacteria, have unique pattern of pigments (chlorophyll *a*, other accessory chlorophylls, carotenoid protein pigments, phycobilins) of intrinsic fluorescence properties, being the basis for the spectral differentiation of phytoplankton communities (Yentsch and Phinney, 1985; Determan et al., 1998; Gregor and Maršálek, 2005). For chlorophyll *a* the frequently reported emission wavelength value is 680 nm with the maximum chlorophyll *a* content at excitation

wavelength 440 nm (430-530 nm). Cyanobacterial pigments are excited at higher wavelengths in the red and orange part of the spectrum, with maximum at 620 nm (550-680 nm) and with emission at 645 nm (640-680 nm). Chlorophyll fluorescence measurements can be carried out in a discrete way (water samples, in vivo fluorescence) or in situ, directly in a water column, with the possibility of online detection and quantification. In situ chlorophyll *a* fluorescence measurements have been performed by many authors (Asai et al, 2001; Pinto et al, 2001; Beutler et al, 2002; Leboulanger et al, 2002; Gregor and Maršálek, 2005).

Previous applications of fluorescence in drinking water treatment include studies of OM reactivity with disinfectant (Korshin et al., 1999; Westerhoff et al., 1999; Świetlik and Sikorska, 2004), prediction of THMs formation (Beggs et al. 2006; Johnstone and Miller, 2009; Roccaro et al., 2009), correlation of fluorescence properties with SUVA and DBPs formation during chloramination (Yang et al. 2008), and more recently OM characterisation in membrane permeates (Peiris et al. 2008). Her et al. (2003) demonstrated combined HPSEC-fluorescence approach for DOM characterisation in bulk water samples without fractionation.

Korshin et al. (1999) investigated the effect of chlorination on OM fluorescence. The presence of three groups of fluorophores of different reactivity with chlorine was ascertained (fast-, medium-, and slow-decaying fluorophores). It was observed that upon chlorination, fast-decaying sites were selectively eliminated, whereas the contribution of medium- and slow-decaying fluorophores increased with the chlorine dosage. Overall, the fluorescence emission bands were contracted and a shift towards lower wavelengths was discerned indicating a breakdown of high molecular weight, aromatic compounds. The changes in structural composition of OM were also reflected in the increase of fluorescence intensity for Cl/DOC ratios less than 2.

Świetlik and Sikorska (2004) corroborated the results obtained by Korshin (1999) and found that the observed changes in fluorescence spectra are oxidant dependent. While the fluorescence emission tends to shift to lower wavelengths upon disinfection with chlorine, ozonation produces the opposite effect and a relative increase in the emission wavelength can be observed. Thus, during ozonation, the structural composition of fluorophores is changed and carbonyl-, hydroxyl-, alkoxy-containing moieties and amino groups become more predominant. A hypsochromic shift in the fluorescence spectra (towards shorter wavelengths) during chlorination can be caused by a reduction of conjugated bonds in aromatic rings and transformation to more aliphatic conformation by elimination of particular functional groups (carbonyl, hydroxyl, and amine). Furthermore, ozonation increased the protein-like fluorescence intensity (comprising small molecular weight components, aromatic and aliphatic amines and amino acids), whereas the oxidation effect on humic-like and fulvic-like fluorescence was more equivocal. Oxidation leads to fractionation of OM into smaller chromophoric fractions enhancing fluorescence, whereas advanced oxidation produces structural changes in fluorophores decreasing fluorescence (Henderson et al., 2009).

Beggs et al., 2006 found that chlorine reaction with DOM resulted in the decrease in fluorescence intensity. A shift to shorter emission wavelengths was observed with chlorine addition, the greater with higher chlorine dosages and reaction times.

Marhaba et al. (2000, 2003, 2009) introduced multiple regression models providing prediction of DOC, concentrations of OM isolates obtained from resin fractionation (Leenheer, 1981) and THMFP from the fluorescence properties. The authors measured the fluorescence signal of different OM fractions (hydrophobic and hydrophilic acid, neutral, and base fraction) and defined the EEM areas of the highest intensity for each isolate. All fractions demonstrated increased emission intensities within UV excitation wavelengths, with hydrophobic fractions

extending from 370 to 430 nm emission wavelength and hydrophilic fractions with emission wavelengths lower than 370 nm. Likewise, the optical properties of the hydrophobic and hydrophilic fractions were evaluated by Chen and Valentine (2007) and Yang et al. (2008). However, in each of those studies, the spectral location of the peaks reflecting the contribution of particular isolates was different and therefore the results are inconclusive. In a similar study the fluorescence properties of the wastewater effluent NOM were related to the THMFP (Zhang et al., 2009). The hydrophobic acid and base contained peak C fluorescence, aromatic proteins absorbing in the UV wavelengths (acid) and peak A fluorescence (base), whereas the hydrophobic neutrals comprised protein-like fluorescence. The composition of the hydrophilic fraction was similar for three isolates, and exhibited the presence of both peak C and A, and protein-like fluorescence (hydrophilic acid).

Chen and Valentine (2007) conducted a comparative study of N-nitrosodimethylamine (NDMA) formation with monochloramine. Monochloramine as a disinfectant demonstrates lower oxidation strength compared to free chlorine and longer residuals in disinfected water. Wu et al. (2003b) showed that small amounts of free chlorine in equilibrium with monochloramine react preferentially with aromatic moieties in humic substances. Free chlorine during the chlorination was found to destroy more effectively cross-linking structures of fused-ring systems of humic acids compared to monochloramine. Thus, chlorination was demonstrated to create more active sites for subsequent oxidation or substitution reaction than chloramination and was hypothesized to form more THMs than HAAs (Wu et al., 2003b).

As a result of the reaction between NOM and disinfectant, fluorescence EEMs revealed changes in the location and relative intensity of fluorescence peaks. Overall, the changes indicated a breakdown of the peak C fluorophore and a shift towards lower emission wavelength suggesting reduction in the degree of aromaticity. No relationship was found

between NDMA formation potential and fluorescence intensity and hence authors concluded that fluorescence is a poor quantitative probe of NDMA formation. However, the study contains several methodological inaccuracies regarding naming and identification of fluorescence EEM peaks and no attempts have been made to utilize the changes in spectral peak locations.

In a related study, Hua et al. (2007) investigated the DBPs formation potential in relation to fluorescence properties for freshwater lakes. The higher fluorescence intensities in all fluorescence regions were found to correlate with higher THMFP. Moreover, the protein-like fluorescence was found to be a good predictor of N-nitrosodimethylamine formation.

Yang et al. (2008) compared the application of SUVA and fluorescence OM properties to DBPs formation during chloramination. Although OM-chlorine interactions in chloramination process are less intensive than in chlorination, they can yield a significant information on overall reactivity of OM with oxidants. It was found by Yang et al. (2008) that SUVA performed better as a THMFP predictor compared with fluorescence-derived prediction. The poorer fluorescence performance was elucidated by the authors and ascribed to the utilization of a fluorescence regional integration (FRI) model for integration of the fluorescence spectra instead of specific peak information. The FRI model provides an evaluation of relative distributions of different fluorophores. In the method, EEM is divided into a few fluorescence regions notionally corresponding to particular OM constituents of distinctive spectrofluorometric properties (Chen et al., 2003). The relative importance of a particular fluorescence region can be quantified by determining the volume of fluorescence beneath given region. It was found that the cumulative EEM volumes at regions II and IV (protein-like fluorescence) correlated with the yields of dichloroacetic acid (DCAA, $R^2 = 0.60$), chloroform ($R^2 = 0.42$), dichloroacetonitrile (DCAN, $R^2 = 0.53$), and TOX ($R^2 = 0.63$). The

corresponding correlations with SUVA were higher (DCAA $R^2 = 0.82$, chloroform $R^2 = 0.73$, DCAN $R^2 = 0.88$ and TOX $R^2 = 0.80$). Thus, the authors concluded that the FRI could possibly have hindered the relationship between fluorescence and DBPs formation and advised investigating the fluorescence peak information.

Recently, Johnstone and Miller (2009) used the EEM technique coupled with fluorescence regional integration model to determine DBPs formation. DBPs formation was related using multifactor linear regression to changes in fluorescence properties and chlorine consumption. The authors tested two models, the first model incorporating only fluorescence regional integration data and the second model with both fluorescence and chlorine consumption predictors. The prediction quality was dependent on the particular DBPs species formed and was significantly improved for the more complex model (CL₃AA $R^2 = 0.59$ for the solely fluorescence model and 0.90 for fluorescence/chlorine consumption model, CHCl₃ $R^2 = 0.33$ and 0.82, Cl₂AA $R^2 = 0.38$ and 0.86 respectively). Moreover, different fluorescence regions were found to be predictors for different DBPs species. In particular, region 2 of simple aromatic proteins correlated with chloroform yields ($R^2 = 0.78$), region 4 of soluble microbial-like OM correlated with dichloroacetic acid ($R^2 = 0.79$), and the combination of region 3 (humic-like fluorescence with UV emission) and 4 was found to correlate with trichloroacetic acid ($R^2 = 0.87$). Similarly to the study of Yang et al. (2008), the results obtained by Johnstone and Miller (2009) refer to the arbitrary selected fluorescence regions not to the real fluorophores. Therefore the true relationships between fluorophores properties and DBPs formation can be obscured.

Likewise, Roccaro et al. (2009) attempted to correlate changes in synchronous fluorescence properties of OM with DBPs formation. They defined a fluorescence index $\lambda^{em}_{0.5}$ as the position of the normalized emission band at its half-intensity for the fixed excitation

wavelength of 320 nm. During chlorination, with increasing the reaction time or chlorine dosage, a blue-shift of the fluorescence index was observed. Similarly, the changes in the magnitude of fluorescence intensity ratio at 500 and 450 nm were examined. Both parameters were also found to correlate with the concentrations of several DBPs species ($R^2 > 0.90$). The changes in the fluorescence parameters were hypothesized to reflect destruction of the reactive aromatic groups in OM. However, the effect of chlorine quenching on the applicability of fluorescence parameters presented in the chlorination study was not addressed by the authors.

Peiris et al. (2008) assessed nanofiltration membranes fouling with the use of fluorescence spectroscopy. Fluorescence analysis discriminated between two types of membranes of different fouling rates. It was found that the membrane of higher fouling rate demonstrated higher levels of protein-like fluorescence in its permeate. Furthermore, the red shift in the emission wavelength of humic-like fluorescence was observed in permeates of different membranes suggesting preferential rejection of highly aromatic fraction.

From this overview of fluorescence spectroscopic applications in drinking water treatment it appears that to date most of the research concentrated on the characterisation of changes in OM induced by chlorination (Korshin et al., 1999; Świetlik and Sikorska, 2004; Beggs et al., 2006) and correlation between OM fluorescence and DBPs formation (i.e. Yang et al., 2008; Johnstone and Miller, 2009; Roccaro et al., 2009). In the majority of studies, fluorescence measurements were carried out on OM fractions derived from other analytical techniques (i.e. HPSEC or resin fractionation) and no reference to bulk OM properties has been made. As discussed in the previous sections, the use of fractionation techniques can significantly alter physico-chemical properties of OM, and hence provide inadequate or misleading information on OM reactivity and treatability. Furthermore, no attempt has been made to utilize

fluorescence spectroscopy for comprehensive OM characterisation of raw and partially-treated waters to provide rapid indication of OM reactivity and treatability. From the presented applications of fluorescence in the environmental sciences, it can be seen that fluorescence spectroscopy can provide thorough characterisation of OM constituents including structural and molecular information, discriminate between different OM sources, and finally outperform existing OM characterisation methods. Thus, the potential application of fluorescence spectroscopy in drinking water treatment does not seem to have been fully recognized and utilized.

Moreover, in the studies presented here, problems with the analysis of fluorescence data have been reported. Although fluorescence excitation-emission matrices (EEMs) comprise valuable information on OM properties, the complexity of the fluorescence data can hamper the interpretation process. Fluorescence data extraction methods reported in the presented studies, solely rely on fluorescence regional integration technique (e.g. Yang et al., 2008) or arbitrary selected spectral parameters (e.g. Roccaro et al., 2009). The former method can only provide quantitative information on the content of pre-defined groups of fluorophores located in the similar areas of the EEM. As different fluorophores can reflect substantially unlike OM properties, the use of FRI can lead to inadequate conclusions regarding OM reactivity and functionality. This can explain the poor correlations found between FRI-derived parameters and OM properties (e.g. THMFP) compared with other analytical methods (e.g. Yang et al., 2008). Therefore, along with the evaluation of application of fluorescence spectroscopy for OM characterisation in drinking water treatment, an application of adequate data analysis techniques should be addressed in future studies.

2.5. CHARACTERISING WATER TREATMENT WORKS PERFORMANCE USING FLUORESCENCE SPECTROSCOPY

OM characterisation in terms of its reactivity with disinfectant and treatability with the use of standard treatment processes is of the greatest importance to drinking water treatment operators. Understanding the organic character of water can help water operators to predict the applicability of coagulation and other OM removal techniques and to estimate the required chlorine demand. Moreover, the rapid assessment of OM constituents can help in determining the expected levels of DBPs for the final water and make adequate adjustments in the operation of the water system prior to disinfection stage (Figure 2.9). Therefore, knowledge about the structural and chemical properties of OM and their reactivity is a key factor in controlling DBPs formation in drinking water treatment.

From the evaluation presented in the previous sections, two groups of methods for characterising OM in drinking water treatment have been commonly utilized; viz. different fractionation techniques (i.e. HPSEC, resin fractionation) and spectroscopic techniques based on several parameters derived from UV-Vis absorbance measurements. The first group of methods provides a comprehensive OM characterisation in terms of molecular weight distribution and the degree of hydrophobicity, which are important characteristics of OM reactivity and treatability. However, the fractionation measurements are time-consuming, require intensive sample pre-treatment that can significantly change the OM properties, and are difficult to implement in the rapid assessment of OM at WTWs. Likewise the selectivity of the fractionation techniques can be limited to particular OM fractions. For instance, in the HPSEC analysis OM is separated according to its molecular weight but the detection of non-chromophores is relatively poor limiting the application of the technique to heterogeneous OM samples (Her et al., 2002). Furthermore, in the fractionation studies the aquatic OM is

decomposed into a set of different fractions and moieties whose reactivity and treatability can be entirely different from the bulk OM properties.

The spectroscopic techniques tend to circumvent the limitations of the fractionation techniques. They provide rapid, non-invasive analysis of aquatic OM samples, with the potential for online monitoring assessment of OM reactivity and treatability. To date, the application of the spectroscopic techniques based on the measurements of the inherent optical properties of OM has been limited to the UV-Vis absorbance spectroscopy. Several spectral properties of the OM absorbance spectra have been proposed as surrogate parameters for characterisation of the OM reactivity and treatability, with UV measured at 254 nm (UV₂₅₄) and SUVA being the most commonly utilized. However, absorbance-based techniques can only determine the content and properties of the aromatic functional groups in OM with no information provided on the presence and relative importance of non-chromophoric OM. Furthermore, absorbance techniques are prone to interference from various impediments present in water (i.e. UV-absorbing impurities, pH, iron, nitrate). Therefore, correlations between absorbance-derived surrogates and OM reactivity and treatability parameters (i.e. reactivity with coagulants and disinfectants, THMFP) may contain limited information and reflect only the site-specific OM composition. Thus, a direct comparison between reactivity of different waters or OM fractions with absorbance technique is limited.

TOC is commonly used in drinking water treatment to evaluate the works performance in OM removal efficiency. Although TOC analysis provides valuable information on the overall quantity of OM present in water, it does not characterise its reactivity and treatability.

Municipal WTWs usually abstract water from multiple sources representing different, temporally variable in time NOM concentrations, physico-chemical properties, reactivity and treatability. Depending on the specific catchment characteristics and combination of

hydrological and climatic factors, NOM undergoes significant seasonal changes (Fabris et al., 2008). The seasonal problems with treatment processes such as coagulation, flocculation, sedimentation, flotation and filtration associated with increased rainfall and DOC concentrations have been previously reported (Parsons et al., 2004).

Furthermore, the increased DOC concentrations in natural and semi-natural catchments Europe and North America observed within the last 20 years reflect global changes in the level of acidification. However, future equivocal rain patterns and anticipated temperature increase can intensify DOM release from wetlands and increase aquatic DOC concentrations. Those quantitative changes would possibly be accompanied by the qualitative changes in OM composition with a relative increase in low molecular weight, hydrophilic fraction, more recalcitrant to conventional treatment. To face those challenges and meet current regulations regarding the formation of DBPs, water treatment companies will have to find adequate analytical techniques for OM characterisation to keep the operation of the water systems adjusted to constantly changing environmental conditions.

Thus, to successfully minimise and control the formation of DBPs, a suitable analytical technique for comprehensive OM characterisation should be facilitated at WTWs. From the prior evaluation, fluorescence spectroscopy appears to be a potential alternative to the existing analytical methods of OM characterisation. The advantages of fluorescence include rapid, sensitive and selective characterisation of aquatic OM, no sample pre-treatment including chemical alterations, small sample volume (a few ml), and the potential for on-line monitoring incorporation. Fluorescence enables the rapid assessment of OM constituents and can potentially provide useful information on OM reactivity and treatability. Additionally, fluorescence can provide an indication of the OM sources (allochthonous versus autochthonous). However, the limited number of studies in drinking water treatment has

applied fluorescence spectroscopy to characterise the changes in OM upon disinfection and to relate the fluorescence properties to DBPs formation. Well established literature relationships between fluorescence spectra and OM physico-chemical properties (i.e. degree of hydrophobicity, microbial content), have not been evaluated for the field of water treatment.

Therefore, this thesis presents the evaluation of the application of fluorescence spectroscopy as a direct analytical method to determine the quantitative and qualitative OM characteristics in drinking water treatment. In particular, the following objectives will be addressed:

Objective 1: To develop an understanding of the ability of fluorescence spectroscopy to characterise intake water quality at various water treatment works at different spatial and temporal resolutions and so identify the presence of pollutants.

Hypothesis H1: That fluorescence peak T and peak C emission wavelength can provide discrimination between water treatment sites and water types and that these parameters can be used to characterise OM in spatial and temporal terms.

Objective 2: To develop an understanding of the ability of fluorescence spectroscopy to characterise unit process performance at water treatment works. In particular, to develop an improved understanding of NOM fractions, and their variations, within raw and partially treated water.

Hypothesis H2: That there are robust and quantifiable relationships between peak C and TOC in raw and clarified water and that those relationships can be used to assess the clarified water potential for DBPs formation.

Hypothesis H3: That peak C and peak T fluorescence properties can characterise the spatial and temporal variation in OM quantity and quality, including the identification of OM fractions of different reactivity and treatability in drinking water treatment.

Objective 3: To develop techniques for the use of fluorescence spectroscopy as a tool to determine the formation and prevalence of DBPs in water treatment works.

Hypothesis H4: That peak C and peak T fluorescence properties can assess the raw and partially-treated water potential for DBPs formation and provide quantitative information on residual OM content post-clarification.

Objective 4: To evaluate data mining techniques for fluorescence data analysis and information extraction.

Hypothesis H5: That peak-picking approach and visual inspection of fluorescence EEMs are not effective tools for analysis of large fluorescence dataset and characterisation of spatial and temporal variability.

Hypothesis H6: That application of different data decomposition techniques can enhance the analysis of fluorescence data and improve the understanding of spatial and temporal variation in fluorescence properties related to OM removal efficiency.

Objective 5: To compare the use of fluorescence spectroscopy with the standard OM characterisation tools in drinking water treatment (e.g. TOC and UV-Vis).

Hypothesis H7: That fluorescence spectroscopy can provide additional information on OM quantity, reactivity and treatability compared with TOC and UV-Vis analyses.

2.6. CHAPTER 2 FIGURES

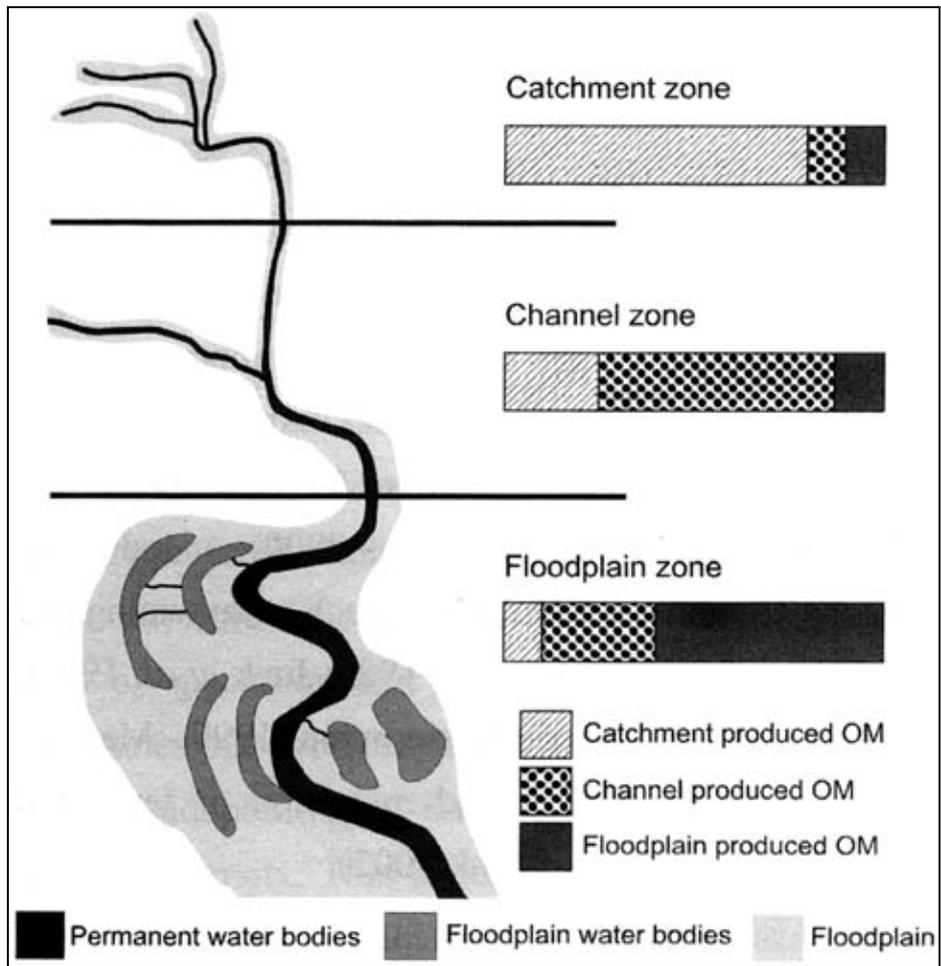


Figure 2.1: The river continuum concept (after Junk and Wantzen, 2003)

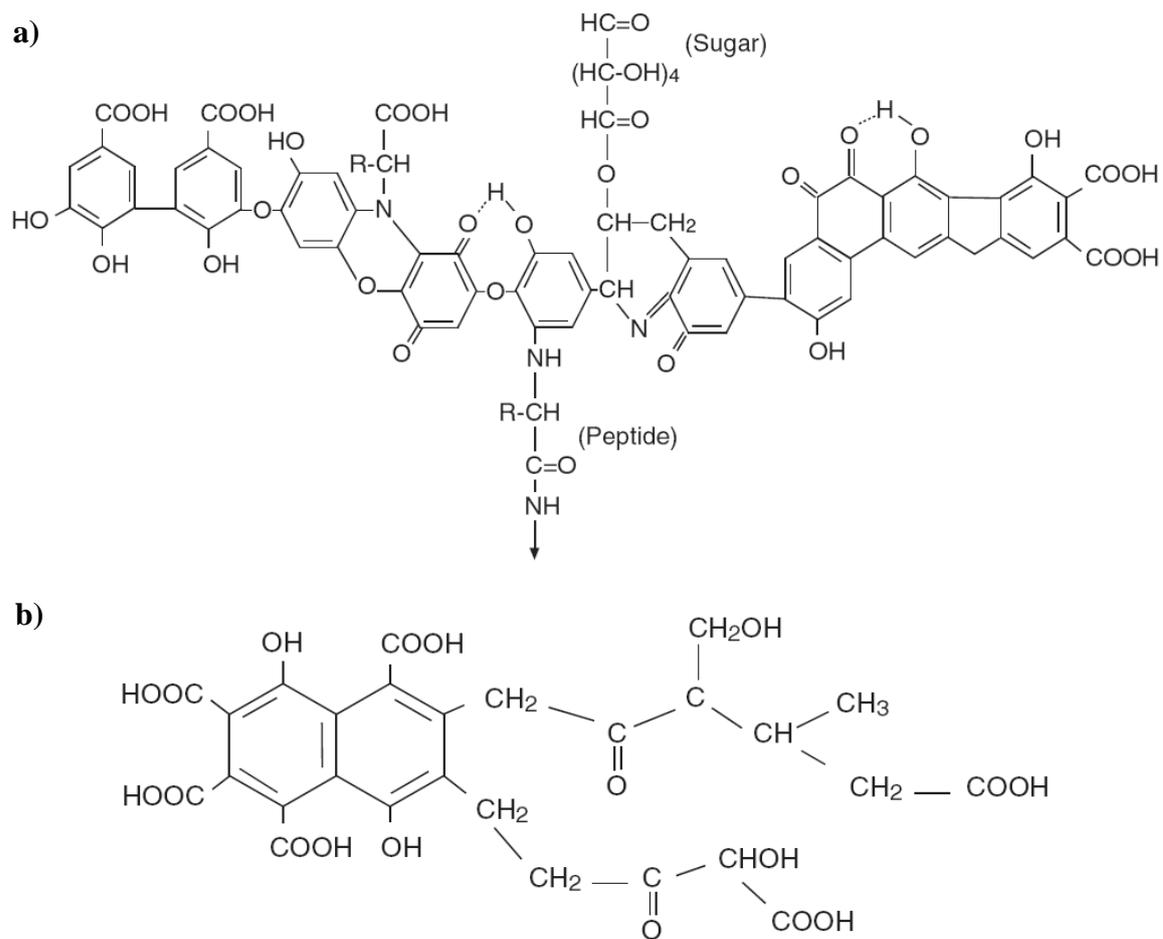


Figure 2.2: Theoretical humic (a) and fulvic acid (b) (after Hudson et al., 2007)

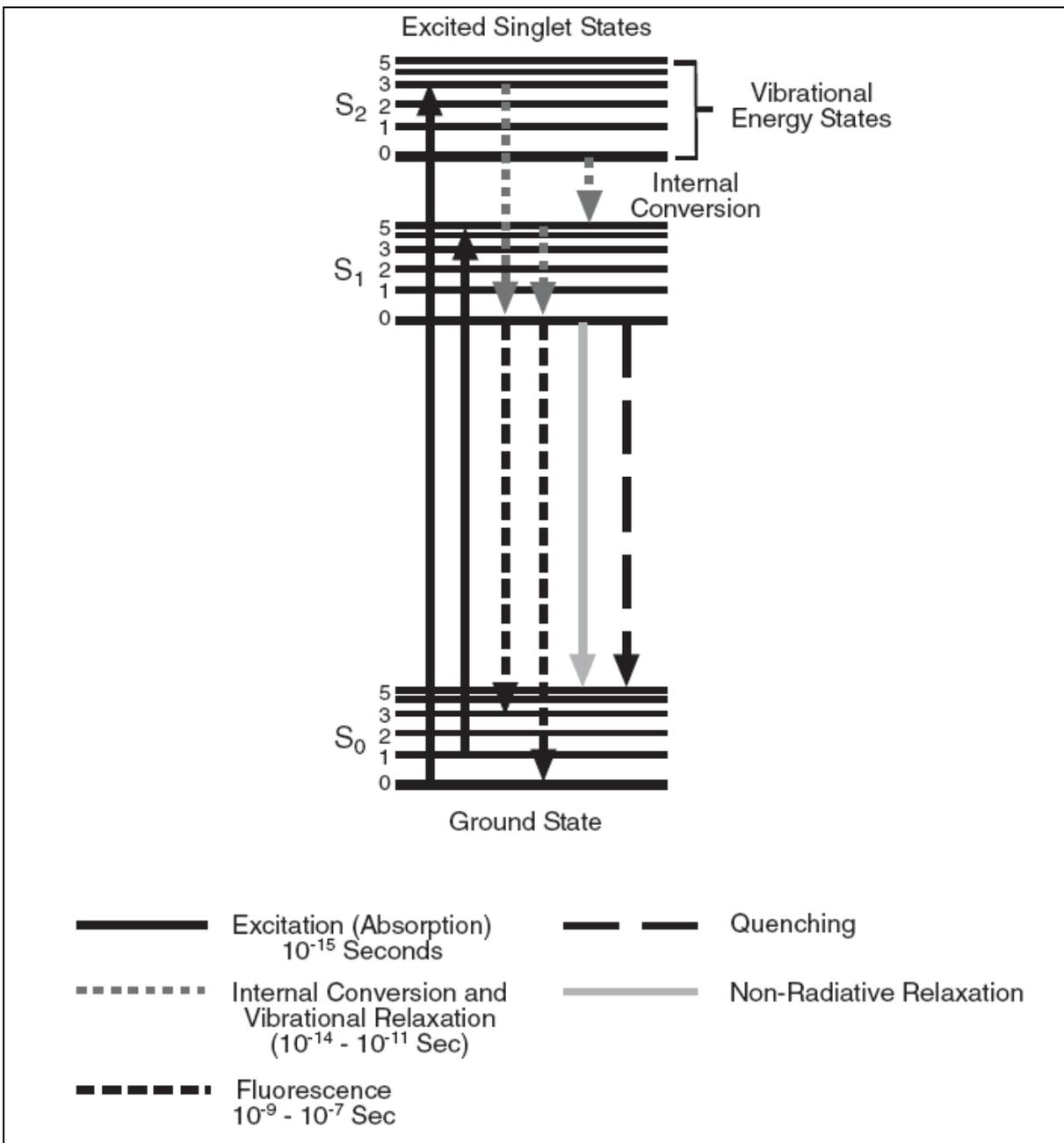


Figure 2.3: Jablonski energy diagram (after Hudson et al., 2007 and Lakowicz, 1999)

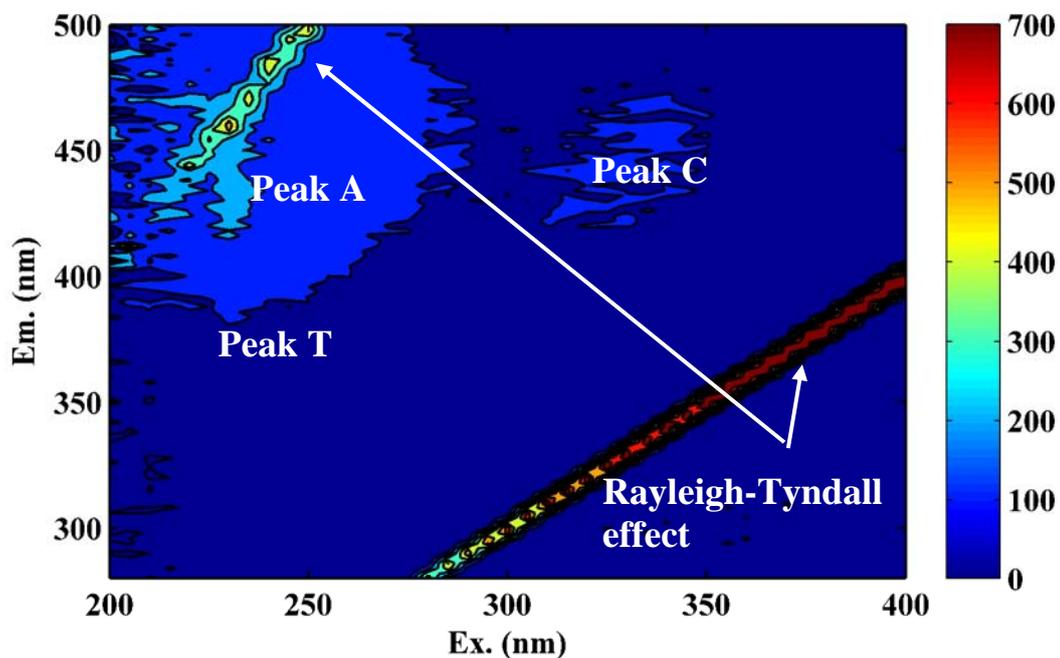


Figure 2.4: Fluorescence excitation-emission matrix with typical fluorescence features (peak A, C, T and Rayleigh-Tyndall effect) (after Coble, 1996)

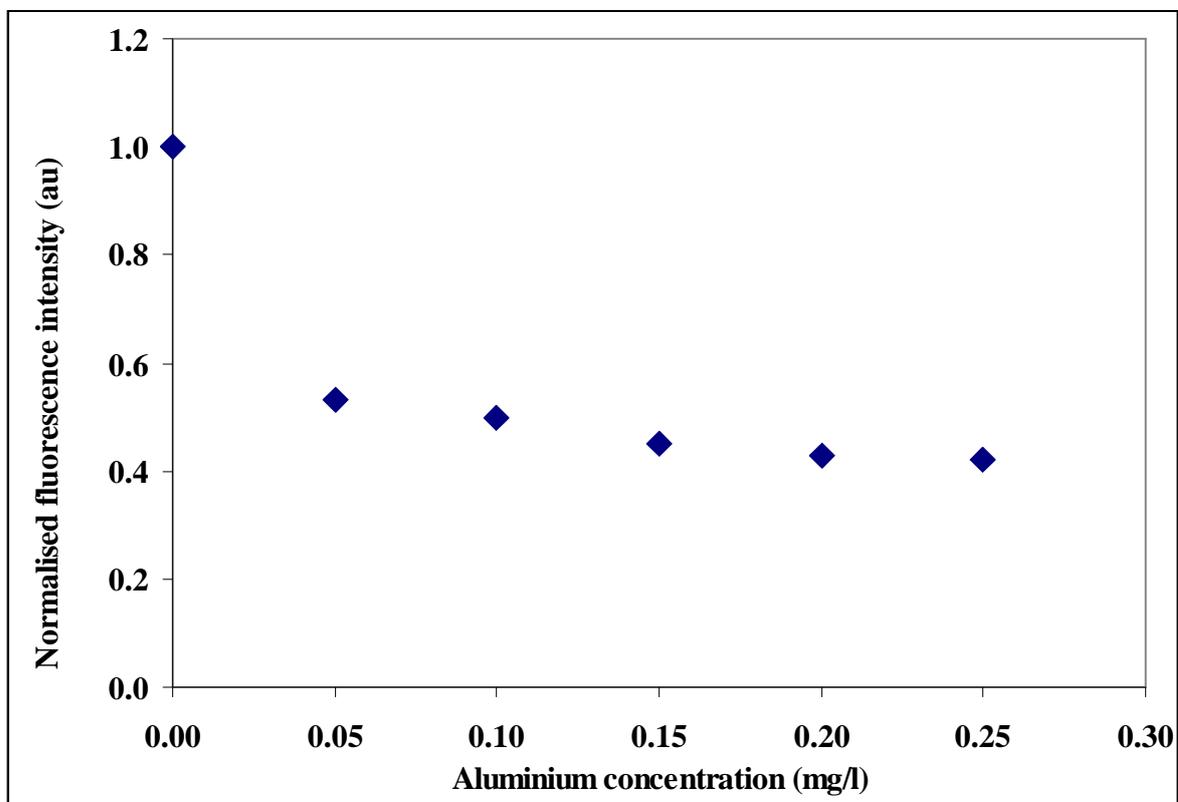


Figure 2.5: The effect of aluminium on fluorescence of humic acid (9 mg/l; Reynolds and Ahmad, 2005)

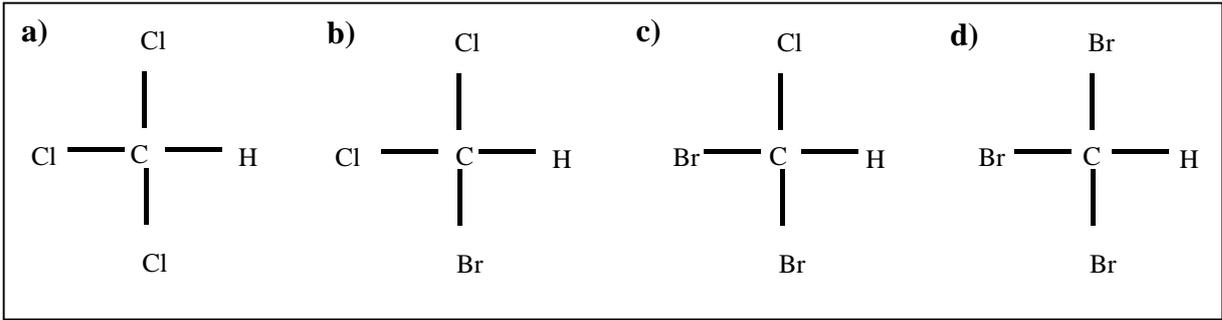


Figure 2.6: Four most common THMs: chloroform (a), dichlorobromomethane (b), dibromochloromethane (c), bromoform (d)

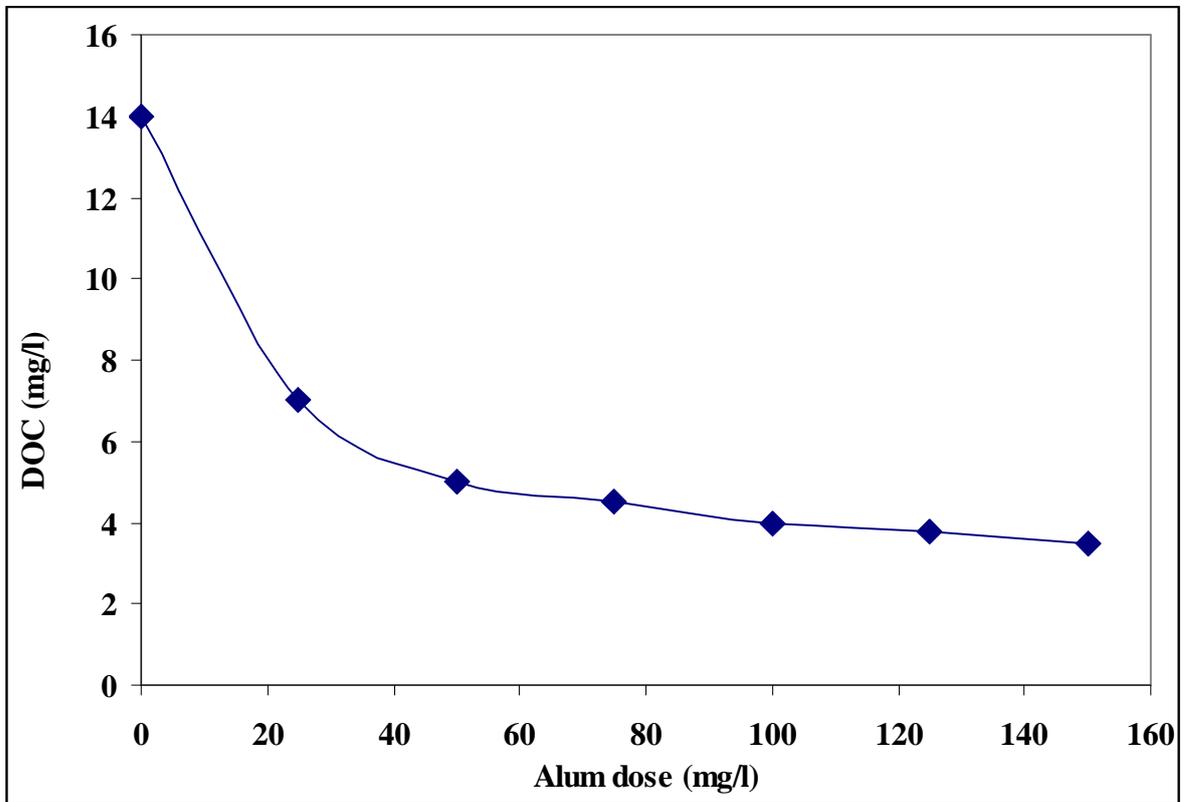


Figure 2.7: Removal of DOC from a sample of Middle River Reservoir (South Australia) water under jar test conditions at pH 5.5 (after van Leeuwen et al., 2005)

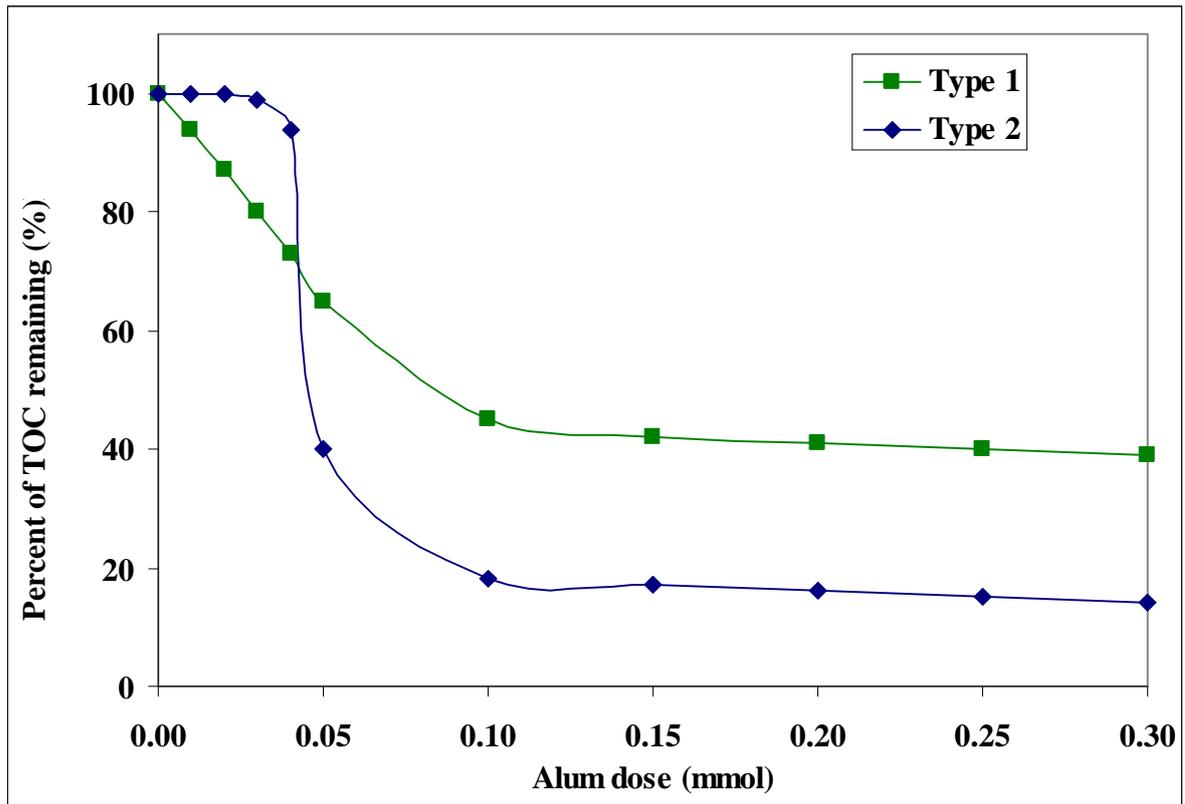


Figure 2.8: The two patterns of relationship between TOC removal and coagulant dosage (after Randtke, 1988)

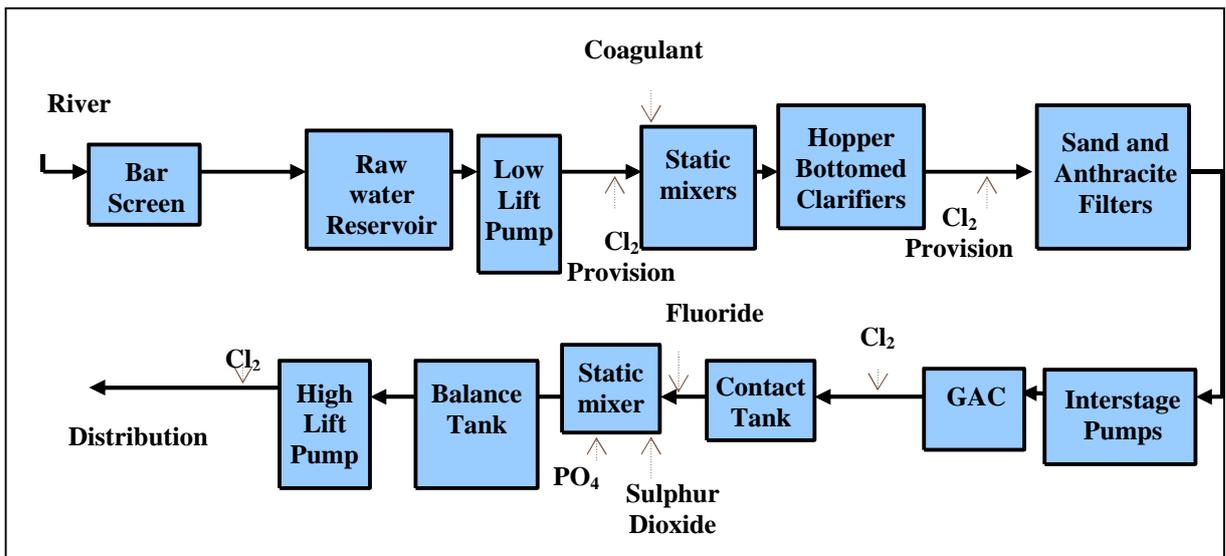


Figure 2.9: Schematic diagram of the drinking water treatment stages

3. MATERIALS AND METHODS

3.1. SOURCE WATERS

Fluorescence spectroscopic analysis was carried out on samples of raw and clarified water from 16 surface water treatment works (WTWs), collected monthly between August 2006 and February 2008. The treatment works are located in the Midlands region in central UK and are owned and operated by Severn Trent Water Ltd (Figure 3.1).

The works are located in four main catchments: Upper Severn (Shelton, Trimpley), Lower Severn (Campion Hills, Draycote, Mitcheldean, Mythe, Strensham), Upper Trent (Frankley, Tittesworth, Whitacre), and Lower Trent (Bamford, Church Wilne, Cropston, Little Eaton, Melbourne, Ogston). Frankley WTW imports water from the Elan Valley Reservoirs system (owned by Dwr Cymru Welsh Water) under gravity via the Elan Aqueduct from Rhayader in Powys to Frankley in Birmingham. The typical volume transferred to Frankley WTW varies from 320 Ml/d during dry summer conditions to 340 Ml/d during increased demands periods (Severn Trent Water, 2008). The treatment works system connects the Derwent Valley system in North Derbyshire with the Mythe WTW near Tewkesbury in Gloucestershire and forms a main strategic line of the system. The connections to all the remaining Severn Trent Water WTWs form the lateral expansions of the treated water grid (Figure 3.2). The grid-like configuration of the treatment system allows for water supply between Derby and Strensham from any of the WTW linked to the grid (Severn Trent Water, 2008). The main drinking water demand centres are located in the cities of Birmingham, Coventry, Leicester, Nottingham, Gloucester, and Derby.

The 16 Severn Trent Water treatment works considered in this study abstract surface water from a wide range of different sources with different factors influencing raw water quality.

The relative proximity of the treatment sites in spatial terms (approximately 100 miles between the two furthest sites to the north and south, Bamford and Mitcheldean respectively) is responsible for similar rainfall and river flow patterns and trends across the study area. However, the local geomorphological conditions in the particular catchments, (i.e. hills elevation and exposure, presence of the rain shadow), can significantly alter the observed rainfall sums and change the rivers regimes. In addition to the rainfall-runoff variation, the local differences in the land cover cause a variation in the OM composition between sites, i.e. the relative contribution of natural OM (fulvic and humic acids) and anthropogenic inputs from domestic and industrial wastes. The upland water sources (e.g. Bamford) demonstrate higher content of humic acids, of higher molecular weight and hydrophobic character, whereas lowland water sources (i.e. Mythe, Strensham) are more affected by the anthropogenic impacts and comprise a more hydrophilic and low molecular weight OM fraction. Thus, the sample sites represent different OM properties, indicative of the prevailing environmental conditions in their catchments.

Table 3.1 summarizes the water treatment sites included in this study with a description of typical catchment land cover. The typical land use was calculated on the basis of the Corine Land Cover 2000 dataset obtained from the European Environment Agency (Corine Land Cover, 2007). The predominant type of catchment land use is arable land and pasture. The largest area covered by arable land, 65 %, is typical of the River Leam catchment (Campion Hills and Draycote), whereas the highest percentage of pastures (76 %) is observed in the River Churnet catchment (Tittesworth). The smallest area covered by cultivated land is found in the Derwent to confluence with Wye catchment (Bamford) with significant contribution of moors, heathland and natural grassland (36.8 %). Likewise, the Derwent to confluence with Markeaton Brook catchment (Little Eaton) exhibits a high percentage of urban features (21 %)

and green urban areas (21 %). Large urban areas are also present in the following catchments: Trent to confluence with Soar (11 %, Church Wilne), Soar to confluence with Kingston Brook (11 %, Cropston), Trent to confluence with Derwent (10 %, Melbourne), and River Amber (10 %, Ogston). Additionally, the Lower Blythe catchment (Whitacre) presents a very high percentage of industrial, transport or commercial units (32 %). Forested areas are observed in the Lower Blythe catchment (13 %, Whitacre), Lower Severn catchment (11 %, Mitcheldean), and River Churnet catchment (10 %, Tittesworth).

The size of the catchment and the location of the treatment work within the catchment are also important factors influencing the OM properties (Table 3.1). In the smaller catchments, the local conditions (i.e. geomorphology, soils, land use) more strictly control the OM properties compared to the larger catchments where the environmental factors are more dispersed. Moreover, the anthropogenic impacts have greater significance to water quality and OM composition in the smaller catchments. Likewise, the location in the upper part of the catchment relatively increases the importance of the environmental factors, whereas in the lower catchments the effects of the anthropogenic impacts can be superimposed and affect the OM quality to a greater extent. The largest catchments are Upper Mid Severn (1,161.5 km², Shelton and Trimpley) and Lower Severn (844.4 km², Mitcheldean), whereas the smallest are Lower Blythe (0.9 km², Whitacre), Trent to confluence with Soar (7.2 km², Church Wilne), and Derwent to confluence with Markeaton Brook (15.2 km², Little Eaton).

Of note is the fact that some WTWs are located in the same catchment, e.g. Champion Hills and Drayctote (River Leam), Mythe and Strensham (Lower Avon), Shelton and Trimpley (Upper Mid Severn). Thus, the OM character and variation at those sites can be expected to be related.

Furthermore, the WTWs differ in the way the raw water is abstracted and stored prior to treatment. The largest group, with abstraction from the river and storage within a raw water reservoir, comprises eight treatment sites (Campion Hills, Church Wilne, Cropston, Draycote, Melbourne, Ogston, Trimpley, and Whitacre). Direct abstraction from the river is present at 5 sites (Little Eaton, Mitcheldean, Mythe, Shelton, Strensham), whereas abstraction from and storage in an impounding reservoir occurs at Bamford, Frankley and Tittesworth. The sites with direct abstraction from the river are prone to a larger variation of OM quantity and quality compared to the reservoir sites. At the latter, due to longer water residence time, the OM character is more stable and uniform over time.

The catchment conditions control both the quantitative (TOC) and qualitative (SUVA) properties of OM at the WTWs studied (Table 3.2). During the study period, the highest TOC concentrations were observed at Cropston, Bamford, Tittesworth and Draycote WTWs (7.8, 7.3, 6.9, and 6.2 mg/l respectively). However, Bamford and Tittesworth demonstrated entirely different OM character measured as SUVA compared to Cropston and Draycote WTWs. The high SUVA values (5.3 and 3.6 mg/l.m respectively) and the predominant hydrophobic OM character at Bamford and Tittesworth sites result from the high percentage of extensively utilized pastures and natural vegetation areas (forests, moors, heathland) and relatively low pressure from the anthropogenic impacts. On the contrary, Cropston and Draycote represent sites with high TOC and low SUVA (2.9 and 2.5 mg/l.m) indicative of the predominance of more hydrophilic OM derived from arable land and anthropogenic pollution. The lowest TOC concentrations were predominant at Frankley, Mitcheldean, and Little Eaton WTWs (2.8, 2.9, and 3.1 mg/l respectively) with relatively high SUVA values (4.2, 3.0, and 3.0 mg/l.m respectively). The OM properties at Frankley demonstrate a high degree of hydrophobicity

and low TOC that can reflect the predominance of the humic substances in the input from the catchment, however, the overall OM quantity is low.

Overall, the low SUVA values correspond with more diverse land use types and higher contribution of urban fabric and arable farmland, whereas high SUVA is observed in the catchments with less transformed land cover, i.e. extensive pastures, natural grasslands and moors.

While the specific catchment properties control the OM concentrations and character on a spatial basis, the rainfall pattern produces a variation in OM quantity and quality over time. The rainfall pattern is an important factor affecting DOM concentrations and character. In previous studies (Maurice et al., 2002; Parsons et al., 2004; Baker et al., 2008) it was found that prolonged rainfall and high river discharge increase the overall DOC concentrations and the degree of aromaticity of OM. The rainfall pattern during the course of the investigation was analysed on the basis of data obtained from the Met Office from Sutton Bonington meteorological station (Figure 3.3). Sutton Bonington station (52.84 N, 01.25 W, 43 m above sea level) is located south of Nottingham in the vicinity of 6 miles from Church Wilne, 7 miles from Cropston, 8 miles from Melbourne, and 12 miles from Little Eaton WTW. Thus, this station represents the meteorological conditions for the treatment sites located in the Lower Trent catchment. In the period between July 2006 and March 2007, the total monthly rainfall was similar to the average monthly sums calculated for years 1961-2008, with the exception of August 2006 and February 2007 when the total rainfall was higher than average (161 % and 167 % of average rainfall sum respectively). In April 2007 recorded the total rainfall was significantly lower compared with the average sums for period 1961-2008 (14 % of average rainfall for April). The period between May 2007 and July 2007 was one of the wettest in the last century, with the monthly totals significantly exceeding average rainfall

sums (228 %, 255 %, and 235 % respectively). In the following quarter (August 2007-October 2007) the total monthly rainfall was distinctively lower than average (83 %, 59 %, and 40 % respectively). Towards the end of the study period, the total rainfall was similar to average values apart from in January 2008 when 187 % of the normal rainfall was observed. The rainfall conditions in the Lower Severn catchment (station Ross-on-Wye, 5 miles from Mitcheldean WTW) were similar to the northern catchments, except slightly higher sums in the period between October and December 2006 (Figure 3.4). Overall, it can be expected that during months with high rainfall, August 2006, February 2007, May – July 2007 and January 2008, higher DOC concentrations and more aromatic OM were predominant, affecting both the reactivity and treatability of OM.

A typical treatment sequence at Severn Trent WTWs consists of screening, pH adjustment (if required), coagulation and flocculation, clarification via dissolved air flotation (DAF) or hopper bottomed clarifiers (HBC), filtration, adsorption and disinfection. Predominantly, the coagulant type used at Severn Trent Water WTWs is Ferripol XL (Bamford, Church Wilne, Frankley, Little Eaton, Melbourne, Mitcheldean, Tittesworth, Trimpley, Whitacre). Other coagulants are less frequent, i.e. Ferripol 125S (Ogston), ferric chloride (Campion Hills), ferric sulphate (Draycote), and aluminium sulphate (Cropston, Mythe, Shelton, Strensham). DAF is used at Cropston, Draycote, Frankley, Melbourne, Ogston (new works), and Tittesworth WTWs, whereas HBCs are used at Bamford, Campion Hills, Mitcheldean, Mythe, Shelton, Strensham, and Whitacre. Both DAF and HBCs are used at Church Wilne and Trimpley WTWs, whereas lamella plate clarifiers are utilized at Little Eaton WTW. Granular activated carbon (GAC) adsorption prior to chlorination is used at all sites but three (Bamford, Frankley, and Tittesworth).

The works' performance did not undergo any significant changes over the study period as illustrated by the minor variations in treatment parameters (coagulant dose, clarification pH, final water chlorine residual and final water pH) shown in Table 3.3. The coagulant dose is adjusted according to the current raw water quality, and in general is higher when TOC concentrations are higher. The highest variation in the coagulant dose was observed at Mythe, Little Eaton, and Trimpley WTWs as indicated by the high standard deviation values (4.7, 4.2, 3.9 mg/l respectively; Table 3.3.). More stable coagulant doses during the study period were recorded at Campion Hills, Draycote, Tittesworth (standard deviation of 0.1, 0.1, 0.3 mg/l respectively).

3.2. LABORATORY METHODS

Organic matter fluorescence was measured using a Cary Eclipse Fluorescence Spectrophotometer (Varian, Surrey, UK), by scanning excitation wavelengths from 200 to 400 nm in 5 nm steps, and detecting the emitted fluorescence in 2 nm steps between 280 and 500 nm. Excitation and emission slit widths were set to 5 nm and photomultiplier tube voltage to 725V. In order to maintain the consistency of measurement conditions, blank scans with a sealed cell containing deionised water and the Raman calibration, were run systematically following the procedure described by Baker (2001). In the Raman calibration, the average intensity of Raman line of water at 348 nm excitation wavelength is measured and provides a basis for quantitative comparison between different fluorescence measurements. The Raman value during the study period changed from 22.2 to 28.2 with the mean value of 25.0 intensity units (S.D. = 0.68) (Figure 3.5). To enable comparison between samples, all the fluorescence intensities were corrected and calibrated to a Raman peak intensity of 20 units at 396 (392-400) nm emission wavelength. Water samples were not filtered prior to analytical analyses to

reduce the perturbation in NOM composition. Organic matter fluorescence was measured on unfiltered samples in 4 ml cuvettes. The water samples demonstrated low turbidity and low colour. Therefore, no dilution was performed on water samples as the overall inner-filter effect was negligible. This is in accordance with the findings of Hudson et al. (2007) who showed that for freshwater samples with TOC concentrations below 25 ppm the inner filter effect is not observed.

In fluorescence analysis, manufacturer generated corrections for instrument-specific wavelength bias are commonly applied to compare results obtained with different spectrophotometers (McKnight et al., 2001; Baker et al., 2008). However, in this study all fluorescence analyses were performed on the same Cary-Eclipse spectrophotometer and hence fluorescence data was not corrected. The instrumental corrections for emission and excitation wavelengths generated for the spectrophotometer used in this study are presented in Figure 3.6. The emission correction factors varied between 2.5 at 280 nm to 1.0 at 420 nm, whereas the excitation correction varied between 1.49 at 270 nm to 1.25 at 344 nm. In addition, Hudson et al. (2009) reported problems with application of the wavelength corrections for the Cary-Eclipse spectrophotometer. They found that the corrections can not be accurately applied at wavelengths shorter than 220 nm and an attempt to apply these correction factors would exclude correction of much of the peak A and peak T2 areas. Moreover, an application of correction factors in peak C and peak T1 regions would possibly result in the following corrections: peak T1 = $\times 1.84 \pm 0.21$ and peak C = $\times 1.37 \pm 0.05$.

Examples of typical EEMs for raw and clarified water are presented in Figure 3.7. Arising from the fluorescence measurements, EEMs were obtained for each water sample, displaying the intensity of fluorescence within the sample against the wavelengths at which excited fluorophores emitted the light. Fluorescence regions can be attributed to both natural

fluorescence (humic- and fulvic-like), defined as peaks A and C and microbial derived organic matter (tryptophan- and tyrosine-like fluorescence, defined as peaks T and B) at shorter emission wavelengths (Coble, 1996). In the standard fluorescence EEM data analysis, the maximum fluorescence intensity for each peak is recorded along with its spectral location: excitation and emission wavelength (“peak-picking”). Here, fluorescence intensities were measured in the following spectral areas: peak A emission <400; 450> and excitation <225; 260>, peak C emission <400; 450> and excitation <300; 355>, and finally peak T was separated into two peaks: peak T1 emission <330; 370> and excitation <225; 235> and peak T2 emission <330;370> and excitation <270; 280>. Peak B was not observed in the samples. A script in Java was written to facilitate basic fluorescence data processing and automated extraction of the fluorescence peak data.

Each month, water samples were collected from the 16 WTWs over a period of two days; the samples being stored cool and in the dark until analysis, between three and seven days from collection. Storage test experiments were undertaken to demonstrate that degradation of water samples was insignificant under these storage conditions. In the measurements, water of different organic matter properties was analysed (Frankley – hydrophobic, low TOC, Draycote – hydrophilic, high TOC, and Campion Hills – intermediate organic matter character). For two weeks samples were kept at a constant temperature of 5 °C and fluorescence properties for both peak C and peak T were measured at daily intervals. The fluorescence intensity demonstrated stability for all samples, regardless of the water properties. Typical variations, including both increase and decrease in peak C and peak T (Figure 3.8) fluorescence intensities were 4.4 % and 5.3 % respectively, whereas the typical change in TOC was less than 5 %. Sample stability assessment was carried out on the basis of

a peak-picking approach, and so the results should be treated as an estimation of real changes in each fluorescence region.

To determine the reproducibility of fluorescence measurements, each sample during the storage experiment was measured in triplicate. It was found that fluorescence intensities of peak C and peak T replicated within a range of $\pm 3\%$ and $\pm 5\%$ respectively. The changes in the excitation and emission wavelengths for all three fluorescence peaks did not exceed 4 nm for the replicated measurements which is in accordance with results published previously (Baker et al., 2008). The reproducibility of fluorescence EEM measurements determined by Baker et al. (2008) indicate that the mean error of triplicate analyses was the highest for tryptophan-like fluorescence (10 %, 3 intensity units), whereas for both peak A and C it was below 5 % (15 and 6 intensity units respectively). The shift in excitation wavelength between repeated samples was of 1 nm for peak A and 3 nm for peak C. Similarly, the difference in emission wavelength of the fluorescence replicates was of 3 nm for peak A and 4 nm for peak C.

TOC was measured using a Shimadzu TOC-V-CSH analyser with auto-sampler TOC-ASI-V. Samples were sparged with 2 M HCl to remove all inorganic carbon prior to catalytically-aided platinum 680 °C combustion. The non-purgable organic carbon (NPOC) determination method was employed and the resulting NPOC was calculated as a mean of the three valid measurements. The typical error of the duplicated analyses was less than 10 % indicating sufficient precision of the TOC measurements. Additionally, TOC data were compared with the results from independent TOC measurements carried out by Severn Trent Water. The results were consistent throughout the study period, with correlation coefficient values of $R^2 = 0.81$ and 0.78 for raw and clarified water respectively.

To provide a comparison between fluorescence and absorbance spectroscopy efficacy and to calculate the SUVA parameter, additional data were obtained from Severn Trent Water. UV absorbance was measured in duplicate with the Biochrom Libra S12 Spectrophotometer between 200 and 800 nm, with deionised water as the blank. Samples were filtered through a 0.45 µm membrane. A quartz cuvette with 1.0 cm path length was used.

Furthermore, supplementary data on algal cell counts of three groups (diatoms, green, and a separate group for other algae species) were also collected from raw water at selected WTWs by Severn Trent Water staff. Although the algae species analyses were carried out on raw water samples taken on different dates than the fluorescence samples (up to one week's difference), this ancillary dataset can enable a better understanding of protein-like fluorescence results. Monthly algal data was provided for the period of August 2006 to November 2007.

All data used in this study including EEMs of raw and clarified water for all 16 WTWs (folder EEMs) and dataset containing fluorescence, absorbance, TOC, and supplementary data (file Dataset.xls) were presented in the Appendix A – Data files.

3.3. DATA PRE-PROCESSING

Although fluorescence data contain detailed information on OM properties, the EEMs also contain areas of a redundant and noisy signal that should be removed. Thus, prior to data analysis, the fluorescence data were pre-processed with scripts written in Matlab® 7.0. Firstly, fluorescence spectra were normalized to the Raman scatter peak (at 348 nm excitation wavelength) of deionised water by subtracting the Raman signal from the raw data (Determann et al., 1998; Stedmon et al., 2003).

Furthermore, fluorescence EEMs demonstrate exhibit particular diagonal features related to the presence of a scatter signal. The scatter should be removed before further data analysis to enhance the EEM modelling efficiency of decomposition models such as PCA and PARAFAC (Bahram et al., 2006). The Rayleigh and Raman scatters were removed, assuming that the position of the Rayleigh scatter occurs at the excitation equal to the emission wavelength (first-order) or double excitation (second-order) and the position of the Raman line is at constant energy shift with respect to the first-order Rayleigh scatter (Bahram et al., 2006) (Figure 3.9). The fluorescence regions containing redundant information in areas where excitation wavelengths are larger than emission wavelengths and of low signal to noise ratio for excitation wavelengths less than 240 nm were removed from further analysis by replacement with NaN (Not A Number) (Bro, 1998; Stedmon et al., 2003). The resultant EEMs ranged from 240 to 400 nm excitation and from 280 to 500 nm emission wavelengths respectively. Therefore, the final dataset used in the fluorescence data analysis comprised 625 samples of raw and clarified water and 2515 fluorescence excitation-emission wavelengths.

All Matlab® and Java scripts used in this study for pre-processing of fluorescence data (folder Pre-processing) and determination of peak fluorescence values (file Peak-picking.java) were presented in the Appendix B – Data processing.

3.4. CHAPTER 3 FIGURES

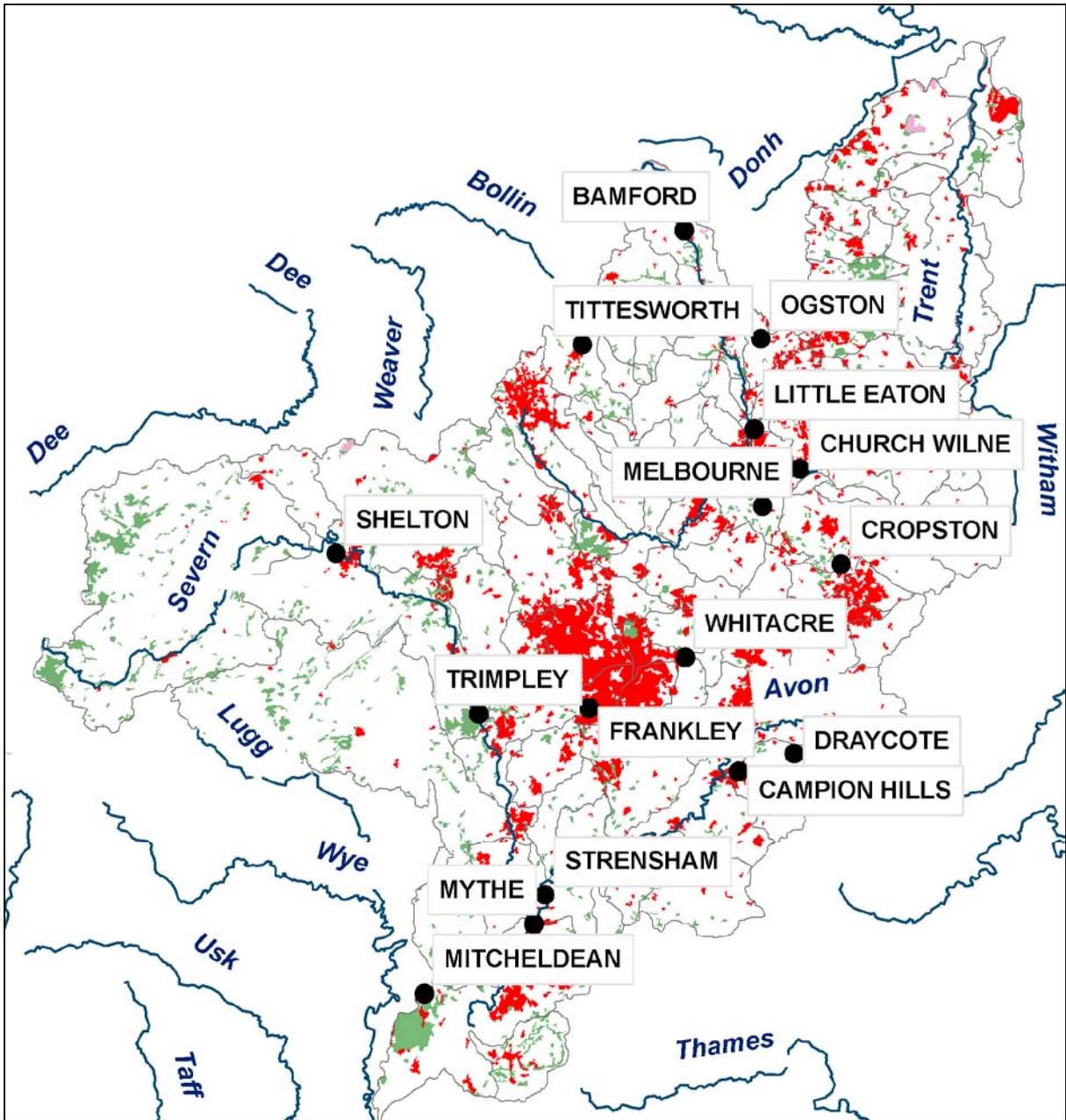


Figure 3.1: Location of the WTWs with the selected land use types: urban fabric (red), forests (green), and wetlands (purple)

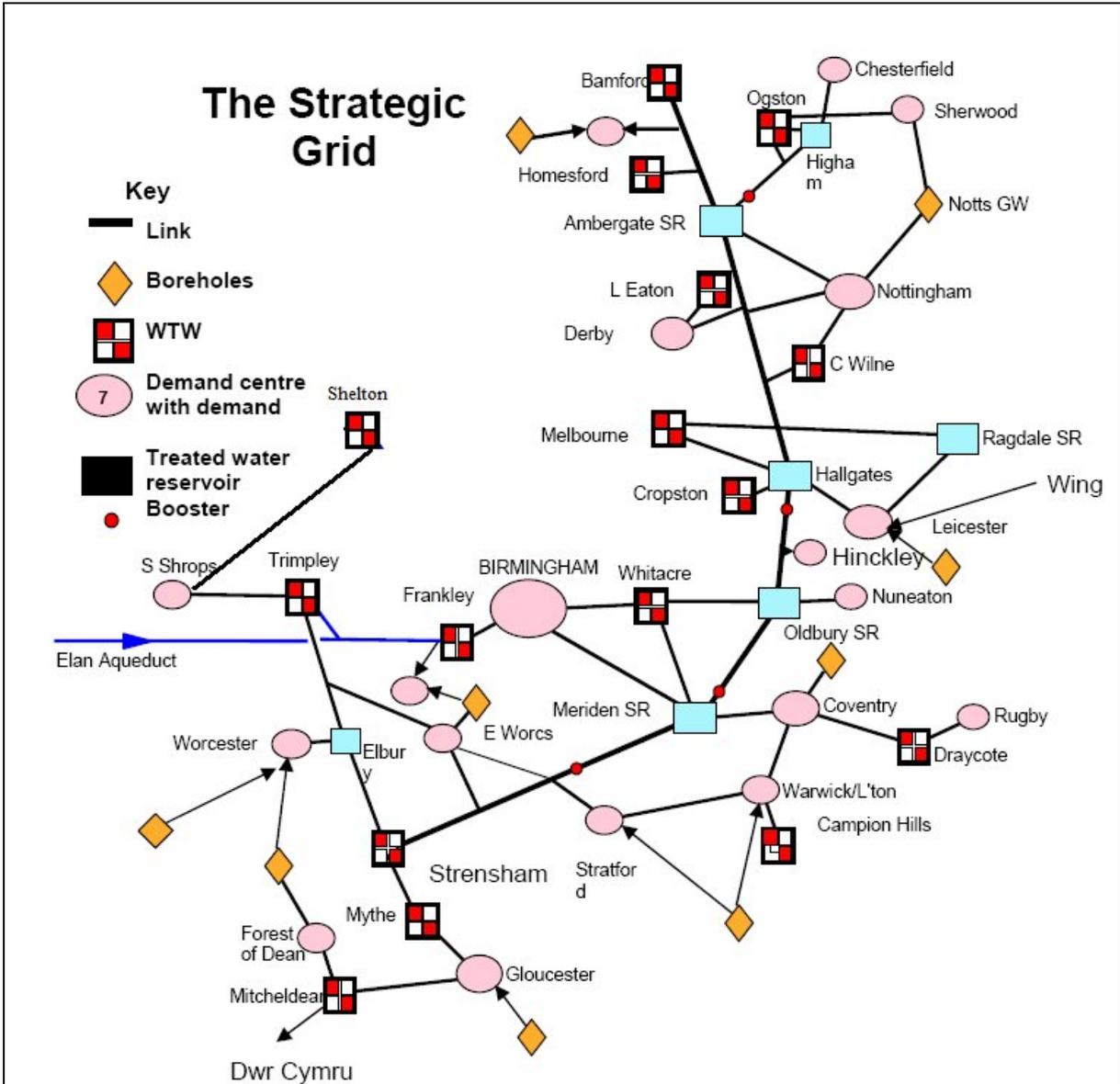


Figure 3.2: Severn Trent Water WTWs network (Severn Trent Water, 2008; modified)

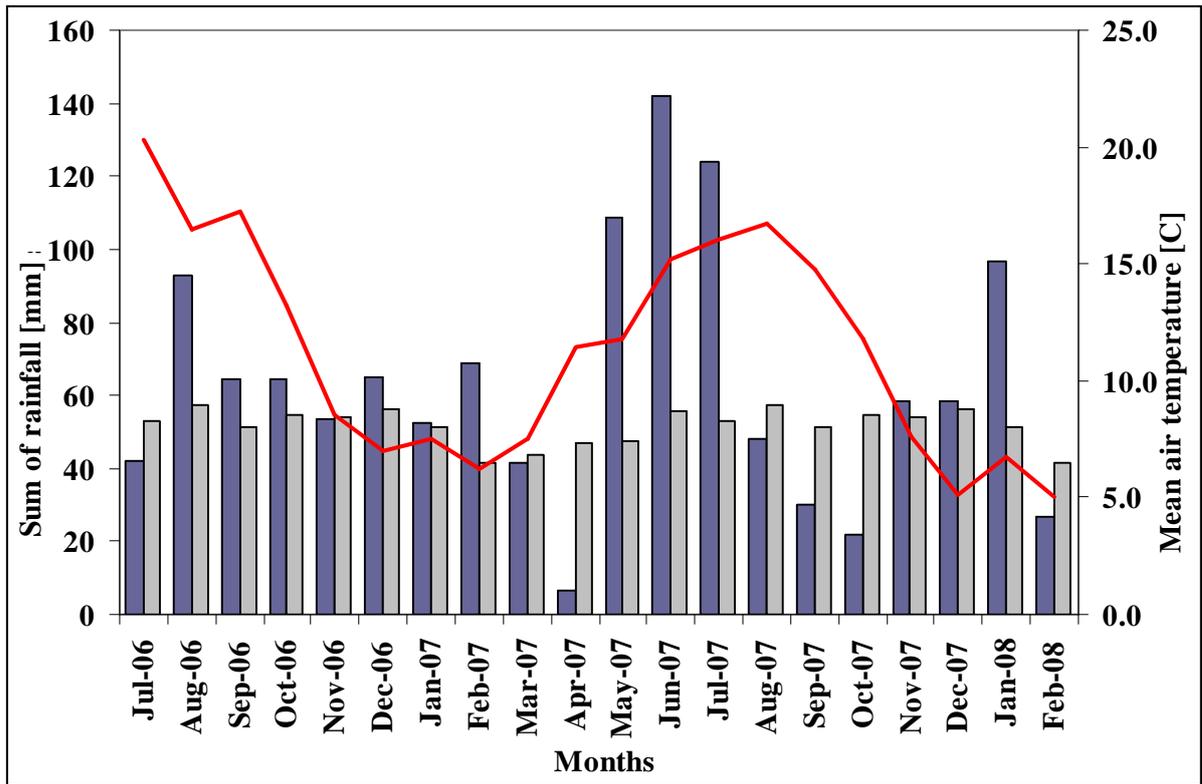


Figure 3.3: Total monthly rainfall at Sutton Bonington station (blue columns), average monthly rainfall (grey columns), and mean air temperature (red line) in the study period (July 06 – February 08) (Met Office data)

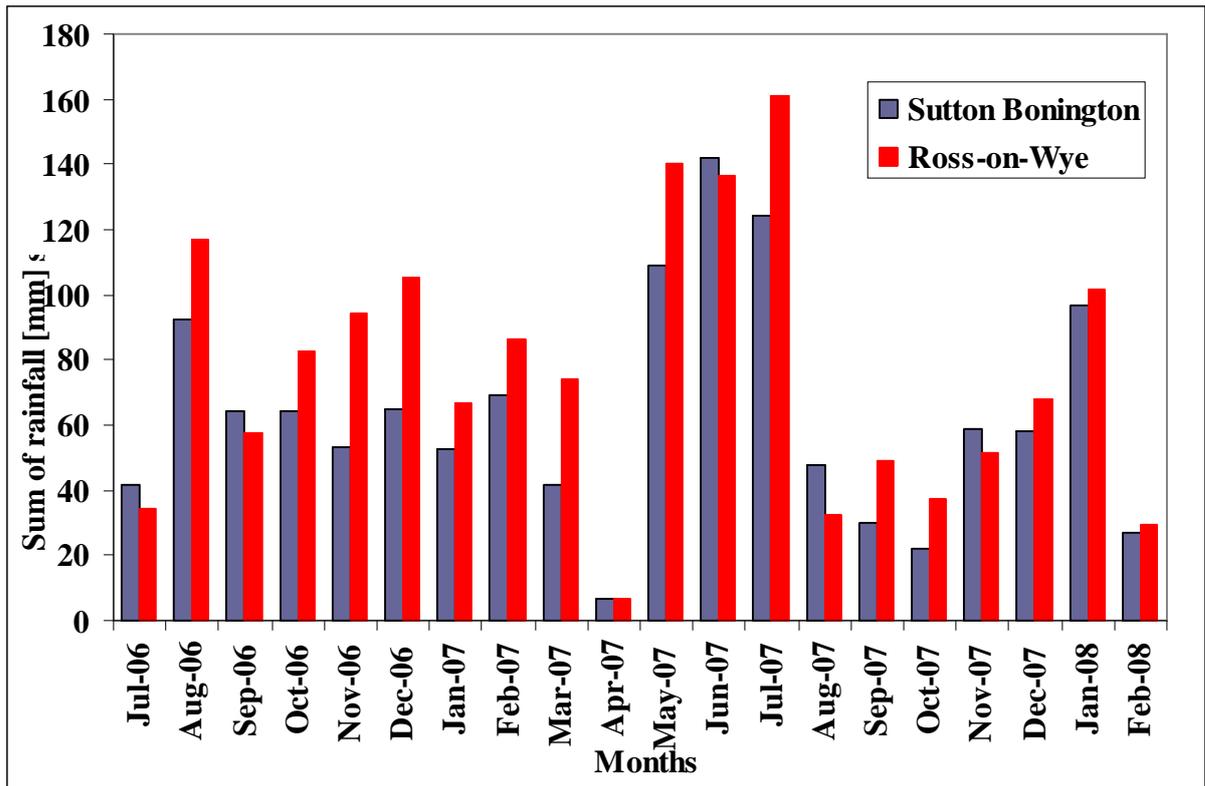


Figure 3.4: Total monthly rainfall at Sutton Bonington station (blue columns), and at Ross-on-Wye (red columns) in the study period (July 06 – February 08) (Met Office data)

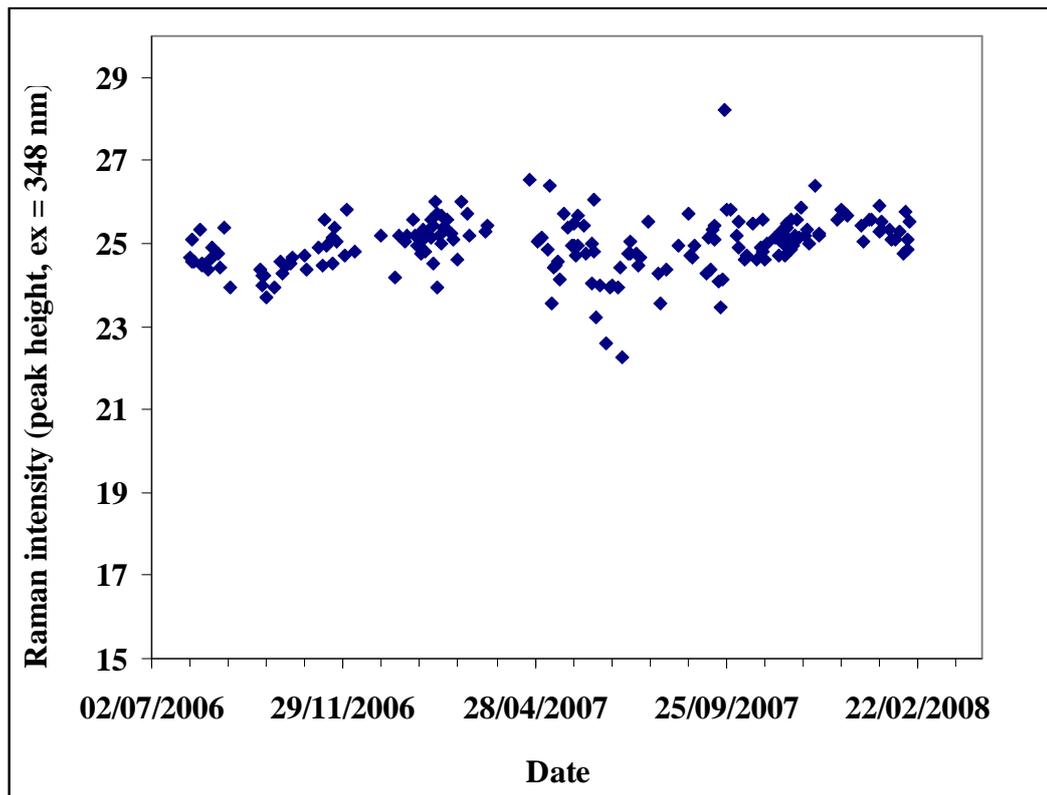


Figure 3.5: Raman peak intensity (au) during the study period

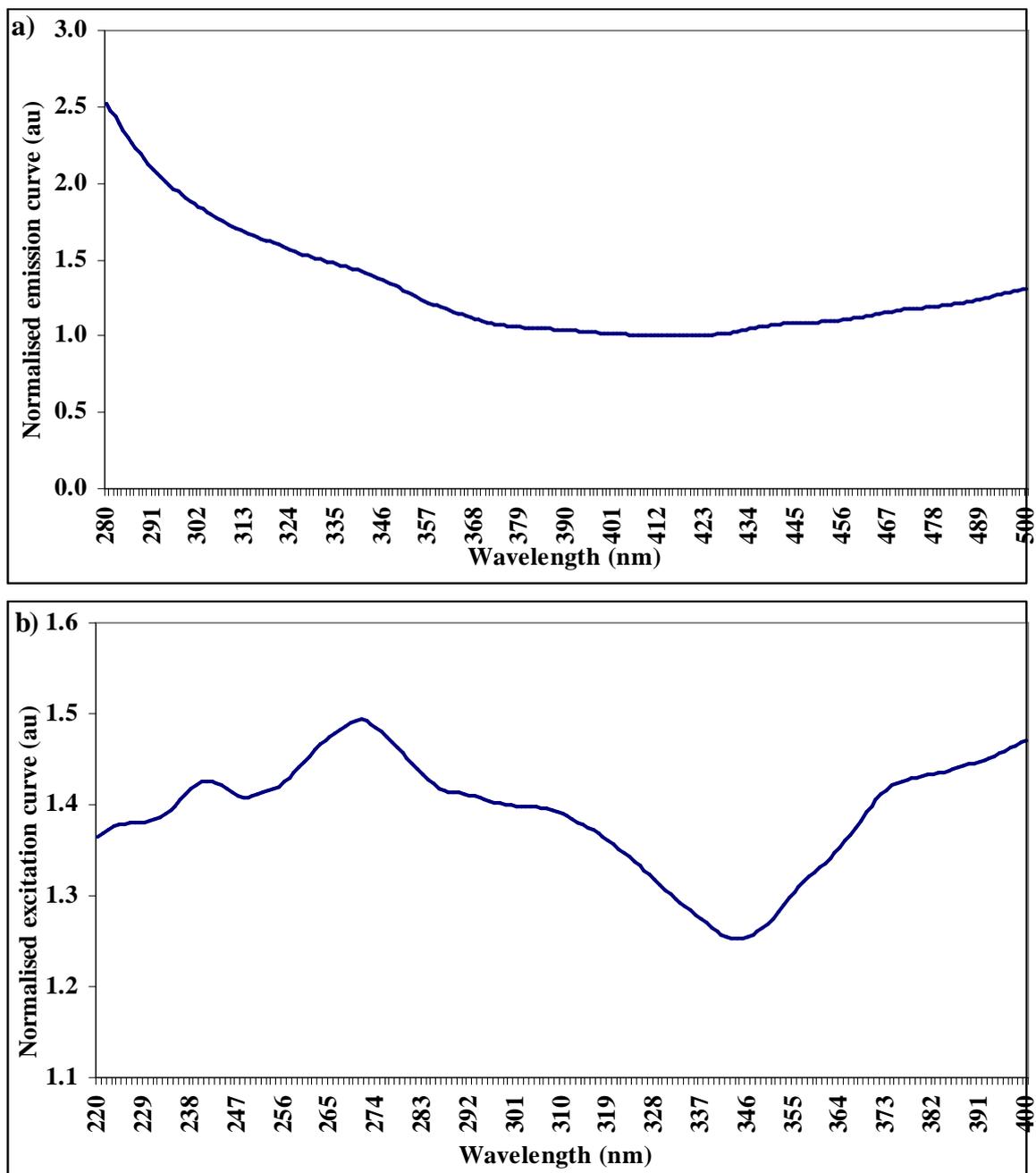
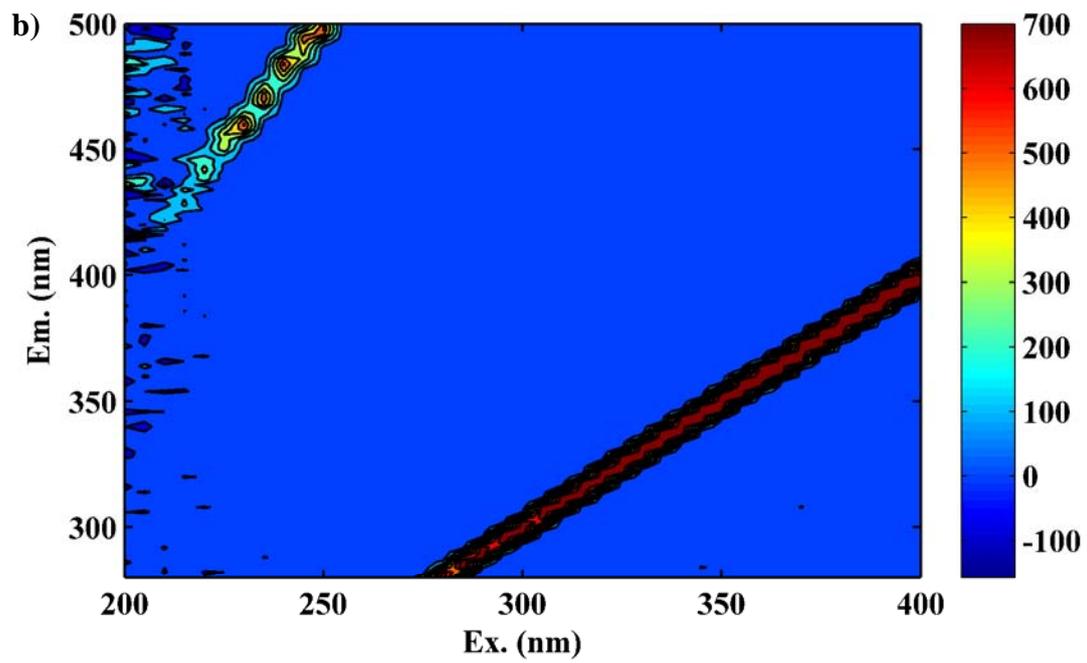
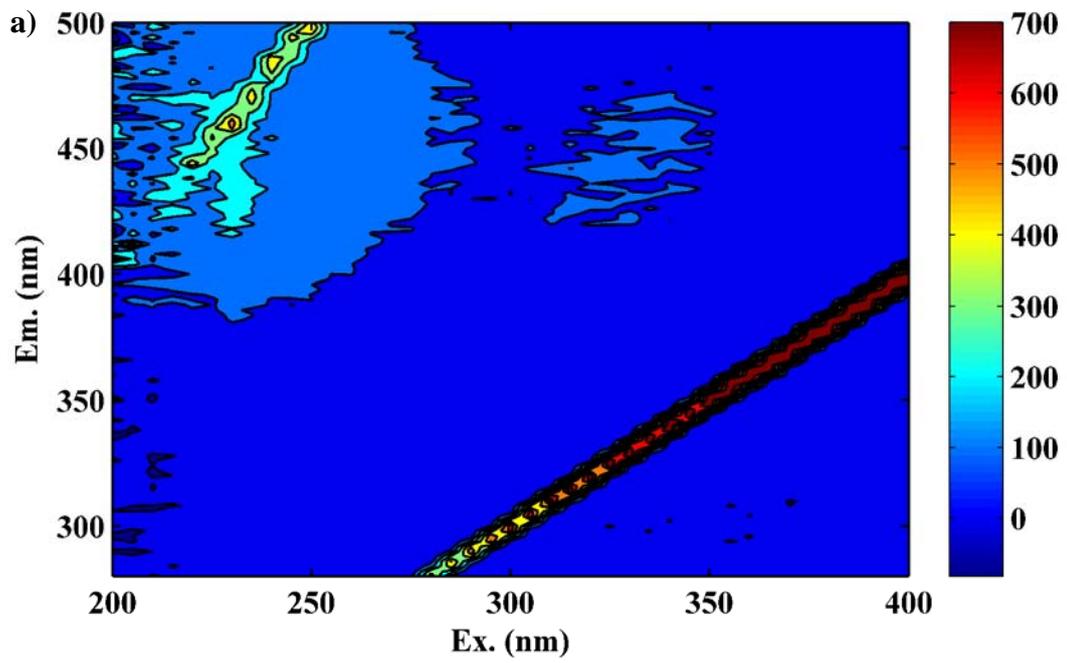


Figure 3.6: Instrumental correction factors for emission (a) and excitation (b) wavelengths



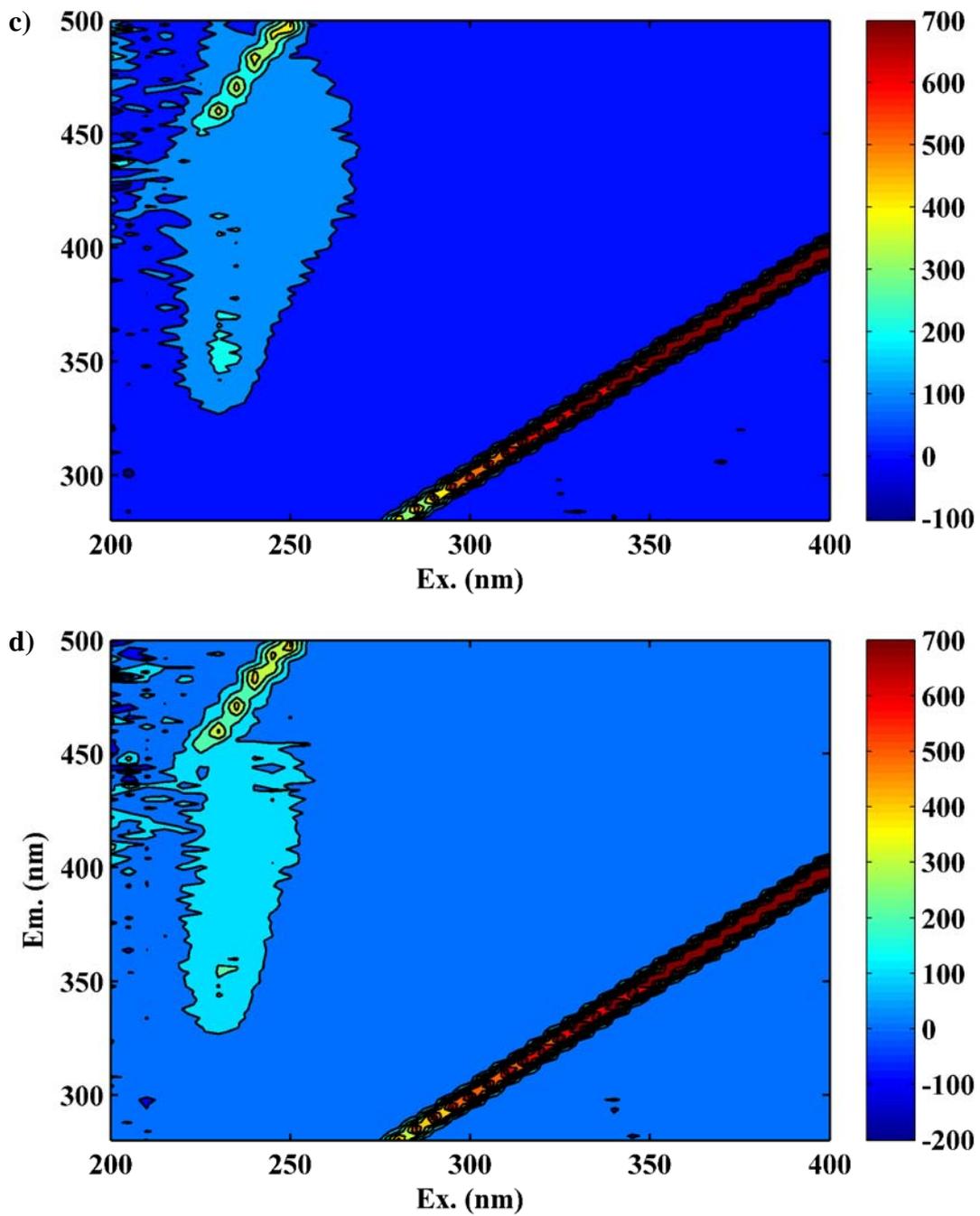


Figure 3.7: Excitation-emission matrices (EEMs) with the Rayleigh-Tyndall scatter for raw and clarified water for Bamford and Little Eaton WTWs. Bamford raw (a) and clarified water (b), Little Eaton raw (c) and clarified water (d)

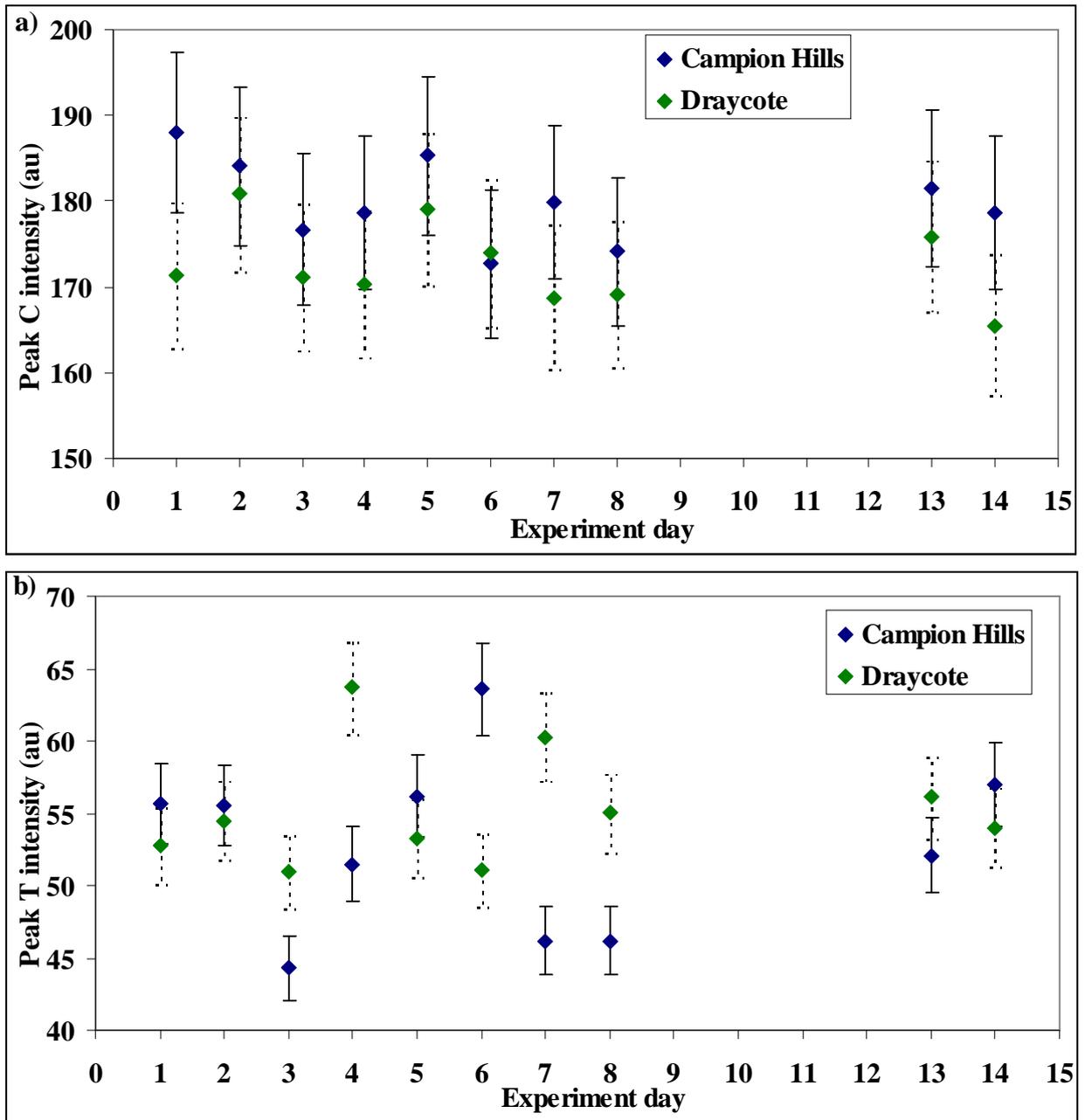


Figure 3.8: Changes in fluorescence intensity of peak C (a) and peak T (b) due to sample storage for Champion Hills and Draycote WTWs. Error bars (dashed line Draycote, solid line Champion Hills)

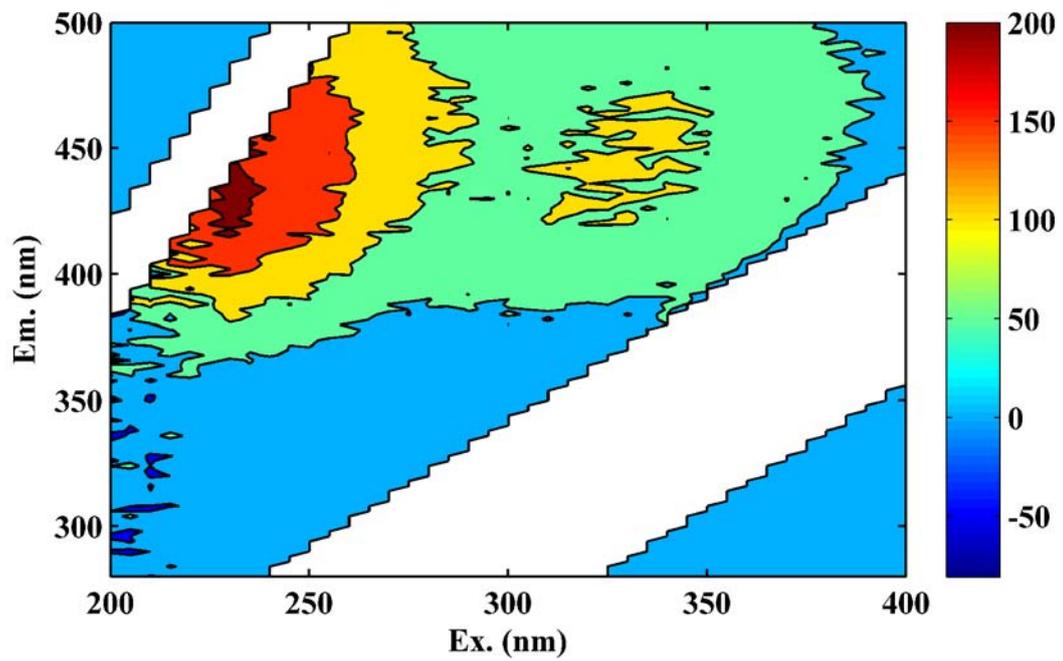


Figure 3.9: Raw water EEM for Bamford WTW after scatter removal (compare with Figure 3.5 a)

3.5. CHAPTER 3 TABLES

Table 3.1: Summary of sampling sites: * - direct abstraction from river to WTW; Typical catchment land use, selected, types, of the largest percentage in total catchment area : A – non-irrigated arable land; P – pastures; C – other cultivated areas; U – urban fabric; I – industrial, transport or commercial units; G – green urban areas; F – forests; O – other areas

No	Site	Catchment	Catchment area (sq km)	Source	Typical catchment land use (main divisions)
1	Bamford	Derwent to confluence with Wye	231.9	Reservoir	P 48%, O 39%
2	Campion Hills	River Leam	372.9	River	A 65%, P 25%
3	Church Wilne	Trent to confluence with Soar	7.2	River	P 44%, C 19%, U 11%
4	Cropston	Soar to confluence with Kingston Brook	283.2	River	A 48%, P 24%, U 11%
5	Draycote	River Leam	372.9	River	A 65%, P 25%
6	Frankley	The Elan Valley	152.1	Reservoir	P 45%, A 33%
7	Little Eaton	Derwent to confluence with Markeaton Brook	15.8	River*	A 26%, U 21%, G 21%, P 21%
8	Melbourne	Trent to confluence with Derwent	265.3	River	A 43%, P 30%, U 10%
9	Mitcheldean	Lower Severn	844.4	River*	P 30%, A 26%, C 13%, F 11%
10	Mythe	Lower Avon	351.1	River*	A 63%, P 24%
11	Ogston	River Amber	145.1	River	P 55%, C 15%, A 11%, U 11%
12	Shelton	Upper Mid Severn	1,161.5	River*	A 38%, P 31%, C 14%
13	Strensham	Lower Avon	351.1	River*	A 63%, P 24%
14	Tittesworth	River Churnet	231.7	Reservoir	P 76%, F 10%
15	Trimpley	Upper Mid Severn	1,161.5	River/ Reservoir	A 38%, P 31%, C 14%
16	Whitacre	Lower Blythe	0.9	River	A 55%, I 32%, F 13%

Table 3.2: Organic matter characteristics: mean TOC: > 6.0 mg/l indicates high TOC, mean TOC < 3.0 mg/l indicates low TOC; mean SUVA > 3.1 mg/l.m indicates hydrophobic OM (HPO), mean SUVA < 3.0 mg/l.m indicates hydrophilic OM (HPI)

Site		Mean TOC		Mean SUVA		
		Mean	S.D.		Mean	S.D.
Bamford	high and variable	7.3	1.8	HPO	5.3	1.2
Campion Hills	high	5.6	1.3	HPI	2.5	0.7
Church Wilne	intermediate	3.3	0.7	HPI	2.9	1.2
Cropston	high	7.8	1.5	HPI	2.9	0.9
Draycote	high	6.2	1.2	HPI	2.5	2.1
Frankley	low	2.8	0.5	HPO	4.2	1.0
Little Eaton	low and variable	3.1	1.0	HPI	3.0	1.0
Melbourne	intermediate	4.1	0.4	HPI	2.9	0.3
Mitcheldean	low and variable	2.9	1.0	HPI	3.0	0.8
Mythe	intermediate	4.5	1.3	HPO	3.1	0.5
Ogston	intermediate	4.8	0.8	HPI	2.8	0.7
Shelton	intermediate	4.0	1.1	HPO	3.3	1.1
Strensham	intermediate	4.4	1.3	HPI	3.0	0.8
Tittesworth	high	6.9	1.0	HPO	3.6	0.7
Trimpley	intermediate	4.7	0.9	HPO	3.3	0.8
Whitacre	high	5.9	1.2	HPI	2.6	0.8

Table 3.3: WTWs' performance parameters during the investigation period

Site	Coagulant dose (mg/l)		Clarification pH		Final water chlorine residual (mg/l)		Final water pH	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Bamford	9.3	0.9	4.5	0.3	0.6	0.0	8.8	0.0
Campion Hills	7.4	0.1	7.4	0.1	0.5	0.2	7.5	0.2
Church Wilne	7.1	0.7	7.3	0.4	0.5	0.0	7.3	0.2
Cropston	8.0	1.3	6.9	0.2	1.0	0.1	7.2	0.2
Draycote	7.9	0.3	7.4	0.1	0.5	0.1	7.2	0.1
Frankley	3.4	0.4	5.1	0.2	0.4	0.0	8.6	0.5
Little Eaton	11.9	4.2	7.0	0.6	0.5	0.0	7.7	0.2
Melbourne	5.8	0.9	7.6	0.1	0.5	0.2	7.6	0.3
Mitcheldean	7.2	2.3	7.3	0.2	0.4	0.1	7.7	0.2
Mythe	4.9	4.7	7.2	0.4	0.5	0.0	7.5	0.1
Ogston	11.0	2.7	7.6	0.9	0.5	0.0	7.7	0.1
Shelton	3.5	0.6	6.7	0.3	0.3	0.0	7.0	0.1
Strensham	3.7	0.5	7.3	0.2	1.2	2.2	7.5	0.1
Tittesworth	11.0	0.3	5.8	0.2	0.6	0.1	7.8	0.3
Trimpley	13.1	3.9	7.0	0.4	0.6	0.0	7.1	0.1
Whitacre	9.9	1.4	7.8	0.2	0.5	0.0	7.2	0.1

4. FLUORESCENCE PROPERTIES OF RAW AND PARTIALLY-TREATED WATER AND ORGANIC MATTER REMOVAL EFFICIENCY

4.1. FLUORESCENCE CHARACTERISATION OF RAW AND CLARIFIED WATER

The fluorescence signature of aquatic organic matter has been shown previously to characterize OM in terms of physico-chemical properties (degree of aromaticity, molecular weight) and contribution of particular OM fractions (microbial OM, humic substances) (Stewart and Wetzel, 1980; Senesi et al., 1991; Kalbitz et al., 1999; Wu et al., 2003a; Belzile and Guo, 2006; McKnight et al., 2001; Baker et al., 2008; Hudson et al., 2007). Different fluorescence EEM properties were found to provide a good discrimination between various OM sources and types. The fluorescence intensity of certain fluorescence peaks was corroborated to characterise quantitative OM properties, whereas the qualitative OM properties have been inferred from the spectral location of fluorescence peaks. In particular, fluorescence peak C emission wavelength was found to indicate the degree of aromaticity of OM and the contribution of highly conjugated aromatic moieties in the overall OM pool (Senesi et al., 1991; Kalbitz et al., 1999; Baker et al., 2008). In the drinking water treatment studies, the OM aromatic fraction was found to demonstrate higher reactivity with disinfectant and better treatability with coagulation compared to the hydrophilic fraction (Fearing et al., 2004; Kim and Yu, 2005; Sharp et al., 2006; Roe et al., 2007). Thus, here the peak C emission wavelength was hypothesized to provide an indication of the hydrophobic/hydrophilic content in drinking water samples and hence provide the discrimination between easier to remove, highly reactive hydrophobic OM and recalcitrant, less reactive hydrophilic fractions. Likewise,

the protein-like fluorescence, particularly in the tryptophan-like region (peak T) was found to correlate with BOD and reflect the presence of autochthonous and low molecular weight OM fraction (Determan et al., 1998; Smith et al., 2004; Nguyen et al., 2005; Elliot et al., 2006; Henderson et al., 2008a). While tryptophan-like fluorescence indicates the predominance of labile organic material, peak C intensity provides information on the quantity of more recalcitrant, allochthonous OM.

Therefore to characterise aquatic OM from 16 WTWs, fluorescence measurements were carried on raw and partially-treated (clarified) water samples. Examples of EEMs of raw and clarified waters for the Bamford and Little Eaton treatment sites are given in Figure 4.1. Those two sites demonstrate contrasting OM properties resulting from different dominant types of land use in their catchments (Bamford – semi-natural with pastures and heathland, Little Eaton – highly urbanised). Bamford WTW abstracts raw water from reservoirs, whereas at Little Eaton water is taken directly from River Derwent. OM derived from Bamford WTW comprises high and variable TOC with the hydrophobic fraction dominating, whereas Little Eaton exhibits low TOC and hydrophilic OM. A visual inspection of raw and clarified fluorescence EEMs for both sites indicated that fluorescence intensity in all spectral regions was significantly reduced during the coagulation and clarification processes. Both raw water fluorescence intensities and the fluorescence intensity reduction between raw and clarified water stages were higher for Bamford WTW. The calculated decrease in peak C fluorescence intensity for Bamford and Little Eaton was 72.6 % and 28.6 %, in peak A 73.6 % and 21.5 %, and in peak T1 45.5 % and 9.2 % respectively. Peak T2 was not observed for Bamford samples, whereas for Little Eaton the reduction in its intensity was 18.2 %. The higher fluorescence intensity decrease for Bamford WTW reflected a higher efficiency of OM removal by coagulation. During the coagulation and clarification treatment stages, the

hydrophobic and high molecular weight OM is preferentially removed. Here, the position of peak C shifted towards lower emission wavelengths, indicating changes in OM functional group composition and a decrease in the degree of aromaticity between raw and clarified water. At Bamford the peak C emission wavelength decreased from 430 to 420 nm, whereas at Little Eaton the corresponding decrease was observed from 420 to 406 nm (Figure 4.1). The higher raw water peak C emission wavelength at Bamford compared to Little Eaton WTW indicated a higher initial degree of aromaticity and the predominance of more hydrophobic organic material. Similarly, a reduction in peak A emission wavelength was also observed, from 440 to 420 nm for Bamford and from 418 to 408 nm for Little Eaton WTW. The location of tryptophan-like fluorescence for both treatment sites did not change significantly between raw and clarified water samples.

The clarified water fluorescence signature demonstrates that OM fraction which is recalcitrant to removal by coagulation. It can be seen that humic-like fluorescence (peak A and C) is removed more effectively than the protein-like fluorescence (peak T1 and T2) (Figure 4.1 b and d). Furthermore, the residual fluorophores are more hydrophilic as indicated by the lower emission wavelengths of the humic-like fluorescence regions.

Figure 4.2 shows the relationships between selected fluorescence properties for raw and clarified waters derived from fluorescence data for all 16 WTWs. The intensities of peak C and T2 were found to be correlated with correlation coefficient $R^2 = 0.54$ for raw water and $R^2 = 0.85$ for clarified water (R^2 values presented in this study are significant at 0.05 level; Figure 4.2 a and b). In related studies, fluorescence intensity has been shown to depend on the composition of functional groups within the OM pool. The electron-donating groups including hydroxyl and amine exhibit higher fluorescence intensities, whereas the electron-withdrawing functional groups have relatively lower fluorescence intensities (i.e. the

carboxylic functional group) (Senesi, 1991; Świetlik and Sikorska, 2004). Thus, the stronger correlation between peak T and peak C fluorescence intensity for clarified water may result from a more uniform composition of fluorophores in the residual OM.

The highest fluorescence intensities in peaks T and C of the raw OM (> 70 au for peak T and > 200 au for peak C) were observed for Whitacre, Cropston, and Tittesworth WTWs, whereas the lowest were typical of Frankley, Bamford, and Shelton (< 50 au for peak T and < 200 au for peak C). The same pattern was shown for clarified water samples albeit at lower fluorescence intensities. The observed significant variation in the clarified peak T intensities at Little Eaton and Church Wilne WTWs (Figure 4.2 b) was related to an increased level of biological activity. In an analysis of supplementary algal count data, for a number of sites, a strong relationship was observed between peak T fluorescence and diatoms (Little Eaton $R^2 = 0.36$, Campion Hills $R^2 = 0.40$, Church Wilne $R^2 = 0.36$), green algae (Campion Hills $R^2 = 0.39$, Church Wilne $R^2 = 0.44$, Cropston $R^2 = 0.68$, Draycote $R^2 = 0.50$), other algae taxons (Little Eaton $R^2 = 0.93$, Cropston $R^2 = 0.38$), and between peak T fluorescence and total algae (Campion Hills $R^2 = 0.44$, Church Wilne $R^2 = 0.38$, Draycote $R^2 = 0.58$).

Two reservoir-fed WTWs with a typical high degree of OM aromaticity, Bamford and Frankley, demonstrated a distinctively different pattern of changes in raw water peak C intensity. The variation in peak C intensity was independent of changes in peak T intensity, as the latter was low and stable throughout the course of the investigation (Figure 4.2 a). Furthermore, Bamford WTW demonstrated higher peak C intensities in relation to the corresponding peak T intensities. This can reflect significantly different functional groups composition in OM derived from the Bamford WTW catchment compared to the other sites. Mobed et al. (1996) demonstrated that highly substituted aromatic moieties of humic acids

have lower fluorescence intensities than fulvic acids derived from similar sources as a result of self-quenching mechanisms.

To indicate to what extent the trends observed in the two-way analysis presented in Figure 4.2 are within or between sites, the nonparametric Kruskal-Wallis ANOVA was carried out. The Kruskal-Wallis ANOVA can determine whether or not the values of particular fluorescence variable differ between the treatment sites. In the Kruskal-Wallis ANOVA, the null hypothesis of no difference between treatment sites was tested against the alternative hypothesis that a significant difference exists between the sites (at 0.05 confidence level). Here, for all fluorescence variables $p = 0.000$ was determined and thus the null hypothesis was rejected in favour of the alternative hypothesis indicating significant differences in the fluorescence properties between the treatment sites (Table 4.1).

Figures 4.2 c and d plot peak C emission wavelength and the ratio of peak T2 to peak C intensity ($T2_{int}/C_{int}$ ratio) for raw and clarified water respectively. As discussed above, peak C emission wavelength provides the indication of OM aromaticity, whereas the $T2_{int}/C_{int}$ ratio denotes the ratio of the labile to recalcitrant OM fractions. The correlations were weak and only significant for raw water ($R^2 = 0.38$, Figure 4.2 c). Furthermore, the discrimination between sampling sites was poorer compared to the relationship between peak C and peak T intensities. Thus, Figure 4.2 c suggests that samples with a higher degree of aromaticity comprise more recalcitrant OM as indicated with the low values of $T2_{int}/C_{int}$ ratio. Moreover, for sites with highly aromatic OM (Bamford and Frankley), the $T2_{int}/C_{int}$ ratio was constant throughout the study period suggesting that the predominant OM character in their catchments was independent of varying meteorological conditions. On the contrary, Little Eaton and Church Wilne WTWs demonstrated a very high content of labile, autochthonous OM in raw water samples (Figure 4.2 c). The fluorescence properties of the clarified water were more uniform suggesting that OM character of the residual fraction after the treatment with

coagulation was similar and independent of raw water properties. These results are in accordance with the results presented in the previous studies (Fabris et al., 2008).

To provide an indication of the average fluorescence OM properties during the study period, a mean peak C emission was plotted against mean $T_{2_{int}}/C_{int}$ ratio for each treatment site and water type (Figure 4.3 a and b). Bamford and Frankley WTWs showed distinctively different raw water OM properties as indicated by the high mean peak C emission wavelength and low mean values of $T_{2_{int}}/C_{int}$ ratio (449 nm, 0.23 for Bamford and 436 nm, 0.32 for Frankley) (Figure 4.3 a). Ten WTWs exhibited similar raw water OM character, with the mean peak C emission wavelength value of ~425 nm and $T_{2_{int}}/C_{int}$ ratio value of ~0.46 (Campion Hills, Cropston, Melbourne, Mitcheldean, Mythe, Ogston, Strensham, Tittesworth, Trimpley, and Whitacre). Shelton WTW demonstrated intermediate raw water OM properties (430 nm peak C emission wavelength and 0.38 $T_{2_{int}}/C_{int}$ ratio). Finally three sites, Draycote, Church Wilne, Little Eaton, showed distinctively higher values of mean $T_{2_{int}}/C_{int}$ ratio (above 0.50), with low mean peak C emission wavelength indicating the predominance of hydrophilic OM fraction. The distinctive OM properties of these WTWs can be related to the increased biological activity of the raw waters. Compared to the raw water OM properties, clarified water demonstrated slightly higher values of the mean $T_{2_{int}}/C_{int}$ ratio and lower mean peak C emission wavelengths for all treatment sites (Figure 4.3 b). This relative increase in the content of more labile OM can be explained by better removal of the allochthonous OM fraction denoted by peak C fluorescence. As observed previously, the clarified water OM properties were similar for most of the treatment sites (peak C emission wavelength <417 nm; 426 nm> and $T_{2_{int}}/C_{int}$ ratio <0.44; 0.56>) with the exception of Bamford, Church Wilne and Little Eaton WTWs. Subsequent to coagulation, Bamford exhibited relatively more hydrophobic OM character, whereas Church Wilne and Little Eaton demonstrated higher

values of the mean T_{2int}/C_{int} ratio. The greatest change in OM properties between raw and clarified water was observed for Bamford (change in mean peak C emission wavelength $\Delta_{Cem} = 19$ nm and change in mean T_{2int}/C_{int} ratio $\Delta_{T2int/Cint} = 0.20$), Frankley ($\Delta_{Cem} = 12$ nm, $\Delta_{T2int/Cint} = 0.12$), Tittesworth ($\Delta_{Cem} = 8$ nm, $\Delta_{T2int/Cint} = 0.10$), and Little Eaton WTW ($\Delta_{Cem} = 7$ nm, $\Delta_{T2int/Cint} = 0.14$). The lowest degree of OM character change between raw and clarified water was typical of Melbourne ($\Delta_{Cem} = 0$ nm, $\Delta_{T2int/Cint} = 0.04$), Mitcheldean ($\Delta_{Cem} = 0$ nm, $\Delta_{T2int/Cint} = 0.07$), and Strensham ($\Delta_{Cem} = 1$ nm, $\Delta_{T2int/Cint} = 0.08$). Virtually no change in the T_{2int}/C_{int} ratio between raw and clarified water was found for Draycote WTW, suggesting that the predominant OM at this site is highly hydrophilic and recalcitrant to removal by coagulation.

4.2. ORGANIC MATTER REMOVAL EFFICIENCY

Organic matter removal by coagulation calculated as a reduction in TOC between raw and clarified water was correlated with quantitative and qualitative fluorescence properties. Both fluorescence and TOC measurements were carried out independently on paired samples. TOC is an accepted indicator of organic matter concentration and of DBPs precursor material in water. There exists the need for a technique which can be used accurately and quickly at WTWs and which will give a satisfactory indication of the THMs concentration leaving the works (Banks and Wilson, 2002). Considering the need to infer and minimise THMs levels in treated water, previous work has used a variety of approaches to optimise treatment and infer TOC removal at WTWs. These approaches included the standard jar test (Uyak and Toroz, 2006), UV-Vis absorbance spectroscopy (Banks and Wilson, 2002; Iriarte-Velasco et al., 2006) and zeta potential measurements (Sharp et al., 2006). However, each of these approaches had limitations in precision or ease of use. For example, Cheng et al. (2004)

showed that fluorescence was more stable than UV in detecting DOC removal from reservoir water. Therefore, to address these limitations, work was undertaken to assess the potential for using fluorescence spectroscopy to predict TOC removal and so infer likely THMs concentrations.

Previously, peak C fluorescence intensity has been shown to exhibit a general correlation with TOC (Hudson et al., 2007; Cumberland and Baker, 2007). Here, a good overall relationship was found between peak C intensity and TOC concentrations with the correlation coefficients $R^2 = 0.68$ for raw and $R^2 = 0.77$ for clarified water (Figure 4.4 a and b). The lowest raw water TOC concentrations and peak C intensities were observed for Frankley, Mitcheldean, and Little Eaton WTWs (< 5 mg/l and < 140 au), whereas the highest were predominant at Bamford, Cropston, Tittesworth and Draycote (> 4 mg/l and > 140 au). With increasing OM concentrations, an increased degree of variability of OM concentrations was observed. For Draycote, Bamford and Cropston, the observed TOC concentrations were higher (or peak C intensities lower) compared to the values resulting from the TOC/peak C intensity relationship (Figure 4.4 a). A different explanation of this tendency can be inferred for the particular WTWs. The predominance of highly substituted aromatic functional groups in the OM pool and the presence of fluorescence self-quenching mechanisms can determine the relatively lower peak C fluorescence intensities for Bamford raw water samples. At Draycote and Cropston WTWs, a weaker relationship between TOC and peak C intensity can result from a higher contribution of autochthonous, microbial fraction, which increases the TOC and peak T intensity but without corresponding changes in peak C intensity. Likewise, for Tittesworth WTW the measured peak C intensities of raw water were higher than the expected corresponding TOC concentrations. This might be a result of the predominance of functional groups increasing the fluorescence intensity, i.e. electron-donating groups

including hydroxyl and amine. For the clarified water, a better relationship between peak C intensity and TOC was obtained suggesting more uniform OM character post coagulation compared to raw water (Figure 4.4 b). A decrease in both TOC concentrations and peak C intensities of clarified water was observed compared to the corresponding raw water. The lowest clarified water TOC concentrations and peak C values (< 3 mg/l and < 60 au) were observed for Bamford, Frankley and Mitcheldean, whereas the highest values (> 3 mg/l and > 100 au) were observed for Campion Hills, Draycote and Whitacre. Similar to the raw water properties, Draycote WTW demonstrated lower peak C intensities than values resulting from the TOC/peak C intensity relationship.

The strong relationship between raw and clarified peak C fluorescence and TOC suggests that the decrease in peak C intensity can be correlated with the OM removal efficiency measured as TOC reduction. The OM removal between raw and clarified water denotes the removal by coagulation of the hydrophobic fraction which has the highest DBPs formation potential. It also demonstrates the content of the hydrophilic fraction which is more recalcitrant to coagulation removal and yet which still has to be effectively removed prior to the disinfection stage to minimise the risk of DBPs formation. Thus, here for each WTW, OM removal was calculated from the decrease in organic matter fluorescence intensity of peak C between raw and clarified water samples and compared with the actual reduction in TOC concentrations. A strong, linear correlation was observed between fluorescence-derived and TOC-measured OM removal between raw and clarified water with the correlation coefficient $R^2 = 0.91$ (Figure 4.5 a). The relationship was developed on the basis of fluorescence data obtained for all sample sites and thus enables direct comparison of the OM removal efficiency between different WTWs. The efficiency of OM removal prediction with fluorescence EEM spectroscopy was compared with the regression model developed from the UV absorbance

measured at 254 nm between raw and clarified water (Figure 4.5 c). The decrease in absorbance in percentage terms was plotted against the TOC removal, however a poorer correlation (coefficient value of $R^2 = 0.82$) was obtained compared with the fluorescence-derived model (Table 4.2). In comparison with fluorescence spectroscopy, UV absorbance has been reported to be a less selective technique (Eaton, 1995; Wang and Hsieh; 2001; Marhaba et al., 2003), and prone to the underestimation of the OM samples where low molecular weight fraction is predominant (Matilainen et al., 2002). As both techniques measure the spectral properties of OM, there is a significant difference in the intrinsic OM properties determined with fluorescence and UV absorbance (Yang et al. 2008). The highest absorptivities are pertinent to aromatic OM structures, however the fluorescence intensities of the aromatic fraction are dependent on the composition of functional groups in molecules (higher for electron-donating groups like hydroxyl and amine, and lower for electron-withdrawing groups like carboxylic) (Senesi, 1991; Świetlik and Sikorska, 2004).

The analysis of the samples' spatial distribution revealed that the highest OM removals (on average 80 %) were observed at Bamford ($OMrem_{FLU} = 72.7$ % based on fluorescence measurements and $OMrem_{TOC} = 77.1$ % based on TOC measurements), Frankley ($OMrem_{FLU} = 58.0$ % and $OMrem_{TOC} = 62.0$ %) and Tittesworth ($OMrem_{FLU} = 60.2$ % and $OMrem_{TOC} = 65.5$ %) (Figure 4.5 a and Table 4.2). Accordingly, the lowest removals were observed at Melbourne ($OMrem_{FLU} = 15.3$ % and $OMrem_{TOC} = 13.7$ %), Draycote ($OMrem_{FLU} = 17.3$ % and $OMrem_{TOC} = 17.5$ %), Whitacre ($OMrem_{FLU} = 17.7$ % and $OMrem_{TOC} = 20.9$ %), and Church Wilne ($OMrem_{FLU} = 18.3$ % and $OMrem_{TOC} = 19.5$ %).

While OM removal discriminated between particular treatment sites, no explicit seasonal relationship with OM removal was discerned (Figure 4.5 c and Table 4.3). The highest overall TOC removals calculated as a mean value for all WTWs were observed in the period from

December 06 to March 07 and in July 07 and January 08. The maxima recorded in July 07 and January 08 can be plausibly attributed to the increased rainfall and flooding events (Figure 3.3), whereas the winter OM removal maxima can result from the relatively higher contribution of autochthonous, more hydrophobic OM as indicated by the seasonal changes in peak C emission wavelength (Table 4.3). The TOC removal efficiency is generally higher during the flood events due to the higher TOC concentration and during the winter months due to the predominance of easier to remove, more hydrophobic fraction. Correspondingly, the lowest OM removals were observed in September 06, September 07, and December 07. However, for the lowest OM removals the effects of rainfall and the relative contribution of aromatic OM fraction were more equivocal (low sums of rainfall in September 07 but average rainfall conditions in September 06 and December 07).

The spatial and seasonal distribution of the samples in the relationship between TOC removal and UV-Vis removal was similar to the fluorescence-derived relationship (Figure 4.5 c and d). However, for all sites and months higher UV-Vis removals were observed compared to the corresponding fluorescence and TOC removals (Table 4.2 and 4.3). The highest absorbance removals were pertinent to Bamford (90.9 %), Frankley (81.6 %), and Tittesworth (80.5 %), whereas the lowest to Draycote (23.6 %), Melbourne (27.0 %), Church Wilne (27.3 %), and Whitacre (35.6 %) (Table 4.2).

Organic matter removal and DBPs formation potential depend on the organic matter content and character (Stevens et al., 1976; Chen et al., 2008), which can be directly derived from fluorescence measurements. Therefore, the relationship between TOC removal and various potential fluorescence predictors was investigated further. To compare OM character derived from the fluorescence measurements and TOC removal, simple Pearson's correlations were carried out for selected fluorescence (peaks intensity and spectral location) and absorbance

properties (Table 4.4). A statistically significant relationship was found between TOC removal and qualitative fluorescence properties of clarified water: peak T2 intensity ($R^2 = -0.54$), peak C intensity ($R^2 = -0.40$), peak T1 intensity ($R^2 = -0.38$), peak A intensity ($R^2 = -0.36$). Here, the R^2 coefficient value of 0.36 was used as a threshold for significance of the relationship assuming that strong relationship between variables exists for correlation coefficient value of $r = 0.60$ (the value commonly reported in the environmental studies). In the study, corrections for multiple significance testing were not applied however e.g. the Bonferroni correction could be used to account for the errors in the simultaneous determination of a set of statistical inferences. From the raw water predictors, only peak C emission wavelength and SUVA parameter correlated with TOC removal (both $R^2 = 0.36$). The negative correlation between fluorescence intensities of clarified water and TOC removal indicates that better OM removal is typical of sites with lower residual OM content, with no account of the OM character (microbial vs allochthonous). Likewise, the positive TOC removal correlation with both raw water peak C emission wavelength and SUVA reflects the better removal of more hydrophobic raw water OM.

Furthermore, the decrease in all fluorescence peaks and absorbance-derived parameters (UV_{254} , SUVA) between raw and clarified water was compared to TOC removals (Table 4.4). The highest correlation coefficients were found for the peak C intensity removal ($R^2 = 0.91$), peak A intensity removal ($R^2 = 0.84$), and UV_{254} removal ($R^2 = 0.82$). Lower correlation coefficients were observed for other predictors (T2 intensity removal [$R^2 = 0.63$], SUVA removal [$R^2 = 0.49$], and T1 intensity removal [$R^2 = 0.46$]). These results suggest that fluorescence intensities of humic-like fluorescence provide a better indication of TOC removal compared to absorbance-derived parameters. The latter tend to overestimate OM

removal as only the hydrophobic fraction (which is easier to remove by coagulation) that is accounted for by absorbance measurements.

The strength of the relationship between fluorescence predictors and TOC removal was also analysed on a site-specific and monthly basis (Table 4.5 and 4.6). Overall, no significant correlation was found between mean fluorescence properties derived for individual sites and TOC removal (Table 4.5). In particular, a strong relationship was observed for quantitative raw water fluorescence predictors for Mitcheldean (peak C intensity $R^2 = 0.74$, peak T1 intensity $R^2 = 0.60$, peak T2 intensity $R^2 = 0.48$, and $T2_{int}/C_{int}$ ratio $R^2 = 0.46$), Mythe (peak C intensity $R^2 = 0.63$), and Campion Hills (peak C emission wavelength $R^2 = 0.44$). For clarified water predictors only Church Wilne demonstrates a correlation between peak C intensity and TOC removal ($R^2 = 0.37$). This would appear to suggest that for particular sites fluorescence properties of raw or clarified water are not good TOC removal predictors although strong relationships exist between fluorescence properties and TOC concentrations for the entire dataset. This also indicates that for the particular sites, the OM removal efficiency depends on both the OM properties and the treatment conditions (i.e. coagulation pH, dose).

Total algae counts (diatoms, green algae and other algae groups) were also correlated with TOC removal, but the correlation was significant only for Cropston, Whitacre and Frankley (correlation coefficients 0.74, 0.47 and 0.59), indicating the important role of algae-produced organic matter in the overall OM pool at those sites. The TOC removal relationship with particular algal groups was insignificant for most of the sites with the exception of site Cropston (correlation coefficient for diatoms 0.78, green algae 0.80 and other algae groups 0.86). It is interesting to note that for these sites the total algae counts were more important TOC removal efficiency predictors than the other fluorescence properties. Nevertheless, the

algal relationship with TOC removal and fluorescence data should be analysed with caution as algal counts data were collected independently of the main fluorescence sampling.

The monthly correlations with TOC removal were significantly stronger compared to site-specific relationships (Table 4.6). On a monthly basis a good correlation was found for raw water peak C emission wavelength, tryptophan-like fluorescence, and $T_{2_{int}}/C_{int}$ ratio, whereas for clarified water the predictors of greatest importance to TOC removal were peak C and tryptophan-like fluorescence intensity. The lack of correlation between clarified water peak C emission wavelength and $T_{2_{int}}/C_{int}$ ratio suggests that clarified water OM properties (degree of hydrophobicity and ratio of labile to recalcitrant OM) are similar for all WTWs. Thus, only the residual OM concentrations denoted as peak C and peak T intensities discriminate between particular sites. Likewise, for raw water, the qualitative OM properties (peak C emission wavelength and $T_{2_{int}}/C_{int}$ ratio) are stronger OM removal predictors compared to quantitative OM properties.

The relationship between TOC removal and the best fluorescence predictors (raw water peak C emission wavelength, clarified water peak C and T2 intensity) was investigated in more detail below (Figure 4.6). A predominance of the hydrophobic fraction, indicated by higher peak C emission values (Wu et al., 2003a; Baker et al., 2008), correlated with stable, high TOC removal, whereas the discernible amplitude of variation in aromaticity characterised lower TOC removal efficiencies (Figure 4.6 a). The higher TOC removal corresponded with the predominance of hydrophobic organic matter (Bamford, Frankley, and Tittesworth) whereas for relatively hydrophilic waters the removal was poorer (Draycote). The changes in peak C emission wavelength were greater for reservoir sites (Bamford, Frankley, and Tittesworth), however the OM character remained hydrophobic throughout the year and thus the OM was independent of seasonal OM aromaticity variation. The effect of

the seasonal peak C emission changes at lower wavelengths on TOC removal efficiency was more significant. This suggests that for sites with predominantly hydrophobic OM, seasonal changes in OM properties are of little consequence to OM removal efficiency compared to sites with predominant OM of lower aromaticity.

The monthly correlations between raw water peak C emission and TOC removal were found generally higher than the corresponding relationship derived for the whole fluorescence dataset (Figure 4.6 b). The highest values of correlation coefficients were observed for December 07 ($R^2 = 0.84$), August 07 ($R^2 = 0.63$), November 07 ($R^2 = 0.53$), February 08 ($R^2 = 0.48$), and June 07 ($R^2 = 0.43$) (Table 4.6).

Peak C and peak T intensity of clarified water in relation with TOC removal provided a good discrimination between treatment sites (Figure 4.6 c and e). Overall, the correlation of peak T intensity with TOC removal was slightly stronger compared to peak C prediction ($R^2 = 0.54$ and 0.40 respectively). The higher OM removal efficiency corresponded with lower peak C and peak T intensity of clarified water. The stable and efficient TOC removal correlated with low peak T intensity of clarified water and low microbial activity (Cammack et al., 2004; Hudson et al., 2008) (Figure 4.6 e). Low TOC removal corresponded to higher and variable microbial content. A pattern of stable TOC removal and low microbial fraction content characterised sites with water abstraction from the reservoirs (Bamford, Frankley, Tittesworth).

For a given peak C or T intensity, sites including Bamford, Tittesworth, Cropston demonstrated relatively higher and sites including Little Eaton, Church Wilne, Melbourne, and Mitcheldean demonstrate relatively lower OM removal efficiency than the results from the relationship equation indicated. The similar location of the sample sites in both relationships with TOC removal suggests that sites with relative overestimation of OM

removal (the first group) demonstrate a different composition of both peak C and peak T fluorophores compared to the second group with underestimated OM removal.

Similarly to raw water peak C emission wavelength, the clarified water peak C and peak T intensity provided a better prediction of TOC removal on a monthly than on a site-specific basis (Figure 4.6 d and f, Table 4.6).

4.3. CHAPTER CONCLUSIONS

In the chapter, a basic analysis of fluorescence data characterising OM properties and removal efficiency was presented. To retrieve the most important information on spatial and temporal OM variability, peak fluorescence values (maximum fluorescence intensity in peak A, C, T1, and T2 along with emission and excitation wavelengths) were recorded and analysed with simple linear regression. The results found in this chapter fully addressed the Objectives 1 and 2 and Hypotheses 1 to 3 (see Section 2.5).

It was found that peak C emission wavelength provides an indication of the hydrophobic content in drinking water samples, whereas peak T intensity reflects the presence of hydrophilic and low molecular weight OM (Objective 1, *Hypothesis H1*, see Section 2.5). Thus, these two fluorescence properties were shown to discriminate between easier to remove highly reactive hydrophobic OM (high peak C emission) and recalcitrant, less reactive hydrophilic fraction (high peak T intensity). For Little Eaton and Church Wilne WTWs a significant variation in the clarified peak T intensity was observed and related to increased levels of biological activity based on supplementary algal data.

During the conventional OM removal by coagulation it was observed that humic-like fluorescence (peak A and C) was removed more effectively than the protein-like fluorescence (peak T1 and T2). The fluorescence properties of the clarified water were found to be more

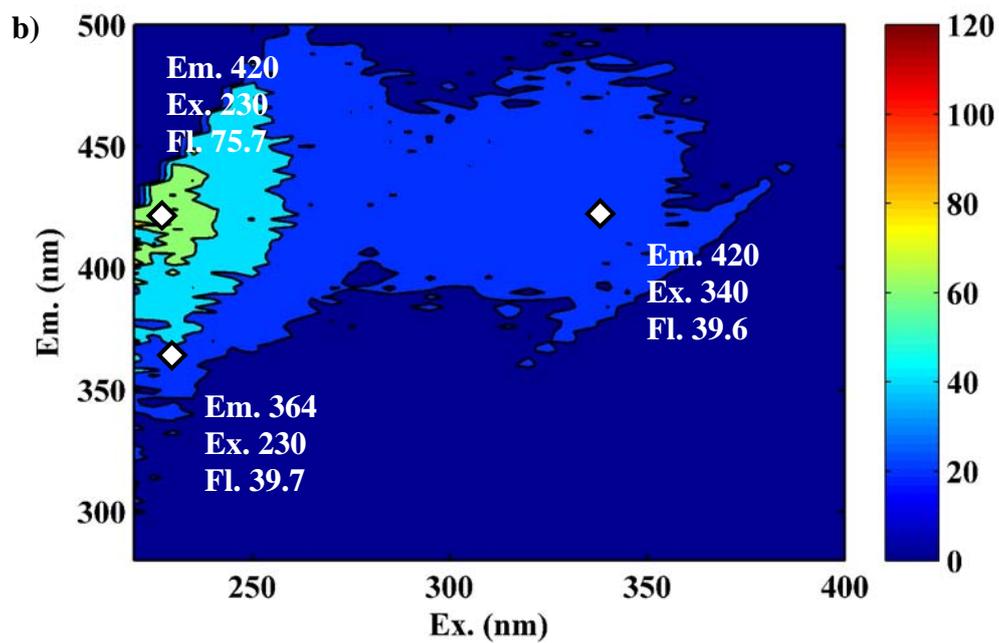
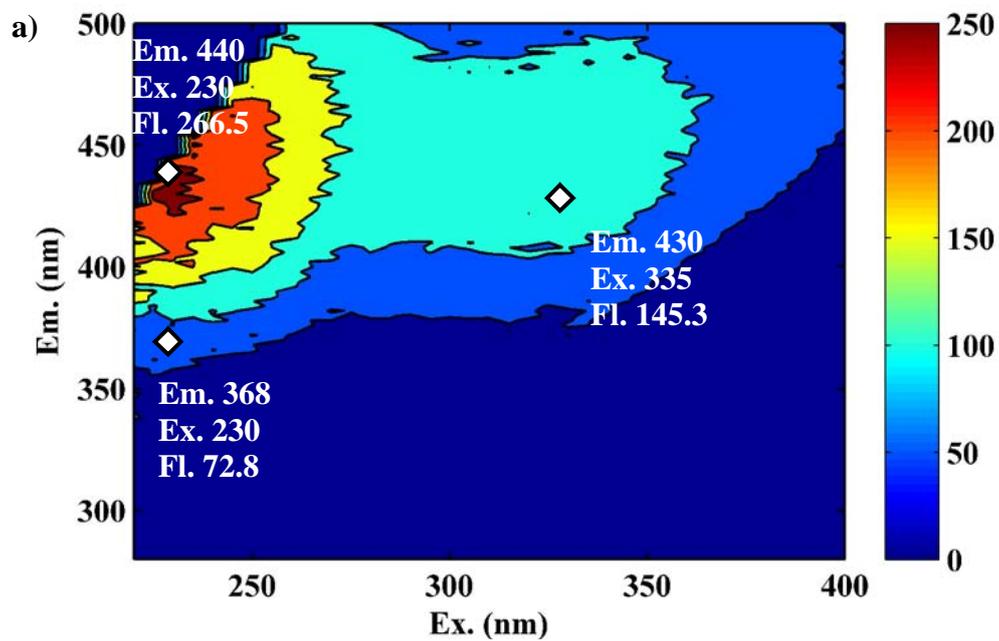
uniform between WTWs in relation to significant variability in raw water properties (Objective 2, *Hypothesis H3*, see Section 2.5). The greatest change in OM properties between raw and clarified water were observed for Bamford, Frankley, Tittesworth, and Little Eaton, whereas the lowest for Melbourne, Mitcheldean, and Strensham WTW.

Peak C fluorescence intensity of raw and clarified water has been shown to exhibit a good correlation with TOC concentrations (Objective 2, *Hypothesis H2*, see Section 2.5). The reduction in peak C and peak A intensity was found to correlate with TOC removal efficiency between raw and clarified water ($R^2 = 0.91$ and respectively 0.84). Corresponding correlation between reduction in UV-Vis absorbance and TOC removal was weaker ($R^2 = 0.82$). The highest OM removals were observed at Bamford, Frankley, and Tittesworth, whereas the lowest at Melbourne, Draycote, Whitacre, and Church Wilne. The better OM removal was observed for WTWs with high peak C emission wavelength and low peak T intensity of raw water.

The results presented in this chapter pertinent to raw water OM character and TOC removal efficiency have been published recently (Bieroza et al., 2009 a).

The peak-picking approach used in this chapter is the easiest way of analysing fluorescence data. However for large dataset analysed here, the results' interpretation was impeded by large amount of data points. Therefore, in the next chapter an application of more robust data mining techniques for fluorescence data analysis is discussed.

4.4. CHAPTER 4 FIGURES



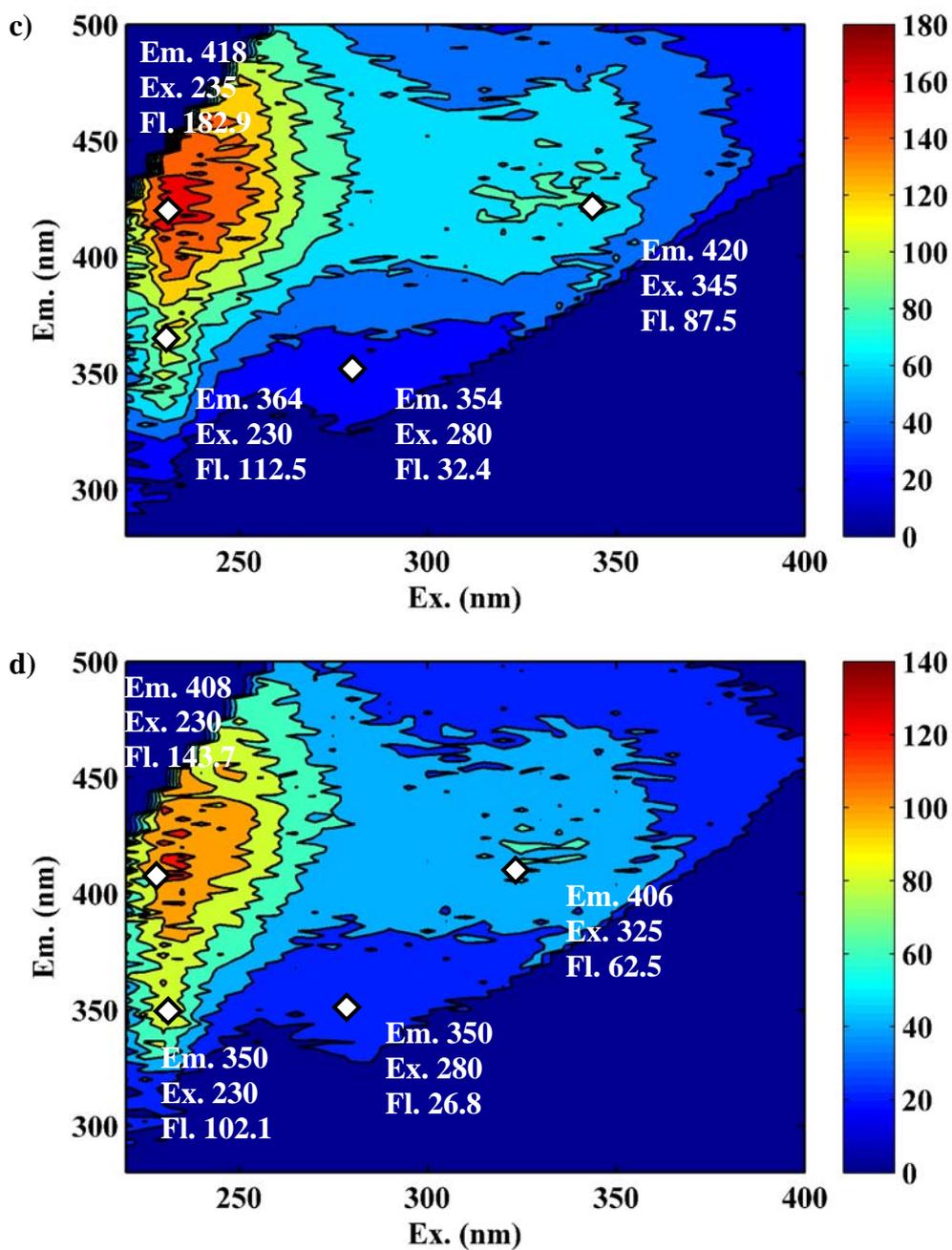
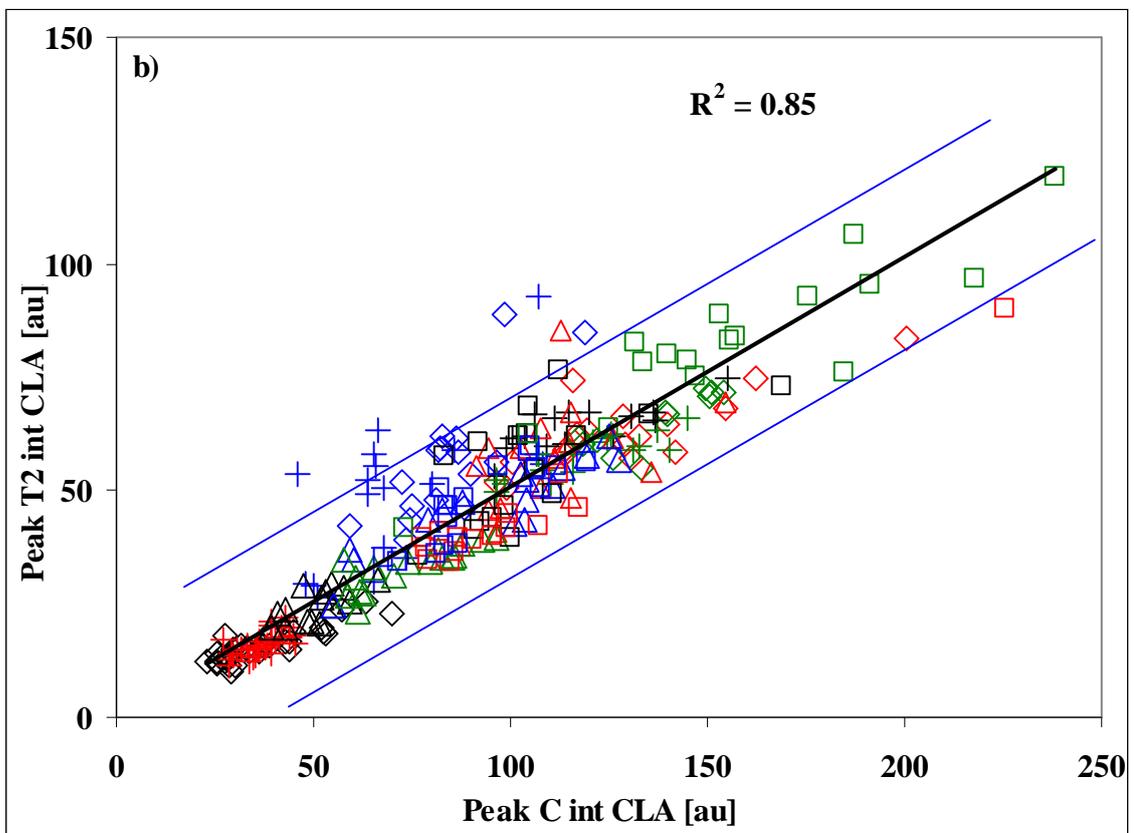
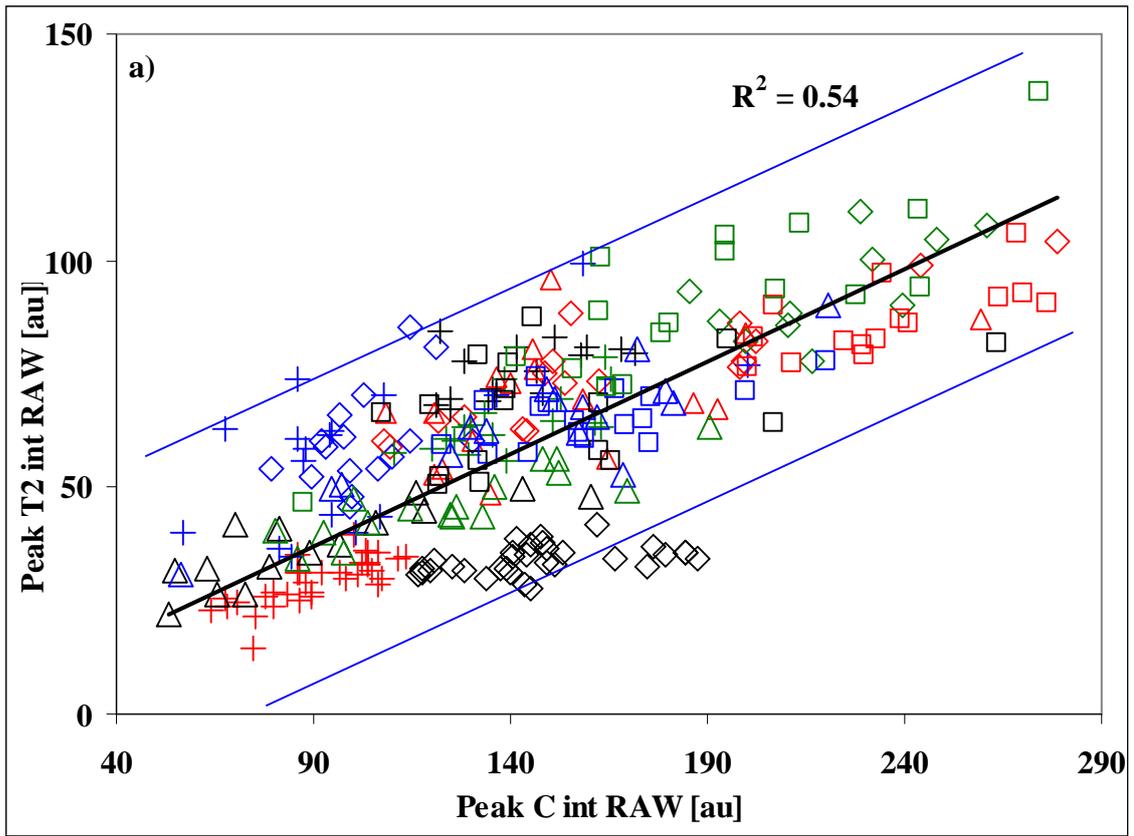
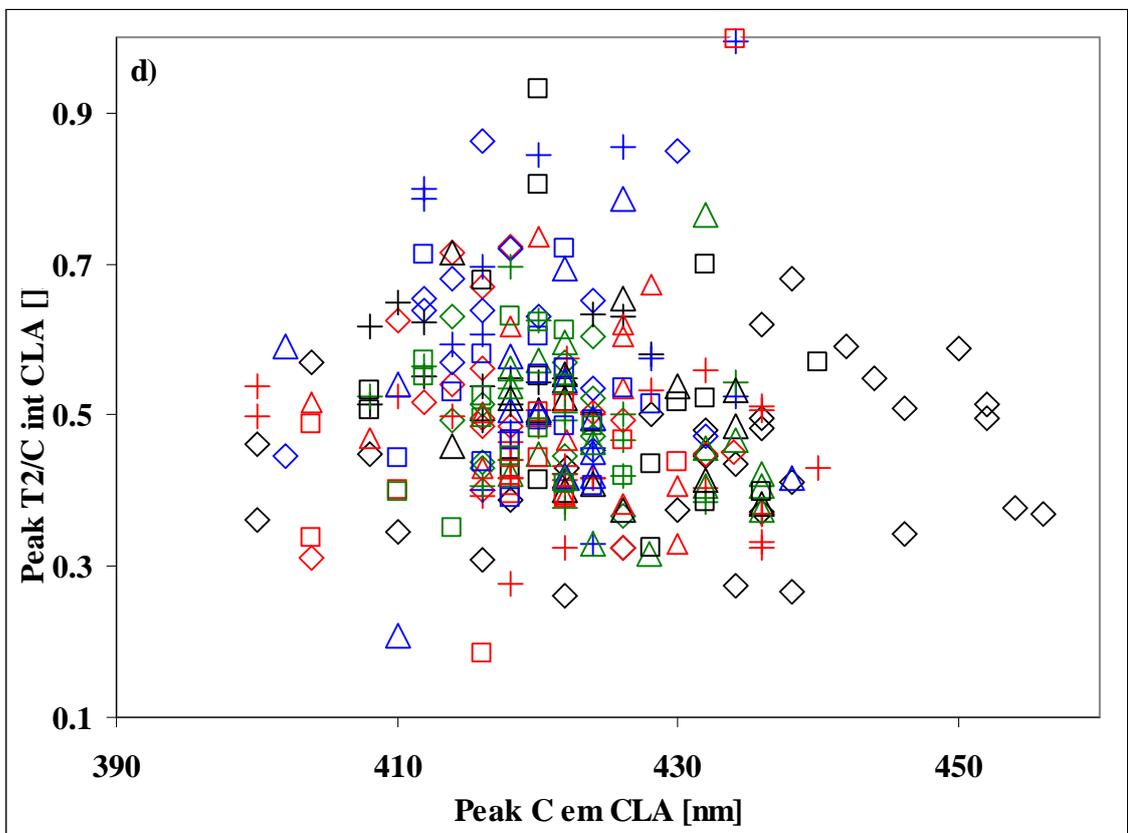
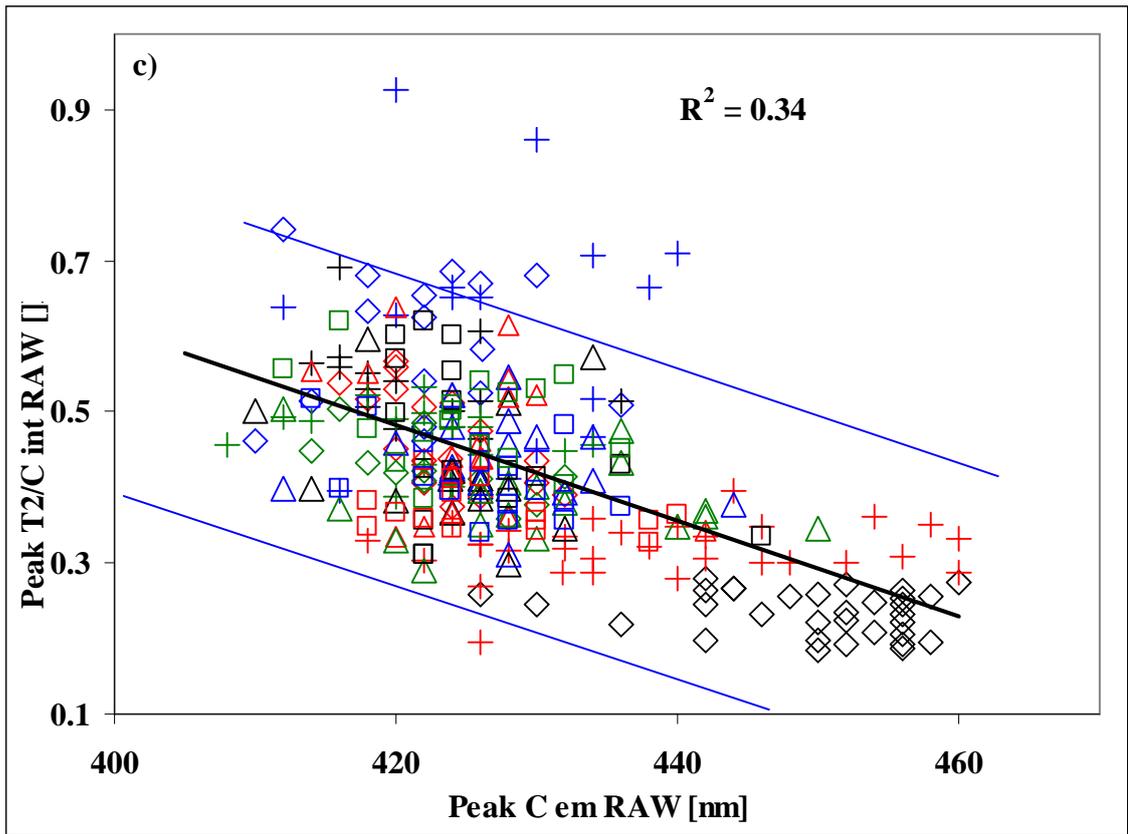


Figure 4.1: Fluorescence EEMs characterising OM changes during the treatment process. Bamford raw water (a), Bamford clarified water (b), Little Eaton raw water (c), and Little Eaton clarified water (d)





◇ Bamford	◇ Campion Hills	◇ Church Wilne	◇ Cropston
+ Draycote	+ Frankley	+ Little Eaton	+ Melbourne
△ Mitcheldean	△ Mythe	△ Ogston	△ Shelton
□ Strensham	□ Tittesworth	□ Trimpley	□ Whitacre

◇ Oct-06	◇ Nov-06	◇ Dec-06	◇ Jan-07
+ Feb-07	+ Mar-07	+ Apr-07	+ May-07
△ Jun-07	△ Jul-07	△ Aug-07	△ Sep-07
□ Oct-07	□ Nov-07	□ Dec-07	□ Jan-08
* Feb-08			

Figure 4.2: Fluorescence properties of raw and clarified water for all WTWs. Peak C intensity vs peak T2 intensity for raw (a) and clarified water (b), peak C emission wavelength vs T2/C intensity ratio for raw (c) and clarified water (d). Fluorescence intensity in arbitrary units (au). Linear regression with 95 % prediction lines

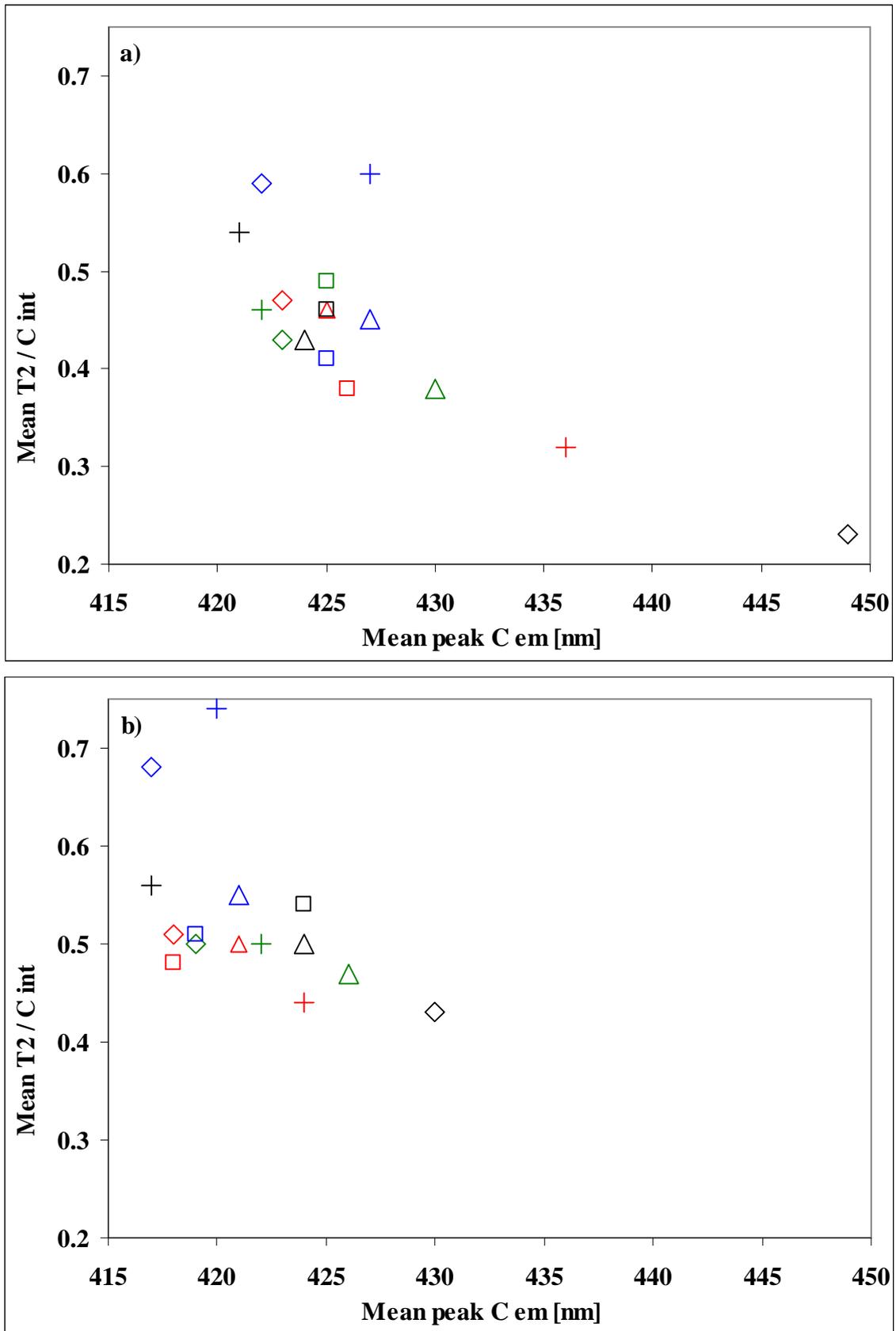


Figure 4.3: Mean peak C emission wavelength vs mean T2/C intensity ratio for raw (a) and clarified water (b). For explanation of the labels see Figure 4.2

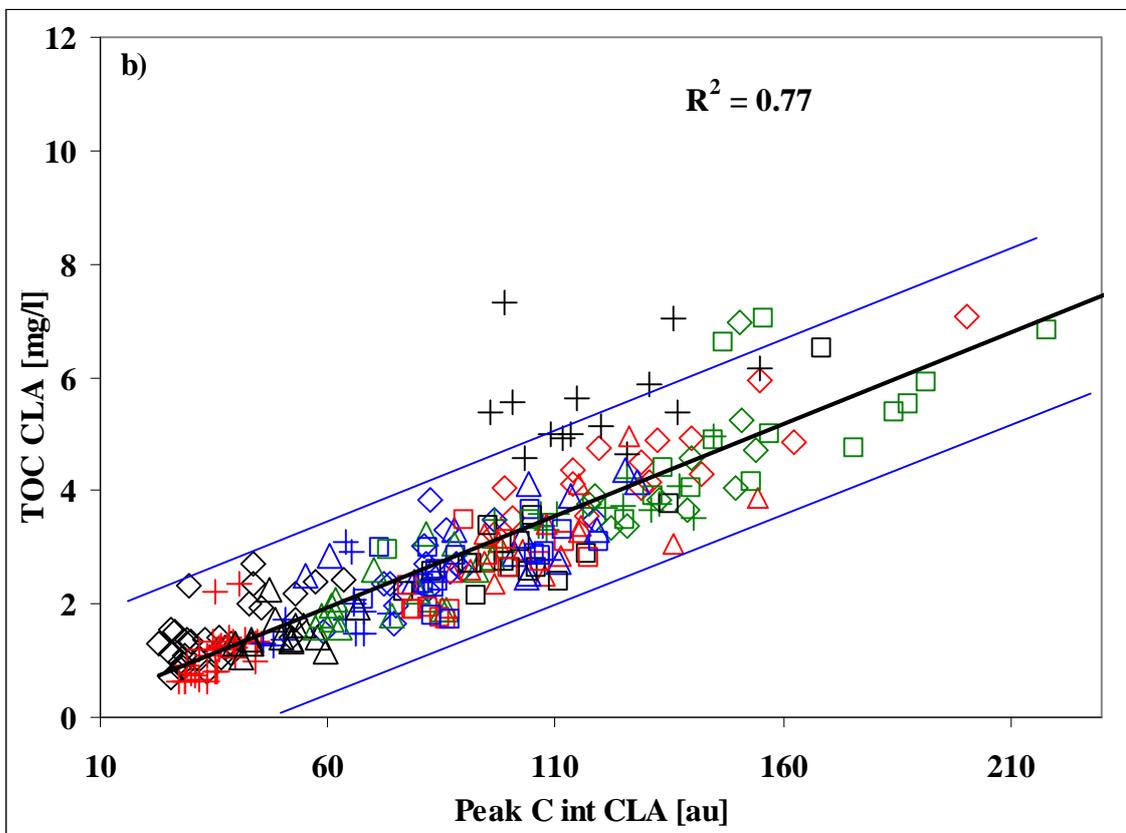
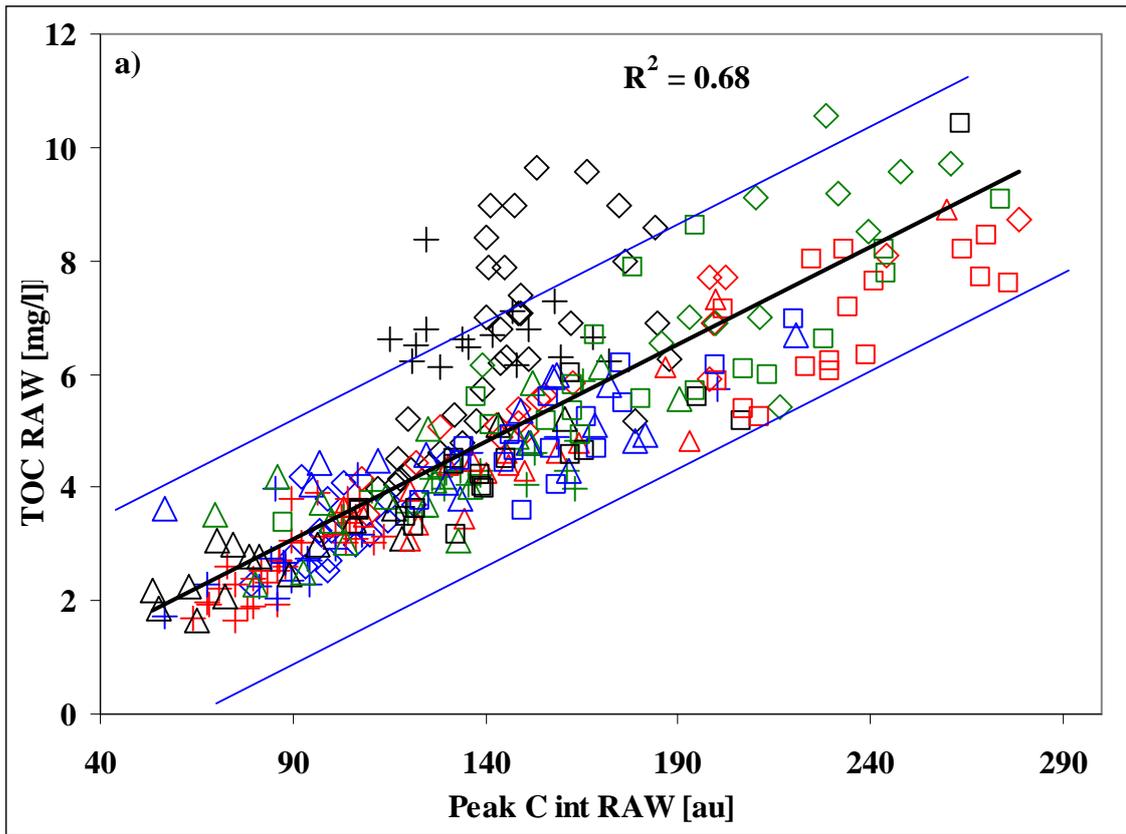
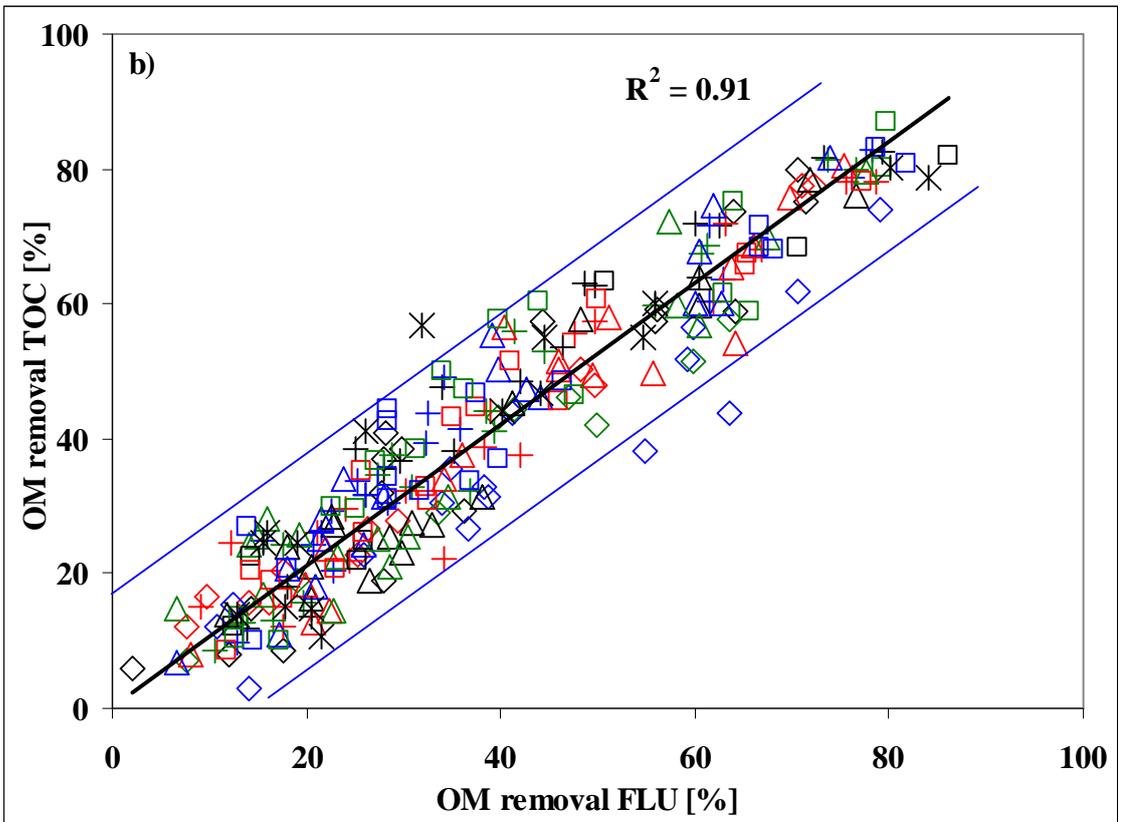
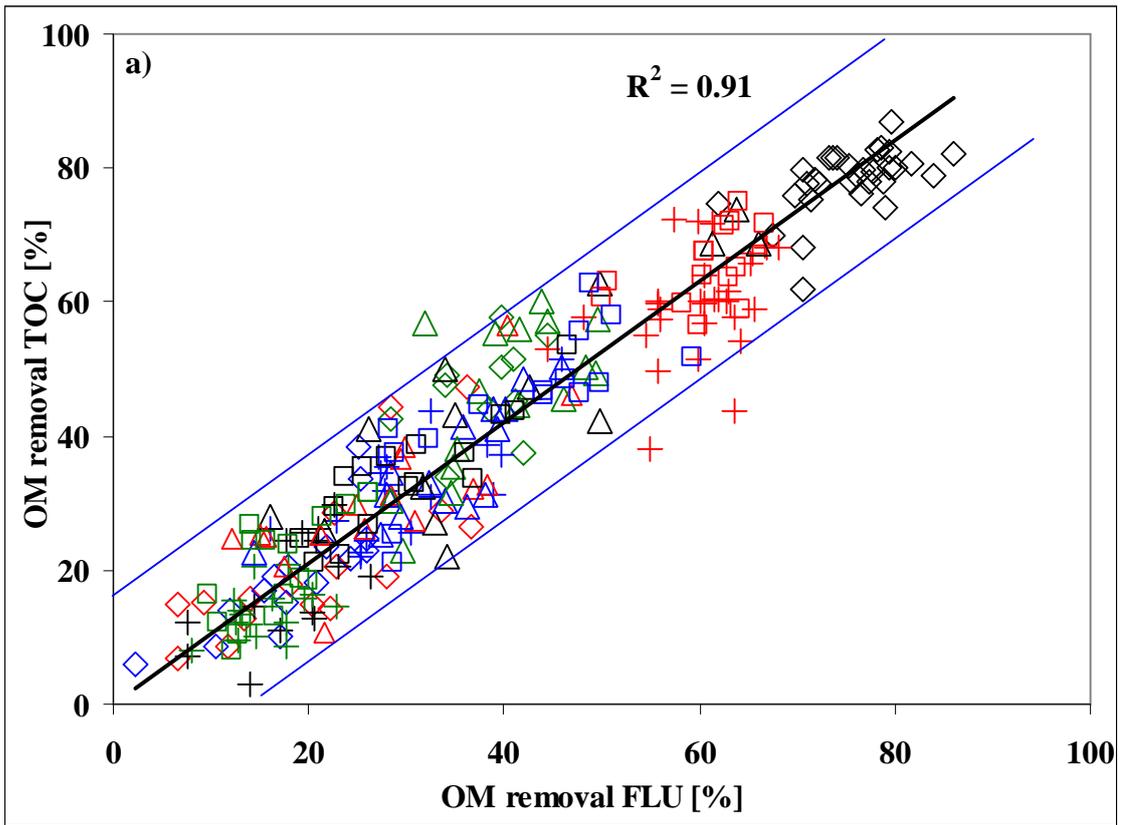


Figure 4.4: TOC concentrations vs peak C intensity for raw (a) and clarified water (b). For explanation of the labels see Figure 4.2. Linear regression with 95 % prediction lines



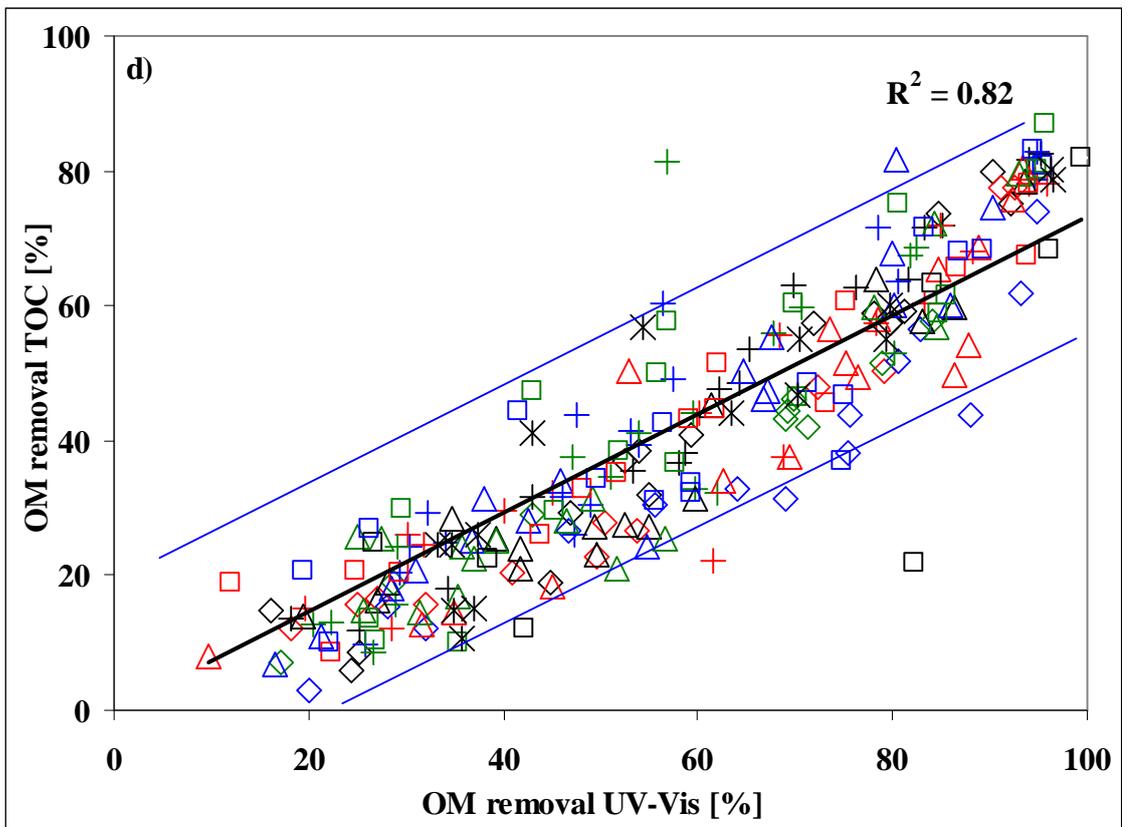
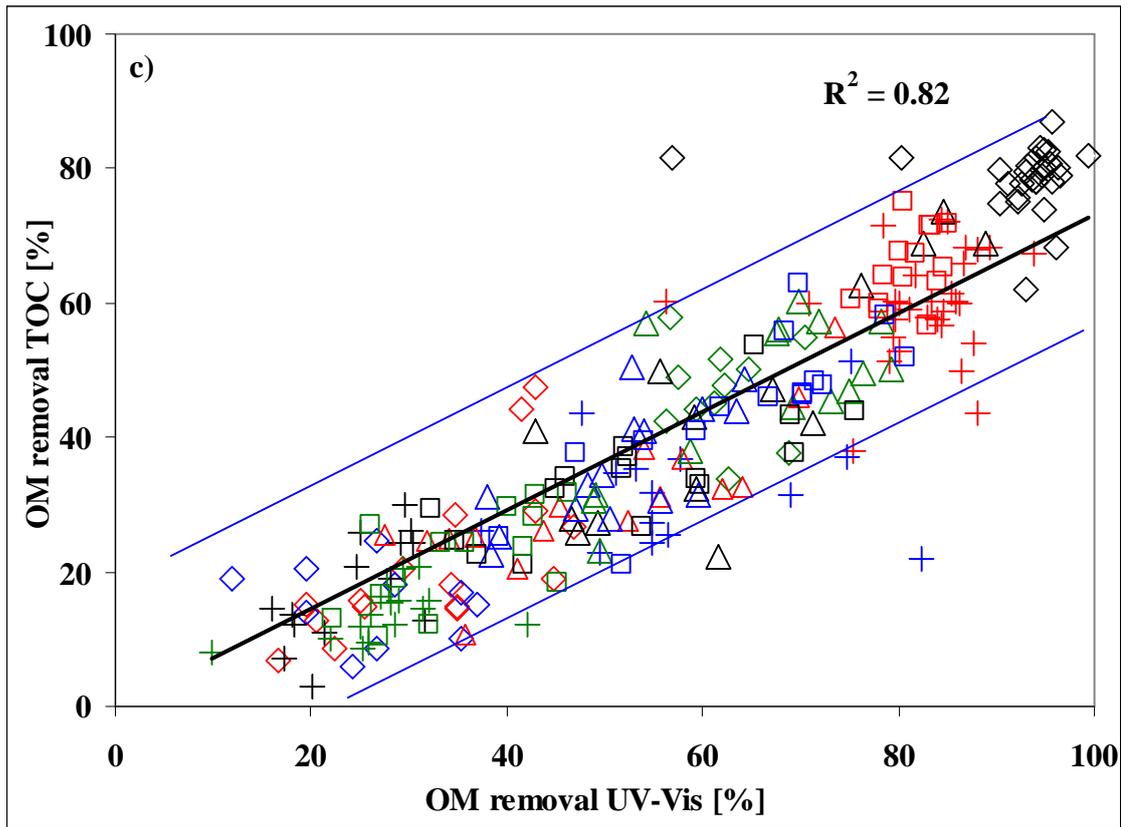
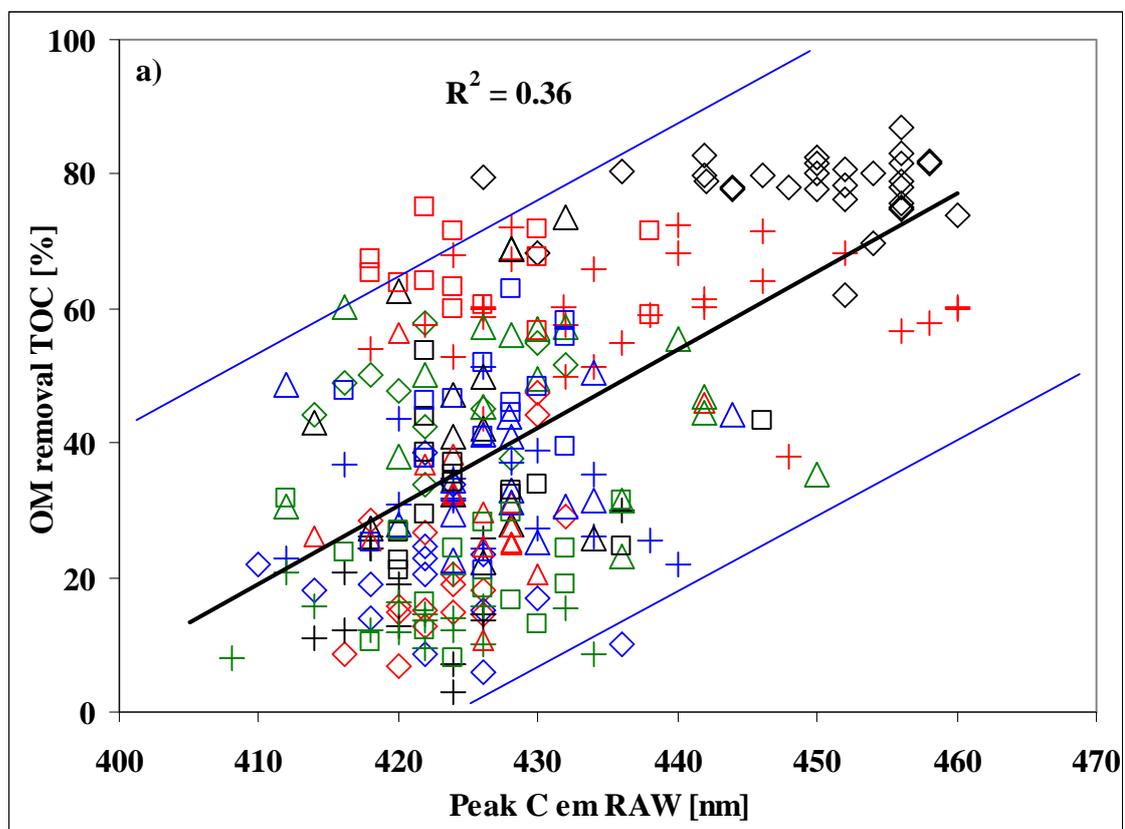
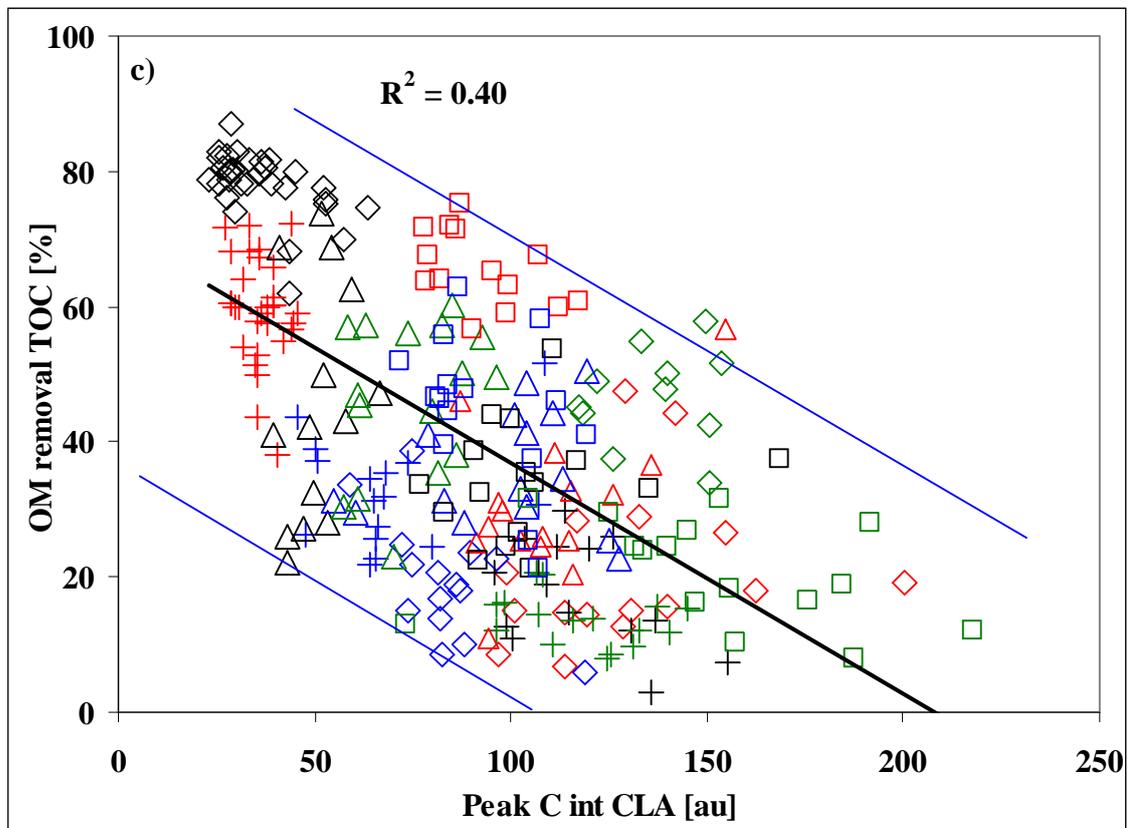
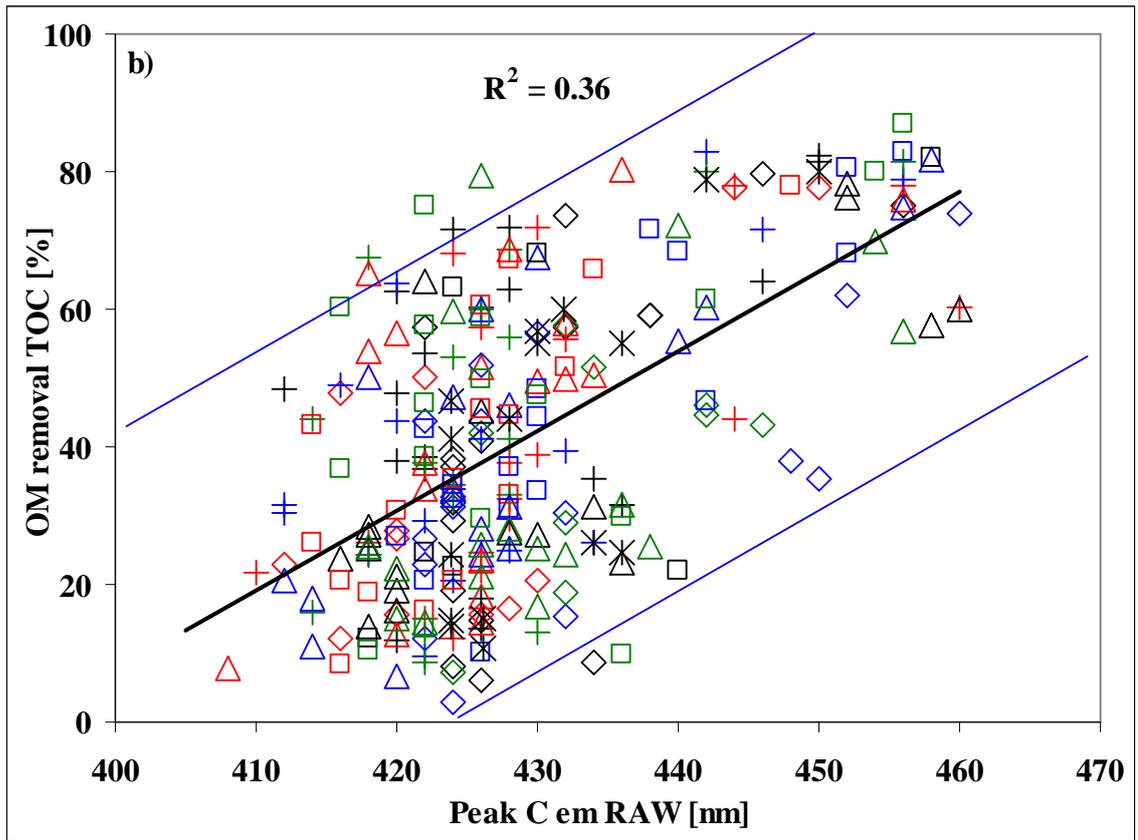
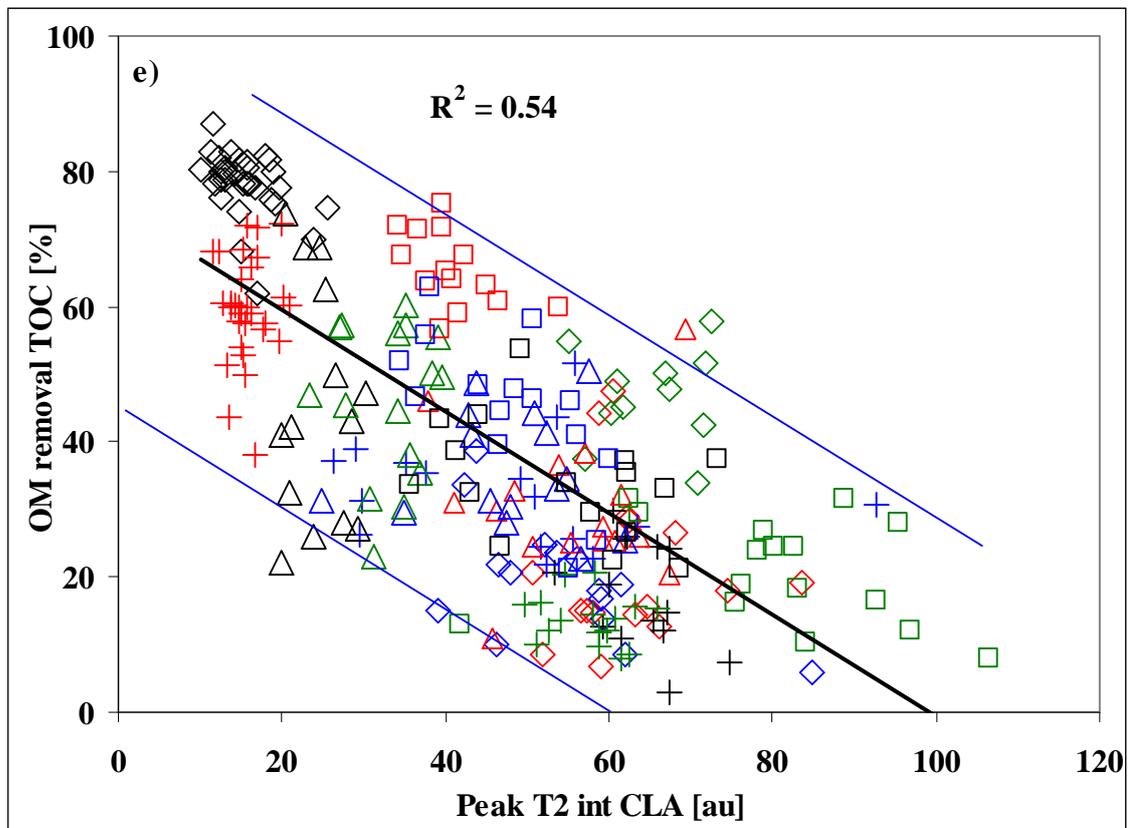
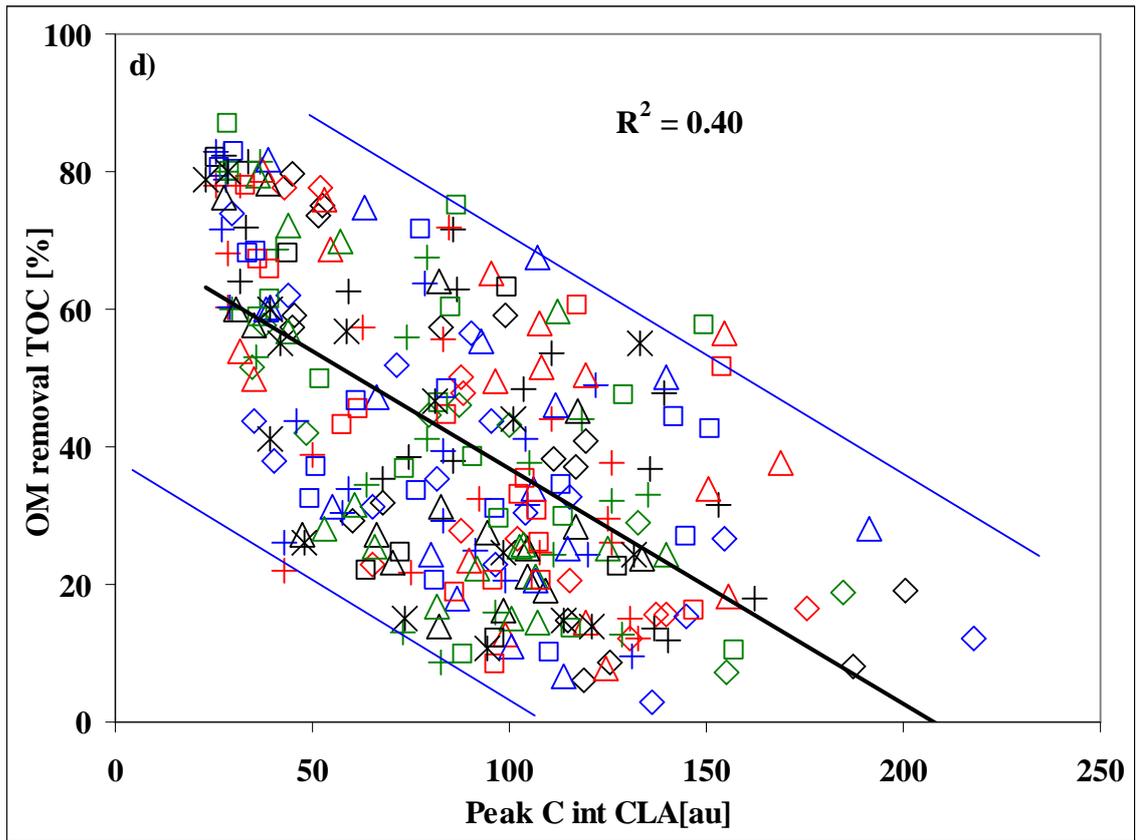


Figure 4.5: TOC removal vs peak C intensity (FLU) removal and absorbance measured at 254 nm (UV-Vis) removal between raw and clarified water. TOC removal vs FLU removal

for sites (a) and months (b). TOC removal vs UV-Vis removal for sites (c) and months (d). For explanation of the labels see Figure 4.2. Linear regression with 95 % prediction lines a, b, c, d







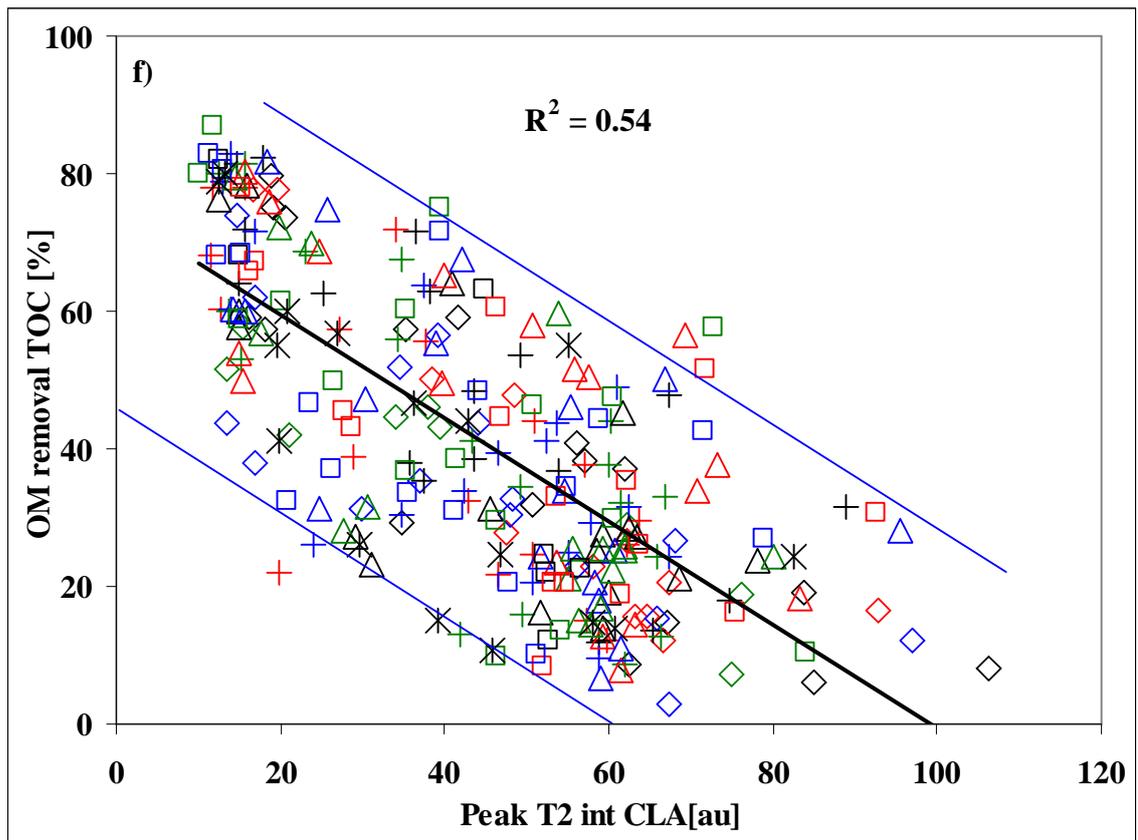


Figure 4.6: Fluorescence OM properties vs TOC removal. Raw water peak C emission for sites (a) and months (b), clarified water peak C intensity for sites (c) and months (d), clarified water peak T2 intensity for sites (e) and months (f). For explanation of the labels see Figure 4.2. Linear regression with 95 % prediction lines

4.5. CHAPTER 4 TABLES

Table 4.1: Kruskal-Wallis ANOVA test results. Mean ranks and test statistics (df – degrees of freedom, Asymp. Sig. – the asymptotic significance)

Site	Number of samples	C em Raw (nm)	C int Raw (au)	T2 int Raw (au)	T/C int Raw ()	C em Cla (nm)	C int Cla (au)	T2 int Cla (au)	T/C int Cla ()
Bamford	19	290.4	187.2	41.1	17.5	235.5	39.3	38.1	99.0
Campion Hills	19	97.1	207.3	209.8	185.7	108.4	260.2	220.2	150.6
Church Wilne	18	98.4	78.6	230.2	249.5	115.3	154.7	249.2	224.7
Cropston	12	111.0	284.0	276.1	151.5	133.9	279.0	248.4	143.0
Draycote	19	82.4	173.6	233.8	235.0	116.6	242.7	227.4	198.2
Frankley	19	234.3	51.8	38.6	59.7	168.6	41.3	39.2	100.5
Little Eaton	18	169.7	75.1	219.7	230.0	130.9	119.5	221.9	213.9
Melbourne	18	109.8	169.0	165.9	185.7	162.7	246.7	200.1	145.3
Mitcheledean	19	154.3	53.9	84.2	147.9	183.6	79.8	95.2	137.7
Mythe	19	140.7	184.7	210.7	175.6	150.8	219.9	217.2	150.2
Ogston	17	171.6	181.6	161.2	169.2	173.8	191.7	151.9	156.5
Shelton	19	172.0	127.1	98.5	106.1	192.6	126.2	103.3	113.8
Strensham	19	126.0	182.9	216.9	170.0	184.9	206.6	216.6	155.0
Tittesworth	17	146.7	303.4	205.9	100.0	121.4	179.6	137.3	96.9
Trimpley	18	145.6	216.8	183.5	141.6	147.4	179.8	198.7	159.8
Whitacre	19	137.7	247.3	274.2	203.2	131.1	281.5	281.5	161.7
Chi-Square		135.3	202.5	239.5	198.7	52.8	261.2	247.4	60.7
df		15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0
Asymp. Sig.		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 4.2: Organic matter removal (%) calculated as a decrease in fluorescence intensity, UV-Vis absorbance and TOC concentrations for 16 WTWs. The highest values in red, the lowest values in blue

Site	Organic matter removal (%)					
	Peak C fluorescence intensity		UV-Vis absorbance (254 nm)		TOC	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Bamford	72.7	5.7	90.9	9.4	77.1	6.5
Campion Hills	21.5	9.7	32.6	9.4	21.0	11.6
Church Wilne	18.3	6.4	27.3	7.5	19.5	8.8
Cropston	36.7	6.2	59.9	8.4	46.8	7.1
Draycote	17.3	5.6	23.6	5.6	17.5	8.0
Frankley	58.0	7.0	81.6	7.7	60.2	8.1
Little Eaton	30.9	7.3	58.8	11.8	32.5	8.0
Melbourne	15.3	3.9	27.0	7.0	13.7	3.7
Mitcheledean	39.8	14.8	62.7	14.4	45.2	17.5
Mythe	26.8	9.4	48.5	13.4	30.5	10.5
Ogston	34.6	7.5	51.0	8.1	35.7	8.6
Shelton	39.5	6.5	65.3	10.4	46.1	11.4
Strensham	29.3	7.8	51.7	12.8	34.2	8.5
Tittesworth	60.2	4.7	80.5	3.2	65.5	5.6
Trimpley	41.8	11.2	64.2	12.2	44.9	11.2
Whitacre	17.7	5.6	35.6	7.7	20.9	7.7

Table 4.3: Organic matter removal (%) calculated as a decrease in fluorescence intensity, UV-Vis absorbance and TOC concentrations for particular months and raw water peak C emission (nm). The highest values in red, the lowest values in blue. “–” indicates no measurements

Site	Organic matter removal (%)							
	Peak C fluorescence intensity		UV-Vis absorbance (254 nm)		TOC		Peak C emission wavelength (nm)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Aug-06	37.6	19.2	52.1	21.2	-	-	425.9	10.7
Sep-06	32.4	16.5	52.0	21.4	-	-	426.4	12.0
Oct-06	38.3	21.8	59.7	24.2	40.9	24.6	430.6	9.4
Nov-06	34.3	22.9	54.9	26.0	34.3	23.6	426.8	12.1
Dec-06	42.1	20.8	64.7	24.7	36.3	18.8	432.2	12.7
Jan-07	42.3	18.3	61.1	23.1	37.8	16.2	434.6	7.1
Feb-07	41.8	19.4	62.8	22.8	48.4	22.1	428.1	11.0
Mar-07	40.9	21.5	61.2	25.5	43.0	21.9	430.9	12.6
Apr-07	37.3	20.9	53.6	21.9	42.3	21.1	426.7	11.7
May-07	38.5	20.3	55.3	22.6	42.4	23.1	426.0	9.9
Jun-07	36.7	19.2	53.0	21.8	37.0	20.9	430.3	15.0
Jul-07	43.9	19.9	67.3	24.4	44.9	22.2	426.7	10.0
Aug-07	37.1	19.7	55.5	23.5	41.2	22.3	428.6	13.0
Sep-07	33.8	21.8	50.0	22.9	36.1	21.9	430.9	11.0
Oct-07	40.8	28.9	67.0	30.2	42.1	28.0	430.8	13.9
Nov-07	36.3	19.3	55.8	26.1	39.2	20.3	424.6	8.6
Dec-07	32.6	6.0	57.8	2.1	38.1	6.2	426.0	5.7
Jan-08	41.3	22.0	59.4	23.4	46.7	23.4	429.0	12.1
Feb-08	36.5	22.6	57.8	23.1	40.5	22.7	430.6	7.5

Table 4.4: TOC removal prediction form raw and clarified water fluorescence and absorbance properties. Significant correlations with coefficients > 0.36 highlighted in red

Predictor	R ²
T1 int Raw	-0.25
T2 int Raw	-0.24
A em Raw	0.03
A ex Raw	0.21
A int Raw	0.00
C em Raw	0.36
C ex Raw	0.12
C int Raw	0.01
T1 int Cla	-0.38
T2 int Cla	-0.54
A em Cla	0.01
A ex Cla	0.34
A int Cla	-0.36
C em Cla	0.07
C ex Cla	0.12
C int Cla	-0.40
UV ₂₅₄ Raw	0.31
UV ₂₅₄ Cla	-0.24
SUVA Raw	0.36
SUVA Cla	0.00
OM removal A	0.84
OM removal T1	0.46
OM removal T2	0.63
OM removal C	0.91
OM removal UV ₂₅₄	0.82
OM removal SUVA	0.49

Table 4.5: TOC removal prediction from different raw and clarified water fluorescence properties - sites. Significant correlations with coefficients > 0.36 highlighted in red

Site	C em Raw (nm)	C int Raw (au)	T1 int Raw (au)	T2 int Raw (au)	T/C int Raw ()	C em Cla (nm)	C int Cla (au)	T1 int Cla (au)	T2 int Cla (au)	T/C int Cla ()
Bamford	0.01	0.13	0.07	0.04	0.28	0.06	0.02	0.00	0.01	0.05
Campion Hills	0.44	0.25	0.09	0.26	0.22	0.00	0.05	0.02	0.00	0.00
Church Wilne	0.05	0.05	0.22	0.26	0.08	0.06	0.37	0.31	0.34	0.32
Cropston	0.01	0.16	0.12	0.05	0.02	0.03	0.01	0.01	0.00	0.20
Draycote	0.04	0.03	0.10	0.00	0.10	0.00	0.20	0.02	0.19	0.04
Frankley	0.05	0.08	0.17	0.07	0.00	0.01	0.00	0.05	0.00	0.00
Little Eaton	0.05	0.19	0.07	0.00	0.15	0.00	0.03	0.11	0.03	0.00
Melbourne	0.01	0.06	0.02	0.00	0.13	0.03	0.12	0.07	0.03	0.00
Mitcheledean	0.02	0.74	0.60	0.48	0.46	0.00	0.08	0.02	0.01	0.01
Mythe	0.01	0.63	0.00	0.10	0.38	0.00	0.27	0.03	0.00	0.05
Ogston	0.01	0.30	0.01	0.07	0.35	0.02	0.01	0.05	0.00	0.08
Shelton	0.02	0.23	0.17	0.10	0.33	0.06	0.07	0.03	0.00	0.21
Strensham	0.13	0.01	0.02	0.03	0.00	0.16	0.05	0.00	0.02	0.00
Tittesworth	0.16	0.07	0.24	0.00	0.07	0.05	0.12	0.13	0.14	0.03
Trimpley	0.00	0.01	0.05	0.01	0.00	0.22	0.21	0.05	0.20	0.21
Whitacre	0.20	0.01	0.13	0.12	0.14	0.16	0.05	0.12	0.09	0.14

Table 4.6: TOC removal prediction from different raw and clarified water fluorescence properties - months. Significant correlations with coefficients > 0.36 highlighted in red

Month	C em Raw (nm)	C int Raw (au)	T1 int Raw (au)	T2 int Raw (au)	T/C int Raw ()	C em Cla (nm)	C int Cla (au)	T1 int Cla (au)	T2 int Cla (au)	T/C int Cla ()
Sep-06	0.22	0.13	0.73	0.65	0.21	0.00	0.05	0.64	0.34	0.31
Oct-06	0.41	0.00	0.57	0.44	0.63	0.31	0.51	0.67	0.73	0.38
Nov-06	0.42	0.06	0.63	0.66	0.73	0.25	0.66	0.50	0.80	0.19
Dec-06	0.34	0.01	0.24	0.19	0.63	0.01	0.43	0.30	0.50	0.57
Jan-07	0.33	0.16	0.53	0.38	0.65	0.26	0.55	0.68	0.61	0.09
Feb-07	0.28	0.16	0.42	0.38	0.44	0.32	0.69	0.52	0.63	0.03
Mar-07	0.41	0.00	0.25	0.12	0.59	0.27	0.49	0.51	0.57	0.18
Apr-07	0.42	0.00	0.11	0.24	0.32	0.01	0.46	0.20	0.55	0.03
May-07	0.37	0.00	0.41	0.33	0.63	0.05	0.55	0.64	0.74	0.14
Jun-07	0.43	0.02	0.17	0.18	0.51	0.06	0.32	0.23	0.43	0.21
Jul-07	0.31	0.04	0.29	0.13	0.73	0.05	0.29	0.41	0.46	0.04
Aug-07	0.63	0.03	0.31	0.17	0.72	0.23	0.28	0.48	0.49	0.02
Sep-07	0.20	0.15	0.20	0.15	0.75	0.24	0.35	0.40	0.55	0.13
Oct-07	0.38	0.61	0.36	0.22	0.76	0.28	0.41	0.55	0.81	0.51
Nov-07	0.53	0.05	0.14	0.08	0.59	0.23	0.26	0.22	0.41	0.08
Dec-07	0.84	0.02	0.36	0.23	0.78	0.01	0.47	0.59	0.58	0.03
Jan-08	0.22	0.06	0.09	0.08	0.80	0.00	0.32	0.46	0.36	0.24
Feb-08	0.48	0.03	0.23	0.15	0.52	0.00	0.35	0.52	0.43	0.03

5. ADVANCED DATA MINING TECHNIQUES IN PATTERN RECOGNITION AND CALIBRATION OF FLUORESCENCE DATA

5.1. APPLICATION OF DATA MINING TECHNIQUES FOR ADVANCED ANALYSIS OF FLUORESCENCE DATA

Fluorescence EEMs can be a substantial source of environmental information on organic matter composition and variability. However, the complexity of the fluorescence data generates difficulties in its interpretation and extraction of useful information. Here, the fluorescence dataset comprised 512 EEMs (i.e. raw and clarified samples from 16 WTWs and 16 months of measurements) to facilitate analysis of spatial and seasonal variability of the sites (data presented in Chapter 4). Assuming that each original EEM contains more than 4000 data points (111 emission wavelengths and 37 excitation wavelengths, reducing to 2515 data points after scatter removal), the peak-picking approach which was discussed in the previous chapter utilizes only small proportion of the available fluorescence data. Therefore, to handle a large amount of fluorescence data, a more robust approach is needed, which both preserves the important topological and metric relationships of the fluorescence data (high-dimensionality and non-linearity) and enables correlation between fluorescence data and various treatment process parameters.

Advanced EEMs analysis techniques can be classified by the algorithms used and by the purpose of the analysis. The most common techniques include different multivariate analysis tools from chemometric analysis (multi-way analysis: principal components analysis (PCA), partial least squares analysis (PLS), parallel factor analysis (PARAFAC); multiple linear regression (MLR), multivariate curve resolution (MCR)) and other computation and modelling techniques, such as artificial neural networks (ANNs) and fuzzy logic. Moreover,

Brunsdon and Baker (2001) demonstrated the analysis of fluorescence EEMs data with the principal filter analysis. Examples of the methods used for EEMs analysis found in the literature are presented in Table 5.1.

In fluorescence analysis, multivariate techniques are commonly used in the exploratory analysis, pattern recognition and multivariate calibration of spectra. Exploratory analysis aims to gather information on the dataset and formulate hypotheses worth testing without prior knowledge of the regularities and patterns in the data. This approach often employs unsupervised decomposition algorithms (e.g. self-organizing map, PCA). The multivariate calibration involves development of a mathematical model relating the fluorescence intensity of a component (fluorophore) with the concentration from a set of reference samples (calibration) and prediction of the component concentration from unknown sample fluorescence spectra (validation) (Martens and Naes, 1989; Bos et al., 1993).

The differences between the calibration methods lie in the way the parameters of those algorithms are optimised (Despaigne and Massart, 1998). ANNs represent the models without any constraints on the parameters, which are fitted to minimise the calibration samples squared residuals. In the decomposition models, such as PCA, PLS and PARAFAC, the parameters are obtained with constraints such as scores orthogonality, maximisation of the variance (PCA) or covariance of the independent and dependent datasets (PLS), and with a non-negativity constraint in the PARAFAC model (Bro, 1998; Despaigne and Massart, 1998).

The multivariate methods are often utilized in combination, where the preliminary analysis aims to reduce the amount of data and extract the most important features of the fluorescence data and the second step involves the calibration with known standards and concentrations. In particular, the outcome of data decomposition methods, e.g. PCA and PARAFAC, is often

used as input to other models, e.g. regression models (MLR, PLS) or ANNs (Bro, 1998; Scott et al., 2003; Lee et al., 2005).

In this chapter, different pattern recognition and calibration algorithms were tested to retrieve fluorescence information on organic matter variability and to correlate fluorescence properties with organic matter removal in drinking water treatment. The common pattern recognition techniques facilitate fluorescence data reduction and decomposition into a set of numerical values representing the most important features of the input dataset. The outcomes of PCA, PARAFAC, and self-organizing map (SOM) are evaluated and compared with standard peak-picking method. Furthermore, different calibration algorithms (MLR, stepwise regression (SR), PLS, and artificial neural network with back-propagation learning (BPNN)) were used to correlate organic matter properties derived from the above fluorescence analysis with measured TOC removal across the clarification stage for 16 WTWs.

In particular, the following questions were addressed:

- can a PARAFAC model outperform other fluorescence data decomposition techniques?
- can decomposition techniques (PCA, PARAFAC, SOM) outperform the standard peak picking technique in the identification of additional spectral components?
- which calibration technique (MLR, SR, PLS, BPNN) best explains the fluorescence – TOC removal relationship?

For the purpose of the computational analyses, the following notation was used. The input fluorescence data can be represented as a vector which length was denoted as m (number of fluorescence intensities measured at m excitation-emission wavelength pairs), and number of observations (raw and clarified water samples) as n .

5.1.1. *Pattern recognition techniques*

Analysis of large fluorescence dataset presents significant computational difficulties. Therefore, a common approach in fluorescence data analysis is the decomposition of the original EEMs into a set of fewer fluorescence parameters (of size p , and $p \leq m$) measured at certain fluorescence regions, e.g. maximum peak C fluorescence intensity. This “peak-picking” method is useful when the spectral properties of organic matter constituents are known or assumed a priori, and the analysis is restricted to the fluorescence regions of particular interest. However, in most cases no preceding assumptions or knowledge on the fluorescence data variability are given, and the aim of the analysis is to extract the most characteristic fluorescence features of the dataset.

To retrieve the most important information on organic matter composition and to reduce the data dimensionality, the raw and clarified water EEMs were processed with three decomposition algorithms: PCA, PARAFAC and SOM; methods which have previously been used tools for the exploratory analysis of fluorescence data (Persson and Wedborg, 2001; Stedmon et al., 2003; Boehme et al., 2004; Lee et al., 2005).

5.1.1.1 Multi-way analysis techniques

Principal components and parallel factor analysis are examples of multi-way analysis techniques used to analyse two- and three-way matrices respectively. In multi-way analysis the fluorescence data are decomposed into sets of scores and loadings, describing the most important features of the original dataset (Bro, 1997).

Fluorescence EEMs are examples of three-way matrices, where for every sample the fluorescence emission intensity is measured at several wavelengths for different excitation wavelengths (Bro, 1997). As a result, fluorescence data can be represented as a three-way

matrix, $I \times J \times K$, where the index I refers to the sample number, J to emission wavelength and K to the excitation wavelength. EEMs can also be transformed to two-way matrices $I \times JK$, where each column corresponds to one emission-excitation wavelength pair.

A detailed description of PCA transformation can be found in the literature (Despaigne and Massart, 1998; Persson and Wedborg, 2001; Spencer et al., 2007; Wolf et al., 2007). Essentially, PCA is a multivariate method of high-dimensional data simplification, where an original data matrix X is reduced to number of principal components calculated as the directions of the maximum variance of combined variables. The calculated principal components are in order of decreasing variance, where the first principal component describes the greatest variance within the dataset and all successive components account for the variance in decreasing order of magnitude. Depending on the overall variability in data explained by the primary components, in PCA just the first few components are analysed to investigate any valid correlations and relate them to input data structure and functions. The linear PCA projection is defined as:

$$Y = X T, \tag{5.1}$$

where: Y is the pattern matrix (scores matrix) of size $n \times p$, X is the input data matrix of size $n \times m$, and T is the transformation matrix (loadings matrix) of size $m \times p$. The loadings matrix expresses the importance of each variable (excitation-emission pair) in the original data matrix, whereas the scores are coordinates of samples in the PCA projection.

The advantage of the three-way PARAFAC model results from the appropriateness of the fluorescence data format (sample by emission by excitation wavelength) and the model structure, which reflects the true underlying parameters (Bro, 1998). PARAFAC is commonly

used in fluorescence data analysis to extract the most important fluorescence components (models of fluorophores) along with their spectral properties (emission and excitation wavelengths) and relative concentration in a sample. Consequently, the components in the three-way models are easier to interpret than in the corresponding two-way models (e.g. PARAFAC vs two-way PCA). Although PARAFAC is commonly described as a three-way version of PCA with data decomposition into scores and loadings, there are significant differences between models regarding the input data structure and model constraints (Bro, 1998; Stedmon et al., 2003; Ohno et al., 2008). Unlike PCA, where an infinite number of models with equal fit exist, the PARAFAC model produces a unique, potentially chemically meaningful solution (Bro, 1998). Nevertheless, the robustness of three-way models is highly dependent on trilinearity of the data and selection of the correct number of components.

The PARAFAC algorithm decomposes the data array $I \times J \times K$ into a trilinear model that minimizes the sum of squares of the residuals E_{ijk} (Equation 5.2) (Bro, 1997; Stedmon et al., 2003; Divya and Mishra, 2007):

$$X_{ijk} = \sum_{f=1}^F A_{fi} B_{fj} C_{fk} + E_{ijk} \quad (5.2)$$

where: X_{ijk} is the fluorescence intensity measured at emission wavelength j and excitation wavelength k for sample i , A , B and C are the loading matrices and F is the number of components (Bro, 1998).

The PARAFAC technique is a commonly used tool for fluorescence data analysis. Examples of fluorescence data analysis using PARAFAC include characterization of the organic matter-metal binding process (Ohno et al., 2008), quantitative determination of the kerosene fraction present in diesel (Divya and Mishra, 2007), classification of ballast water

(Hall et al., 2005), estuarine water (Stedmon et al., 2003) and edible oils (Guimet et al., 2005). Bro (1998) utilized PARAFAC component scores for modelling process parameters and quality of sugar production, where fluorescence data were used to monitor the sugar production process. More recently, Fellman et al. (2008) used coupled EEM-PARAFAC approach to characterise OM from Alaskan soil and riverine water. Murphy et al. (2008) used an EEM-PARAFAC approach to distinguish between terrestrial and autochthonous OM sources in marine environments.

5.1.1.2 Self-organizing map and artificial neural networks

The SOM is a powerful pattern recognition algorithm based on the two-layered artificial neural network (ANN), consisting of number interconnected single processing units called neurons or nodes. To understand the mechanism of SOM network training and its ability for data reduction, some fundamental concepts of ANNs were summarized below.

Artificial neural networks are powerful computational tools, frequently used in environmental modelling studies (Andrews and Lieberman, 1994; Häck and Köhne, 1996; Daliakopoulos et al., 2005; Alunkaynak, 2007). An ANN can be described as a mathematical model of a specific structure, consisting of a number of single processing elements (nodes, neurons), arranged in inter-connected layers (Figure 5.1). Each active neuron multiplies each input vector by its weight, sums the products and passes the sum through a transfer function to produce the output.

ANNs are powerful, non-parametric, parallel computational tools frequently employed for data classification and calibration (Bos et al., 1993; Despagne and Massart, 1998; Basheer and Hajmeer, 2000) due to their ability to model nonlinearity, and properties such as fault and noise tolerance (ability of processing noisy, uncertain data), self-modelling, self-learning (by

example) and generalization capabilities (Basheer and Hajmeer, 2000). ANN functions are facilitated by a special structure, consisting of the number of processing units (neurons) arranged in interconnected layers. In a typical ANN, the input layer neurons are responsible for presenting the data to the network, whereas the output layer neurons generate the overall network response to the input data. Neurons in the hidden and output layers are active and perform computational transformations by summing up their input connection weights multiplied by the output of corresponding neurons from the preceding layer and generating the output that is passed to the successive layer. The iterative process of adjusting connection weights between neurons of different layers is called network training. The trained network needs to be validated on data not used in the training phase. If the trained network subsequently returns the appropriate results for the independent dataset, it can be used as a calibration or classification model (Hammerstrom, 1993).

There is a substantial amount of published work which employed multifarious ANNs in modelling of fluorescence and water quality data. Scott et al. (2003) used ANNs for pattern recognition of olive oil fluorescence spectra. Li et al. (2000b) applied ANNs for classification of the fluorescence properties of multicomponent mixtures of fluorescent dyes, whereas Wolf et al. (2007) associated the fluorescence measurements of biofilm reactors with process performance parameters. A similar approach was presented by Häck and Köhne (1996) to estimate wastewater process parameter concentrations (e.g. NH_4) using ANNs (input data included pH value, conductivity, redox potential, turbidity). Gontarski et al. (2000) utilized back-propagation neural networks (BPNN) to predict the reduction of TOC in wastewater treatment using standard process parameters (e.g. inlet wastewater TOC, the concentration of suspended solids in reactors).

The self-organizing map (or Kohonen ANN) approach provides the conversion of nonlinear statistical relationships between high-dimensional data into simple geometric relationships on a low-dimensional map, while keeping the most important topological and metric relationships of input data (Kohonen, 1998; Rhee et al., 2005; Garcia et al., 2007). This type of neural network comprises two, fully-connected layers: input vector of high dimension and an output layer.

SOMs are often used in pattern recognition and feature extraction, data compression and as a pre-processing tool for other networks (Hammerstrom, 1993). Examples of SOMs use include classification of fluorescence data to monitor the fermentation processes (Lee et al., 2005; Rhee et al., 2005), and exploratory analysis of metalloproteins based on X-ray fluorescence analysis of the metal ions (Garcia et al., 2007).

Like PCA, SOM is an example of an unsupervised clustering algorithm in which any existing pattern is assigned to one of the categories, not specified or not known a priori by a domain-expert. In PCA and SOM, the feature extraction from the input domain is performed with linear (PCA) or nonlinear (SOM) transformation of the input data on the lower dimension of k principal components (PCA) or k -dimensional map (SOM). The feature describes an elementary pattern of information that represents partial aspects or properties of an item (Kohonen, 1998). In fluorescence data analysis with SOM, the extracted features can be referred to the presence of a particular fluorophore (or group of fluorophores) or its spectral properties.

In a SOM network, the connection weights of size of the input data m are stored in input neurons and during training are projected onto k -dimensional output space (Kohonen, 2001). The neurons in the input layer of the SOM are connected with each input sample (EEM converted to a single vector) and have an associated reference vector that contains SOM

weights. The reference vector (also called the codebook or weight vector) can be defined as $d_i = [d_{i1} \ d_{i2} \ \dots \ d_{im}]$, where m is equal to the dimension of the input vectors (number of fluorescence excitation-emission pairs). The analysis of the network output provides the basis for extraction of relationships and regularities from the original data. Neurons in the output layer can be arranged into hexagonal or rectangular lattices and different map shapes can be implemented (grid, cylinder or toroid; Kohonen, 2001).

SOM training is an iterative process in which for each input sample comprising an unfolded EEM, the neuron with reference vector weights most similar to the input vector is first identified (winner or best-matching unit, BMU). Thus, for each input EEM presented to the SOM network, the output neuron with the reference vector most similar to the vector representation of EEM is selected. Once the best-matching reference vector for each input EEM vector is found, its weights and the weights of its neighbouring neurons are modified and moved towards the input vector (self-organization feature of the algorithm) (Equation 5.3):

$$w_i(k+1) = w_i(k) + \varepsilon(k)h_p(i, k)\{x_j(k) - w_i(k)\} \quad (5.3)$$

where: $w_i(k)$ is the previous weight of neuron, $w_i(k+1)$ is the new weight of neuron, $\varepsilon(k)$ is the learning rate, $h_p(i, k)$ describes the neighbourhood of the winning neuron, k is the number of epochs (a finite set of input patterns presented sequentially) and p is the index of the winning neuron. The learning rate describes the speed of the training process ($0 < \varepsilon(k) < 1$) and decreases monotonically during the training phase. The topological neighbourhood can be described as a neighbourhood set of array points N_c around the given node c . During the training of the map, the radius of the N_c (a size of the neighbourhood set) decreases monotonically to enable the global ordering of the map. Thus, the projection of the input data

is done in two phases: the main training (large N_c radius) and fine adjustment of the map (small N_c radius) (Kohonen, 2001).

The trained network activates the appropriate output neurons of the network according to the input samples without prior knowledge of the process that produced the data and its distribution. Consequently, the analysis of the networks output provides the basis for extraction of relationships and regularities from the original data.

Here, for the purpose of the SOM and PCA analyses, fluorescence EEMs of raw and clarified water for 16 WTWs were deconvoluted to two-dimensional vectors, where each column corresponds to one emission-excitation wavelength pair.

5.1.2. *Calibration techniques*

In the multivariate calibration of fluorescence spectra, a mathematical model relating the fluorescence properties of fluorophores (e.g. intensity) to an actual, measured quantity (e.g. TOC concentration) is developed. The set of reference samples containing both independent variables (explanatory variables) and target values is firstly used to calibrate the model (pattern learning), and then the model's prediction accuracy is tested (validation) for an unknown set of samples.

Here, the fluorescence spectra were calibrated with measured TOC removal for 16 WTWs. The initial fluorescence dataset decomposed into PCA, PARAFAC scores and SOM normalized weights was used as an input for different calibration models. To evaluate the efficacy of selected decomposition methods in recognition of the most important features of fluorescence spectra, the complete EEMs and peak-picking results were used in the calibration. The fluorescence dataset (290 paired raw and clarified water EEMs) was divided into calibration and validation sets by selecting 25% (70 samples) of the validation samples

covering the whole range of data variation in the projection of the first two principal components. The details of the input datasets for the calibration models can be found in Table 5.2.

Four different regression algorithms were tested: multiple linear regression (MLR), stepwise regression (SR), partial least squares analysis (PLS), and artificial neural network with back-propagation learning (BPNN).

In a standard MLR regression model, all dependent variables are simultaneously regressed onto all the independent variables to minimise the squared error of the predictions according to equation 5.4:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (5.4)$$

where: Y is the predicted dependent variable, b_0 to b_n are partial regression coefficients, and X_1 to X_n are independent variables.

Both stepwise regression and partial least squares analysis techniques can be considered as extensions of the standard MLR model.

The SR analysis aims to select a statistically significant subset of independent variables in order of importance in predicting the dependent variable. It involves sequentially testing the independent variables and exclusion from the regression variables of p -values greater than 0.05. Although the stepwise regression model has clear advantages in selecting and grading of independent variables, it is important to be aware of some of the method limitations. Some authors (Thompson, 1995; Hopkins, 2005) underline three possible misconceptions with regard to the application of the stepwise regression method: lack of theory or model to support the stepwise-derived equation and possibility of incorporating accidental variables, the

conceivable inclusion of false-best variables, and finally, the possible violation of the model specification due to the addition of extra variables to the analysis.

The PLS algorithm enables sequential decomposition of an array of independent variables into a multi-linear model in which the scores have the maximal covariance with the yet unexplained variation of the dependent variable. In the PLS model, the two sets of variables, X and Y , are simultaneously decomposed into a sum of k variables (Equation 5.5 and 5.6):

$$X = TP^T + E = \sum t_k p_k^T + E \quad (5.5)$$

$$Y = UQ^T + F = \sum u_k q_k^T + F \quad (5.6)$$

where: T and U are the score matrices, P and Q are the loading matrices, E and F are the residual matrices. The matrix Y , containing, for example, concentration data of a particular fluorophore, can be derived from independent data (matrix X , containing for example the fluorescence spectra) according to equation 5.7:

$$Y = TBQ^T + F \quad (5.7)$$

where: B is the regression coefficient matrix for scores T and U .

The artificial neural network with back propagation learning (BPNN) is the most common optimisation algorithm of neural networks (Rumelhart and McClelland, 1986). For the given input, the actual output is compared with the desired output and weights are adjusted iteratively to minimize the error of the entire network (Zupan and Gasteiger, 1991; Bos et al., 1993; Hammerstrom, 1993). The weight adjustment and error calculation is propagated

backwards from the output layer until the input layer is reached. In the design of BPNN, three parameters are defined: the learning rate, the momentum factor and the range in which the initial weights are randomized (Bos et al., 1993). The learning rate describes the speed of the training process, while the momentum coefficient is used in weight updating to maintain the optimum search stability (Basheer and Hajmeer, 2000). The determination of both parameters is a trade-off between the speed of the training and likelihood of finding the global rather than a local minimum. Therefore, a trial-and-error procedure is commonly used to adjust back-propagation algorithm parameters (Zupan and Gasteiger, 1991).

In this work, a three layer feed-forward type of neural network with back-propagation learning, the most popular type of ANN, was used to calibrate fluorescence and TOC removal data (Figure 5.1). Firstly, the network topology (number of layers and number of nodes in the hidden layer) was assessed to produce a network large enough to train the fluorescence data but small enough to generalize well the regularities within the data.

The input layer comprised six nodes, while the output had only one node, denoting the TOC removal. One hidden layer with a sigmoid transfer function was used in the model, as recommended for the purpose of multivariate calibration by other authors (Smits et al., 1994; Despagne and Massart, 1998). Initially, the cross-validation algorithm was used as a convergence criterion to optimize the learning epoch size and avoid overtraining and to select the optimum number of hidden nodes (Li et al., 2000b; Wolf et al., 2001). Additionally, the calibration error for different settings of networks was compared. The topology corresponding to the lowest error determined in five trials with different sets of initial weights was retained for the validation.

The best test network architecture (6 input nodes, 4 hidden nodes, 1 output node) was set to train until convergence was achieved. The Levenberg-Marquardt algorithm (Marquardt, 1963)

with early stopping for improving generalization and preventing data overfitting was employed for training the networks. The Levenberg-Marquadt technique is faster and less easily trapped in local minima than other optimization methods, as the BP momentum coefficient is decreased during the training in relation to error gradient information (Marquardt, 1963; Basheer and Hajmeer, 2000). The early stopping algorithm monitors the validation error and the training stops when data begin to overfit as indicated by the increasing validation error.

5.2. PATTERN RECOGNITION OF DRINKING WATER FLUORESCENCE

DATA

5.2.1. Principal components analysis

Principal components analysis was performed on the normalized and mean-centered fluorescence dataset, containing samples of both raw and clarified water. It was found that the first ten principal components explained solely 74.6 % of the total variance (Figure 5.3), while the first three components accounted for 69.3 %. The first principal component can be attributed to the most important spectral features of the dataset, however its contribution in explaining the total variance was less than 50 %. Successive components diminished substantially in significance, with the residual variance distributed uniformly between a large number of components. Thus, the variation in the spectral properties determined the presence of many distinctively different fluorescence features that can not be effectively decomposed into set of just a few principal components.

Figure 5.4 presents the loading plots for the first four components. The first principal component (PC1) presents broad peak with the excitation wavelengths between 250 and 300 nm and emission between 300 and 350 nm, obscured by the region of the removed

fluorescence of the Raman line. The spectral position of this fluorescence peak can be attributed to protein-like fluorescence, particular to microbially-impacted waters.

On the other hand, PC2 represents natural, humic-like fluorescence identified as a peak A with emission maximum of 400 nm and excitation between 260 and 300 nm. The shift towards higher excitation wavelengths in the loadings of PC2, suggests the presence of the secondary, visible fraction of humic-like fluorescence, designated peak C. PC3 loadings exhibit increased intensities in the excitation wavelengths typical of peak C fluorescence and blue-shifted emission wavelengths obscured by the Raman line. Consequently, this component can indicate the presence of the humic-like fraction with a lower degree of hydrophobicity, implying an increase in the hydrophilic fraction which is the most difficult to remove during the treatment process. The relative change in this fluorescence between raw and clarified water appears to correlate with TOC removal and thus can indicate the organic matter removal efficiency. The signature of the loading plot of PC4, demonstrates the background, residual fluorescence and reveals the noise character. Therefore, this component appears to be invalid and was excluded from the further analysis.

Figure 5.5 and 5.6 present score plots of the first three components. PC1 of raw water (Figure 5.5 a) negatively highlights sites Bamford, Frankley and Tittesworth with water abstraction from reservoirs. For those sites, the raw water organic matter exhibits a higher degree of hydrophobicity and lower microbial fraction content (tryptophan-like fluorescence). Therefore, the raw water characteristics enhance organic matter removal efficiency (better removal of more aromatic organic matter) as indicated by the greatest change in PC1 scores between raw and clarified water. On the contrary, sites including Church Wilne and Little Eaton positively correlate with PC1, and the relative change in PC1 scores between raw and clarified water is lower compared with the reservoir sites. For these sites, poor organic matter

removal is related to the predominance of highly-variable, riverine organic matter with a significant contribution of highly-microbial and hydrophilic fractions. Hence, PC1 reflects the combined effect of the protein-like fluorescence content and the degree of hydrophobicity on organic matter removal efficiency. The variation in PC2 scores between raw and clarified water are less distinctive and marked with a slight decrease (Figure 5.6). The positive scores on the second component can be attributed to sites with high and stable TOC concentrations (Campion Hills, Draycote, Melbourne, Whitacre), whereas negative scores account for low and variable TOC of sites Frankley and Little Eaton. However, the similar signature of the raw and clarified water indicates that PC2 also corresponds to the more resistant, hydrophilic organic matter. The explicit changes in the score of PC3 are demonstrated by two sites with predominantly hydrophobic organic matter (Bamford and Frankley), while the remaining sites exhibit only a minor increase. Consequently PC3 reflects the overall change in the quantity of peak C fluorescence, and thus could indicate the TOC removal. However, the opposite trend can be discerned for Campion Hills WTW, where an increase in PC3 scores can be observed between raw and clarified water. This anomaly cannot be simply explained with distinctively different organic matter or spectral properties, thus PC3 effect on the sample distribution is equivocal.

5.2.2. *Parallel factor analysis*

A PARAFAC model with a non-negativity constraint on all modes (samples, emission and excitation) was implemented in Matlab® (Bro, 1997). PARAFAC models were calibrated independently for each water type (raw and clarified) and for the whole dataset.

An important step in PARAFAC modelling is the model validation, with the use of several available diagnostics. Here, core consistency analysis (CORCONDIA) was performed to

determine the optimum number of components (Bro, 1997; Andersen and Bro, 2003). The appropriate models with the highest number of components, highest variance explained and a valid CORCONDIA value (raw 84.6%, clarified 92.1 %, all data 90.7 %) were selected (Table 5.3). Additionally, the residual sum of squares was analysed and plotted against the number of components. The curves flatten out for three components indicating the optimum number of PARAFAC components, which is in accordance with the results obtained from core consistency analysis.

However, to derive a valid number of PARAFAC components, not only were the statistical diagnostics evaluated, but also the emission and excitation spectra were analysed (Figure 5.7). The emission loadings demonstrate well-defined, single peaks, whereas in the excitation loadings, double-peaks can be discerned for the first two components, together with an increased signal in the loading of the third component between 245 and 260 nm. This inferior appearance of the excitation loadings could be attributed either to the presence of some variance related to another component (visible as shoulders in the loadings) which is unexplained by the trilinear model and/or model impediment due to the occurrence of non systematic noise in the fluorescence signal at lower excitation wavelengths. Therefore, the validity of additional PARAFAC models was evaluated, adopting approach presented by Andersen and Bro (2003). The original PARAFAC solution was found with the application of a non-negativity constraint on all three modes (samples, emission and excitation wavelengths). As suggested by Andersen and Bro (2003), the unimodality constraint can significantly reduce the interference from the minor artefacts in the data and thus can enhance the interpretation and selection of the valid components. For raw, clarified and all-dataset fluorescence spectra, three- and four-component PARAFAC models with the unimodality constraint in the emission and excitation modes were tested. The application of the

unimodality constraint in the case of the three-component model did not improve the percentage of the explained variance and core consistency diagnostic, however the shoulder observed in the loading of the third component decreased. For the four-component constrained model the variance explained remained unchanged compared with the results of unconstrained models, with a minor increase in core consistency value for the raw water model (from 4.5 to 8.5 %). For the clarified water and all-dataset models a corresponding decrease in CORCONDIA percentage was observed. More explicit were the differences in the emission and excitation loadings in four-component PARAFAC models (Figure 5.8).

In the four-component model, the second and fourth components (Figure 5.8 c, d, and g, h) exhibited similar spectral properties to the corresponding second and third components from the unconstrained models (Figure 3 c, d and e, f). However, the initial first component from the three-component model was decomposed in two separate components, the first and third, with similar emission loadings but varying excitation loadings (maxima below 245 nm and at 350 nm). In the clarified water four-component model, the emission and excitation spectra of the first and second components demonstrated different signatures, with lower emission in the first component loading and a distinctively unique spectral position in the excitation profile of the second component. Hence, the PARAFAC analysis revealed a variation in fluorescence character over the range of samples, pertinent to various organic matter sources and water types, which reflected the removal of specific organic matter fractions during the treatment process.

From the visual appearance of the four-component clarified water model it appears that the first three components might be a linear combination of two real components, and thus this solution appeared to be invalid. The instability of the four-component was inferred from the

CORCONDIA analysis as indicated by the negative value of this diagnostic (-6.5 %) for clarified water and low value (8.5 %) for raw water.

A valid model in terms of statistical diagnostics, with consistent emission and excitation loadings for raw, clarified and all fluorescence data was obtained for three components. However, from the analysis of the loadings it can be seen that two first components produced double-excitation maxima indicative of the presence of more components. An evaluation of more complex models (four and five-components) confirmed this hypothesis. Both components with emission wavelength around 400 nm and 460 nm can be partitioned into two more components of different excitation spectra (less than 245 nm and at 350). Despite this inherent complexity of the components, the higher-component models were unstable as derived from CORCONDIA analysis (low and negative values) and from the repeated PARAFAC analyses. For the same datasets, different PARAFAC models were found, which contradicts the PARAFAC model assumption of producing a unique solution.

Therefore, on further analysis, only the three-component PARAFAC model was validated with the split-half analysis and the residuals evaluation. For three components, the models were found to be valid with the variance explained for raw, clarified and all data being 99.5 %, 99.2 % and 99.3 % respectively. The PARAFAC loadings and scores are given in Figures 5.7 and 5.9.

The number and order (variance explained) of the three components derived independently from raw, clarified and all data models exhibited the same pattern with three components, where emission and excitation loadings resembled each other (Figure 5.7).

Component 1 exhibited a wide emission spectra with maxima at 460 nm and 470 nm for clarified and raw water respectively (Figure 5.7 a), and the two excitation maxima at 250 nm and 345 nm (Figure 5.7 b). The emission and excitation maxima of component 2 occurred at

shorter wavelengths compared to component 1 (402 nm and below 245 nm; Figures 5.7 c and d). The emission spectrum of component 3 was maximum at 350 nm wavelength with excitation spectra consisting of two peaks (below 245 nm and 285 nm; Figures 5.7 e and f). Results of earlier studies (Coble, 1996; Bro, 1997; del Vecchio and Blough, 2002; Stedmon et al., 2003; Hall et al., 2005; Fellman et al., 2008), indicate that the three PARAFAC components can be identified as a visible humic-like fluorescence associated with peak C (Component 1), UV humic-like fluorescence corresponding to peak A (Component 2) and protein-like fluorescence (peak T, Tryptophan, Component 3). When compared with the four-component constrained PARAFAC model, the fourth component (Figure 5.8 g and h) corresponds to tryptophan-like fluorescence and its emission and excitation spectra are similar to the third component in three-component model (Figure 5.7 e and f). However, the application of the unimodality constraint reduced the 'shoulder' in tryptophan excitation loading between 245 and 260 nm. The difference between water types can be discerned from the analysis of the first three components. The raw water and all data four-component models exhibited the presence of two UV humic-like fluorescence peaks with different emission maxima (the first and second component; Figure 5.8 a, b, c, d), and one visible humic-like fluorescence as a third component (peak C; Figure 5.8 e and f). For the clarified model, the first component corresponded to UV humic fluorescence with lower emission wavelength (430 nm; Figure 5.8 a and b), and the second and third to visible humic fluorescence with emission maximum at 400 nm (second; Figure c and d) and at 470 nm (third; Figure 5.8 e and f). From this comparison it appears that, in the clarification process, the UV humic-like fluorescence of the higher emission wavelength is removed more than other fractions.

The PARAFAC score plots (Figure 5.9) reveal interesting correlations between humic-like and tryptophan-like fluorescence. The raw water characteristics of reservoir sites (Bamford,

Frankley and Tittesworth) exhibit minor variations in the microbial fraction compared to Church Wilne and Little Eaton, which are prone to significant changes in tryptophan-like fluorescence intensity related to the distinctive contribution of algae. This division into two groups of sites, the first with stable tryptophan-like fluorescence (variation in humic-like fluorescence which can be indicative of the presence of different fractions) and the second with stable peak C fluorescence (predominance of the one fraction of humic organic matter with variable microbial inputs) is even more discernible for clarified water scores (Figure 5.9 b). Furthermore, for clarified water a significant relationship exists between scores of component 1 and 2 ($R^2 = 0.92$) which is in accordance with the peak values results (Figure 5.9 c).

For the majority of sites, a similar pattern of temporal variation in PARAFAC scores was inferred. Figure 5.10 shows the scores of the three PARAFAC components for raw and clarified water for two sites with contrasting organic matter properties sites (Bamford and Little Eaton). For both sites the score variation was greater for clarified water. Components 1 and 2, corresponding to fulvic and humic-like fluorescence (peak C and peak A fluorescence respectively), exhibited increased scores in the winter with maxima in January, whereas the lower values were predominant over the summer months (Figure 5.10). The opposite applied to component 3, which is indicative of the microbial fraction of organic matter content. The winter and late summer months (July) can be characterised by lower scores on component 3 and elevated values were observed in April and May.

These OM composition PARAFAC results were in accordance with visual inspection of the EEMs and the peak-picking approach suggesting that there were three main fluorophores present in the water samples. Moreover, an analysis of PARAFAC models with a higher number of components suggested the presence of more fluorophores which cannot have been

validated for the entire dataset or even water types (raw and clarified). Therefore, to derive a valid model containing all variation in fluorescence spectra, a solution with fewer components had to be chosen. A lack of good, overall diagnostic for the selection of the number of valid components impedes interpretation of the PARAFAC model and makes an analysis a time-consuming process.

5.2.3. *Self-organizing map analysis*

The input matrix, comprising 625 samples and 2515 excitation-emission pairs, was presented to the nodes of the SOM input layer simultaneously (batch mode), and neuron weights were initialized using linear initialization along the two greatest eigenvectors of the input matrix (Kohonen, 2001). The size of the output layer was determined by finding the ratio of the two greatest eigenvalues of the input matrix. The final map contained 120 nodes (size 15 x 8).

SOM training is an unsupervised process of adjusting the connection weights (reference vectors) between nodes in the input and output (map) layers. The training is completed once for each input sample its representation in the form of the reference vector with the weights most similar to the input data (best-matching unit) is found. The analysis of the SOM output requires the use of various visualization and clustering tools, providing substantial information on the input data distribution and relationships with the measured variables.

Figure 5.11 presents some basic SOM visualization methods, including a unified distance matrix algorithm (U-matrix, Figure 5.11 a), sample distribution on the map with cluster borders determined (Figure 5.11 b) and single hit histograms (Figure 5.11 c). The U-matrix (Ultsch et al., 1993) is the most common graphical representation of the SOM structure, in which distances between neighbouring map units are calculated and visualized using a grey or colour scale on the trained map (Park et al., 2003). Compared with the original map size

(15 by 8 neurons), the U-matrix comprises additional map units to visualize the distances between neurons. High values on the U-matrix (light areas) indicate large distances between neighbouring units and hence can be helpful in determining the cluster borders as clusters typically form uniform areas of low values (dark areas). From Figure 5.11 it can be observed that the cluster structure of the fluorescence data is not well-defined as the light and dark areas on the U-matrix are not easily partitioned. However, the presence of a few clusters can be discerned (i.e. in the lower right-hand side corner of the map). As the U-matrix provides the information on the cluster structure, the number of valid clusters can be derived from the k-means algorithm (Jain and Dubes, 1988). The k-means clustering algorithm is used for minimizing the sum of squared Euclidean distances between the input (unfolded EEMs) and the SOM reference vectors. The best clustering minimizes the sum of the squared distances (and also the Davies-Bouldin index) between each input data vector and its nearest cluster center (Davies and Bouldin, 1979). Here the optimum number of ten clusters was determined (Figure 5.11 b). To correlate the cluster pattern on the map with sample distribution, for each map node the most frequent best-matching unit (BMU) with assigned site number was found. Each cluster contained sites with similar organic matter properties measured with excitation-emission spectra (Figure 5.11 b). Distinctively unique spectral properties of raw water were discerned for Bamford (1) and Frankley (6), which are both reservoir abstractions (raw water is stored in reservoirs prior to uptake for treatment), whereas the distribution of other sites was more complex. A good discrimination between raw and clarified water fluorescence properties occurred for sites located at the bottom of the map, sites Bamford (1), Frankley (6), Mythe (10), Ogston (11), Tittesworth (14). The opposite was observed for sites clustered at the top of the SOM map (Campion Hills (2), Church Wilne (3), Draycote (5)), with similar properties for raw and clarified water, as indicated by the short distances on the map. Finally,

the distribution of samples on the SOM map can be portrayed with hit histograms (Figure 5.11 c). For each neuron, the hit characteristic is calculated on the basis of the map response to the input data. The size of the marker indicates how many times each map unit was the BMU for the dataset. It can be seen that data is uniformly distributed over the map, with a number of neurons located at the edges of the map being the most frequent BMUs (neuron 1 – 16 hits, neuron 5 – 15 hits, neuron 9 – 20 hits, neuron 120 – 18 hits) (Neurons numbered sequentially from top left to bottom right).

While the U-matrix, cluster structure and hit histograms reveal a pattern of sample distribution on the map, the reference vectors of selected neurons and component planes exhibit the significance of particular fluorescence variables. From the hit histogram in Figure 5.11 c it can be inferred that some neurons represent the greatest number of fluorescence samples. Thus the spectral properties derived from the reference vectors of those neurons can provide important information on the dominant fluorescence features of the dataset. The EEMs of two neurons located at the left-hand side of the SOM map (Figure 5.12 a and c), indicate the predominance of humic-like fluorescence at the lower, UV excitation wavelengths. Additionally, a shift towards higher emission wavelengths can be observed for neuron 9. Samples projected onto the upper SOM neurons demonstrate a substantial contribution of protein-like fluorescence as the tryptophan-like centre located at excitation-emission wavelengths of 280/350 nm can be discerned for both neuron 1 and 109 (Figure 5.12 a and b). Distinctively different fluorescence properties can be observed in the excitation-emission spectra of neuron 120 with the humic-like fluorescence peak shifted towards higher excitation and emission wavelengths (Figure 5.12 d). These results are in accordance with the analysis of the response of each fluorescence variable (component) in the form of component planes (Figure 5.13). The component planes depict the values of the reference vectors for

different fluorescence variables and allow correlation between the samples distribution and particular excitation-emission wavelengths and hence main fluorophores. Thus, for each excitation-emission wavelength pair, a corresponding component plane can be obtained that enables correlation between sample location on the map and fluorescence properties. In Figure 5.13 the component planes at excitation wavelength of 280 nm are shown for different emission wavelengths (300, 350, 400, 450 and 500 nm). The high values in the component plane denote higher fluorescence intensity. As the emission wavelength increases, the centre of the highest values moves from the top to the bottom of the map with relative increase in fluorescence intensities (compare maximum values for each component plane). For example, in Figure 5.13 b, the component plane for an excitation wavelength of 280 nm and emission wavelength of 350 nm is presented; this represents the spectral area related to tryptophan-like fluorescence. Higher values on this component plane indicate the predominance of highly-microbial organic matter for sites located in the upper part of the map.

In Figure 5.14 the hit histograms for sites of different organic matter properties and efficiency of organic matter removal are shown. The geometric distance between raw and clarified water samples correlates with the organic matter removal, with increased removal represented by increased distance between raw and clarified samples. Thus sites Bamford (1) and Melbourne (8) exhibit better organic matter removal than Draycote (5) and Church Wilne (3). The greater the spread of water samples of particular type on the map, the greater the variation in spectral properties can be discerned. The varying location of the water samples in the horizontal plane indicates qualitative changes in OM character (e.g. Mythe (10), Strensham (13)), whereas changes in the vertical plane correspond to quantitative changes in OM (e.g. Draycote (5), Melbourne (8)). Thus, the raw water OM character and concentrations at Mythe and Strensham undergo significant seasonal changes (raw water samples spread over

the entire map), whereas at Draycote and Melbourne the OM character is stable but with varying TOC concentrations. Furthermore, the location of the raw water samples in the left part of the map, suggests that water at Draycote and Melbourne demonstrates hydrophilic OM character.

Stable qualitative (degree of aromaticity, content of microbial fraction) raw water properties are typical of Bamford (1), Draycote (5), Frankley (6), and Tittesworth (14). For each of these sites except Draycote, the treatment process can be readily optimised as throughout the year similar OM character and concentrations can be expected. Draycote WTW, however, demonstrates stable OM character but varying OM concentrations; hence seasonal treatment adjustment is needed to account for changing raw water TOC. Likewise, variable qualitative and quantitative raw water OM properties are observed at Campion Hills (2), Church Wilne (3), Little Eaton (7), Mitcheldean (9), Mythe (10), Shelton (12), and Strensham (13). Stable qualitative clarified water properties can be observed for Bamford (1), Church Wilne (3), Cropston (4), Draycote (5), Frankley (6), Mitcheldean (9), Tittesworth (14), and Trimpley (15) with highly variable TOC typical of Bamford (1), Draycote (5), Frankley (6). Variable clarified water properties suggest presence of variation in the raw water OM properties and are observed for Mythe (10), Ogston (11), and Strensham (13).

The qualitative OM properties can be directly compared with OM removal efficiency on the map. Site 1 (Bamford) represents uniform raw water properties, whereas a greater variation is typical for sites Draycote (5), Church Wilne (3) and Melbourne (8). It can be concluded that Draycote (relatively hydrophilic raw water organic matter with lower emission wavelength) has poorer removal compared to site 1 (Bamford), where the hydrophobic character of the organic matter enhances the efficiency of the treatment process.

As stated above, the location in the upper part of the SOM (lower emission wavelengths) corresponds to the increased inputs of the microbial fraction. Thus, distinctive spectral properties can be attributed to sites Church Wilne (3) and Draycote (5), where microbial fraction related to tryptophan-like fluorescence has a significant contribution in the raw water fluorescence signature. The presence of microbial organic matter indicates that those sites could be more prone to algal outbreaks which can further deteriorate drinking water quality.

5.2.4. *Comparison of pattern recognition methods*

In the data decomposition algorithms, the original dataset containing n samples and m variables is projected onto k -dimensional space and new coordinates (numerical transformation of m variables) are calculated for each sample. From the three data decomposition methods evaluated in this study, PCA and PARAFAC represent multi-way algorithms, where the original dataset is projected onto k components coordinate plane with the components scores being the numerical representations of the variables. In SOM, the original data is projected on k neurons arranged in a two-dimensional map, where the original variables are transformed into neurons weights. Therefore, the pattern recognition involves analysis of the samples distribution in the new coordinate planes calculated with each algorithm (compare Figures 5.5, 5.6, 5.9 and 5.11).

The relationship between the original space and the PCA, PARAFAC component space or SOM map is expressed by loadings in multi-way analysis and SOM component planes. In fluorescence data decomposition, PCA and PARAFAC loadings and SOM component planes demonstrate the importance of particular excitation-emission combinations (fluorophore(s)) for the model and therefore provide a basis for the interpretation of variation in organic matter properties between samples.

To provide meaningful information on the fluorescence data patterns using decomposition methods, the calculated models have to be valid in terms of the optimum number of components (PCA), optimum number and chemical interpretation of components (PARAFAC), and optimum number of neurons comprising the map (SOM). In the SOM approach the map size is determined by the size of the input data, and the ratio of two greatest eigenvalues is commonly used to calculate the map size. The importance of a given component in multi-way analysis is calculated as a variance explained by this component. The first component can be attributed to the most important spectral features of the dataset and successive components are less important as indicated by the decreasing explained variance.

Here, the first three PCA components explained only 69.3 % of the total variance (Figure 5.3 and Table 5.4). Thus, PCA revealed the presence of many distinctively different fluorescence features of similar importance (explained variance). From the fluorescence data decomposition algorithms considered in this study, determination of the optimum number of PARAFAC components was the most complex and involved analysis of both statistical diagnostics (variance explained by the model, core consistency analysis (CORCONDIA) (Bro, 1997; Hall et al., 2005) and component loadings.

PCA and PARAFAC components and SOM components planes are models of groups of fluorophores with similar fluorescence characteristics. On the basis of spectral properties of PCA and PARAFAC components derived from component loadings, three components were recognized as peak T, peak A and peak C fluorescence which is in accordance with visual inspection of EEMs and fluorescence peaks discerned (Table 5.5).

The order and specific spectral properties of each component varied between PCA and PARAFAC models. The peak T component in PCA exhibited a broader peak with excitation wavelengths between 250 and 300 nm and emission between 300 and 350 nm, obscured by

the region of the removed fluorescence of the Raman line. The spectral position of this fluorescence peak demonstrated more complex properties compared to a well-defined peak T PARAFAC component and thus it was attributed to protein-like fluorescence rather than simply peak T fluorescence.

The peak A component in both models had a similar emission spectra with distinctively different excitation wavelengths. The shift towards higher excitation wavelengths in the loadings of PC2 suggested the presence of a secondary fluorophore, the visible fraction of peak C fluorescence. The PCA peak C component indicated the presence of the OM fraction with a lower degree of hydrophobicity, which is difficult to remove during the treatment process. The relative change in this fluorescence between raw and clarified water appeared to correlate with TOC removal and thus indicated the organic matter removal efficiency. On the contrary, the PARAFAC peak C component exhibited higher emission wavelengths indicative of the more hydrophobic OM fraction, which is easier to remove in conventional water treatment.

In the SOM approach, component planes depict the values of weights vectors for different fluorescence variables and allow correlation between the samples distribution and excitation-emission wavelengths. Three major directions of fluorescence properties change were observed. Horizontal and vertical axes of the map were found to correspond with fluorescence emission and excitation wavelengths, with increasing values from the top to the bottom and from the left to the right respectively. Moreover, the diagonal that joins the upper left with lower right corner of the map was the line of the greatest changes in variance within the dataset and discriminated the sites of radically different organic matter spectral properties.

The evaluation of the results of the decomposition models provided significant information on the differences between WTWs in organic matter character and removal. Moreover,

differences in organic matter composition between raw and clarified water and the relationship between organic matter character and the efficiency of its removal were discerned.

All methods revealed the presence of two groups of WTWs of distinctively unique spectral properties, the reservoir sites (Bamford and Frankley), and lowland sites Campion Hills, Church Wilne, Draycote, Little Eaton. The first group exhibited a good discrimination between raw and clarified water fluorescence properties, whereas the opposite was observed for the second cluster. For Bamford and Frankley, (reservoir sites) raw water organic matter exhibited a higher degree of hydrophobicity and lower content of the microbial fraction (tryptophan-like fluorescence). Therefore, raw water characteristics enhanced organic matter removal efficiency (better removal of more aromatic organic matter) as indicated by the greatest changes in PC1 scores and distances on SOM map between raw and clarified water. Conversely, sites Church Wilne (3) and Little Eaton (7) demonstrated poor organic matter removal which was related to the predominance of highly-variable, riverine organic matter with a significant contribution of highly-microbial and hydrophilic fractions.

From the above analysis it can be seen that the decomposition models provided both quantitative and qualitative information on organic matter characteristics and removal at WTWs. Both PCA and PARAFAC models decomposed the fluorescence data into a set of components related to fluorophores. In the PARAFAC model, supervised evaluation of the components obtained based on the analysis of performance statistics, emission and excitation loadings is required. However, as shown here, equivocal results of different diagnostic tools impeded the interpretation of the PARAFAC components. PCA and SOM provide an unsupervised data decomposition which also can pose difficulties to the interpretation process. PARAFAC components can be related in principle to the real fluorophores, whilst PCA

components are more complex and represent statistical representations of groups of fluorophores or particular OM properties.

In the SOM approach, the characteristic features of the data are selected automatically. This is a distinct advantage when analyzing samples from similar sources and with a uniform pattern of fluorophores. Unlike PCA and PARAFAC, the SOM enables easier interpretation of the samples distribution and fluorescence variables due to the availability of several visualization techniques (e.g. component planes, hit histograms; for details see Kohonen (2001)). The geometric distances between raw and clarified water samples on the SOM map indicated the degree of similarity in spectral properties of organic matter and correlated with removal efficiency. The greater the spread of water samples of particular type on the map, the more variation in spectral properties was observed.

5.3. CALIBRATION OF DRINKING WATER FLUORESCENCE DATA

5.3.1. Comparison of calibration methods

In the previous sections, the application of PCA, PARAFAC and SOM methods for pattern recognition of fluorescence data was investigated. Here, the quantitative outcomes of decomposition methods were used for calibration of fluorescence data with TOC removal. Different calibration methods were evaluated: multiple linear regression (MLR), stepwise regression (SR), partial least squares analysis (PLS), and artificial neural network with back-propagation learning (BPNN) (see Section 5.1.2). To evaluate the ability of the decomposition methods to reduce the initial dataset to the most important fluorescence features, complete excitation-emission matrices and peak fluorescence parameters (tryptophan- and fulvic-like fluorescence of raw and clarified water) were also used in calibration models (Table 5.2).

By comparing the performance of the selected decomposition models, it can be seen that PARAFAC and SOM approaches provided better calibration of fluorescence data than PCA (Table 5.6). The accuracy and residual error of prediction was similar for EEM, PARAFAC and SOM models, indicating that those decomposition techniques can successfully facilitate fluorescence feature extraction from the whole EEMs. However, the peak-picking approach (PEAK model) was equally good in BPNN modelling and only slightly poorer for SR and PLS models. Thus, for the fluorescence datasets with known or similar patterns of fluorophores, the peak-picking approach can produce similar calibration results to PARAFAC and SOM models without time-consuming data pre- and post-processing. However, PARAFAC and SOM provide additional tools for advanced fluorescence data analysis, where the sample distribution can be directly correlated with the most important fluorescence features, both in quantitative and qualitative terms.

In all cases except SR analysis, PCA performed poorer than the PEAK model, hence the use of standard peak-picking approach is recommended prior to PCA analysis.

The regression models used in this study produced consistent results as indicated by correlation coefficients and prediction errors (Table 5.6). For the independent validation fluorescence data, the poorest results were obtained for MLR model which accounted for 63-87 % of the total variance explained. The SR, PLS, and BPNN regression models produced similar results; however, for all decomposition methods the latter was slightly more efficient (higher correlation coefficients and lower prediction errors).

The BPNN algorithm is considered as being more flexible than the standard regression methods and therefore more difficult to implement as its flexibility can pose a danger of overfitting the calibration data and producing unreliable results. However, the advantages of this approach include fault and noise tolerance (ability to process noisy, uncertain data), self-

modelling, self-learning (by example) and generalization capabilities. Prior to modelling, the BPNN network requires the definition of a topology (number of nodes in a hidden layer) and of parameters that describe the speed of the training process and maintain the optimum search stability (the learning rate and the momentum factor). There are several rules of thumb that facilitate BPNN network design, but a trial-and-error procedure is commonly used to adjust back-propagation algorithm parameters and obtain a feasible network topology. However, once appropriately designed, trained and validated, the BPNN network can be a robust predictive tool provided that a substantial amount of training data is available.

Compared to BPNN, the PLS algorithm is simpler as it does not require definition of a set of parameters in the training phase. However, the validation procedure involves the selection of an appropriate number of components (latent variables), which is the crucial step in generating a valid and robust model. The leave-one-out cross-validation and the root mean squared error of prediction (RMSEP) are the most common techniques and statistical diagnostic used in components selection. The PLS algorithm explicitly incorporates dimension reduction and generates components explaining the most important features of the given dataset. Therefore this regression tool can be also successfully used in exploratory data analysis. Here, four PLS components were selected. The analysis of PLS scores provided a useful information relating the organic matter properties (degree of hydrophobicity) with efficiency of organic matter removal.

From the calibration methods examined in this study, three produced consistent results in TOC removal prediction modelling from fluorescence data (stepwise regression (SR), partial least squares analysis (PLS), and artificial neural network with back-propagation algorithm (BPNN)). However, the SR model is often criticised for a lack of theory or statistical model to

support the stepwise-derived model. Thus, for the purpose of the prediction of organic matter removal in drinking water systems, two models were found useful, PLS and BPNN.

5.4. CHAPTER CONCLUSIONS

In chapter 4 an analysis of fluorescence data with standard peak-picking approach was demonstrated. It was shown that the peak-picking approach uses solely a small portion of the available fluorescence data. Furthermore, an evaluation of large number of fluorescence samples (here 512 EEMs) with the peak-picking approach coupled with simple regression technique was problematic. Therefore, in this chapter application of different data mining techniques for analysis of large fluorescence dataset was discussed.

Three data reduction algorithms were evaluated: PCA, PARAFAC, and SOM. In PCA and PARAFAC analyses, three components were found to represent the fluorescence data and they were related to peak C, peak A, and peak T fluorescence respectively. These results were in agreement with the visual inspection of fluorescence EEMs and peak-picking approach presented in chapter 4. Moreover, PARAFAC modelling revealed differences in the composition of fluorophores in raw and clarified water samples and possible presence of more than three components. However, the higher-component PARAFAC models were unstable and invalid as it was assessed with the use of CORCONDIA diagnostic and split-half validation.

The results of calibration of decomposed fluorescence data with TOC removal efficiency suggested that peak fluorescence values and PCA scores yielded similar prediction accuracy and prediction error. PARAFAC model of the fluorescence data performed slightly better in the calibration. Thus, fluorescence data decomposition with PARAFAC is more robust and adequate compared to peak-picking approach and PCA analysis. However, for samples with

similar OM composition, the peak-picking approach can provide satisfactory indication of fluorophores properties. In addition, the peak-picking approach is more feasible in analysis of small fluorescence datasets as it does not involve lengthy data pre-processing required to use more advanced data analysis techniques (i.e. PCA, PARAFAC or SOM). Thus, for the purpose of fluorescence analysis of optimised coagulation data (54 EEMs) presented in the next chapter, the peak-picking approach was utilized.

Furthermore, in the chapter a novel approach to the analysis and interpretation of fluorescence data with SOM was presented. The SOM facilitated pattern recognition of the fluorescence data and revealed linkages between sample distribution and the importance of the particular spectral properties. With reference to the fluorescence differences between raw and clarified water, the SOM enabled correlation of the organic matter removal efficiency with the organic matter properties derived from the fluorescence EEMs. These results demonstrate that SOM can be a powerful decomposition tool for fluorescence data analysis and, with the use of available toolboxes the implementation and interpretation process can be easier compared to other data analysis techniques.

The results on evaluation of data mining techniques for fluorescence data analysis presented in this chapter, were found to fully address the Objective 4 and hypotheses H5 and H6 (see Section 2.5).

The results presented in this chapter have been published in *Journal of Geophysical Research – Biogeosciences* (Bieroza et al., 2009 b) and presented at *First International Conference on Soft Computing Technology in Civil, Structural and Environmental Engineering* (Bieroza et al., 2009c). In addition, a paper presenting a comparison between different fluorescence analysis techniques was submitted to *Journal of Environmetrics*.

5.5. CHAPTER 5 FIGURES

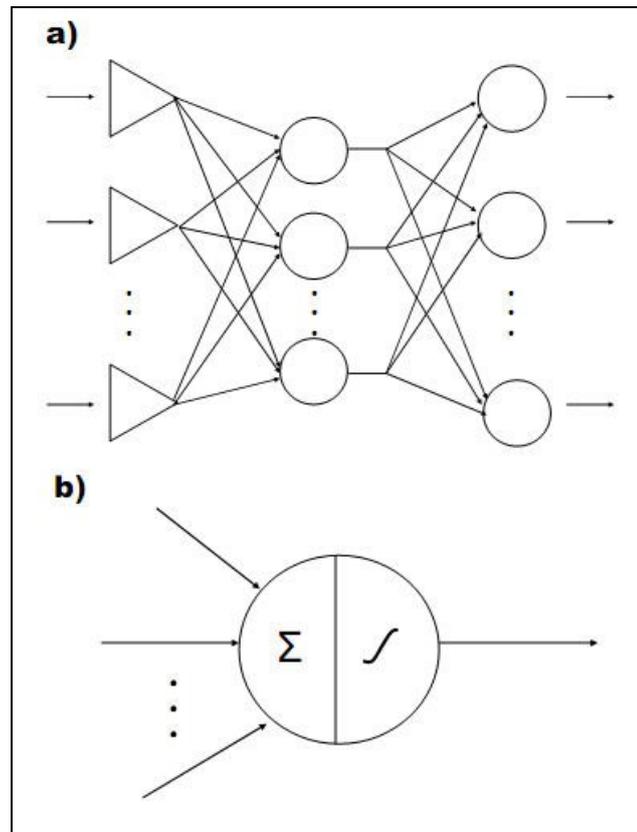


Figure 5.1: Three layer feed-forward neural network consisting of input, hidden and output layer (a). Triangles indicate passive neurons, circles – active neurons. Active neuron with connection weights (arrows) that sums up the inputs (Σ) and transforms the the sum with a sigmoid function (\mathcal{S}) (b)

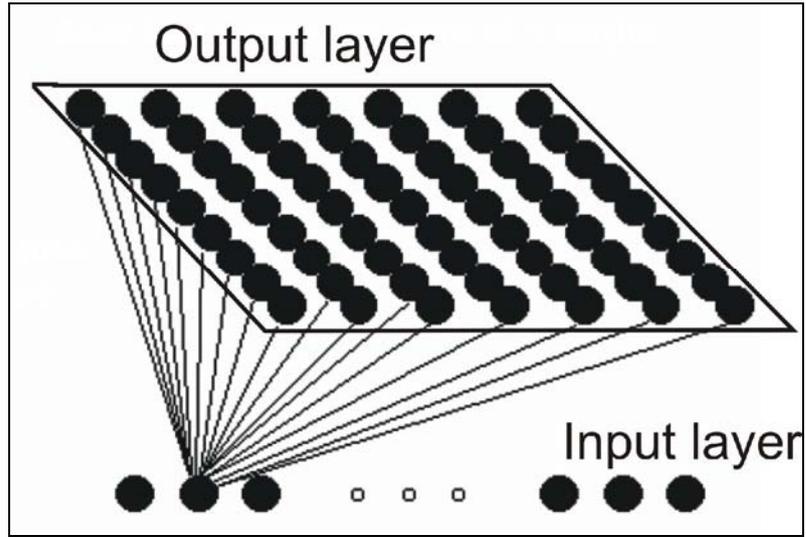


Figure 5.2: Self-organising map. Black dots indicate neurons

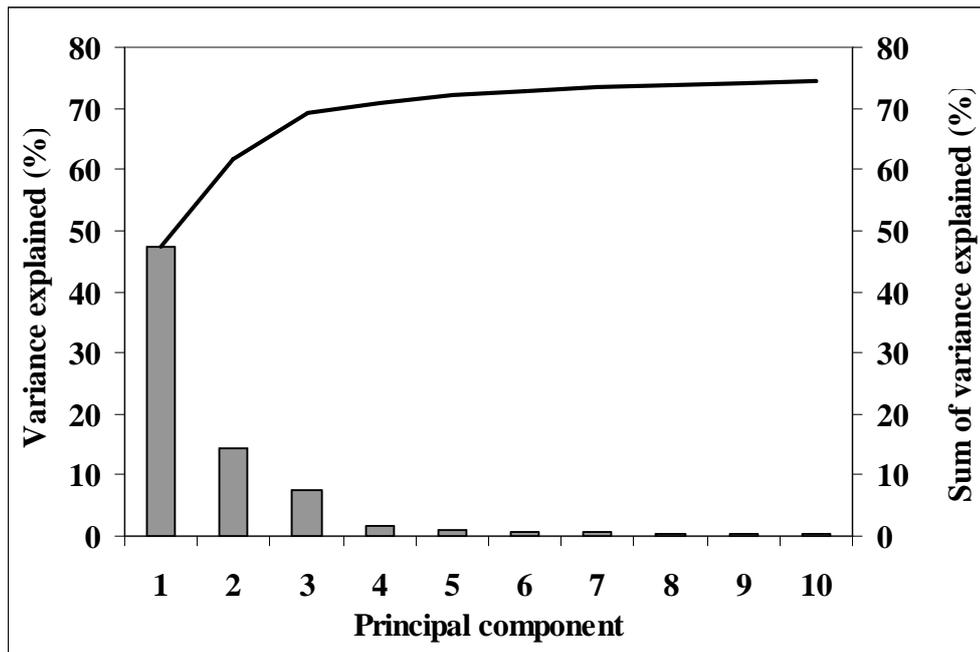
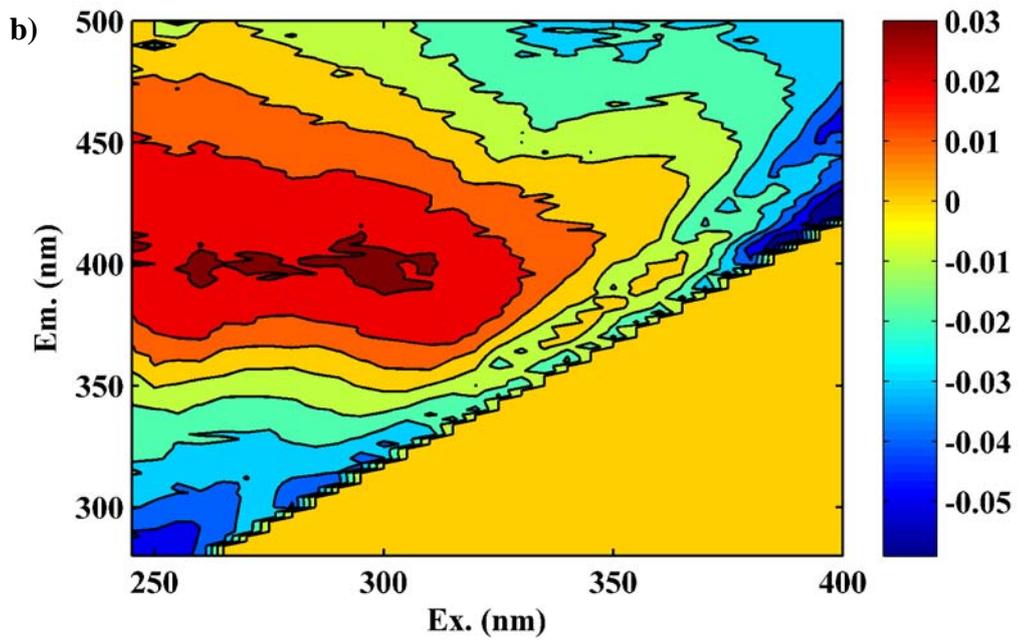
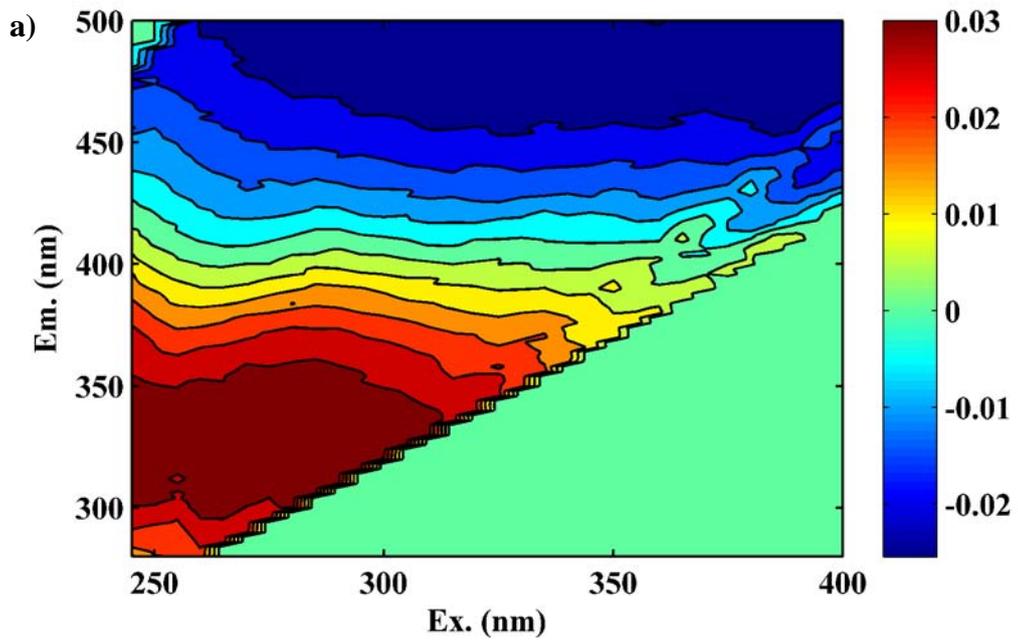


Figure 5.3: Total variance (%) and sum of variance (%) explained by the first ten PCA components



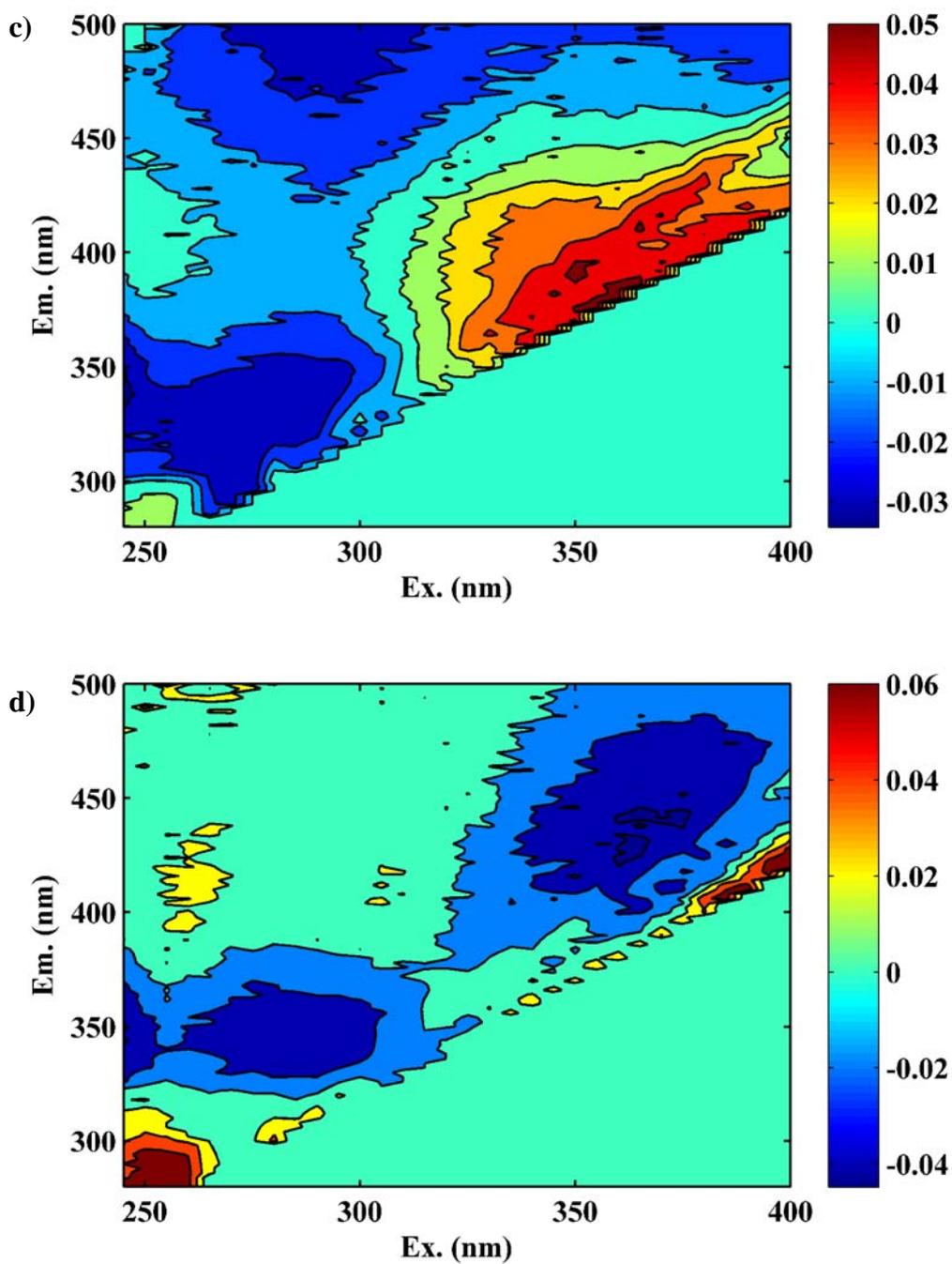


Figure 5.4: Loading plots for the first four principal components. PC1 (a), PC2 (b), PC3 (c), PC4 (d)

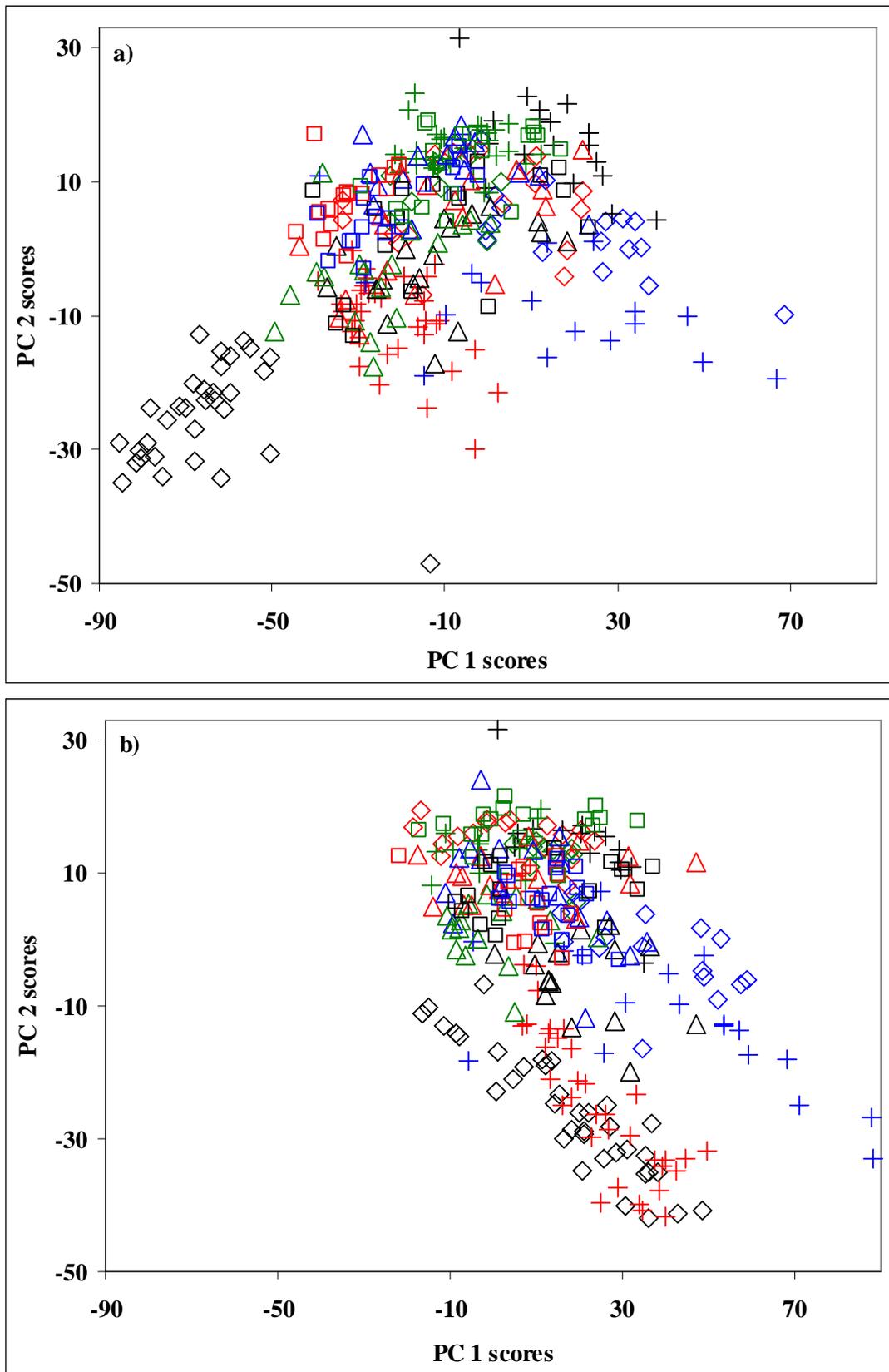


Figure 5.5: Score plots for the first two principal components from PCA analysis of raw (a) and clarified water (b) fluorescence excitation-emission spectra. For explanation of the labels see Figure 4.2

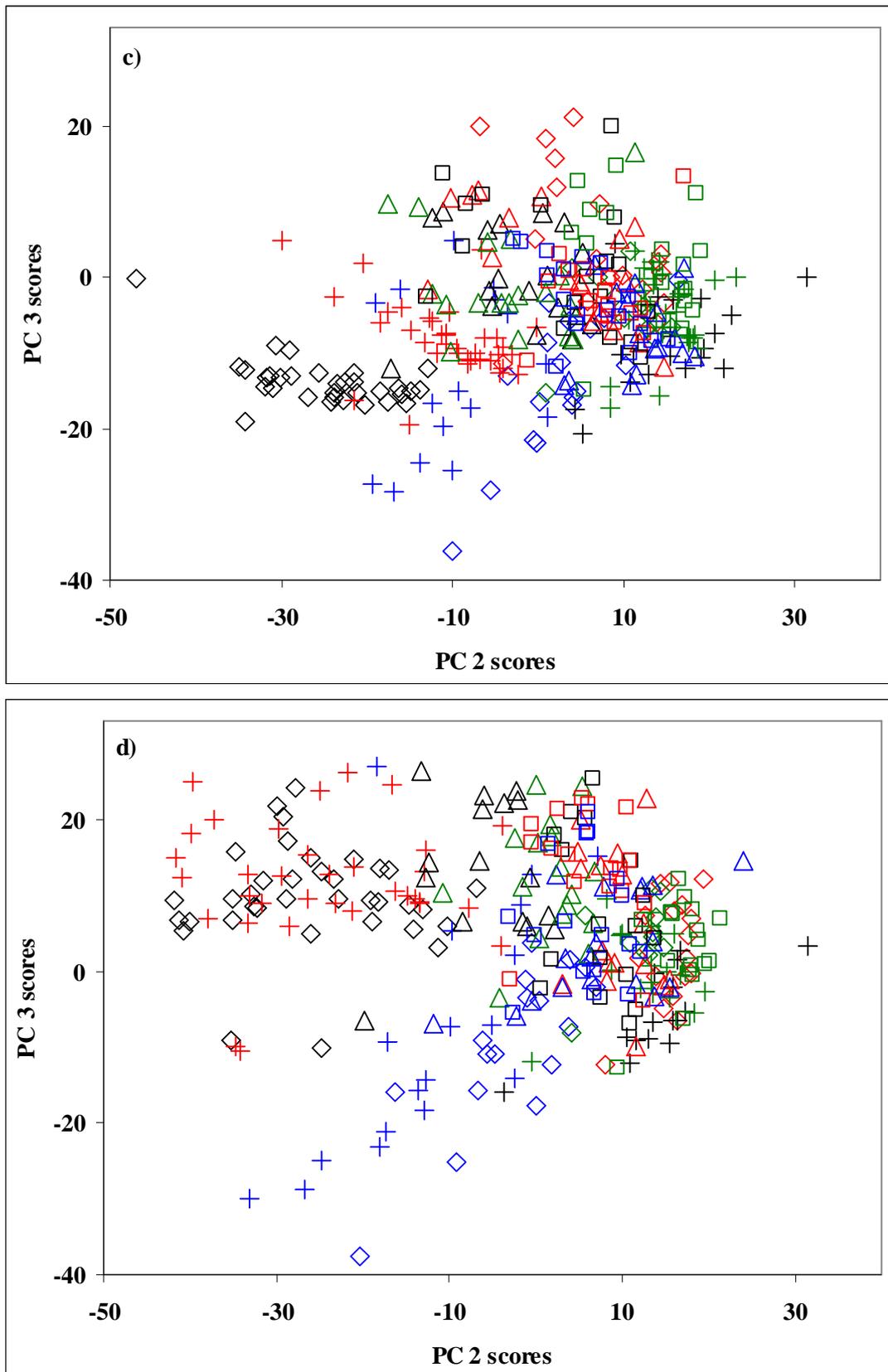
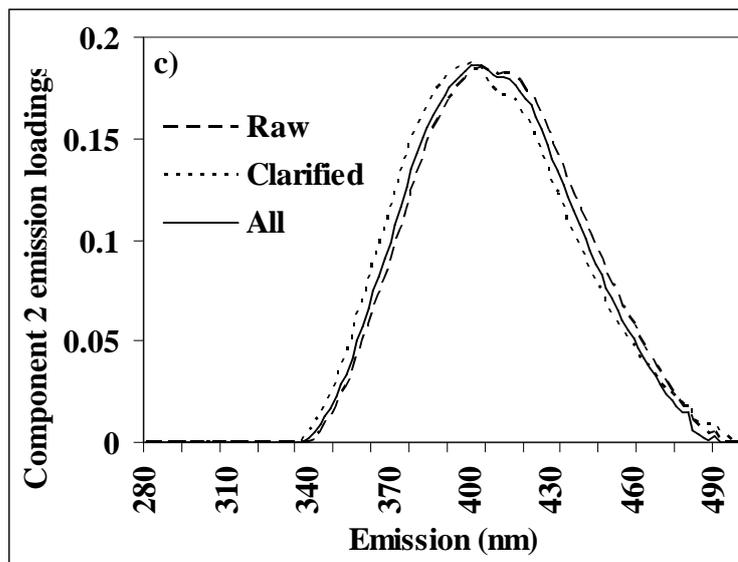
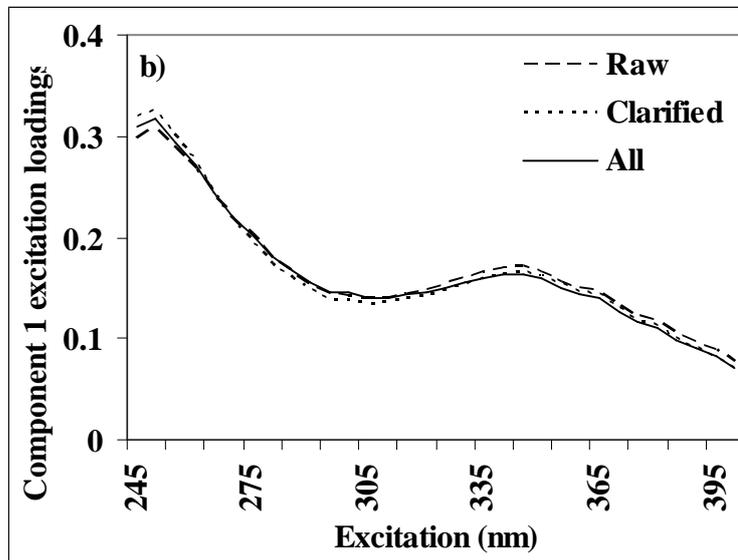
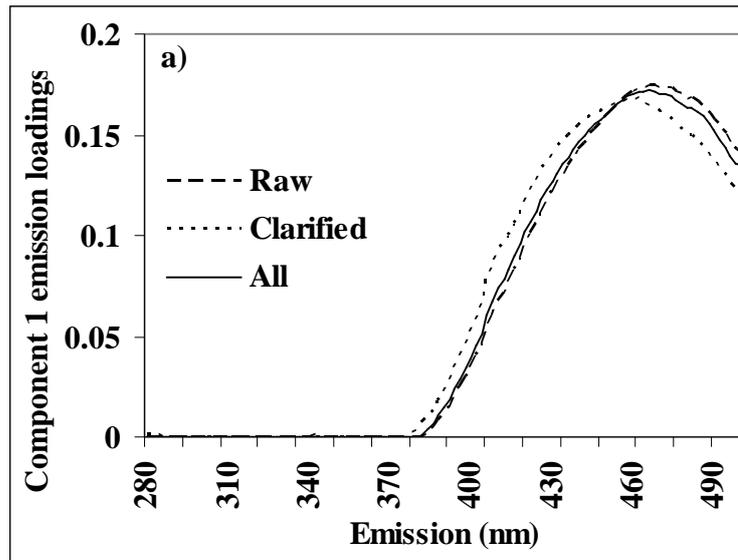


Figure 5.6: Score plots for the second and third principal components from PCA analysis of raw (a) and clarified water (b) fluorescence excitation-emission spectra. For explanation of the labels see Figure 4.2



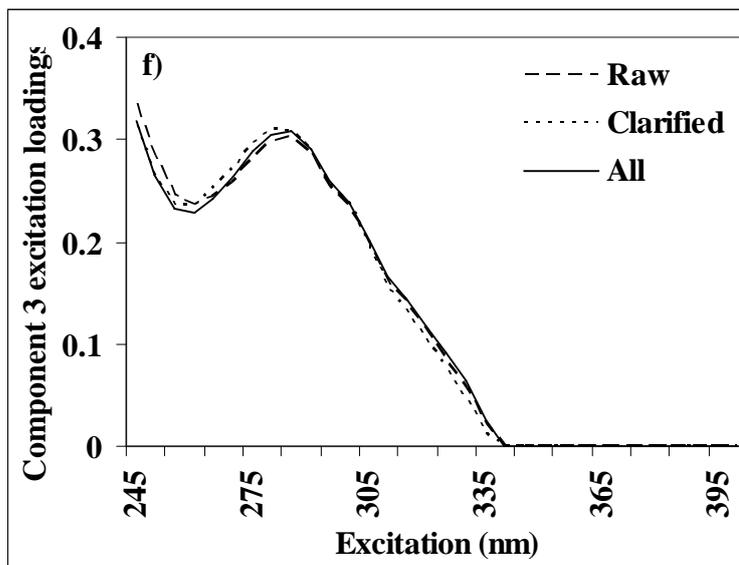
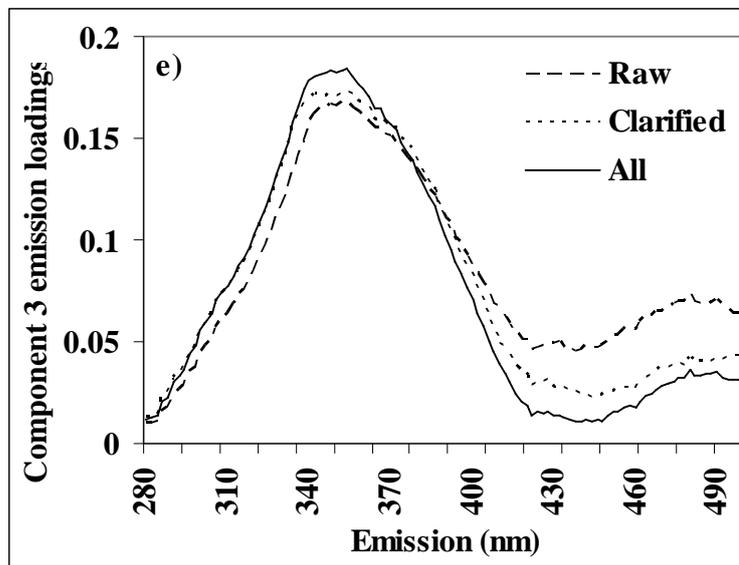
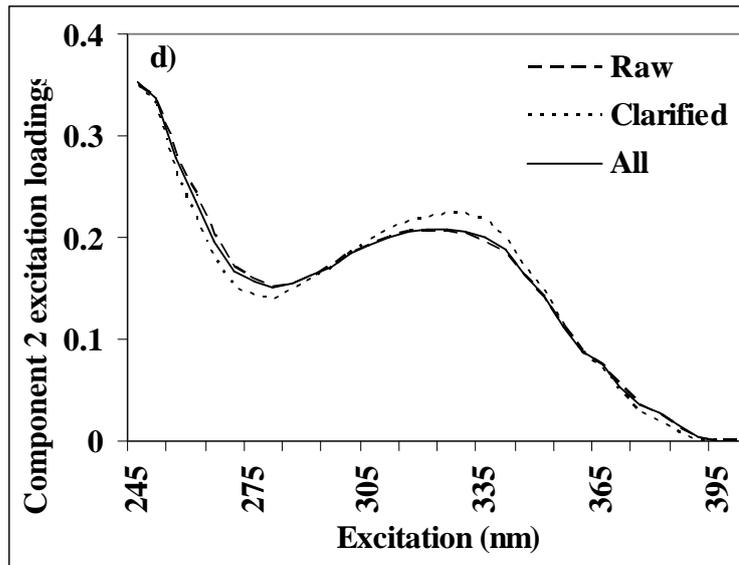
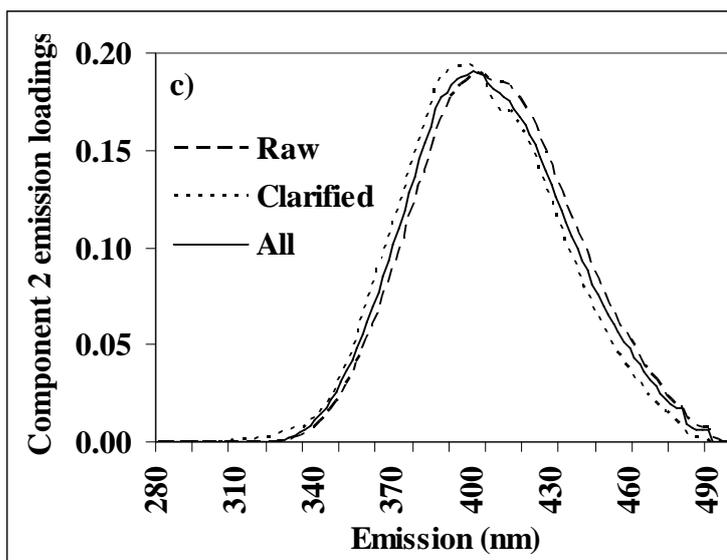
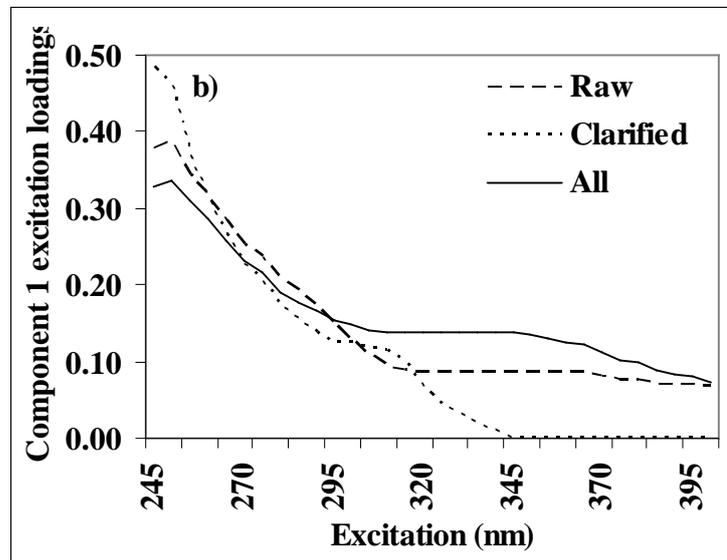
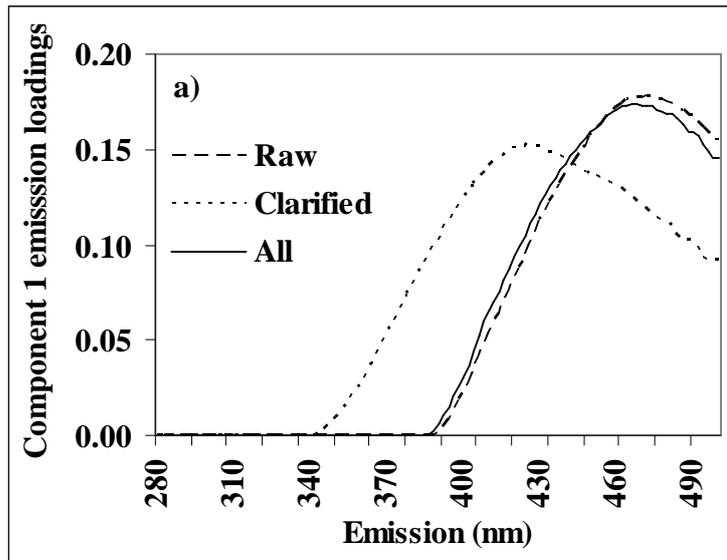
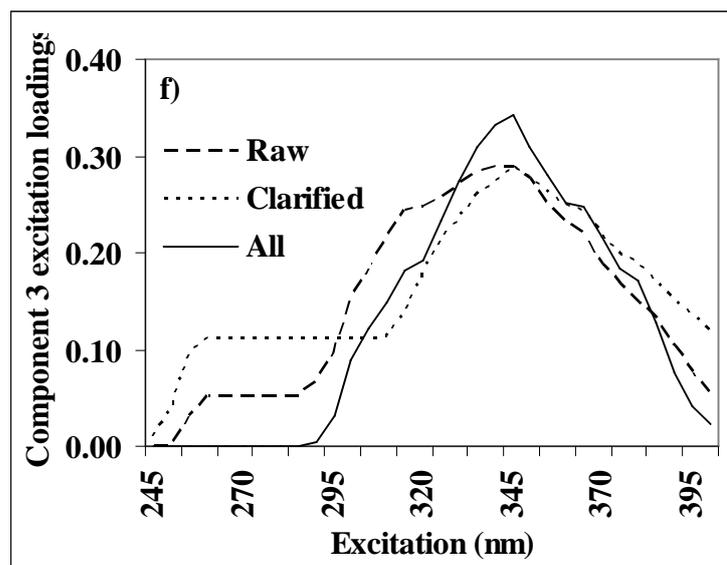
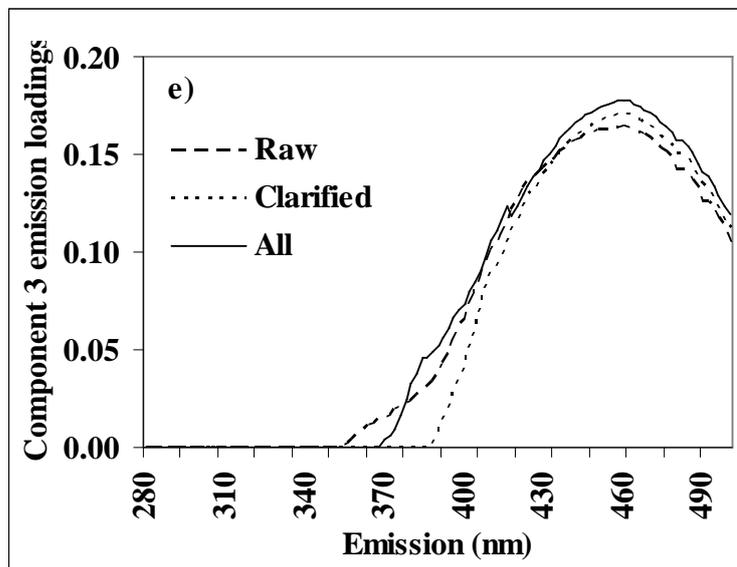
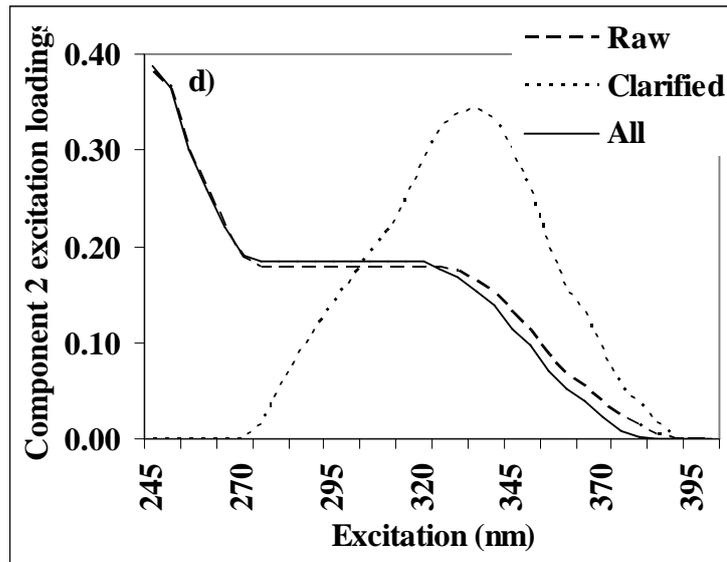


Figure 5.7: Three-component PARAFAC model excitation and emission loadings of raw, clarified and all data. Component 1: emission a), excitation b); Component 2: emission c), excitation d); Component 3: emission e), excitation f)





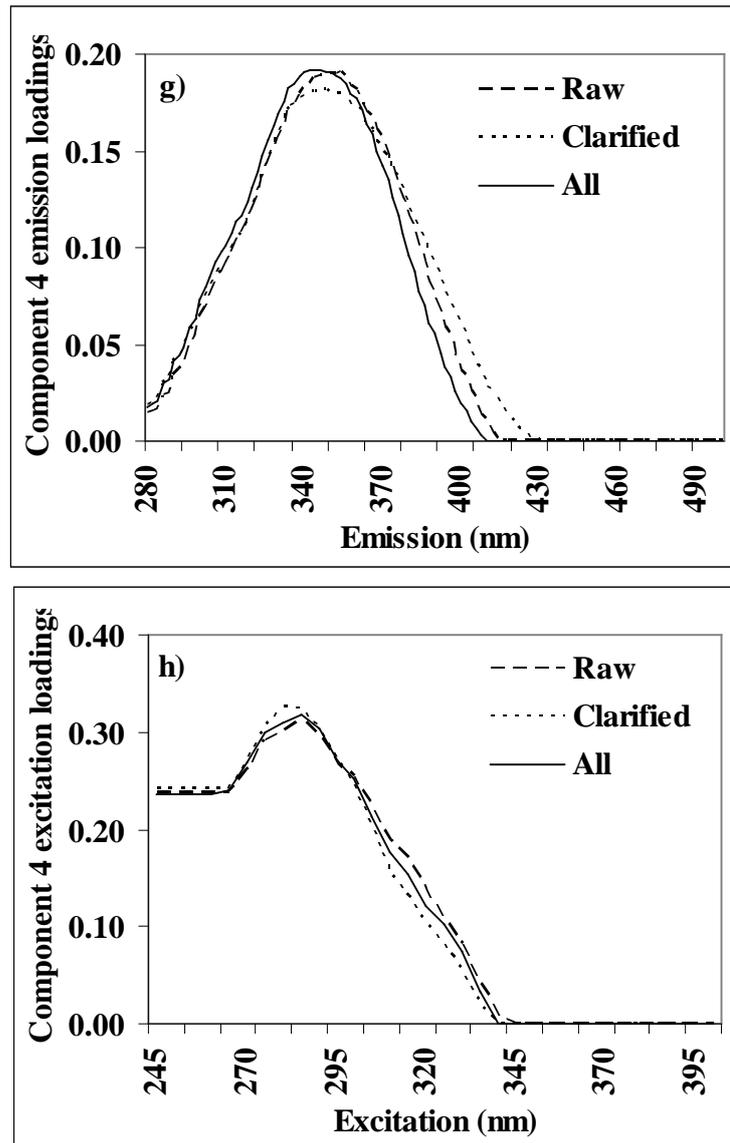
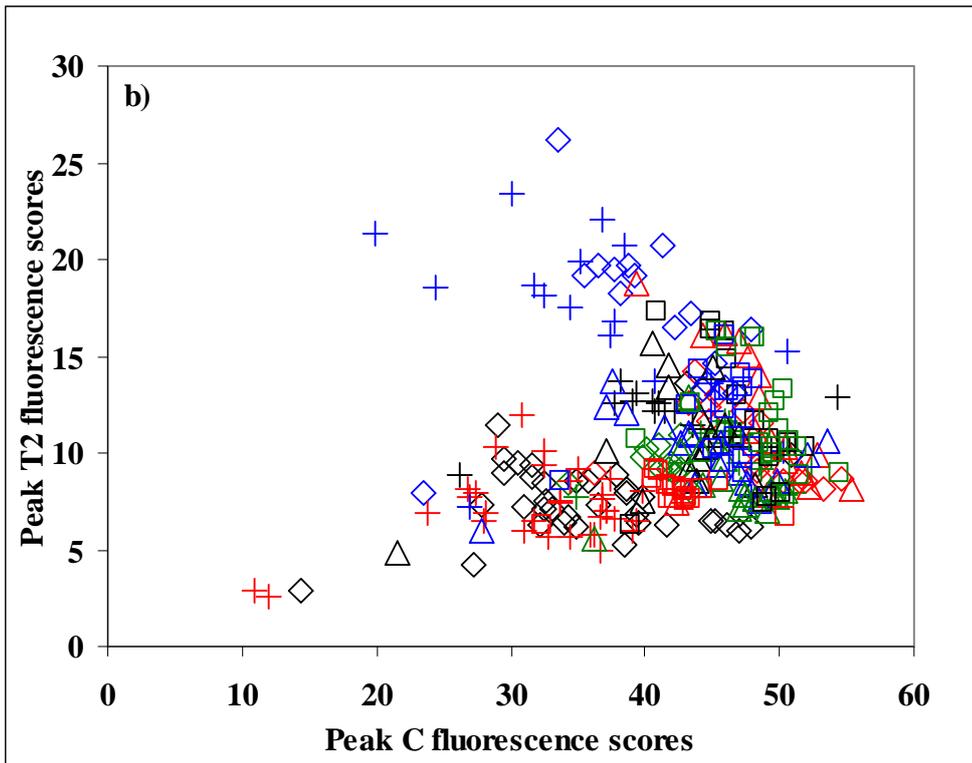
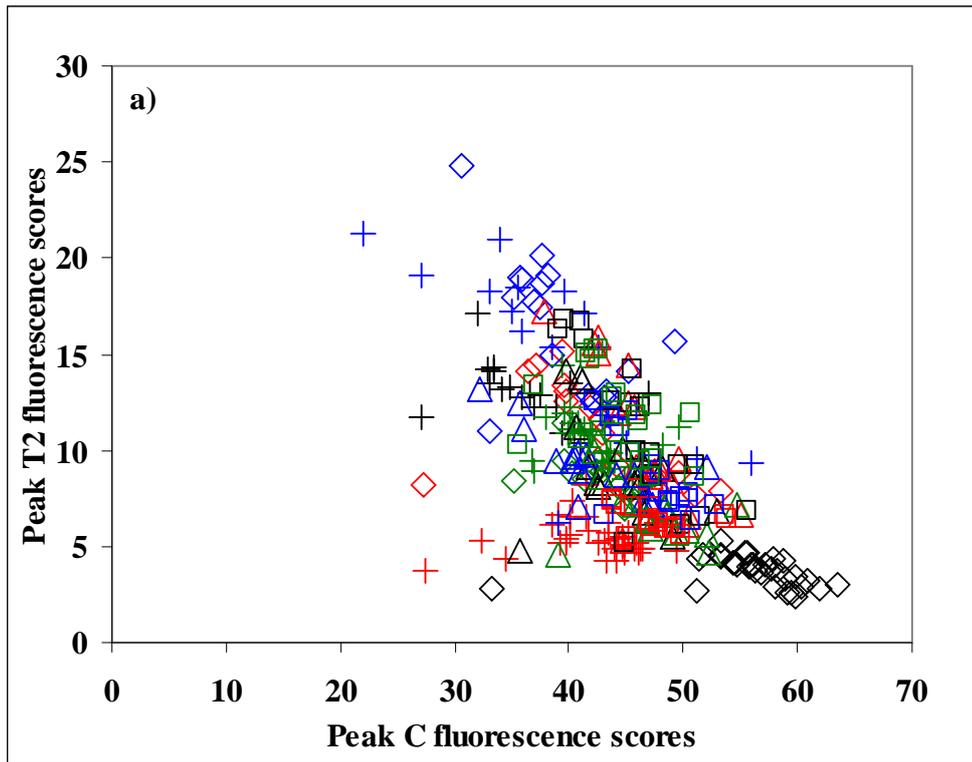


Figure 5.8: Four-component PARAFAC model excitation and emission loadings of raw, clarified and all data. Component 1: emission a), excitation b); Component 2: emission c), excitation d); Component 3: emission e), excitation f); Component 4: emission g), excitation h)



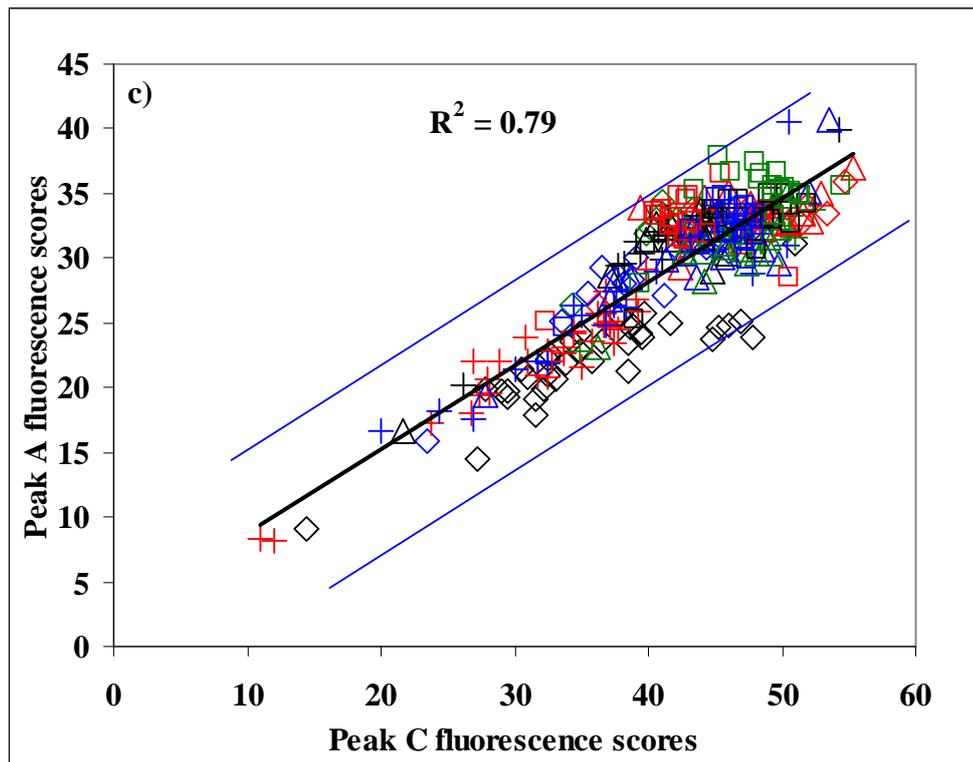
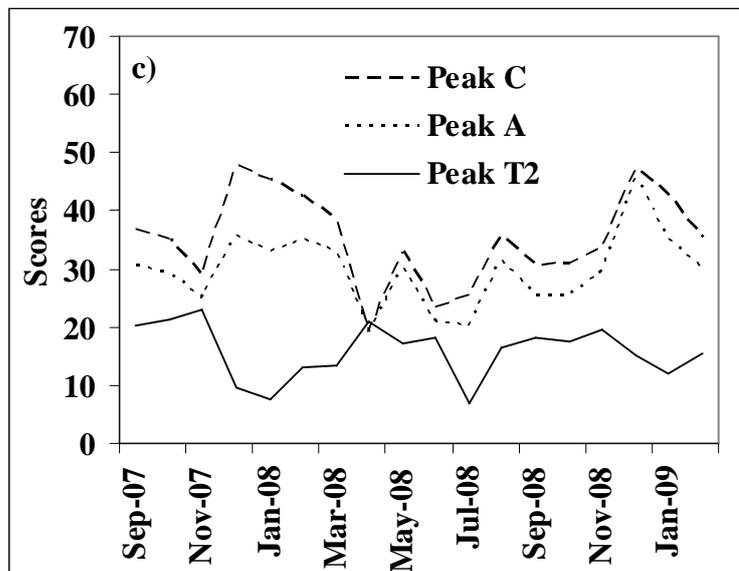
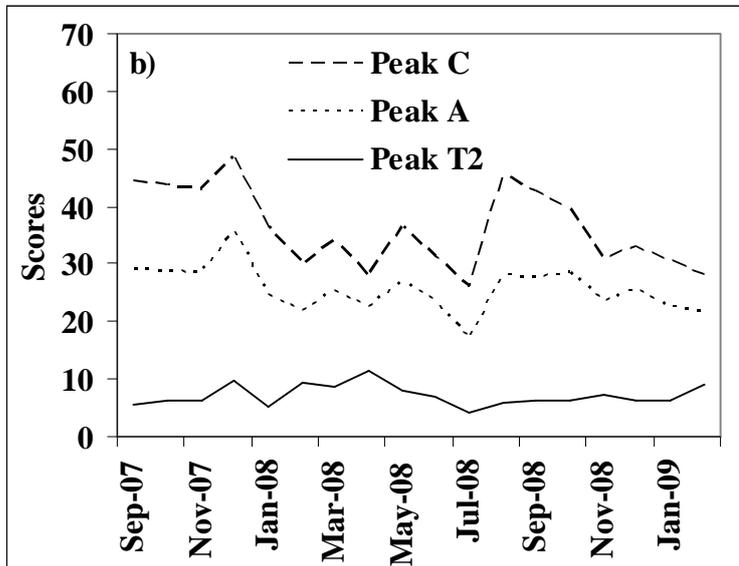
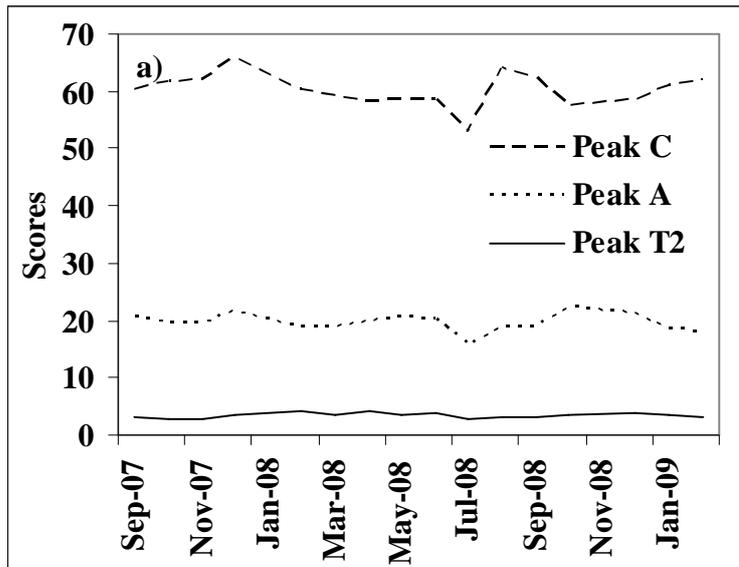


Figure 5.9: PARAFAC scores of raw and clarified water. a) peak C fluorescence (first component, X axis) and peak T2 fluorescence (third component, Y axis) – raw water, b)) peak C fluorescence (first component, X axis) and peak T2 fluorescence (third component, Y axis), c) peak C fluorescence (first component, X axis) and peak A fluorescence (second component, Y axis) – clarified water. For explanation of the labels see Figure 4.2



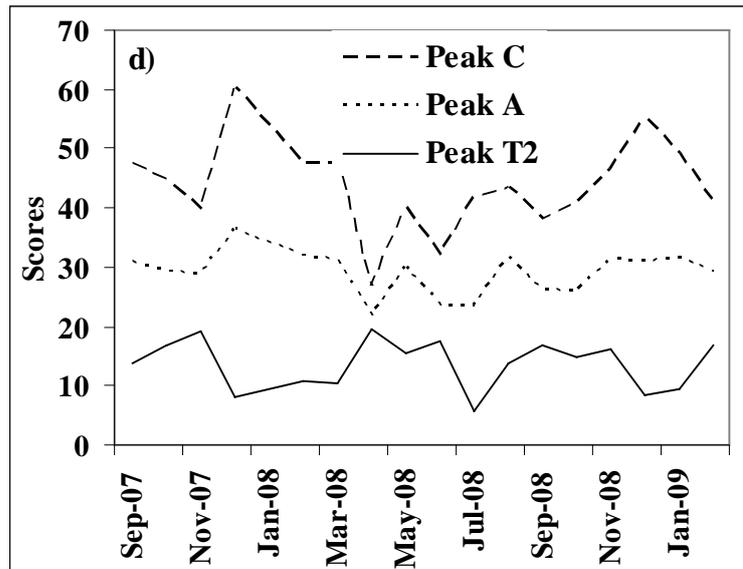


Figure 5.10: PARAFAC scores of raw and clarified water for two contrasting WTWs Bamford (a) raw, (b) clarified and Church Wilne (c) raw, (d) clarified

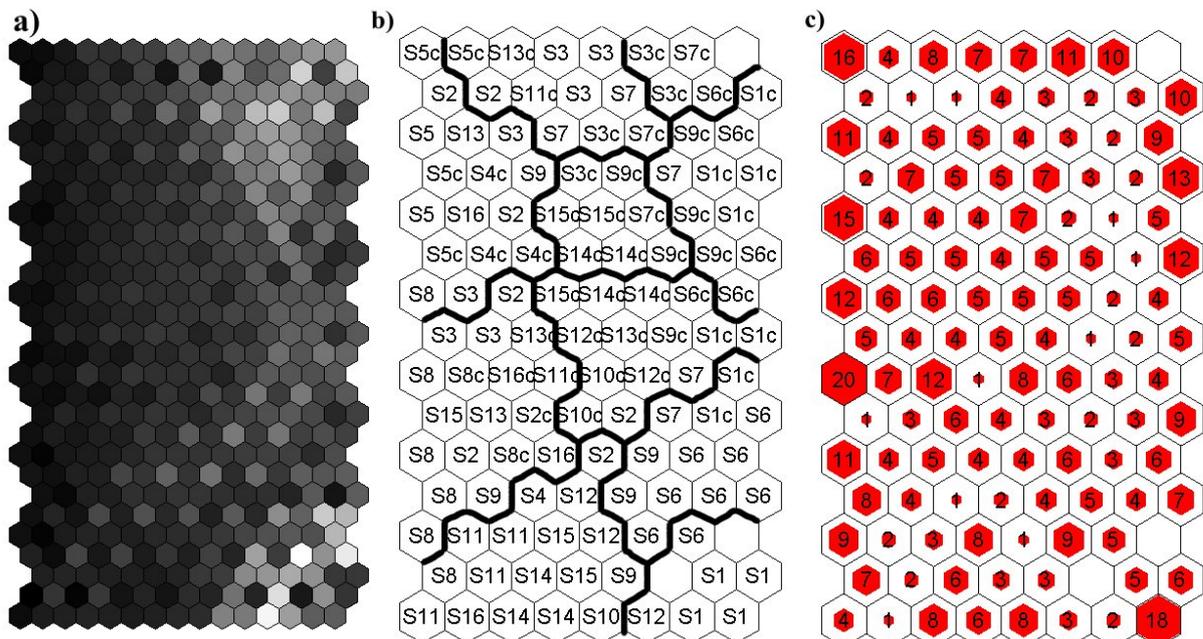


Figure 5.11: Visualisation of the SOM map for fluorescence data: U matrix (a), sample distribution and clusters (b), hit histogram (c). Notation used, e.g. S 1 and S 1c denotes raw and clarified water of site 1 – Bamford (sites numbers can be found in table 3.1)

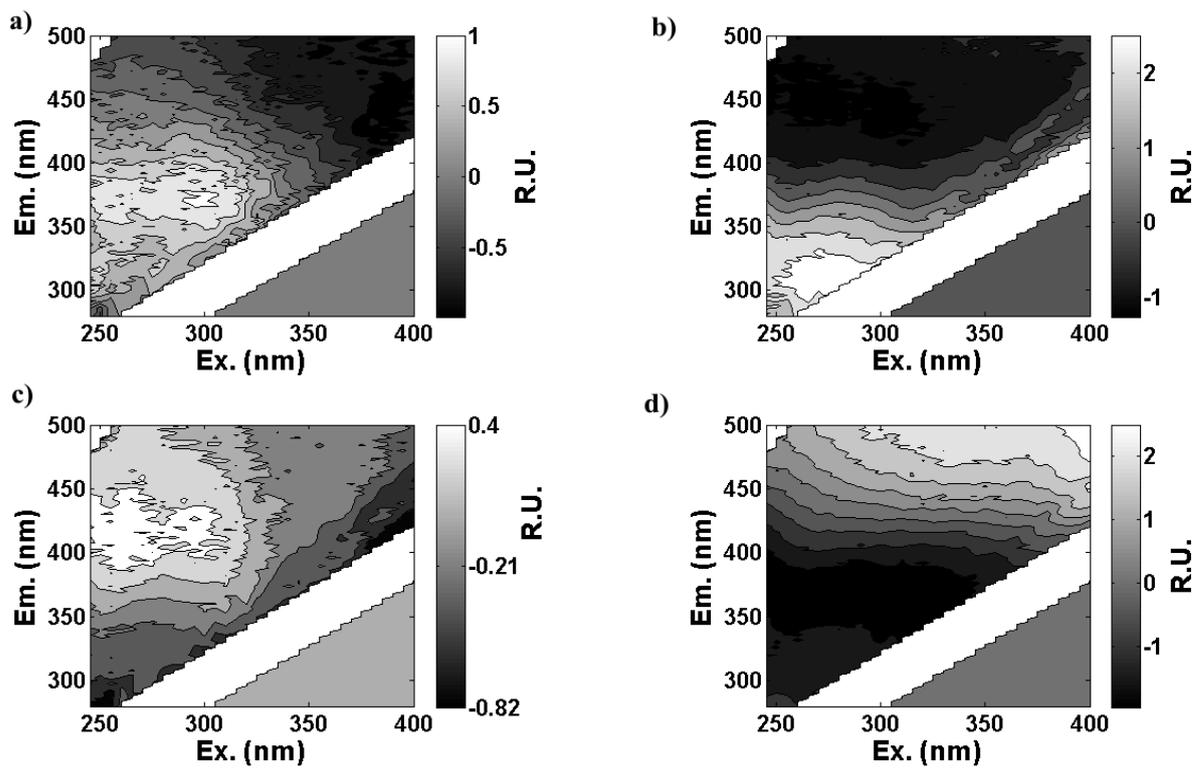


Figure 5.12: Reference vector plots for selected SOM neurons with the highest number of hits. Neuron 1 (a), neuron 109 (b), neuron 9 (c) and neuron 120 (d)

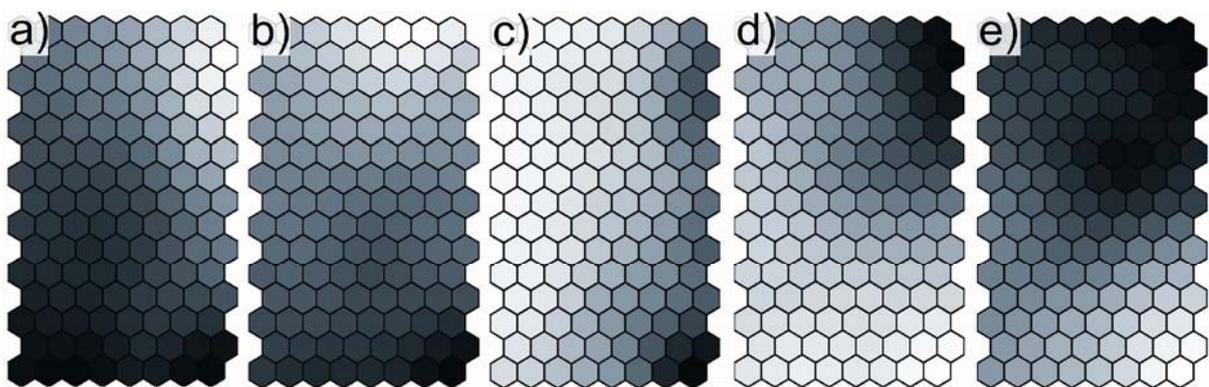
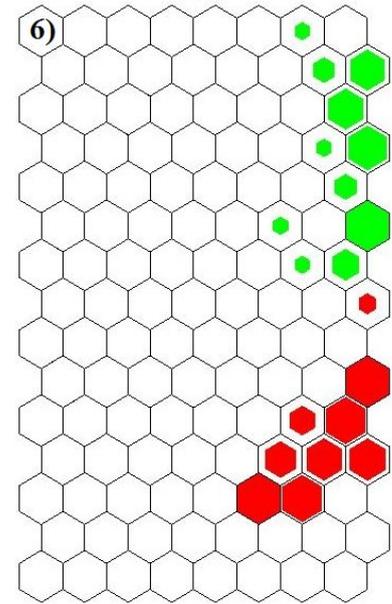
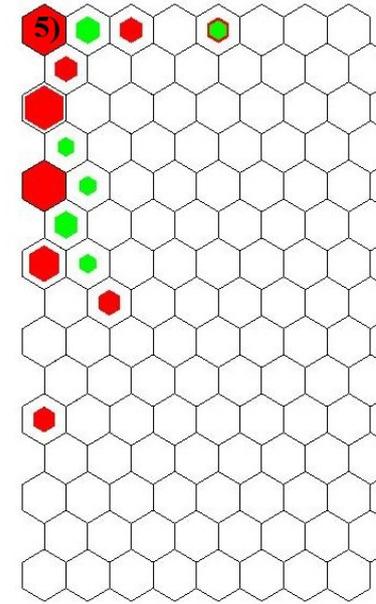
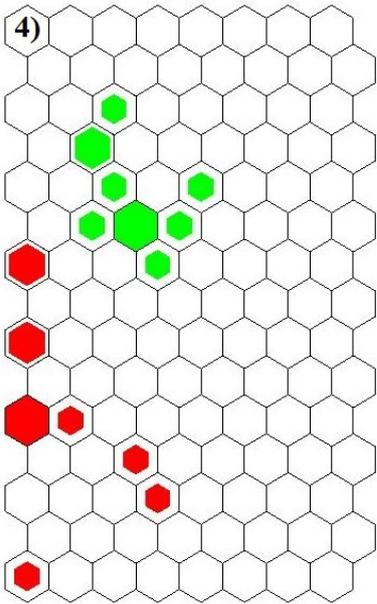
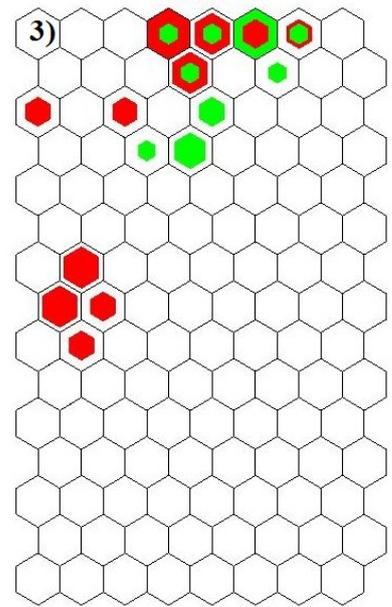
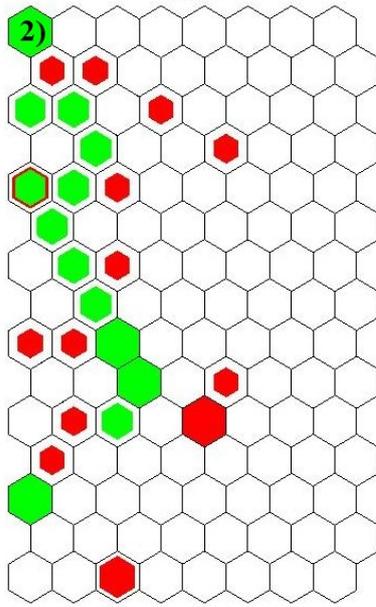
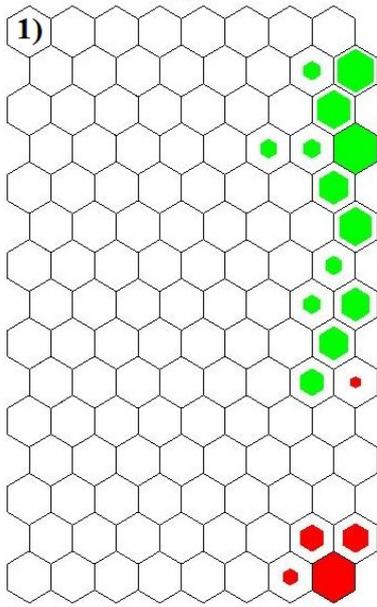
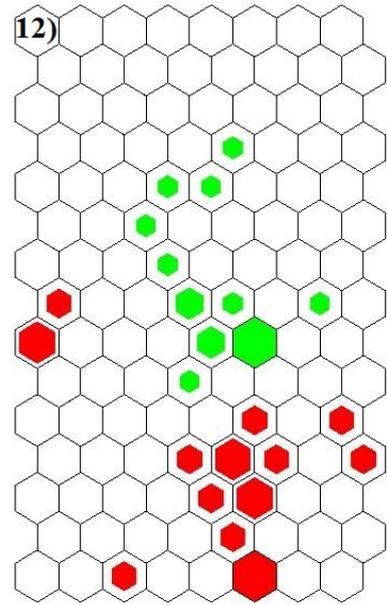
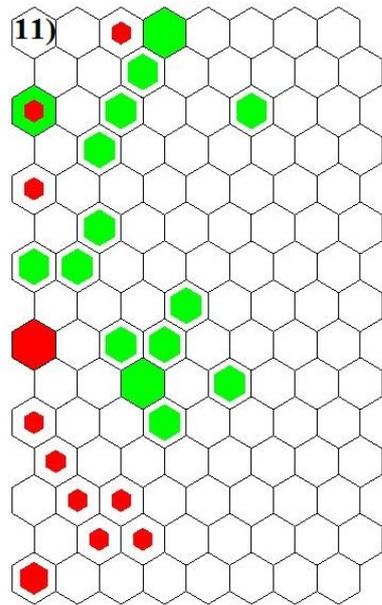
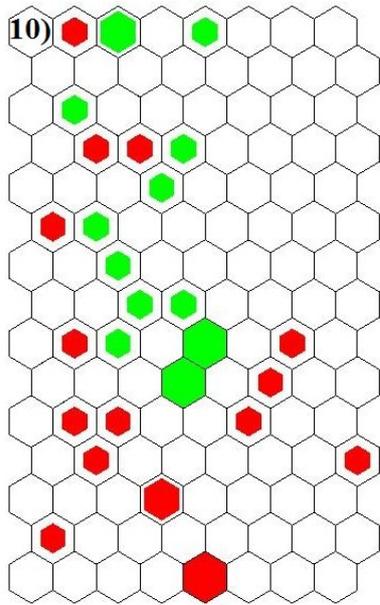
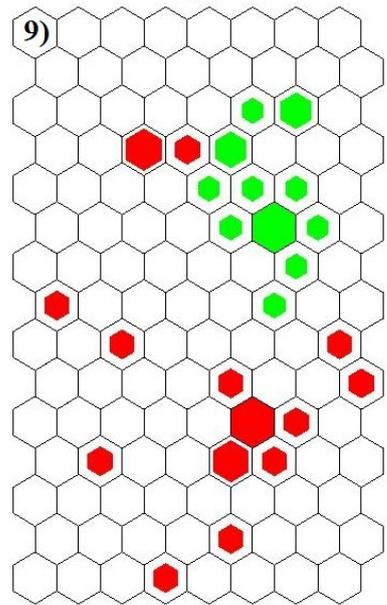
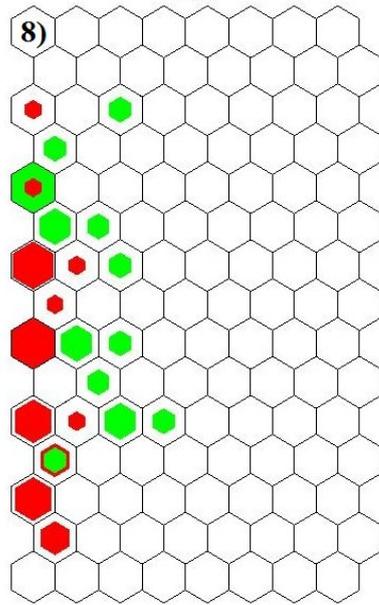
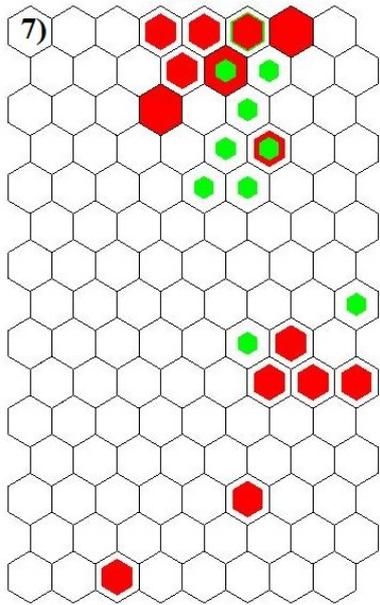


Figure 5.13: Component planes for the excitation-emission wavelengths with fixed excitation at 280 nm. Emission wavelength of 300 (a), 350 (b), 400 (c), 450 (d), and 500 (e) nm





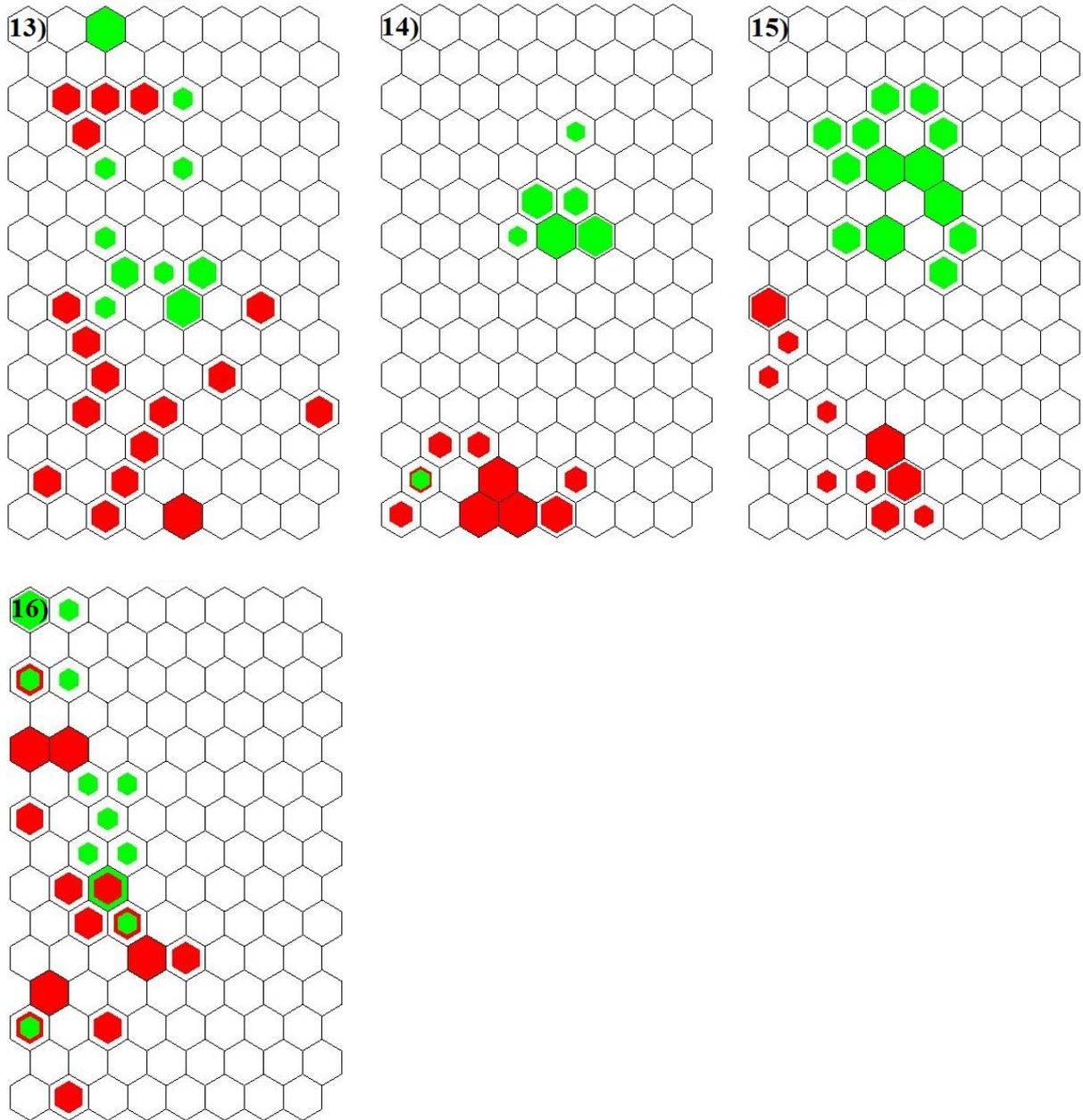


Figure 5.14: Hit histograms for raw (red) and clarified (green) water. Sites: 1 – Bamford, 2 – Campion Hills, 3 – Church Wilne, 4 – Cropston, 5 – Draycote, 6 – Frankley, 7 – Little Eaton, 8 – Melbourne, 9 – Mitcheldean, 10 – Mythe, 11 – Ogston, 12 – Shelton, 13 – Strensham, 14 – Tittesworth, 15 – Trimpley, 16 – Whitacre

5.6. CHAPTER 5 TABLES

Table 5.1: Examples of application of multivariate techniques to EEMs analysis (CAL – calibration, CLAS - classification)

Reference	Topic	Approach	Method	Aim of analysis
McAvoy et al, 1992	DOM	ANNs / Multi-way analysis	ANNs (BP) vs. PLS	CAL
Henrion et al, 1997	algae	Multi-way analysis	3-way PCA	CLAS
Marhaba et al, 2000	DOM	Multi-way analysis	PCA	CLAS
Marhaba et al, 2000	DOM	EEMs exploratory analysis	spectra analysis	CAL
Saurina et al, 2000	water quality	Chemometrics	MCR-ALS	CLAS
Persson and Wedborg, 2001	DOM	Multi-way analysis	PCA vs. PLS	CLAS
Wolf et al, 2001	fermentation monitoring	ANNs	ANNs (BP)	CLAS / CAL
Bengraïne and Marhaba, 2003	DOM	Multi-way analysis	PCA vs. PLS	CAL
Chen et al, 2003	DOM	EEMs exploratory analysis	Regional integration	CLAS
Guimet et al, 2003	olive oils	Multi-way analysis	3-way PCA vs. PARAFAC	CLAS
Marhaba et al, 2003	DOM	Multi-way analysis	PLS	CAL
Scott et al, 2003	olive oils	ANNs	ANNs (BP, SFAM, RBF) vs. PCA	CLAS
Stedmon et al, 2003	DOM	Multi-way analysis	PARAFAC	CLAS / CAL
Boehme et al, 2004	DOM	Multi-way analysis	PCA	CLAS
Antunes and Esteves da Silva, 2005	humic-like fluorescence	Chemometrics	MCR-ALS	CAL
Guimet et al, 2005	olive oils	Multi-way analysis	3-way PCA vs. PARAFAC vs. N-PLS	CLAS
Hall et al, 2005	ballast waters	Multi-way analysis	PARAFAC vs. N-PLS	CLAS
Lee et al, 2005	fermentation monitoring	ANNs / Multi-way analysis	ANNs (SOM, BP) vs. PCA, PLS	CLAS / CAL
Rhee et al, 2005	fermentation monitoring	ANNs	ANNs (SOM)	CLAS
Divya and Mishra, 2007	fuels	Multi-way analysis	PARAFAC vs. N-PLS vs. 3-way PCA	CLAS
Spencer et al, 2007	DOM	Multi-way analysis	PCA vs. DA	CLAS
Ohno et al, 2008	DOM	Multi-way analysis	PARAFAC	CLAS

Table 5.2: Details of the fluorescence dataset used in the calibration analysis

Model	Data	Calibration matrix size	Validation matrix size
EEM	whole EEMs of raw and clarified water	220 x 5030	70 x 5030
PEAK	fulvic-like fluorescence intensity and emission wavelength, tryptophan-like fluorescence intensity of raw and clarified water	220 x 6	70 x 6
PARAFAC	scores of three PARAFAC components of raw and clarified water	220 x 6	70 x 6
PCA	scores of first three PCA components of raw and clarified water	220 x 6	70 x 6
SOM	normalized weights of three SOM clusters of raw and clarified water	220 x 6	70 x 6

Table 5.3: Variance explained (VARE (%)), residual sum of squares (SSQ) and CORCONDIA (COR (%)) result of three PARAFAC models for first seven PARAFAC components (C)

C	RAW			CLARIFIED			ALL		
	VARE	SSQ	COR	VARE	SSQ	COR	VARE	SSQ	COR
1	97.0	34005.5	100.0	97.8	29913.9	100.0	97.7	70243.9	100.0
2	99.2	13982.5	95.4	98.7	18230.5	98.9	98.9	33747.4	96.9
3	99.4	9546.2	84.6	99.1	13175.8	92.1	99.2	24001.5	90.1
4	99.5	7645.6	4.5	99.2	11518.2	-6.7	99.3	20623.0	14.1
5	99.6	6937.0	5.0	99.2	11112.7	-5.4	99.4	19533.2	3.3
6	99.6	6699.3	7.4	99.2	10950.7	8.7	99.4	18085.9	-1.9
7	99.6	6320.9	1.0	99.3	10489.7	1.2	99.4	17267.9	1.1

Table 5.4: Statistical details of PCA and PARAFAC models. VARE – variance explained (%), SSQ – residual sum of squares, COR – CORCONDIA (%)

Component	PCA		SSQ	PARAFAC	
	VARE	Sum of VARE		Sum of VARE	COR
1	47.3	47.3	70243.9	97.7	100.0
2	14.4	61.7	33747.4	98.9	96.9
3	7.6	69.3	24001.5	99.2	90.1
4	1.7	71.0	20623.0	99.3	14.1
5	1.0	72.0	19533.2	99.4	3.3
6	0.8	72.8	18085.9	99.4	-1.9
7	0.6	73.4	17267.9	99.4	1.1

Table 5.5: Characteristics of three PCA and PARAFAC components (C) identified for drinking water fluorescence dataset. Excitation and emission wavelengths maxima and identified fluorophores. Wavelength in brackets denotes secondary maximum

C	PCA			C	PARAFAC		
	Exc. max (nm)	Em. max (nm)	Fluoroph.		Exc. max (nm)	Em. max (nm)	Fluoroph.
1	250 - 300	300 - 350	peak T	1	(<250) 345	460	peak C
2	260-300	400	peak A	2	< 245	400	peak A
3	350	400	peak C	3	285	350	peak T

Table 5.6: Prediction accuracy and prediction error for selected decomposition and calibration methods: MLR – multiple linear regression, SR – stepwise regression, PLS – partial least squares analysis, BPNN – artificial neural network with back-propagation

Decomposition model	Prediction accuracy (R^2)				Prediction error (RMSEP)			
	MLR	SR	PLS	BPNN	MLR	SR	PLS	BPNN
EEM	0.75	0.94	0.94	0.92	1.90	0.52	0.52	0.56
PEAK	0.87	0.82	0.92	0.95	0.73	0.78	0.78	0.45
PARAFAC	0.63	0.93	0.93	0.94	1.29	0.55	0.54	0.51
PCA	0.78	0.85	0.85	0.90	0.94	0.77	0.77	0.64
SOM	0.74	0.93	0.93	0.93	1.21	0.56	0.64	0.60

6. FLUORESCENCE IN CHARACTERISATION OF WATER TREATMENT OPTIMISATION PROCESSES

6.1. OPTIMISATION OF DRINKING WATER TREATMENT

Water treatment works optimisation involves changes in the treatment processes and treatment infrastructure that increase the water treatment efficiency, secure the supplies of potable water, reduce the production of undesirable treatment by-products (e.g. sludge, DBPs), prolong the life of treatment infrastructure, and lower the operational costs. Optimisation of organic matter removal is of key importance in drinking water optimisation, as OM interferes with most of the treatment processes. In particular, OM reduces the operational life of treatment infrastructure (e.g. membranes), increases the operational costs due to increased coagulant and disinfectant demands, and directly (unpleasant odour and turbidity) or indirectly (DBPs formation) impairs the drinking water quality. Organic matter removal can be facilitated by enhancement of the existing OM removal process at the particular WTW or by introduction of another, advanced technologies of OM removal, e.g. advanced oxidation processes, ion exchange, membranes. The facilitation of novel, advanced treatment processes can be expensive (ion exchange, membranes) and requires time-consuming laboratory and full scale experiments to evaluate their efficacy for different OM character and treatment conditions. Thus, in conventional drinking water treatment OM removal optimisation can be achieved by optimisation of the existing treatment processes, e.g. coagulation.

The efficiency of OM removal by coagulation is dependent on both OM properties and treatment conditions (Randtke, 1988; Kim and Yu, 2005; van Leeuwen et al., 2005; Fabris et al., 2008). The hydrophobic, high molecular weight fraction was found to be preferentially removed, whereas the low, molecular weight hydrophilic OM was reported to be recalcitrant

to removal by coagulation. The relative contribution of the more recalcitrant OM fraction is specific to the particular WTW, and demonstrates seasonal changes induced by the natural and anthropogenic impacts. Consequently, the optimisation of the coagulation process can be achieved by appropriate adjustment of the coagulation parameters (coagulant type and dosage, pH, prior to coagulation pre-treatment processes, i.e. ozonation). Coagulation pH was found to be the most significant treatment parameter affecting the efficiency of OM removal by coagulation (Edzwald, and Tobiason, 1999; O'Melia et al., 1999; Yu et al., 2007). Coagulation pH controls the amount, speciation, charge of organic matter, and the availability of the coagulant hydrolysis products for neutralisation and adsorption of organic matter. For each coagulant type, an optimum pH conditions exist that maximise the interactions between anionic organic matter and cationic coagulant hydrolysis products. Thus, at the optimum pH, OM removal by coagulation is more effective compared with conditions higher or lower than the optimum pH. The optimum coagulation pH for organic matter removal was reported to be in range 5-6 for alum and 4-5 for ferric salts (Randtke, 1988; US EPA, 1999). However, in practice the coagulation pH is a trade-off between high costs of lowering the pH and the achieved improvement in OM removal. Thus, at many WTWs the operational coagulation pH is higher than the optimum pH, and the OM removal is poorer compared to the potential optimum pH conditions.

Edzwald and Tobiason (1999) defined the term of optimum coagulation as the overall coagulation conditions of dosage and pH that achieve maximum particle and turbidity removals, maximum TOC and DBPs precursor removals, and minimum residual coagulant, sludge production and operating costs. Unlike optimum coagulation, enhanced coagulation has the single objective of maximizing TOC removal.

Enhanced coagulation should be implemented at WTWs with poor overall TOC removals. In the UK efficiency of OM removal by coagulation is not yet regulated. However, the required TOC removal by enhanced coagulation is specified and regulated in the US by the US Environmental Protection Agency (US EPA, 1999). US WTWs are obliged to adopt a two-step assessment of TOC removal by coagulation. Step 1 of US EPA guidelines involves comparison of actual TOC removal with the required TOC removal specified for particular raw water TOC concentrations and alkalinity levels (Table 6.1). The required TOC removal ranges between 15 and 50 %. Lower TOC removal is accepted for raw waters of low TOC concentrations and/or for high alkalinity. Likewise, higher TOC removal is required for low alkaline, high TOC raw waters. This suggests that waters with low TOC concentrations are relatively more difficult to treat by coagulation. In higher alkalinity waters, pH lowering to a level at which TOC removal is optimal is more difficult and can not be achieved simply through the addition of coagulant alone (US EPA, 1999). WTWs treating waters that contain low OM concentrations and demonstrate scant propensity for DBPs formation (raw and clarified water TOC < 2 mg/l or SUVA < 2.0 mg/l.m or TTHM <40 µg/l and HAA5 <30 µg/l with only chlorine for disinfection) are exempted from the above regulations.

Compliance with the TOC removal requirement is based on a computed quarterly running annual average. If the actual average TOC removal is lower than the required TOC removal, WTWs are obliged to adopt Step 2 of the guidelines. Step 2 involves a series of jar or pilot-scale tests to set an alternative TOC removal requirement for a particular WTW.

6.2. CAMPION HILLS TOC REMOVAL OPTIMISATION INVESTIGATION

The relationship between coagulation pH and organic matter removal efficiency is well established in the literature (Randtke, 1988; Gregor et al., 1997; Volk et al., 2000). Using data

from the 16 WTWs previously described in section 3.1, a coagulation pH – TOC removal relationship was derived with a correlation coefficient value of $R^2 = 0.78$ (Figure 6.1). The highest removal was observed at sites with low coagulation pH, e.g. Bamford. However, the low coagulation pH at Bamford WTW is a result of natural low raw water pH related to acidic upland sources. The lowland treatment sites with predominantly hydrophilic raw water character and pH above 7.0 demonstrated a much higher coagulation pH (Whitacre, Strensham, Church Wilne and Campion Hills). The highest coagulation pH value of 8.5 was typical of Ogston WTW.

To determine the impact of pH adjustment on organic matter removal efficiency, a full scale coagulation enhancement experiment was carried out at Campion Hills WTW in the period November 08 - February 09. For four successive weeks the coagulation pH was decreased from the current operational level of 7.5 to 5.5, and in February the coagulation pH was further decreased from 6.5 to 5.0. For each week of pH change, three samples were collected (Monday-Wednesday) of raw, clarified and final water. For all samples, fluorescence and absorbance measurements were performed along with determination of TOC. The data was used to assess the impact of pH lowering on organic matter removal efficiency (Table 6.2).

In table 6.3 the fluorescence properties characterising raw and clarified water organic matter at Campion Hills WTW were presented and compared with the average values for all 16 sites. High and seasonally variable peak C intensity and TOC indicate the presence of substantial amounts of organic matter in Campion Hills raw water. The average raw water peak C emission wavelength for Campion Hills WTW was slightly lower than the average for all sites, and indicates the predominance of the hydrophilic fraction of organic matter. High tryptophan-like fluorescence intensity suggests the presence of a microbial organic matter fraction including algae-derived material.

As a result of the organic matter composition and non-optimised coagulation process, organic matter removal at Campion Hills was poor, as calculated both on the basis of fluorescence analysis (reduction in peak C intensity between raw and clarified water, 21.5 %) and from the reduction in measured TOC concentrations (20.0 %). The OM removal at Campion Hills WTW was distinctively lower compared with the average value for all 16 Severn Trent Water WTWs. The coagulation pH was distinctively higher for Campion Hills compared with all sites average, resulting from the initial high raw water pH (the average for the coagulation experiment was 8.1).

According to the US EPA regulations (Table 6.1), Campion Hills WTW (average raw water TOC value of 6.1 mg/l and low alkalinity < 60 mg/l as CaCO₃) should achieve TOC removals no lower than 45 %. At Campion Hills the low TOC removal is the result of high content of hydrophilic and microbial organic matter. It was believed that improved organic matter removal efficiency could be achieved at Campion Hills by the implementation of enhanced coagulation, i.e. by manipulating the pH and/or coagulant dosage. High TOC concentrations observed for the raw Campion Hills water indicate the presence of higher molecular-weighted fractions, favourably removed by coagulation (Chow et al., 1999). Thus the efficiency of the coagulation process could be potentially improved. However, the coagulation enhancement at Campion Hills WTW can be costly as the high baseline coagulant pH (average 7.4) is determined by high raw water pH. Thus, to investigate the potential improvements in TOC removal efficiency with lowering the coagulation pH, OM removal based on TOC, fluorescence and absorbance measurements were evaluated.

Figure 6.2 plots the raw water fluorescence properties against TOC removal for Campion Hills data derived from the main fluorescence dataset (August 06 – February 08) and from the coagulation experiment (November 08 – February 09). It can be observed that raw water

hydrophobicity (peak C emission wavelength) and microbial fraction content (tryptophan-like intensity) of coagulation experiment samples correspond to the baseline organic matter properties. However, the TOC removal of the experimental samples was significantly higher than TOC removal obtained for baseline conditions. Thus, by lowering coagulation pH, the TOC removal was improved with no change in raw water character.

For each coagulation pH an average TOC removal was calculated and compared with the baseline (at the beginning of experiment) removal at pH value of 7.0 (34.0 %). A decrease in coagulation pH enabled an additional average organic matter removal of 5.7 % at pH 6.5, 16.9 % at pH 6.0, 20.7 % at pH 5.0, and 32.2 % at pH 5.5. The greatest TOC removal was observed for pH 5.5, not for the lowest pH 5.0. As the coagulation pH experiment was carried out in full scale over several weeks, the raw water OM concentrations varied significantly in the study period (from 6.2 mg/l 08.02.09 to 10.8 mg/l 10.12.08). The higher TOC concentrations of raw water imply higher coagulant doses required to mitigate the effect of increased DBPs precursor input. Thus, the direct comparison of the lowering coagulation pH conditions is impaired by the effect of changing coagulant doses. Thus, the difference in TOC removal efficiency between two lowest coagulation pH may result from higher TOC concentrations of raw water and higher coagulant doses applied (average TOC for pH 5.5: 10.3 mg/l, pH 5.0: 6.8 mg/l) (Table 6.2).

In Figure 6.3 fluorescence EEMs for raw, clarified and final water collected during coagulation optimisation experiment are presented. To compare the effect of enhanced coagulation on OM character and removal efficiency, EEMs corresponding to baseline, operational (18.11.08, coagulation pH 7.0) and optimised conditions (12.02.09 coagulation pH 5.0) were evaluated (Table 6.2). The raw water properties (pH and TOC concentrations) were similar for both sets of samples: pH 8.0 and TOC 7.0 mg/l for baseline and pH 8.2 and

TOC 7.2 mg/l for optimised conditions respectively. This suggests that raw water properties were similar prior to coagulation. The visual inspection of raw water EEMs for both sampling episodes indicated very similar OM composition with slightly higher fluorescence intensities in the humic-like region for the baseline raw water EEM (15 %; Figure 6.3 a and d; Table 6.4). Fluorescence spectra for clarified water indicated a greater reduction in all fluorescence peak intensities for the optimised samples of 74 units for peak A, 30 units for peak C, and 27 units for peak T2 compared to the baseline samples. Furthermore, the difference in peak C emission wavelength between raw and clarified water was also found to be greater for the optimising samples (of 20 nm in peak A and 6 nm in peak C). Thus, coagulation pH reduction impacted both the quantitative and qualitative OM properties, while there was a greater removal of fluorescence intensities for all fluorophores and the residual OM is more hydrophilic. The final water for both baseline and optimised conditions demonstrated similar pattern with fluorescence intensities slightly higher for the former (Figure 6.3 c and f). This indicates that additional removal of OM in baseline conditions needs to be facilitated by the downstream treatment processes, i.e. GAC. However, a high OM load on GAC adsorbers significantly reduces the adsorption capacities. Thus, lowered residual OM content post-coagulation in optimisation process can increase the operational life of the GAC adsorbers and their adsorption capacities. Additionally, as the GAC preferentially removes hydrophilic OM, optimised coagulation can enhance adsorption of OM by retention of more hydrophilic OM compared to the baseline, non-optimised conditions.

In Figure 6.4 the relationship between actual and predicted TOC (raw and clarified water model) removal for all sites is shown. Campion Hills data from the period August 06 – February 08 (baseline coagulation conditions) are shown as red dots, whereas the optimized coagulation organic matter removals are plotted with blue dots. With decreasing pH and

increasing TOC, TOC removal increased from the baseline value of 20% to 50% for the lowest pH 5.0 and even 70% for the low pH 5.5 and high raw water TOC concentrations. Those results are in accordance with the results of pH adjustment tests reported by other authors. Improved organic matter removal with low pH coagulation has been observed in several studies, except for specific conditions of low temperature, low alkalinity or low TOC water (alum and ferric chloride coagulants; Lind, 1995). Chowdhury et al. (1995) observed an increase of 15% in TOC removal when lowering the coagulation pH from 8.0 to 7.3. Similar results were obtained by Volk et al. (2000), who tested three different coagulants: alum, ferric chloride and poly-aluminium chloride.

The impact of low pH coagulation on organic matter removal is shown in Figure 6.5. Organic matter removal was calculated from the reduction in TOC concentrations between raw and clarified water. Additionally, fluorescence peak C intensity and UV absorbance measured at 254 nm were used as a surrogate for TOC. For a given coagulation pH, the fluorescence-derived organic matter removal strongly correlated with TOC removal. The percentage removal of UV absorbing material was higher (average 15%) than TOC removal, indicating that coagulation preferentially removes the aromatic organic matter fraction (Chow et al., 1999; Volk et al., 2000).

Furthermore, two previously developed fluorescence-derived TOC removal models were tested using data obtained in the coagulation experiment. The TOC removal prediction models were calculated with the simple linear regression for whole fluorescence dataset containing peak C emission wavelength and peak T fluorescence intensities for raw and clarified water, for all 16 WTWs (August 2006 – February 2008) (see Chapter 4). In the first model, raw water properties only were used (Equation 6.1), whereas the second model included fluorescence variables characterising both raw and clarified water (Equation 6.2). The

combined raw and clarified water regression explained greater variance within the dataset ($R^2 = 0.86$ compared to $R^2 = 0.55$ for the raw water regression model). The standard errors for the coefficients in the regression models were 30 % for the raw water and 13 % for the raw and clarified model respectively.

$$TOCrem (\%) = -206.71 - 0.68T_{intRAW} + 0.52C_{emRAW} + 0.11C_{intRAW} \quad (6.1)$$

$$TOCrem (\%) = -62.77 - 0.15T_{intCLA} + 0.24C_{emRAW} + 0.26C_{intRAW} - 0.40C_{intCLA} \quad (6.2)$$

where: T_{int} - tryptophan-like fluorescence intensity, C_{em} - peak C emission wavelength, and C_{int} - and peak C fluorescence intensity.

From the Figure 6.4 it can be seen that the model based on raw water fluorescence properties provided a good prediction of TOC removal only for the baseline coagulation conditions for which the model was developed (pH 7.0), and the quality of prediction deteriorated with changing pH. For the lowest pH 5.5 (Figure 6.6), the difference between actual and predicted TOC removal value exceeded 20%. For the model incorporating both raw and clarified water fluorescence properties (Figure 6.7), the accuracy of the prediction was not affected by the coagulation enhancement, and the maximum prediction error was less than 5%. In the experiment, the OM removal improvement was simply the result of optimized treatment conditions, without corresponding changes in raw water OM character. Thus, the raw FLU model demonstrated the baseline OM removal efficiency for the predominant type of OM for non-optimized WTWs. Furthermore, a fluorescence-based organic matter removal model (raw and clarified) can be a robust and accurate tool in TOC removal prediction, including the changing treatment conditions.

Finally, the impact of lowering coagulation pH on organic matter removal between different stages of water treatment is demonstrated in Figure 6.8. The organic matter content was significantly reduced after the coagulation and filtration stages, and the degree of the removal was dependent on the coagulation pH. Coagulation pH optimisation (red dashed-line) compared with the baseline conditions with coagulation pH value of 7.5 (blue line) enabled greater reduction in OM between raw and clarified water stages. Moreover, the residual, post-clarification OM content was significantly reduced. Thus, the optimised coagulation can reduce the DBPs formation potential, disinfectant demand and biofilm growth in the distribution system.

6.3. CHAPTER CONCLUSIONS

Previously, the application of fluorescence spectroscopy for the characterisation of OM and removal efficiency for 16 WTWs was presented. The results reflected operational coagulation conditions and were analysed with the standard peak-picking approach (Chapter 4) and advanced data mining techniques (Chapter 5). To fully address the objective 2 (see Section 2.5), the ability of fluorescence spectroscopy to characterise unit process performance at WTWs should be tested for both the operational (baseline) and optimised treatment conditions.

Thus, in chapter 6 the effect of coagulation adjustment on ability of fluorescence spectroscopy to characterise OM properties and removal efficiency was investigated, based on optimisation experiment carried out at Campion Hills WTW. Campion Hills WTW demonstrates low OM removal efficiency as a result of a predominance of hydrophilic OM and high coagulation pH. It was found based on analysis of fluorescence EEMs that lowering of the coagulation pH from baseline pH 7.5 to 5.5 can yield significant improvements in OM removal efficiency (from 30 % for baseline to 60 % for optimised conditions). For similar raw

water fluorescence OM properties, higher reduction in fluorescence intensity / TOC removal between raw and clarified water was observed for optimised coagulation conditions. Moreover, the fluorescence signature of clarified water OM suggested that lowering the coagulation pH decreases the degree of hydrophobicity of residual OM compared to baseline conditions. Thus, from the fluorescence analysis it was concluded that optimised coagulation impacts both the quantitative and qualitative OM properties and enhances OM removal efficiency in the entire treatment system. Fluorescence spectroscopy was shown to be a robust OM characterisation technique in a range of different treatment conditions (Objective 2, see Section 2.5). So far, the fluorescence analysis was related to the OM removal efficiency by coagulation. To test the ability of fluorescence spectroscopy to characterise the overall OM removal at WTW (between raw and final water treatment stages), the characterisation of OM character and removal through entire treatment process is presented in the next chapter.

6.4. CHAPTER 6 FIGURES

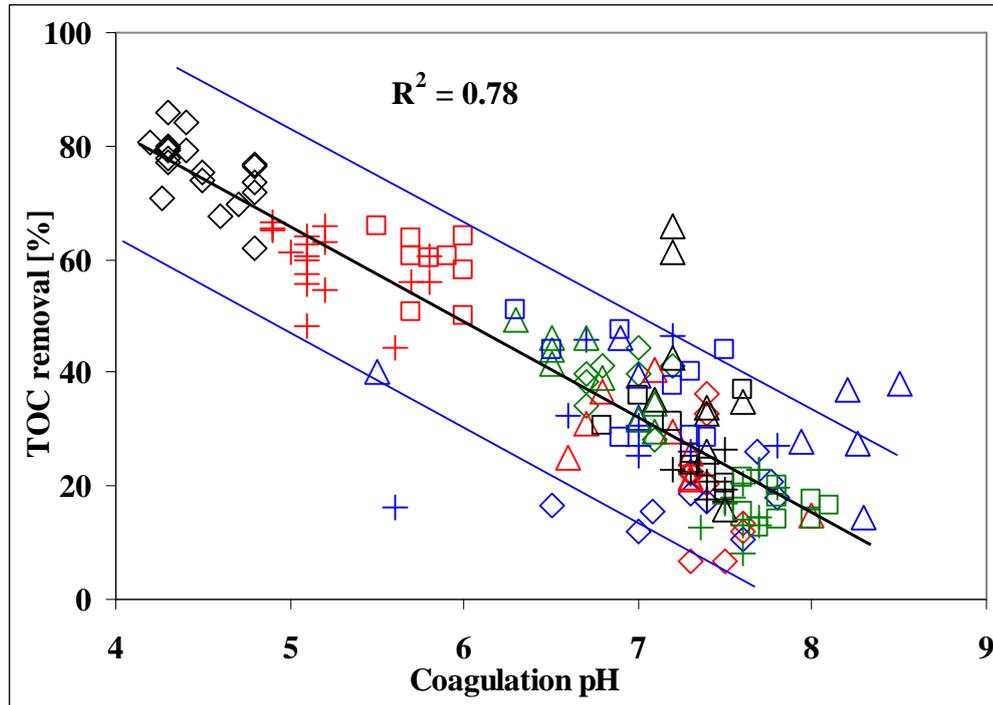


Figure 6.1: A relationship between coagulation pH and TOC removal efficiency (May 2007 – February 2008). For explanation of the labels see Figure 4.2

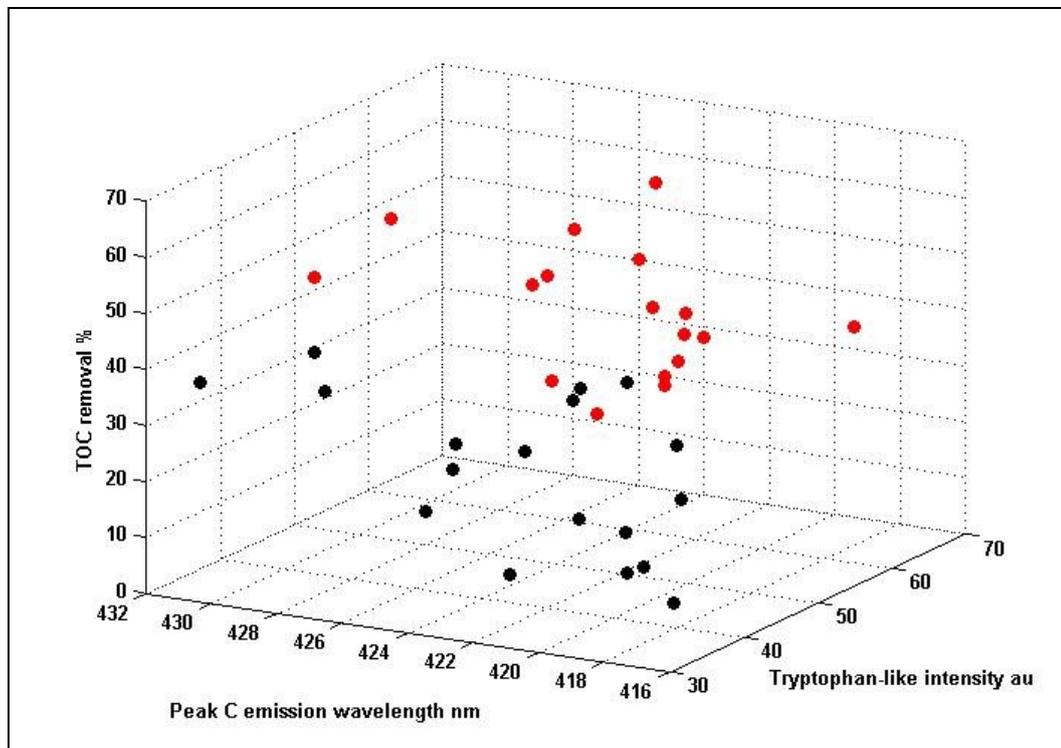
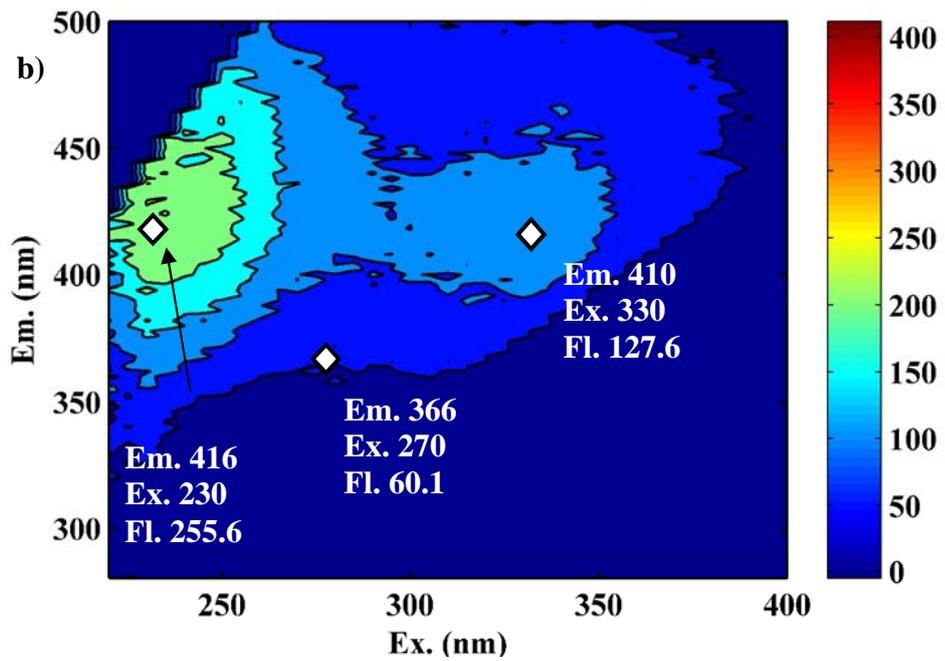
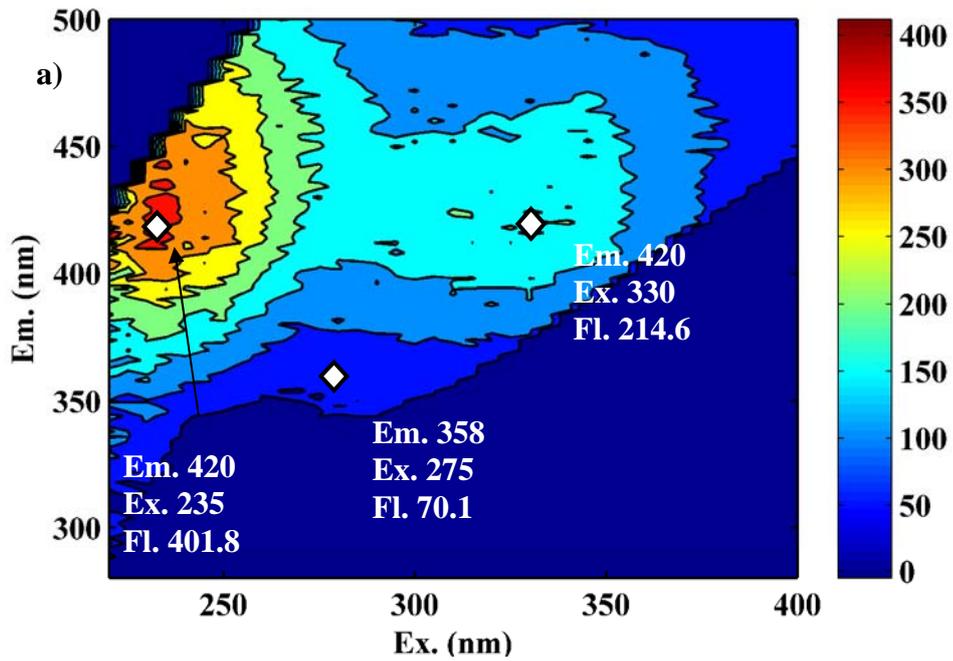
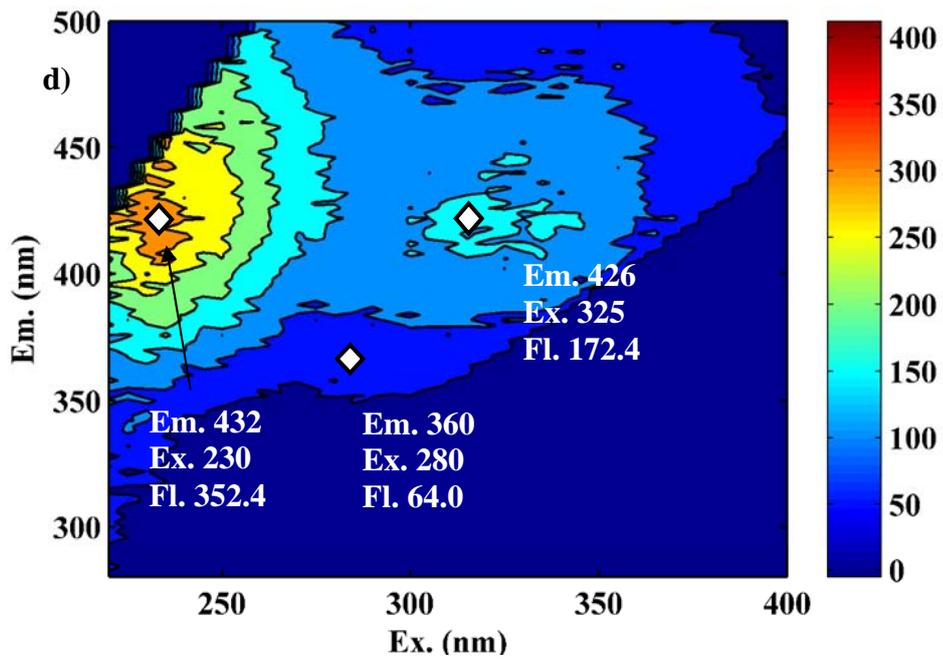
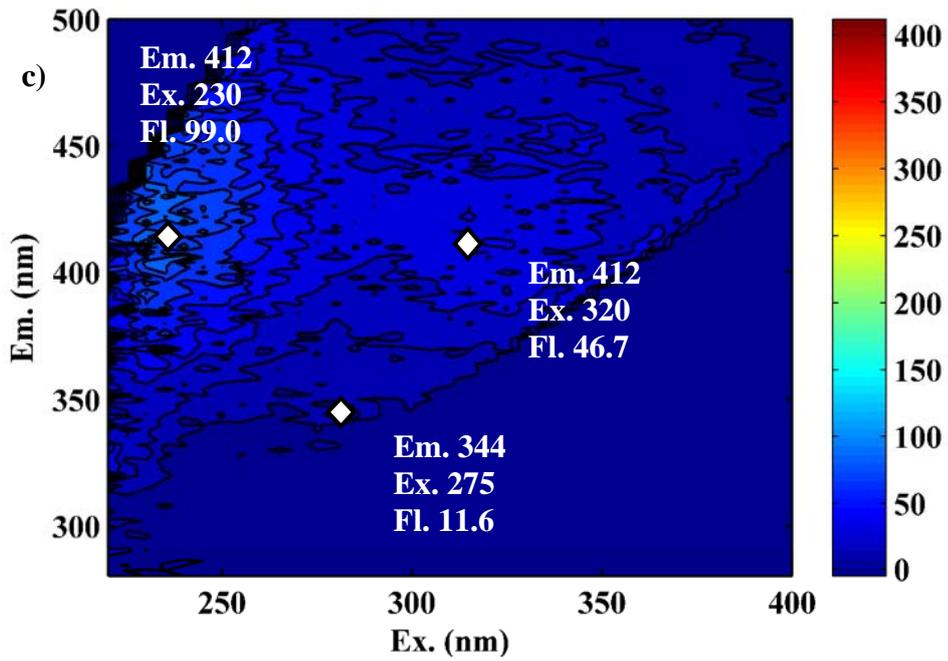


Figure 6.2: Campion Hills raw water spectral properties, peak C emission wavelength and peak T intensity vs TOC removal for all data August 06 – February 08 (black dots) and coagulation experiment data (red dots)





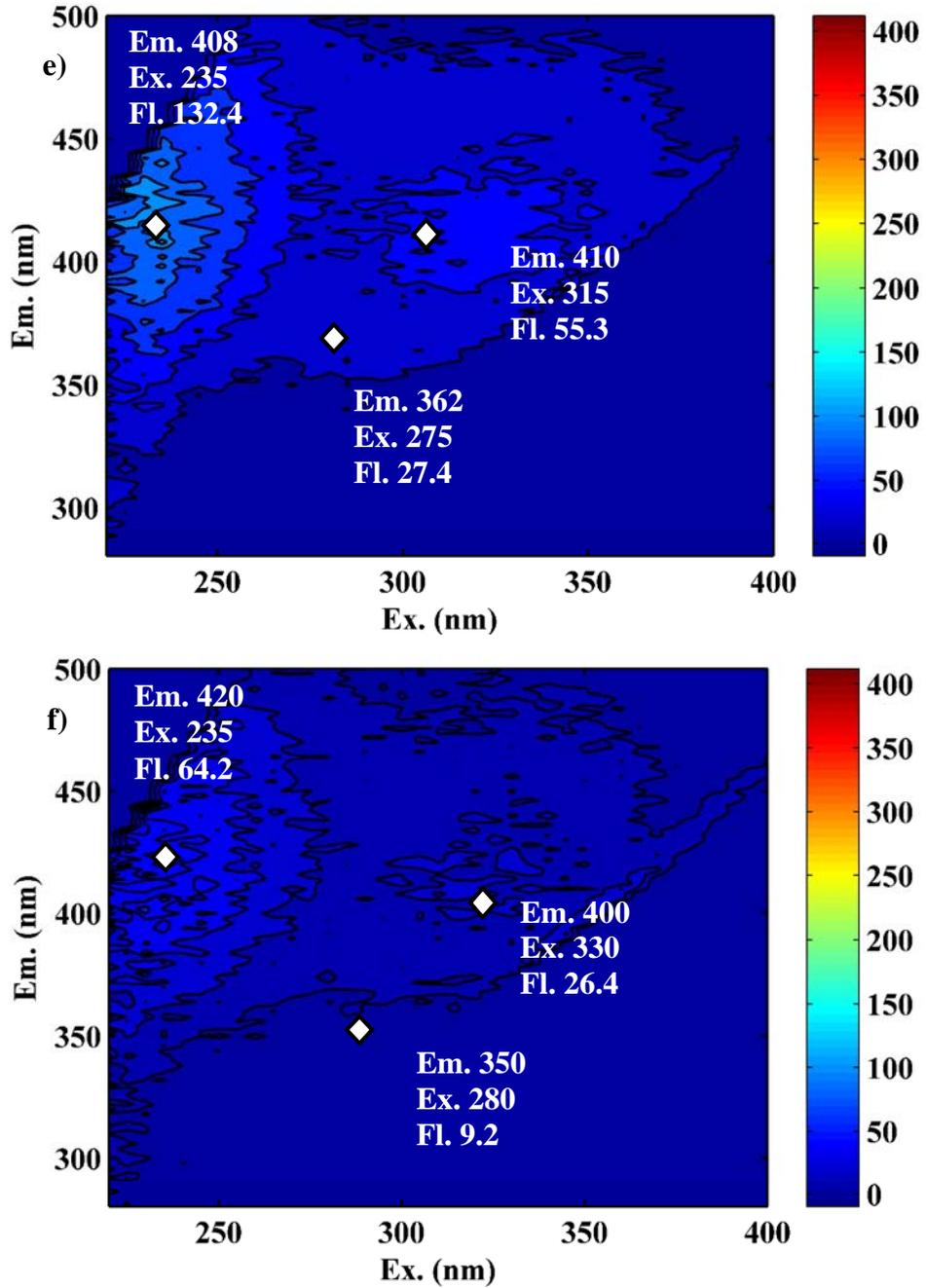


Figure 6.3: Raw, clarified, and final water EEMs for the two coagulation optimisation episodes

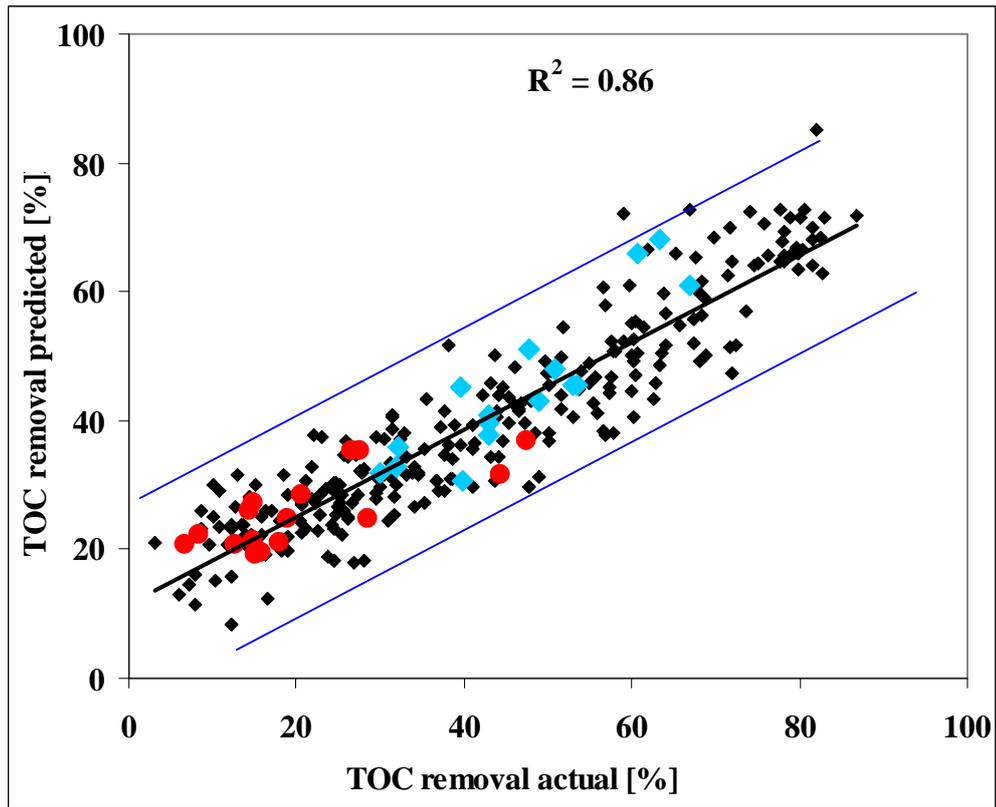


Figure 6.4: TOC removal predicted (raw-clarified water fluorescence model) vs actual TOC removal. CH data August 06 – February 08 (red dots), coagulation experiment data (blue dots)

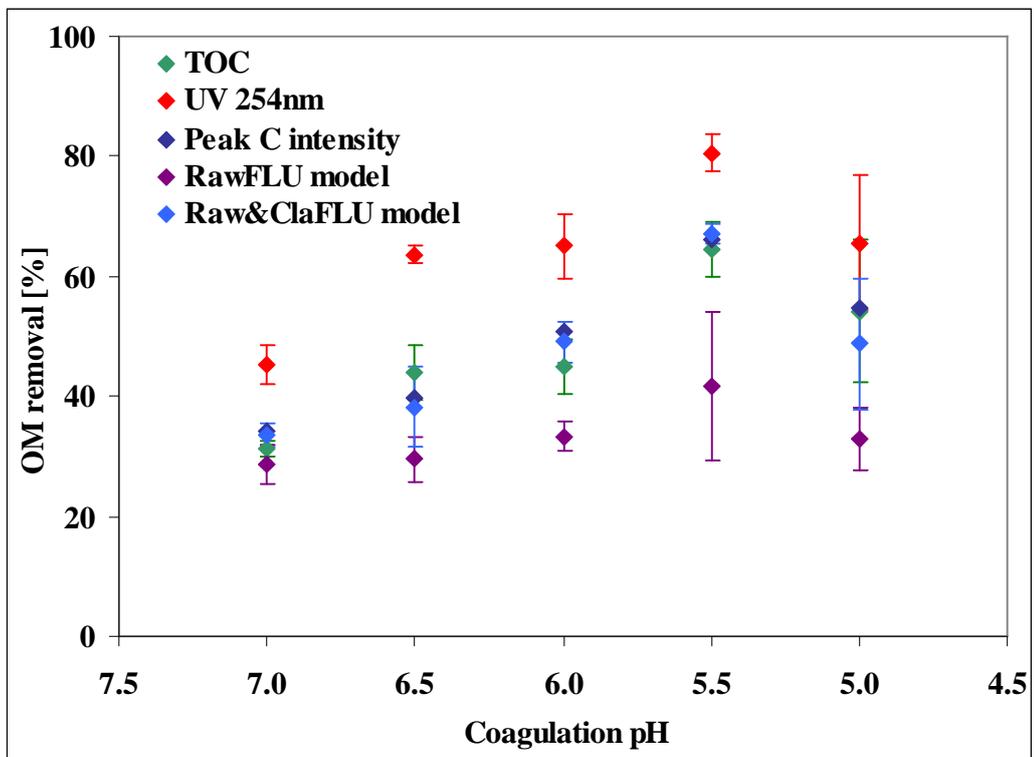


Figure 6.5: Organic matter removal vs pH during the coagulation experiment

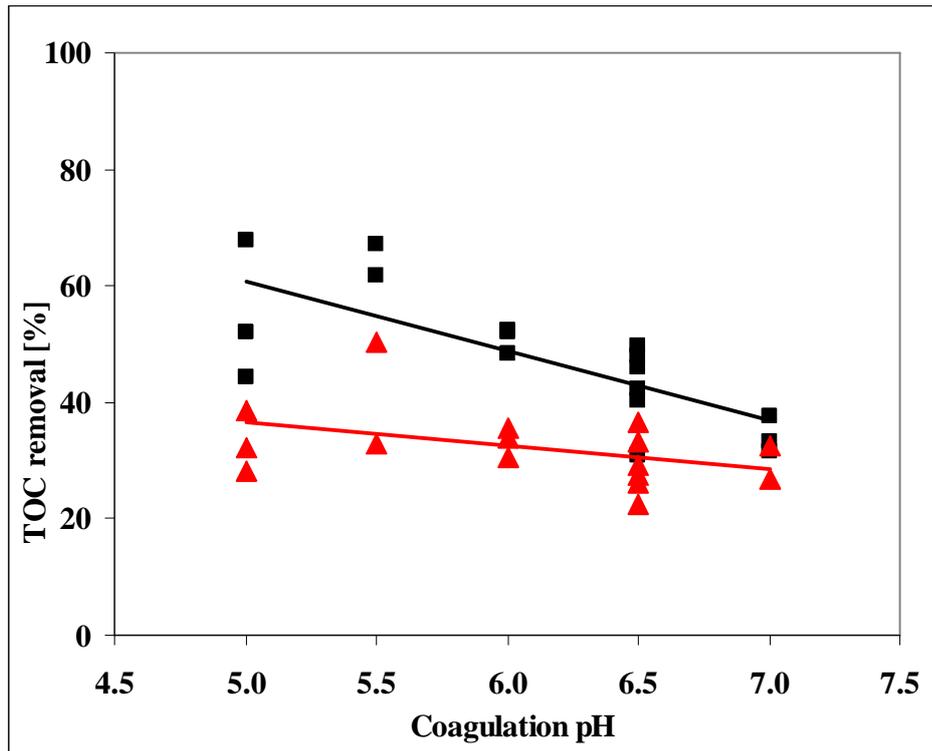


Figure 6.6: TOC removal actual (black squares) and predicted from the fluorescence data of raw water (red triangles) vs coagulation pH during Campion Hills coagulation experiment

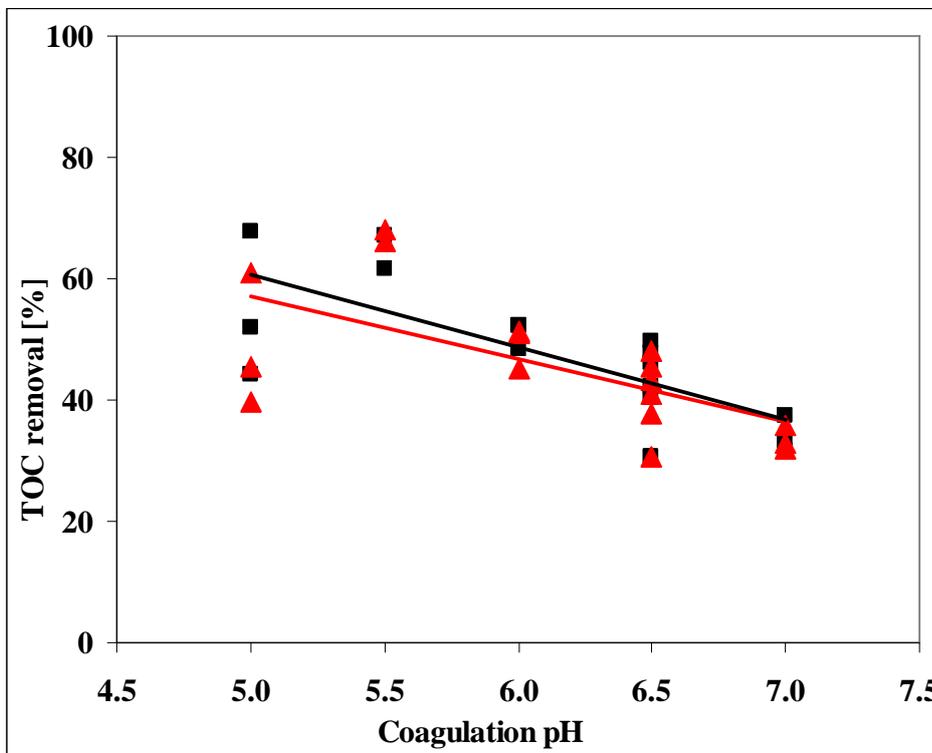


Figure 6.7: TOC removal actual (black squares) and predicted from the fluorescence data of raw and clarified water (red triangles) vs coagulation pH during Campion Hills coagulation experiment

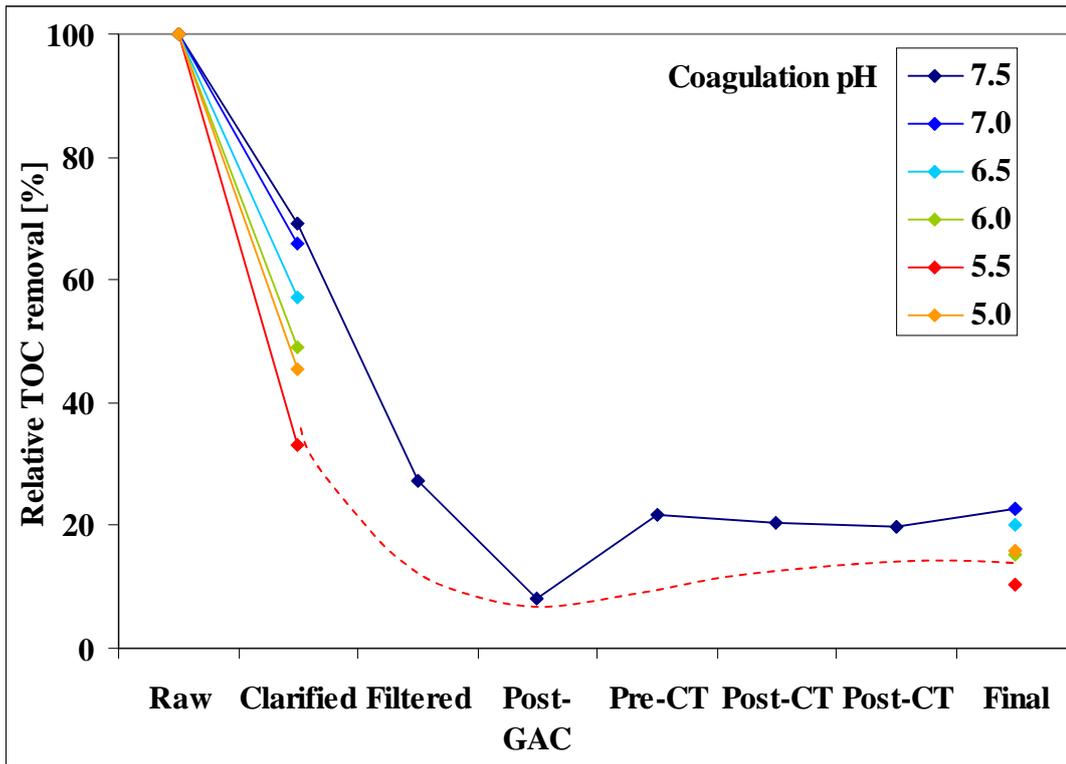


Figure 6.8: Relative TOC removal vs coagulation pH through Champion Hills WTW. A blue line (July 07), a red dashed-line demonstrates hypothesized TOC removal for pH value of 5.5

6.5. CHAPTER 6 TABLES

Table 6.1: Required TOC removals by enhanced coagulation (%) for plants using conventional treatment (Step 1 of US EPA guidelines) (US EPA, 1999)

Raw water TOC (mg/l)	Raw water alkalinity (mg/l as CaCO ₃)		
	0 - 60	>60 – 120	>120
> 2.0 – 4.0	35	25	15
>4.0 – 8.0	45	35	25
>8.0	50	40	30

Table 6.2: TOC (mg/l) and SUVA (mg/l.m) of raw (R), clarified (C), and final water (F). Organic matter removal prediction (%) from fluorescence analysis (peak C int, raw water model Flu1, raw and clarified water model Flu2), TOC removal and UV-Vis absorbance reduction

Date	Coagulant dose mg/l Fe	pH	TOC			SUVA			Peak C int	OM removal			UV-Vis
			R	C	F	R	C	F		Flu1	Flu2	TOC	
18.11.08	7.5	7.0	7.0	4.8	3.3	4.4	3.4	1.7	37.4	26.8	35.9	32.0	47.8
19.11.08	6.4	7.0	6.9	4.7	3.3	3.9	3.1	1.9	33.1	32.5	32.8	32.0	46.3
20.11.08	5.4	7.0	6.8	4.7	3.3	4.1	3.4	1.9	31.6	26.8	31.9	29.9	41.5
25.11.08	6.1	6.5	7.4	4.5	3.2	3.5	2.2	2.6	30.8	33.3	30.6	39.9	62.7
26.11.08	6.6	6.5	6.6	3.7	2.7	4.0	2.6	0.7	42.4	29.1	40.9	43.0	62.6
27.11.08	7.2	6.5	7.2	3.7	2.6	3.7	2.5	0.3	46.1	26.1	43.2	49.0	65.2
02.12.08	8.3	6.0	6.3	3.8	2.4	3.5	2.3	1.2	48.3	30.6	45.2	39.7	61.0
03.12.08	7.5	6.0	6.3	3.3	2.4	3.8	2.7	0.9	52.3	33.9	51.1	47.6	63.0
04.12.08	7.4	6.0	6.3	3.3	2.5	3.7	2.1	0.7	52.1	35.5	51.1	47.6	71.0
09.12.08	12.8	5.5	10.1	3.7	2.0	5.4	2.4	1.6	67.0	33.0	68.2	63.4	83.9
10.12.08	14.1	5.5	10.8	3.3	1.7	5.6	3.7	1.8	70.0	-	-	69.4	79.8
11.12.08	14.1	5.5	9.9	3.9	1.6	4.1	2.3	1.8	61.7	50.4	66.1	60.6	78.1
06.02.09	9.2	6.5	5.8	3.3	2.7	3.3	2.2	1.3	40.3	27.5	37.8	43.0	61.7
07.02.09	7.6	6.5	6.0	2.9	1.9	3.2	2.7	1.2	49.6	36.7	48.1	50.8	58.9
08.02.09	6.4	6.5	6.2	2.9	1.9	2.9	2.8	1.0	47.8	22.6	45.6	53.5	55.9
10.02.09	5.4	5.0	6.5	3.7	2.2	3.1	2.3	0.9	44.1	32.1	39.7	42.9	58.1

11.02.09	6.6	5.0	6.6	3.1	2.0	3.1	2.6	0.7	52.0	28.3	45.6	53.1	60.3
12.02.09	9.9	5.0	7.2	2.4	1.8	3.3	2.1	0.6	67.9	38.7	60.9	66.6	78.5

Table 6.3: Raw and clarified water properties for Campion Hills and all 16 water treatment sites (August 06 – February 08)

Parameter	Campion Hills		All sites	
	Mean	S.D.	Mean	S.D.
Tryptophan-like intensity RAW (au)	42.2	6.8	34.9	15.8
Tryptophan-like intensity CLA (au)	36.1	5.7	27.8	15.8
Peak C emission wavelength RAW (nm)	422.5	5.4	428.4	11.0
Peak C emission wavelength CLA (nm)	418.2	7.6	422.0	9.5
Peak C intensity RAW (au)	162.9	46.9	141.1	48.0
Peak C intensity CLA (au)	127.8	25.9	86.2	40.4
TOC RAW (mg/l)	6.1	1.3	6.4	16.6
TOC CLA (mg/l)	5.1	1.0	4.1	9.1
OM removal fluorescence peak C (%)	21.5	9.7	37.8	20.2
OM removal TOC (%)	20.0	11.8	40.6	21.7
Coagulation dose (mg/l Fe)	6.8	1.5	7.6	3.6
Coagulation pH	7.4	0.1	6.6	1.1

Table 6.4: Fluorescence peak emission (nm), excitation wavelength (nm) and intensity (au) for raw, clarified, and final water for baseline (18.11.08) and optimised (12.02.09) coagulation conditions

Date	Water type	Peak A			Peak C			Peak T2		
		Em.	Ex.	Int.	Em.	Ex.	Int.	Em.	Ex.	Int.
18.11.08	RAW	420	235	401.8	420	330	214.6	258	275	70.1
	CLA	416	230	255.6	410	330	127.6	366	270	60.1
	FIN	412	230	99.0	412	320	46.7	344	275	11.6
12.02.09	RAW	432	230	352.4	426	325	172.4	360	280	64.0
	CLA	408	235	132.4	410	315	55.3	362	275	27.4
	FIN	420	235	64.2	400	330	26.4	350	280	9.2

7. VARIATION IN FLUORESCENCE PROPERTIES THROUGH WATER TREATMENT WORKS

7.1. INTRODUCTION

The relative change in quantitative and qualitative properties of EEMs collected for different water treatment stages can provide useful information on the degree of OM removal in relation to the physicochemical properties of the removed fractions. To evaluate the fluorescence changes through different stages of water treatment, additional fluorescence measurements were performed for selected WTWs.

Fluorescence measurements were carried out on raw, partially-treated and final water from five WTWs: Campion Hills (18.08.07), Draycote (11.07.07 and 24.08.07), Melbourne (18.07.07), Strensham (05.08.07) and Whitacre (09.08.07). In order to reveal the relative change in fluorescence properties, and hence organic matter removal across different treatment stages, the fluorescence intensities of peak A, peak C and peak T fluorescence regions were assessed. Previously, the intensity of peak T fluorescence was measured as a maximum intensity value in a specific wavelength range (see Section 3.2). However, it was found that the spectral location of peak T does not yield any significant information on OM character (see chapter 4). Thus, to account simply for overall quantitative changes in peak T region through WTW, peak T intensity was measured at the fixed emission and excitation wavelengths representing the average location of peak T (350 nm emission wavelength and 280 nm excitation wavelength). For each WTW, eight water samples were analysed: raw, clarified, filtered, post-GAC, pre-contact tank (post chlorine addition), post-contact tank prior to and after de-chlorination (i.e. with sulphur dioxide), and final water.

For Champion Hills and Strensham, pre-contact tank fluorescence was the lowest (10-20 % of raw water fluorescence), with 30 % for Draycote and 40-50 % for Melbourne. Through the final stages of treatment (pre-contact tank to final water), the fluorescence intensity stabilized with a slight increase of 10 % compared with the pre-contact tank minimum.

Final water fluorescence in the peak T region was similar for all sites (30 % of raw water fluorescence), indicating uniform effectiveness in microbially-derived organic matter removal, whereas for peak A and peak C fluorescence, the WTWs responses differed significantly (Figure 7.1 and 7.2): 20% of raw water fluorescence at Champion Hills, 30% at Strensham, 40% at Draycote and Whitacre, and 60% at Melbourne.

7.2. EFFECTS OF CHLORINE ADDITION ON FLUORESCENCE

The fluorescence signature prior to chlorination can be indicative of quantitative and qualitative changes in OM related to its removal. Post chlorination, the effect of fluorescence intensity decrease due to OM removal is masked by the increase induced by the complex reactions of OM with chlorine and redistribution of particular moieties and functional groups.

Throughout the treatment process, in all fluorescence regions (Figures 7.1, 7.2 and 7.3), a steady decrease in fluorescence intensity for all sites was observed between raw and post-GAC waters. Once chlorinated, an increase in fluorescence intensity was discerned, related to fluorescence quenching with chlorine. Overall, an increase in fluorescence intensity post chlorination does not necessarily imply the presence of quantitative changes in organic matter concentration. The precise mechanisms of chlorine quenching of organic matter are poorly understood (Korshin et al., 1999; Świetlik and Sikorska, 2004; Beggs et al., 2009). Upon the reaction of chlorine with organic matter, several transient and stable DBPs species are formed depending on the initial character of OM. The reaction with chlorine and formation of DBPs

changes the chemical structure of the organic matter and therefore its fluorescence properties. Thus, the observed increase in fluorescence intensity can indicate conformational and structural changes in functional group composition. In addition to dynamic quenching, a static quenching mechanism can occur, whereby the free electrons in organic matter interact with free chlorine, rather than having spare energy to emit as fluorescence (Lakowicz, 1999). Unlike dynamic quenching, static quenching occurs without any distinctive changes in chemistry of either chlorine or organic matter.

During the chlorination due to complexity of chlorine/OM reactions, the fluorescence signature becomes more equivocal, and both increases and decreases in fluorescence intensity can be discerned. Depending on the primary reaction pathway (substitution or oxidation), the type and amount of intermediate chlorination species generated, different functional groups are selectively transformed and removed. During halogenation, a common selective removal of high molecular weight, highly conjugated aromatic structures containing carbonyl, hydroxyl and amine functional groups occurs (Senesi, 1990). As a result of the breakdown of active aromatic structures in humic molecules into smaller compounds, the fluorescence intensity increases (Korshin et al., 1999, Świetlik and Sikorska, 2004). This can explain the higher fluorescence intensities for final water stages compared with the chlorination stage. Moreover, the depletion of chlorine residual post-chlorination enhances the microbial activity and results in higher amounts of autochthonous OM ascribed to tryptophan-like fluorescence. Beggs et al. (2009) observed small initial increase in fluorescence intensity, followed by a levelling off and more dramatic decrease after 2 hours of contact time and 1 mg/l consumption of chlorine.

The pattern of fluorescence changes post chlorination depends on the qualitative OM properties including the presence of particular functional groups and overall degree of

aromaticity and the chlorination conditions (contact time, applied dosage of chlorine). The high molecular weight, OM structures containing carbonyl, hydroxyl and amine functional groups were found to be the sites preferentially attacked by chlorine molecules (Reckhow et al., 1990; Senesi, 1990; Gallard and von Gunten, 2002). Therefore, this OM fraction demonstrates rapid initial rate of chlorine consumption, whereas less reactive OM structures (i.e. containing methyl group) yield to a slow and prolonged chlorine demand (Gang et al., 2002). Here, post chlorination changes in all fluorescence regions varied significantly between treatment sites indicating differences in OM composition prior to chlorination (Figure 7.1-7.3).

The higher chlorine dosage and longer reaction times determine more extensive changes in OM structure and composition as more free chlorine molecules are present in the solution. Korshin et al. (1999) suggested that during chlorination, fluorophores demonstrate different decay times (fast-, medium-, and slow-decaying fluorophores). Thus, with increased chlorine dosages, not only are the fast-decaying OM sites eliminated, but the contribution of medium- and slow-decaying fluorophores increases.

7.3. RESULTS

7.3.1. Visual inspection of EEMs

To visualise the overall change in fluorescence properties through various treatment stages, whole EEMs for Strensham WTW with fixed fluorescence intensity scales are shown in Figure 7.4. During the initial treatment stages the fluorescence EEMs indicate the changes in OM properties induced by its selective removal. Clarification can be seen to reduce significantly the fluorescence intensity via removal of organic matter, especially within the areas of the highest initial fluorescence intensity (DOC concentration), in the humic-like and

fulvic-like region (Figure 7.4 a and b). The cumulative reduction in OM concentration measured as a peak C intensity between raw water and subsequent treatment stages increased from 25 % for post-clarification, to 64 % for post-GAC, and to 70 % for final water. The highest OM removal was observed for pre-contact tank stage, 80 % of initial raw water peak C intensity. The pre-contact tank stage demonstrated also the lowest fluorescence intensities in peak A and peak T fluorescence regions, 66.9 % and 85.7% of the initial raw water fluorescence in peak A and peak T respectively (Figure 7.4 e). However, the disinfected water showed an increase in fluorescence intensity that cannot be directly related to the OM removal efficiency due to quenching mechanisms as discussed earlier. Therefore, the qualitative properties of fluorescence spectra can be related to OM removal efficiency just prior to chlorination. Comparison of pre-contact tank and final water (Figure 7.4 e and h) indicates changes in the fluorescence properties induced by the chlorine addition (30.1 % increase in peak C intensity, 67.3 % increase in peak T intensity, and 8.6 % decrease in peak A fluorescence intensity). The relatively higher increase in peak T fluorescence region can result from the depletion of chlorine residual post-chlorination. Lower chlorine residual can enhance the microbial activity and result in higher amounts of autochthonous OM ascribed to tryptophan-like fluorescence.

The OM removal throughout water treatment is accompanied by changes in structural and chemical OM properties. The breakdown of high molecular weight OM compounds is reflected in the fluorescence emission and excitation shift towards lower wavelengths prior to chlorination (Figure 7.4 a-d). The peak C emission wavelength decreased from 432 nm for raw water to 422 nm for clarified and post-GAC water indicating a decrease in OM degree of hydrophobicity. The reduction in aromaticity between raw and subsequent treatment stages reflects the preferential removal by coagulation of aromatic, high-molecular weight OM.

Similar changes were demonstrated by the location of peak A fluorescence maximum. Post chlorination (Figure 7.4 f, g, h) both peak C and peak A emission wavelengths were the lowest compared to preceding treatment stages (raw water peak C 432 nm and peak A 430 nm, final water peak C 416 nm and peak A 416 nm). However, the lowest emission (peak C 408 nm and peak A 406 nm) were observed for post contact tank (de-chlorination; Figure 7.4 g).

7.3.2. *Fluorescence regional integration*

Additionally, to account for the overall quantitative changes in fluorescence regions through the treatment process, a fluorescence regional integration (FRI) approach was adopted. Although the FRI technique is very common in fluorescence analysis, it may produce unreliable results compared to a simple peak-picking approach. In the FRI method, the relative importance of particular fluorescence regions can be quantified by determining the volume of fluorescence beneath this region. Such an approach is easily quantified and visualised and therefore often used in fluorescence analysis for characterisation of OM fluorescence changes through different processes, i.e. DBPs formation, fermentation (Marhuenda-Egea et al., 2007; Yang et al., 2008; Johnstone and Miller, 2009). Thus, the efficacy of FRI in characterising OM fluorescence changes through water treatment was tested on the Strensham WTW data.

Fluorescence regional integration enables the evaluation of relative changes in distribution of different fluorophores through following treatment stages (Chen et al., 2003). In the method, the excitation-emission matrix is divided into five fluorescence regions corresponding to particular organic matter constituents of distinctive spectrofluorometric properties (Figure 7.5). Regions I and II contain peptides with aromatic residues such as tyrosine exhibiting fluorescence at shorter excitation wavelengths and emission wavelengths

(< 250 nm and < 380 nm) (Ahmad and Reynolds, 1999; Determann et al., 1994). Soluble microbial by-products (region IV) demonstrate fluorescence peaks at intermediate excitation wavelengths (250–280 nm) and shorter emission wavelength (< 380 nm) (Coble, 1996; Reynolds and Ahmad, 1997). Finally, fluorescence at longer emission wavelengths (> 380 nm) is related to humic acid-like (region V) and fulvic-like fluorescence (region III) (McKnight et al., 2001).

For each EEM region i the volume of fluorescence beneath this region Φ_i was calculated with the following equation:

$$\Phi_i = \sum_{\lambda_{ex}} \sum_{\lambda_{em}} I(\lambda_{ex} \lambda_{em}) \Delta \lambda_{ex} \Delta \lambda_{em} \quad (7.1)$$

where: $\Delta \lambda_{ex}$ and $\Delta \lambda_{em}$ denote the excitation and emission wavelength interval (here 5 and 2 nm respectively) and $I(\lambda_{ex} \lambda_{em})$ is the fluorescence intensity at each excitation-emission wavelength pair.

Calculated fluorescence volumes were normalized to account for the variations in the number of data points within each regional integration area by using a multiplication factor M_{fi} (Table 7.1).

The results of regional integration analysis for Strensham WTW are shown in Table 7.2. Similar patterns of fluorescence changes through treatment stages can be observed for regions I, II, and III (with the excitation in UV light). The cumulative fluorescence intensity in these regions initially decreased between raw and filtered water stages and then increased until pre-contact tank stage. At the pre-contact tank stage, fluorescence intensities were higher compared to the raw water properties of 7.4 % in region III, 2.1 % in region I, and 8.0 % in region II respectively. For the final water, the cumulative fluorescence intensity in region I, II,

and III were similar to raw water contributions in the total EEM volume. The opposite pattern of changes can be observed for fluorescence regions with absorption in the visible light range (region IV and V). Here, the cumulative fluorescence intensity increased until filtered water stage (increase of 4.6 % in region IV and 7.0 % in region V compared with raw water) and then decreased to reach the minimum for pre-contact tank stage (decrease of 1.2 % and 16.3 % respectively). Post chlorination, the cumulative fluorescence intensity in all regions increased, which is in agreement with previous results (e.g. Korshin et al., 1999). However, the pattern of fluorescence changes prior to chlorination in FRI is distinctively different from the results of EEMs visual inspection and analysis of changes in peak fluorescence. For all fluorescence peaks, peak A representing region III, peak C representing region V, and peak T representing region IV, a steady decrease in fluorescence intensity between raw and pre-contact tank stage was observed. The more complex pattern of fluorescence changes in particular regions compared with the peak values for selected fluorophores indicates that each fluorescence region comprises several different fluorophores. This suggests that particular fluorescence region demonstrates a cumulative fluorescence response that can be different from single fluorophore response. Thus, the overall changes in fluorescence calculated on region-basis do not reflect the changes in fluorophores composition during the treatment. Hence, the FRI method does not provide useful information on qualitative changes in OM composition.

The cumulative fluorescence intensities for each EEM characterising particular treatment stage were compared with the OM removal efficiency measured as a reduction in peak C intensity. For Strensham WTW (Table 7.2), the total fluorescence volume of EEMs decreased through treatment, from 100 % for raw water to 26.7 % for final water. Thus, the final water fluorescence volume (OM quantity) was 26.7 % of original raw water. The relative

change in fluorescence volume was the highest between raw water and clarification stage (30.7 % of raw water fluorescence volume removed). For filtered and post-GAC waters a removal of further 20.4 and 18.3 % of fluorescence volume was observed. These results are in accordance with OM removal measured as a reduction in peak C, whereby removals of 25 % (between raw and clarified water), 64 % (between raw and filtered water), 80 % (between raw and pre-contact tank), and 70 % (between raw and final water) were observed. Thus, it can be inferred that fluorescence spectroscopy can be useful tool in determining the contribution of particular OM fractions and the OM removal efficiency between treatment stages.

7.3.3. *Self-organizing map*

Finally, the fluorescence OM changes through Strensham WTW were analysed with a self-organizing map (Figure 7.6). A map containing 16 nodes (8 by 2) was developed. The U-matrix (Figure 7.6 a) indicated the presence of two data clusters demonstrating distinctive OM fluorescence properties. The cluster at the bottom of the map contained samples of raw and clarified water, whereas the second cluster at the top of the map comprised post-GAC, pre- and post-contact tank, and final water samples. The filtered water fluorescence properties demonstrated intermediate fluorescence properties. Thus, the SOM discriminated between fluorescence properties of particular treatment stages.

The samples' distribution on the map suggests that fluorescence intensity decreases from the bottom to the top of the SOM map. The highest fluorescence intensities were observed for raw water sample (F1), whereas the pre-contact tank (F5) and post-contact tank (prior to de-chlorination, F6) exhibited the lowest fluorescence intensity. Two final water stages (post-contact tank post de-chlorination and final water) showed higher fluorescence intensities

compared to the two subsequent stages. These observations were in agreement with the peak-picking results. Thus, the vertical distance on the SOM map indicated the efficiency of OM removal measured as a reduction in fluorescence intensity.

The changes in the horizontal plane of the SOM map can be referred to the qualitative changes in OM. From the location of samples prior to chlorination, it can be observed that degree of aromaticity measured as a peak C emission wavelength decreased from the left to the right. However, for post-chlorination samples this pattern was less explicit. This can suggest that changes in OM composition during the chlorination and de-chlorination are complex and involve significant structural transformations. Thus, to provide a better understanding of changes in aquatic OM fluorescence induced by chlorination, the detailed fluorescence analysis should be performed in future.

7.4. CHAPTER CONCLUSIONS

In the previous chapter, the application of fluorescence spectroscopy for OM characterisation during WTW optimisation process was demonstrated. It was shown that fluorescence spectroscopy can yield significant information on quantitative and qualitative changes in OM throughout different conditions of coagulation process. Here, the efficacy of fluorescence spectroscopy to analyse the changes in OM through the entire treatment process was addressed. For the eight EEMs representing different treatment stages from raw to final, the relative changes in OM character and quantity were assessed using peak-picking approach, the FRI, and the SOM techniques. It was found that prior to the chlorination, OM fluorescence demonstrates steady OM removal and increasing contribution of hydrophilic OM fraction. Once chlorinated, the fluorescence OM signature was more complex and exhibited superimposed effects of OM removal and fluorescence increase due to quenching processes.

In addition, the ability of three fluorescence analysis techniques (peak-picking approach, the FRI, and the SOM) to determine OM character and removal from small fluorescence dataset was tested. Overall, visual inspection of EEMs coupled with fluorescence peak data analysis and SOM approach provided the best quantitative and qualitative analysis of changes in fluorescence properties through the water treatment process. Although the FRI technique has been reported to characterise changes in fluorescence through different processes, here it was found to be less robust. A more careful determination of the fluorescence regions is recommended for future studies.

7.5. CHAPTER 7 FIGURES

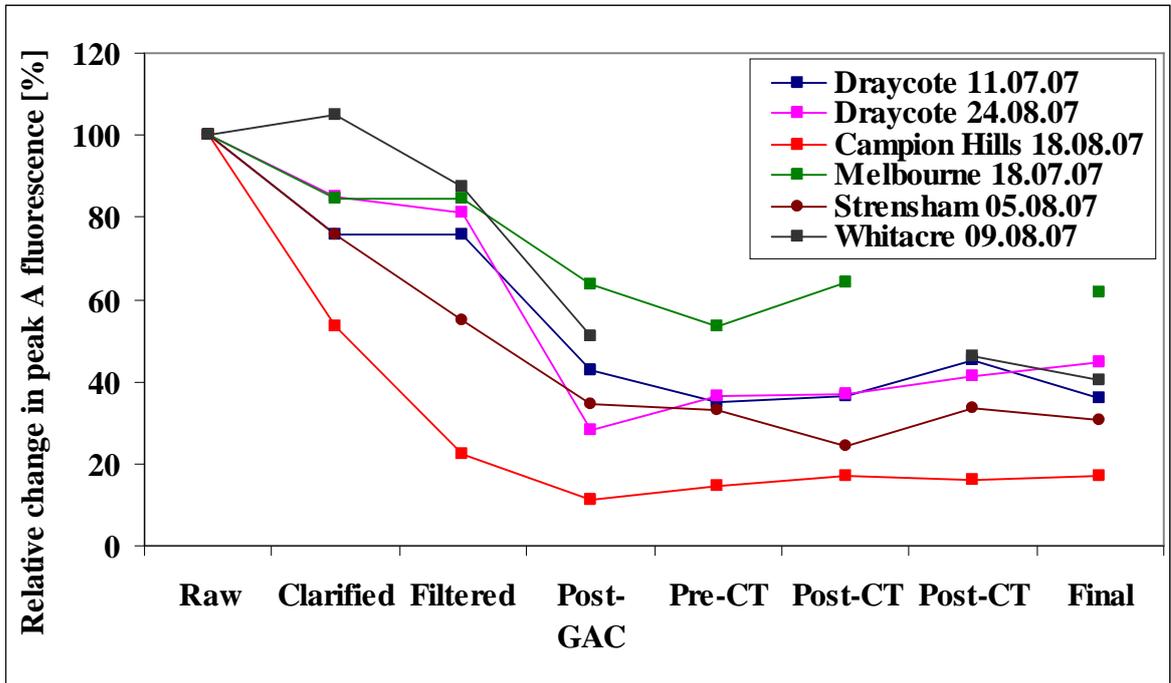


Figure 7.1: Relative change in peak A intensity (%) through different WTWs

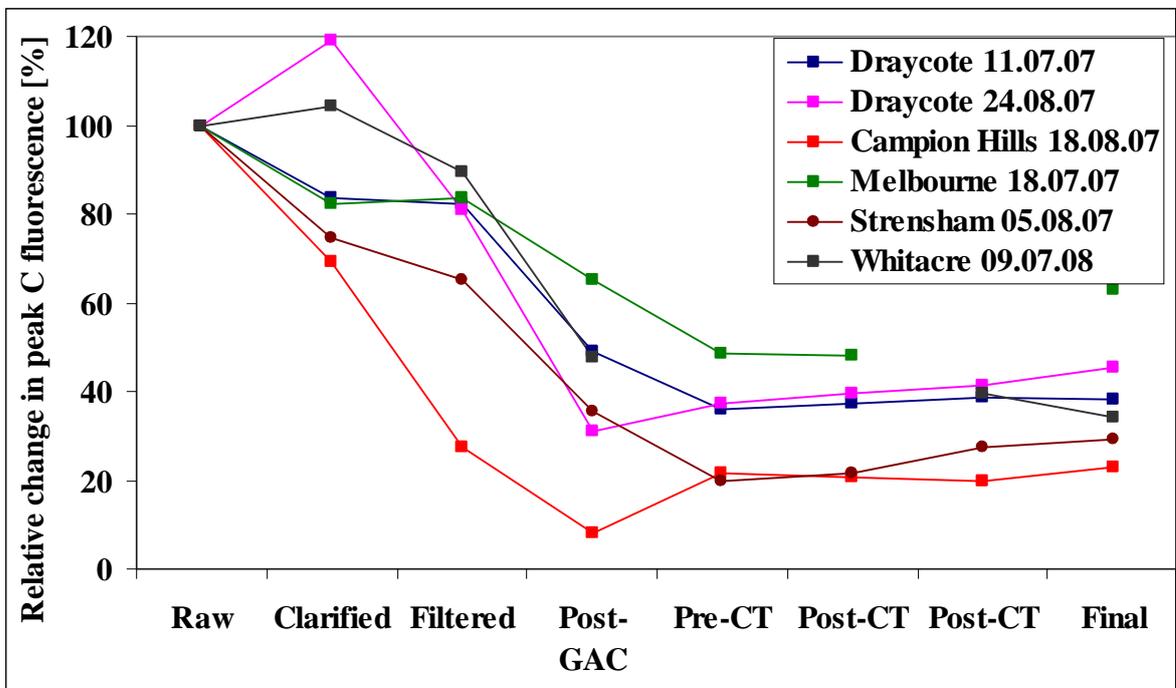


Figure 7.2: Relative change in peak C intensity (%) through different WTWs

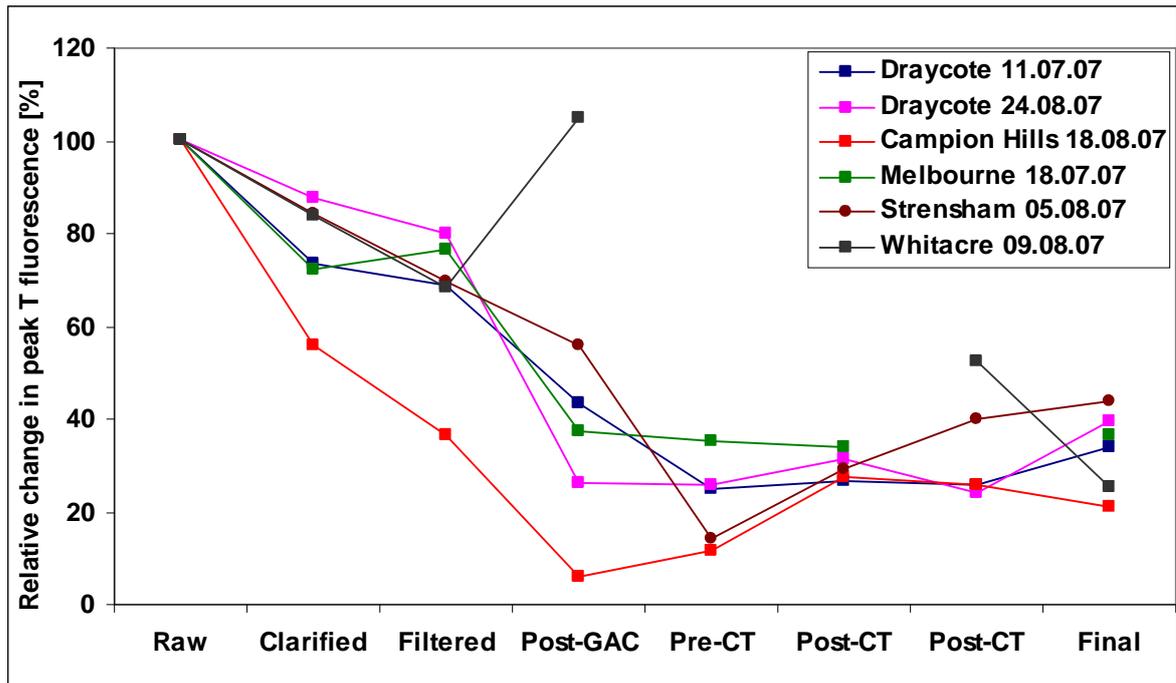
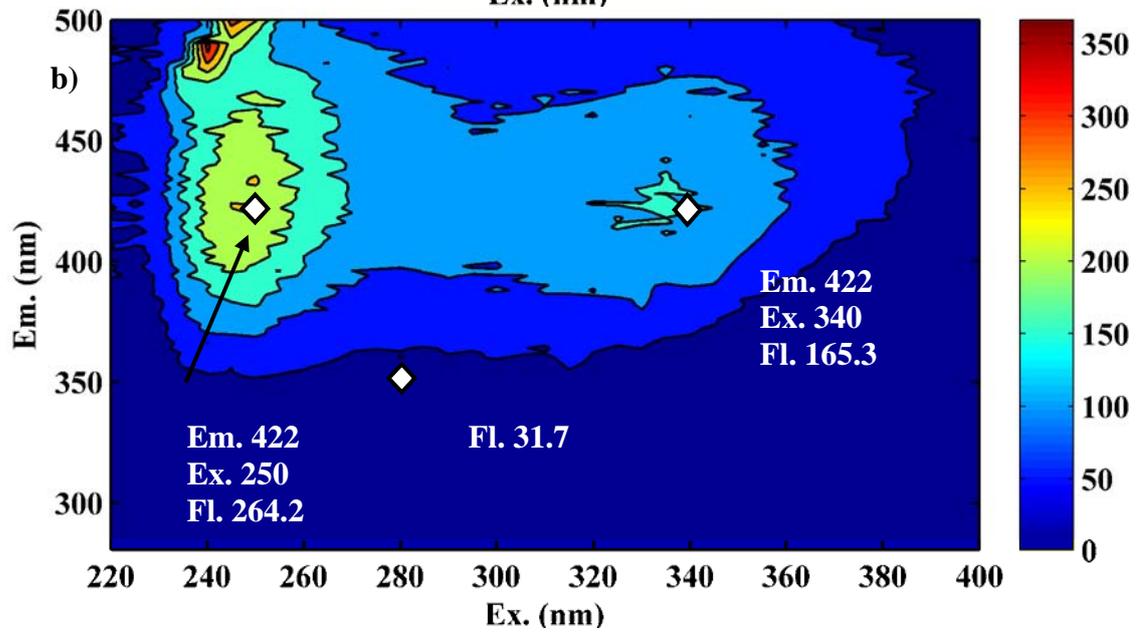
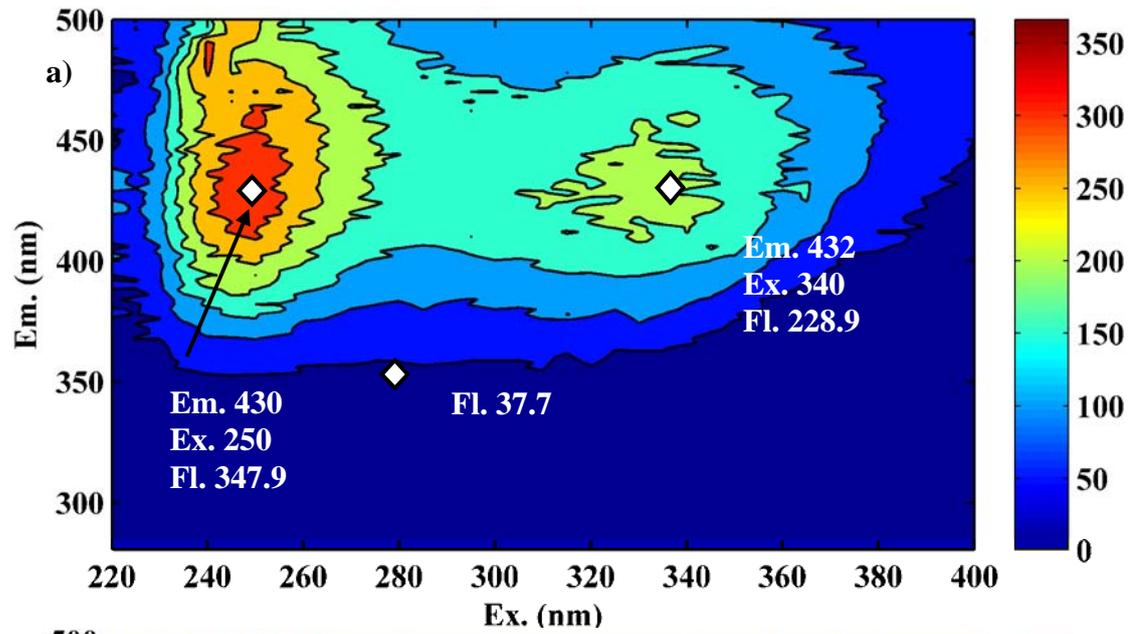
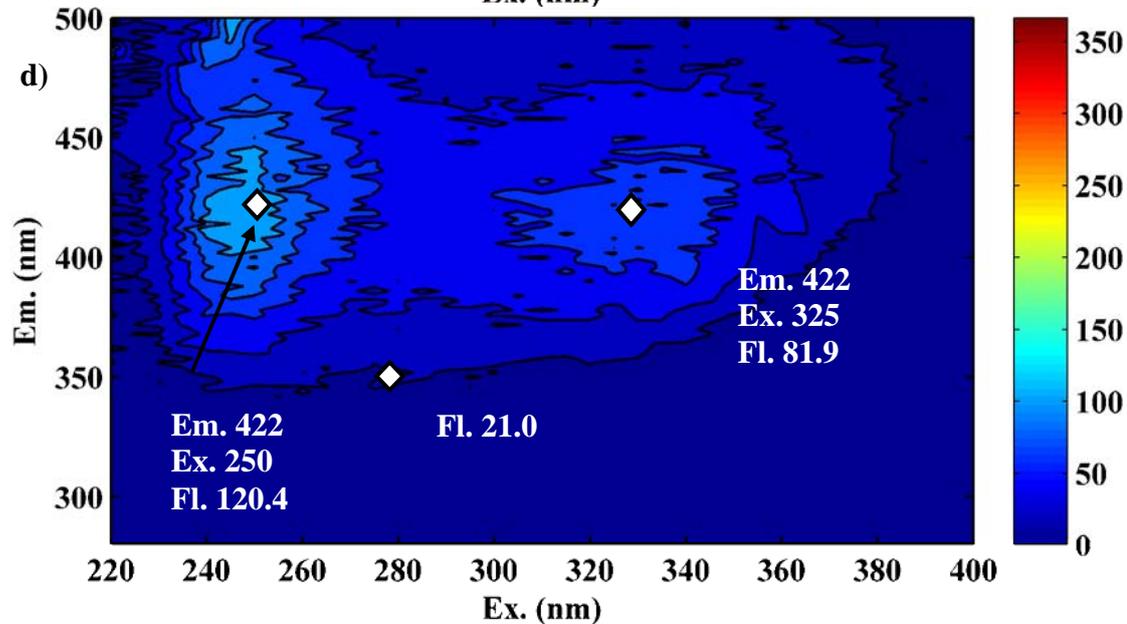
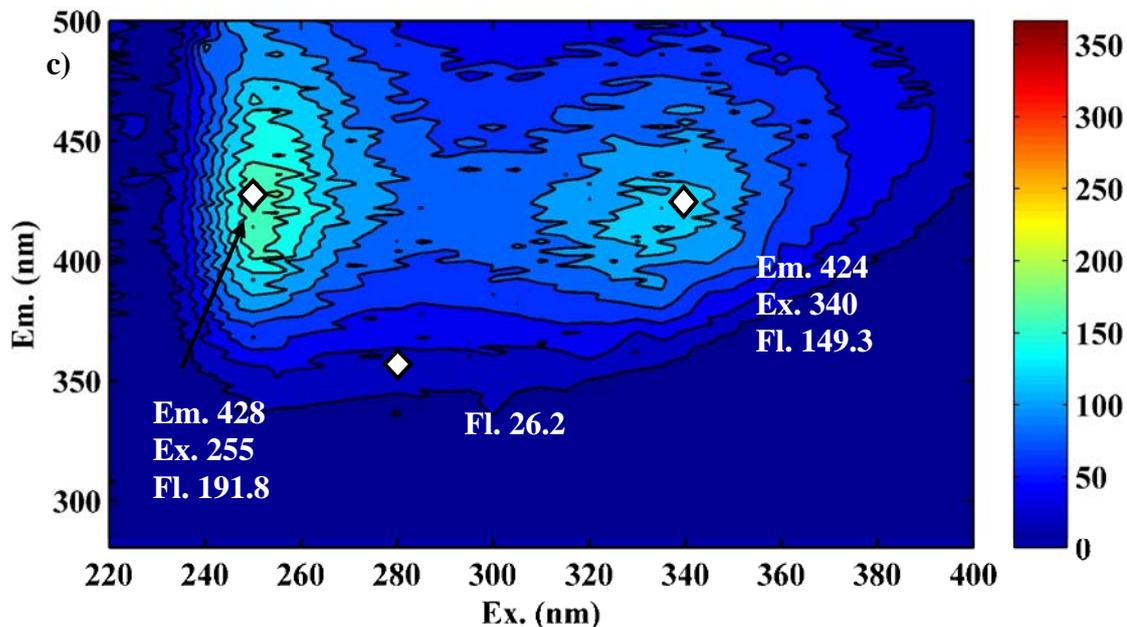
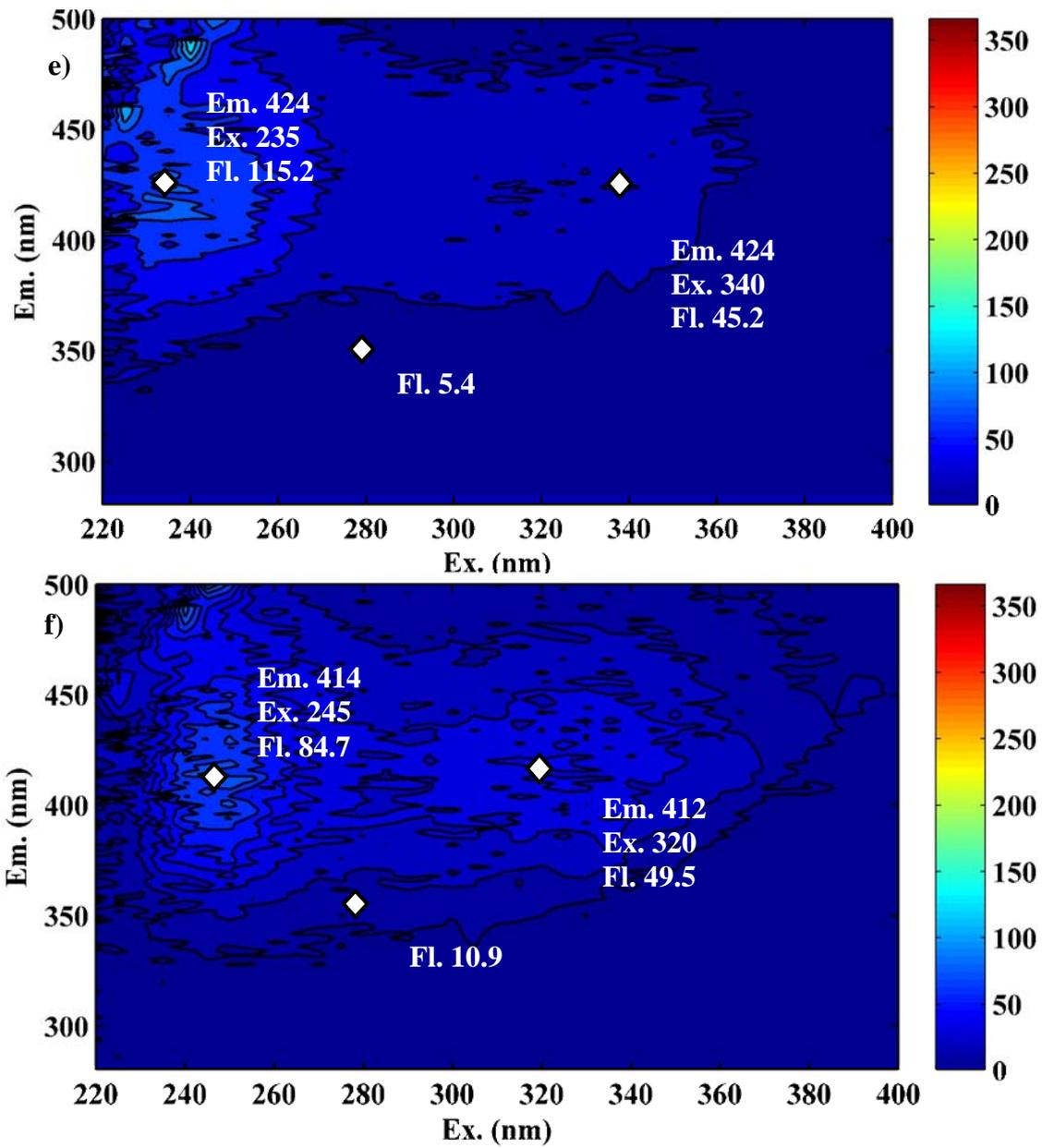


Figure 7.3: Relative change in peak T intensity (%) through different WTWs







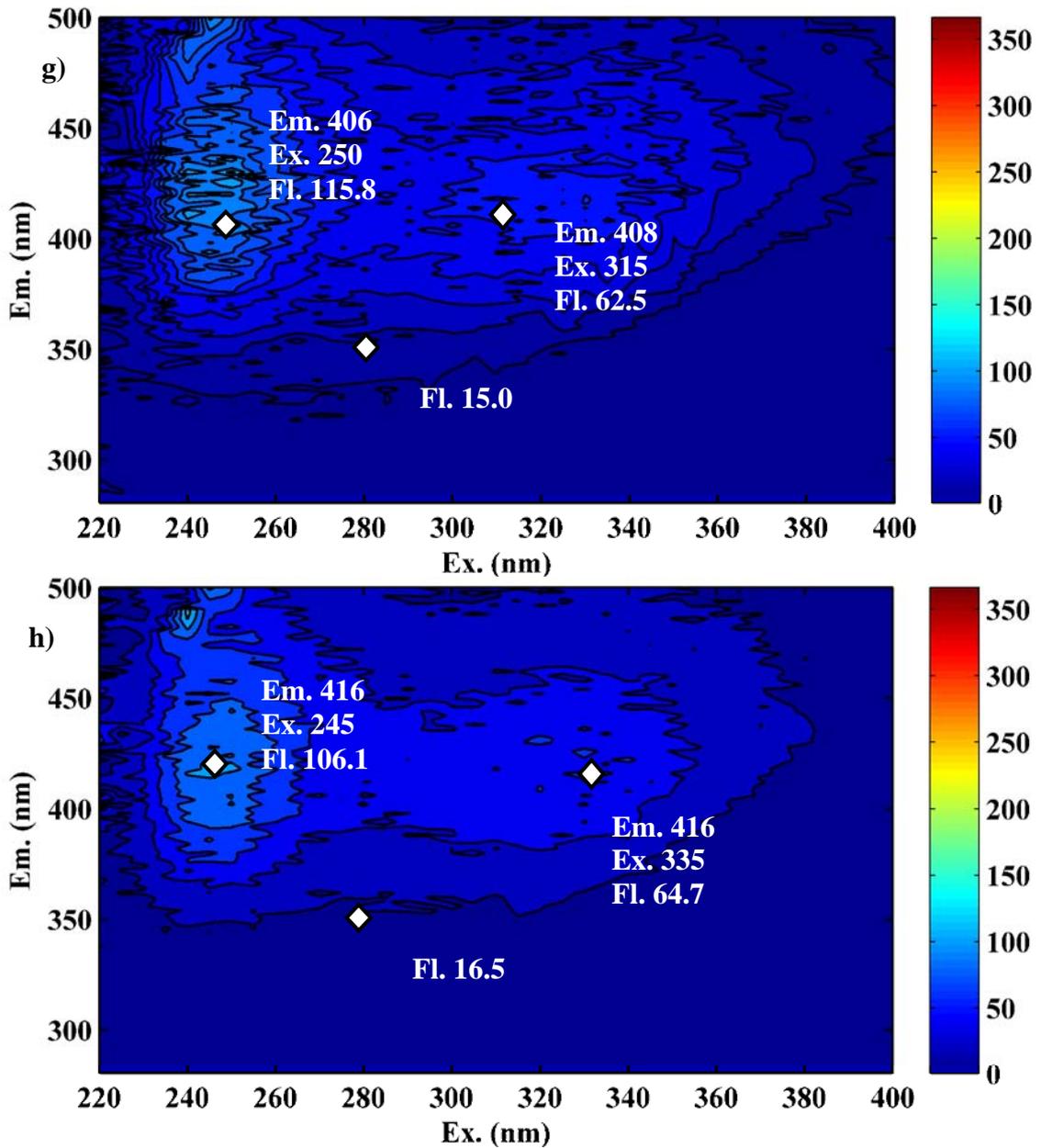


Figure 7.4: Excitation-emission matrices of raw (a), clarified (b), filtered (c), post-GAC water (d), pre-contact tank (e), post contact tank pre de-chlorination (f), post contact tank post de-chlorination (g), and final water (h). Strensham WTW, 05.08.07

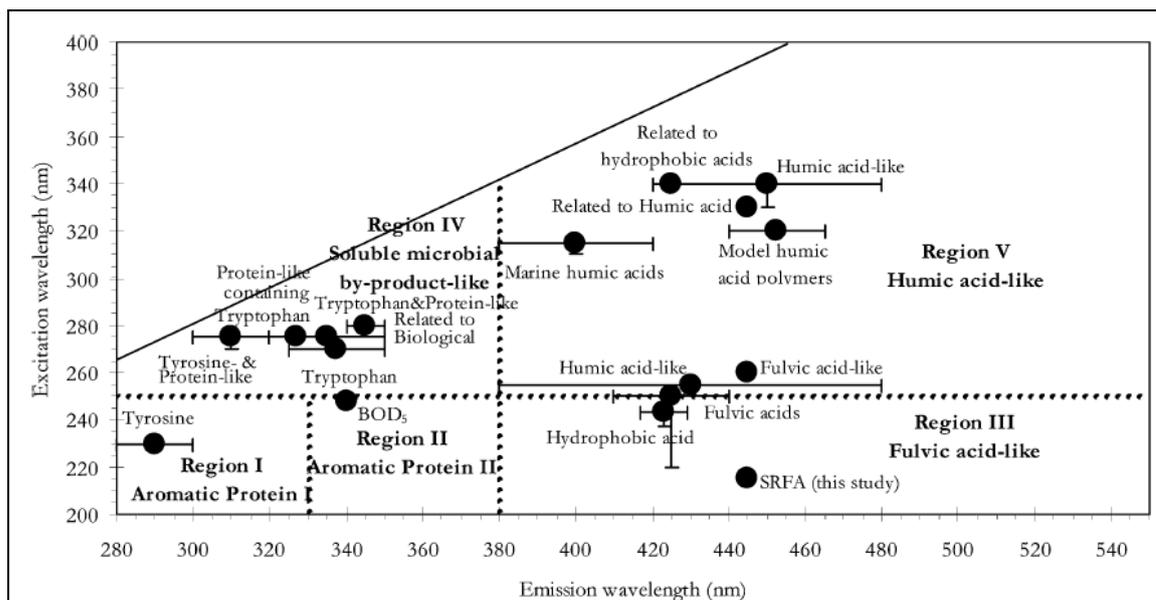


Figure 7.5: Location of EEM peaks and five EEM regions (after Chen et al., 2003)

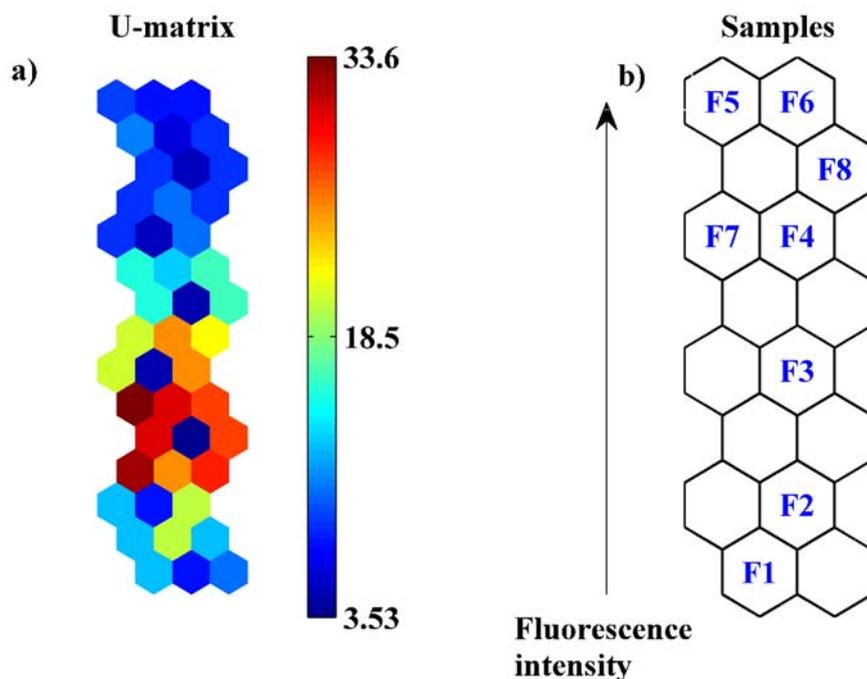


Figure 7.6: SOM analysis of Strensham fluorescence data through WTW. The U matrix (a) and samples distribution (b): raw (F1), clarified (F2), filtered (F3), post-GAC water (F4), pre-contact tank (F5), post contact tank pre de-chlorination (F6), post contact tank post de-chlorination (F7), and final water (F8). Strensham WTW, 05.08.07

7.6. CHAPTER 7 TABLES

Table 7.1: Fluorescence Regional Integration parameters

EEM region	Number of EEM data points per region	Projected excitation-emission area (nm ²)	Fractional projected area per region	M _{fi}
I	182	455	0.06	15.6
II	175	438	0.06	16.2
III	420	1050	0.15	6.8
IV	591	1455	0.21	4.9
V	1716	3696	0.52	1.9
Total	3084	7094	1.00	-

Table 7.2: Relative distribution of normalized fluorescence volumes through Strensham WTW (05.08.07)

Sample	Region					Total fluorescence volume (raw = 100 %)	Relative change (%)
	I	II	III	IV	V		
Raw	1.9	11.2	39.2	9.5	38.2	100.0	-
Clarified	2.2	13.2	38.9	11.2	34.5	69.3	30.7
Filtered	1.7	10.1	28.9	14.1	45.2	48.8	20.4
Post-GAC	2.8	14.9	36.6	12.9	32.8	30.5	18.3
Pre-CT	4.0	19.2	46.6	8.3	21.9	21.5	9.0
Post-CT, pre-SO₂	3.8	18.3	37.3	13.0	27.5	18.6	2.9
Post-CT, post-SO₂	3.1	16.5	38.8	12.0	29.5	26.7	-8.1
Final	3.4	16.0	38.7	12.2	29.7	26.7	0.0

8. DISCUSSION OF THE RESULTS

8.1. RELATING FLUORESCENCE OF RAW AND CLARIFIED WATER TO ORGANIC MATTER REMOVAL EFFICIENCY

Fluorescence spectroscopy has been successfully utilized in environmental studies, providing a robust, holistic and simple source of information on OM function and structure in a variety of environments, from marine (Mopper and Schultz, 1993; Coble, 1996; de Souza-Sierra et al., 1997; Clark et al., 2002; Stedmon et al., 2003; Boehme et al., 2004; Murphy et al., 2008) to freshwater ecosystems (Battin, 1998; Mounier et al., 1999; McKnight et al., 2001; Paterl-Sorrentino et al., 2002; Cammack et al., 2004; Hudson et al., 2007).

Recently, attempts have been made to facilitate the use of fluorescence analysis in water quality studies for deriving surrogates for organic matter concentrations (Cumberland and Baker, 2007; Hudson et al., 2008), and for organic matter characterisation in terms of various physico-chemical properties, i.e. molecular weight and degree of hydrophobicity (Stewart and Wetzel, 1980; Kalbitz et al., 1999; Belzile and Guo, 2006; Baker et al., 2008). However, to date the efficacy of fluorescence spectroscopy has not been evaluated for application in drinking water treatment for the characterisation of OM and its treatment efficiency. Organic matter assessment is a growing concern for water companies responsible for providing reliable and clean drinking water supplies. During the treatment processes insufficient OM removal can result in formation of carcinogenic DBPs as a side effect of reaction with disinfectant. Commonly available techniques of organic matter quantification in drinking water treatment are time- and labour-consuming, therefore alternative and novel approaches, like fluorescence spectroscopy, should be explored and utilized to enable a faster and more robust solution to DBPs detection. The increased interest of drinking water companies in the

potential application of fluorescence spectroscopy to operational monitoring of WTWs results from the method's simplicity, sensitivity, speed of analysis and low cost, combined with potential for incorporation into an on-line monitoring system.

This is the first study summarising the application of fluorescence EEM spectroscopy to characterise organic matter properties and removal in the field of drinking water treatment. In the study, fluorescence spectroscopy was used for the determination of organic matter at 16 WTWs across the Midlands region of the UK. The work concentrated on characterising organic matter removal across different stages of municipal water treatment and providing qualitative information on organic matter properties relevant to treatment processes' efficiency. Specifically, the use of fluorescence spectroscopy as a rapid screening tool for assessing the presence of carcinogenic DBPs was evaluated.

The results obtained indicate that fluorescence spectroscopy can be effectively utilized at WTWs for characterisation of both OM properties and its removal efficiency. Fluorescence EEM spectroscopy enables characterisation of the different OM fractions, including those of key importance to TOC removal efficiency. Fluorescence properties of peak T and peak C were found to characterise OM in spatial and temporal terms and to discriminate between water treatment sites and water types (*Hypothesis H1*, see Section 2.5).

Peak C emission wavelength was found to indicate the presence of a hydrophobic, high-molecular weight OM fraction comprising highly substituted humic acids. The predominance of this fraction can be linked to the upland water sources and low anthropogenic pressure on the catchment (i.e. Bamford, Frankley). Thus, the OM fraction denoted by a high peak C emission wavelength is typical of natural and semi-natural catchments with a high percentage of extensively used pastures, heathland, and wetlands. Bamford and Frankley WTWs demonstrated high peak C emission wavelength, whereas low peak C emission

wavelength can be attributed to the lowland catchments and more aliphatic OM (Draycote, Church Wilne, Melbourne, Campion Hills). In drinking water treatment, hydrophobic OM fraction is preferentially removed by conventional treatment with coagulation. Thus, peak C emission wavelength provides a discrimination between the easier to remove, highly reactive hydrophobic OM and the recalcitrant, less reactive hydrophilic fraction (*Hypothesis H3*).

Conversely, the presence of more difficult to remove, low-molecular weight, hydrophilic acids was related to the peak T (tryptophan-like) and peak C fluorescence at lower emission wavelengths. Peak T fluorescence was also shown to indicate the presence of autochthonous microbial OM, originating from various aquatic microorganisms. Thus, high peak T fluorescence reflects the presence of microbial pollution and is predominant in highly urbanised catchments (Little Eaton, Church Wilne, Cropston, Melbourne).

While tryptophan-like fluorescence indicates the predominance of labile organic material, peak C intensity provides information on the quantity of more recalcitrant, allochthonous OM. Thus, the $T_{2_{int}}/C_{int}$ ratio was found to be a good indicator of overall water quality, with low values for natural upland sources and high values for anthropogenically-impacted lowland waters. Sites with higher raw water peak C emission wavelength and a higher degree of aromaticity comprised more recalcitrant OM as indicated with the low values of $T_{2_{int}}/C_{int}$ ratio.

The effect of OM composition on its removal efficiency was evaluated by correlating the raw water peak C emission wavelength, peak T2 intensity and $T_{2_{int}}/C_{int}$ ratio with OM removal measured as a reduction in TOC between raw and clarified water. It was found that fluorescence properties of clarified water differed significantly compared to corresponding raw waters. The clarified water fluorescence signature demonstrated the character of the OM

fraction recalcitrant to removal by coagulation. The humic-like fluorescence (peak A and C) was removed by coagulation more effectively than the protein-like fluorescence (peak T1 and T2). Thus, the clarified water $T2_{int}/C_{int}$ ratio was higher than corresponding raw water values. The residual fluorophores were more hydrophilic as indicated by lower emission wavelengths of humic-like fluorescence regions. Overall, the clarified water properties were similar for the majority of WTWs except Bamford, Church Wilne and Little Eaton. Bamford exhibited relatively more hydrophobic OM character, whereas Church Wilne and Little Eaton demonstrated higher values of the mean $T2_{int}/C_{int}$ ratio for clarified water.

A good overall relationship was found between peak C intensity and TOC concentrations with the correlation coefficients $R^2 = 0.68$ for raw and $R^2 = 0.77$ for clarified water (*Hypothesis H2*). Furthermore, the decrease in peak C intensity between raw and clarified water was correlated with the OM removal efficiency measured as TOC reduction ($R^2 = 0.91$). Thus, fluorescence measurements can provide a good indication of OM concentrations and therefore enable the rapid determination of OM removal efficiency across the coagulation stage as fluorescence analysis is faster than TOC analysis.

Fluorescence properties were also related to the OM removal efficiency. It was found that OM removal efficiency is positively correlated with raw water peak C emission wavelength and negatively with clarified water peak T2 intensity. This suggests that the OM fraction with a higher degree of aromaticity, denoted by high peak C emission wavelength, is preferentially removed during coagulation. Furthermore, clarified water peak T2 intensity represents the content of a low molecular weight, hydrophilic OM fraction. Thus, sites with higher OM removal efficiency tend to have lower content of hydrophilic OM.

Analysis of fluorescence data with SOM enabled the classification of WTWs on the basis of raw and clarified water OM properties and OM treatment efficiency (Table 8.1). For each

water type, the predominant OM properties were presented based on the interpretation of SOM hit histograms (Figure 5.14). In WTWs classification, four OM fractions of different treatability were characterised: hydrophobic, hydrophilic, microbially-derived, and intermediate OM. Additionally, the concentration of OM of raw and clarified water and the overall OM removal were also characterised. Five groups of WTWs with varying OM properties of raw and clarified water were selected. Groups 1 to 3 demonstrated stable raw water OM properties, whereas groups 4 and 5 were characterised by variable raw water properties. Group 1 comprised sites with stable raw water OM character and concentrations (hydrophobic and high TOC) and similar stable hydrophobic clarified water OM with varying TOC concentrations (Bamford, Frankley, Tittesworth). The OM removal efficiency for this group was the highest within the entire dataset. Tittesworth WTW demonstrated OM properties characteristic for group 1, however the raw and clarified water OM was less hydrophobic and was classified as intermediate OM. The opposite OM properties were typical of group 2 (Cropston, Draycote, Melbourne, Whitacre). For this group OM character of raw and clarified waters was hydrophilic, with varying TOC concentrations and predominantly low OM removal efficiency. Group 3 exhibited similar OM character with additional high contribution of microbially-derived OM in both raw and clarified water (Church Wilne) or solely in clarified water (Ogston). Group 4 comprised sites with variable raw water OM character and TOC concentrations (raw and clarified), and stable hydrophilic OM for clarified water (Campion Hills, Little Eaton). Furthermore at Little Eaton WTW, the OM demonstrated a highly microbial character. Finally, group 5 (Mitcheldean, Mythe, Shelton, Strensham, Trimpley) showed variable raw water OM character with a variable or high TOC concentration (Shelton, Trimpley) and stable, intermediate clarified water OM properties with

varying (Mythe, Strensham) or stable intermediate concentrations. The OM removal efficiency in this group was intermediate.

8.2. RELATING ORGANIC MATTER FLUORESCENCE TO DISINFECTION- BY PRODUCTS FORMATION

It was hypothesised that fluorescence spectroscopy can be effectively used as a tool to determine the formation and prevalence of DBPs in water treatment works (*Hypothesis H4*). It can be corroborated that the poorer OM removal by coagulation implies the greater residual OM and DBPs concentrations in post-clarification waters. Thus, for WTWs with high OM removal (group 1, Bamford 72.7 %, Frankley 58.0 % and Tittesworth 60.2 %) the expected DBPs formation is lower compared to WTWs with low OM removal (group 2 and 3, Melbourne 15.3 %, Draycote 17.3 %, Whitacre 17.7 %, Church Wilne 18.3 %). However, different OM fractions were found to demonstrate different reactivity with chlorine. In general, the hydrophobic OM fraction is more reactive compared to the hydrophilic OM. Consequently, the DBPs yields are significantly higher for the hydrophobic OM fraction. Moreover, the hydrophobic fraction was found to be a more significant HAA precursor than the corresponding hydrophilic fraction that plays a major role in THMs formation (Stevens et al., 1976; Reckhow et al., 1990; Gang et al., 2002; Liang and Singer, 2003; Parsons et al., 2004). Thus, it can be elucidated that group 1 (Bamford, Frankley, Tittesworth) demonstrates the presence of highly reactive DBPs precursors. However, very efficient OM removal by coagulation of this OM fraction significantly reduces the risk of exceeding the permitted DBPs concentrations. The greatest potential for increased levels of DBPs post-clarification can be attributed to groups 2, 3, and 4 with predominantly hydrophilic OM character. Poor OM removal (below 20 %) at these sites indicates that the high content of residual OM can

contribute to DBPs formation. Furthermore, Little Eaton and Campion Hills (group 4) demonstrate variable raw water OM character, which impedes the optimisation of OM removals and poses a greater risk of non-compliance with DBPs regulations.

The raw water potential for formation of different DBPs was determined during the independent measurements carried out in October 06, April 07, June 07, October 07, and January 08 by Severn Trent Water Ltd. (Table 8.2). Although the DBPs analyses were carried out on raw water samples taken quarterly, this ancillary dataset enabled a better insight into linking fluorescence OM properties and DBPs formation. The wider characterisation of NOM and its links to DBPs is the focus of another project (Roe, PhD in progress) and preliminary characterisation results can be found in Roe et al. (2008).

For all 16 WTWs formation potentials of chloroform (CHCl_3), dichlorobromomethane (CHCl_2Br), dibromochloromethane (CHBr_2Cl), bromoform (CHBr_3), total THMFP (TTHMFP) and total HAAFP (THAAFP) were determined. As it was suggested before, WTWs from group 1 demonstrated the highest TTHMFP and THAAFP values (Bamford 830.6 $\mu\text{g/l}$ and 1013.7 $\mu\text{g/l}$, Tittesworth 455.0 $\mu\text{g/l}$ and 441.9 $\mu\text{g/l}$ respectively). High DBPs formation potentials were also shown by Cropston (group 2, TTHMFP = 410.9 $\mu\text{g/l}$, THAAFP = 473.79 $\mu\text{g/l}$), Whitacre (group 2, TTHMFP = 307.3 $\mu\text{g/l}$, THAAFP = 354.7 $\mu\text{g/l}$), Trimpley (group 5, TTHMFP = 305.2 $\mu\text{g/l}$, THAAFP = 302.0 $\mu\text{g/l}$), Shelton (group 5, TTHMFP = 303.9 $\mu\text{g/l}$, THAAFP = 291.0 $\mu\text{g/l}$), and Mythe (group 5, TTHMFP = 301.6 $\mu\text{g/l}$, THAAFP = 225.9 $\mu\text{g/l}$). For Bamford, Cropston, Whitacre WTWs the THAAFP values were higher compared to the corresponding TTHMFP. It can be noticed that the highest DBPs formation potentials were typical of WTWs from group 1, 2 and 5, exhibiting different raw water fluorescence properties. This indicates that the OM character is a less important predictor of DBPs formation compared to the quantitative OM properties. A good correlation

was found between the TTHMFP and THAAFP values and raw water TOC concentrations for Bamford, Cropston, Tittesworth, Whitiacre, Trimpley, and Shelton WTWs (Table 3.2).

The lowest values of TTHMFP and THAAFP can be observed for sites with the lowest raw water TOC concentrations: Little Eaton (group 4, TTHMFP = 220.4 and THAAFP = 166.8), Mitcheldean (group 5, TTHMFP = 216.4 and THAAFP = 129.7), Melbourne (group 2, TTHMFP = 215.8 and THAAFP = 225.5), and Church Wilne (group 3, TTHMFP = 199.1 and THAAFP = 137.5) (Table 8.2).

The combined effect of OM quality and quantity on DBPs formation potential can be observed for Frankley WTW (Tables 3.2 and 8.2). For this WTW, raw water demonstrated high peak C emission wavelengths and low peak T fluorescence intensities indicating the predominance of a highly reactive with chlorine, high-molecular weight, hydrophobic OM fraction. However, the raw water TOC concentrations were very low (2.8 mg/l). Overall, Frankley WTW showed intermediate TTHMFP and THAAFP values as a result of low concentrations of highly reactive DBPs precursors in raw water.

As OM concentration in raw water was found to be the most important predictor of DBPs formation potentials, the OM quality controls the contribution of individual THMs species. For WTWs with highly hydrophobic OM (group 1, Bamford, Frankley, Tittesworth) most of TTHMFP resulted from the contribution of chloroform (97.9 %, 95.3 %, and 94.4 % respectively). Likewise, for WTWs with distinctively hydrophilic raw water OM the contribution of chloroform was lower (Church Wilne 76.3 %, Draycote 76.6 %, Whitacre 77.0 %) and the importance of other THMs species (dichlorobromomethane, dibromochloromethane) increased. These results suggest that raw water peak C emission wavelength can be a predictor of chloroform formation, whereas peak T intensity can be hypothesised to indicate the formation of other THMs species. To test this hypothesis, simple

Pearson's correlations were determined for various fluorescence and absorbance parameters and DBPs formation potentials (Table 8.3). The best predictors of chloroform formation were raw water UV measured at 254 nm (UV_{254}) ($R^2 = 0.73$), raw water SUVA ($R^2 = 0.65$), OM removal peak C ($R^2 = 0.54$), and raw water peak C emission wavelength ($R^2 = 0.46$). As chloroform is the main THM compound, the same type and order of predictors were found valid for TTHMFP prediction with the correlation coefficients $R^2 = 0.72$, $R^2 = 0.60$, $R^2 = 0.48$ and $R^2 = 0.41$ respectively. These results suggest that absorbance-derived parameters (UV_{254} , SUVA) provide a better indication of chloroform and TTHMFP in raw water compared to fluorescence properties. However, for the other THMs compounds and THAAFP, fluorescence provided similar (THAAFP) or better formation prediction (dibromochloromethane, dichlorobromomethane, bromoform). OM removal peak A ($R^2 = 0.70$), OM removal UV_{254} ($R^2 = 0.69$), OM removal peak C and T2 ($R^2 = 0.68$), and clarified water peak T2 intensity ($R^2 = 0.64$) were found to be the best predictors for dibromochloromethane. For dichlorobromomethane, the second important THM compound, several fluorescence properties related to the tryptophan-like fluorescence were found to be the best predictors (peak T2 intensity clarified ($R^2 = 0.72$), peak T1 intensity raw ($R^2 = 0.68$), peak T1 intensity clarified ($R^2 = 0.67$)). None of the absorbance-derived surrogates correlated well with the formation of bromoform and only one significant correlation was found with raw water $T2_{int}/C_{int}$ fluorescence ratio ($R^2 = 0.47$). The best predictor of THAAFP was UV_{254} raw ($R^2 = 0.73$), however correlations was also found with SUVA raw, OM removal peak C (both $R^2 = 0.50$), and $T2_{int}/C_{int}$ ratio raw ($R^2 = -0.49$).

From these results it can be inferred that WTWs with high clarified peak T2 intensity demonstrate higher concentrations of dichlorobromomethane and dibromochloromethane OM precursors (i.e. Whitacre, Church Wilne and Draycote; Figure 8.1). High raw water peak

C emission wavelength reflects higher yields of chloroform and TTHMFP (Bamford, Frankley, Tittesworth). High raw water $T_{2_{int}}/C_{int}$ ratio characterises WTWs with a higher contribution of bromoform (Whitacre, Melbourne), whereas lower values of $T_{2_{int}}/C_{int}$ ratio highlight sites with higher values of THAAFP (group 1).

As the application of fluorescence spectroscopy for OM characterisation in water treatment is a novel approach, only a small number of studies reports the results of fluorescence spectra correlation with DBPs (Yang et al., 2008; Johnstone and Miller, 2009; Roccaro et al., 2009). Furthermore, the application of fluorescence regional integration technique (Yang et al., 2008; Johnstone and Miller, 2009) and empirically-defined parameters (Roccaro et al., 2009) for fluorescence spectra quantification impedes the direct comparison with the results obtained herein for peak fluorescence values. Yang et al. (2008) correlated the cumulative EEM volumes at regions II and IV (protein-like fluorescence) with the yields of dichloroacetic acid (DCAA, $R^2 = 0.60$), chloroform ($R^2 = 0.42$), dichloroacetonitrile (DCAN, $R^2 = 0.53$), and TOX ($R^2 = 0.63$). The corresponding correlations with SUVA were higher (DCAA $R^2 = 0.82$, chloroform $R^2 = 0.73$, DCAN $R^2 = 0.88$ and TOX $R^2 = 0.80$). These results were obtained for treated waters disinfected with monochloramine (chloramination). Monochloramine as a disinfectant demonstrates lower oxidation strength compared to free chlorine and longer residuals in disinfected water (Goslan et al., 2009). Wu et al. (2003) showed that small amounts of free chlorine in equilibrium with monochloramine react preferentially with aromatic moieties in humic substances. Therefore in the study of Yang et al. (2008), poorer correlation with DBPs observed for fluorescence spectra compared to SUVA results from lower reactivity of protein-like fluorophores towards monochloramine.

In studies of Johnstone and Miller (2009) and Roccaro et al. (2009) raw riverine water was subjected to disinfection with chlorine. Johnstone and Miller (2009) found that fluorescence

in region II (simple aromatic proteins) correlated with chloroform yields ($R^2 = 0.78$), region IV (soluble microbial-like OM) correlated with DCAA ($R^2 = 0.79$), and the combination of region III (humic-like fluorescence with UV emission) and IV was found to correlate with trichloroacetic acid ($R^2 = 0.87$). Likewise, Roccaro et al. (2009) found strong correlations between fluorescence index $\lambda^{\text{em}}_{0,5}$ defined as the position of the normalized emission band at its half-intensity for the fixed excitation wavelength of 320 nm and chloroform ($R^2 = 0.80$), dichlorobromomethane ($R^2 = 0.79$), and dibromochloromethane ($R^2 = 0.72$).

More recently, Beggs et al. (2009) correlated PARAFAC components with DBPs formation. They found that 11 of the 13 PARAFAC components showed a strong negative correlation with DBPs concentrations over the chlorination. The authors concluded that fluorescence can be a useful tool in tracking both OM oxidation and DBPs formation during chlorination.

Overall, the results presented in this study relating tryptophan-like fluorescence to formation of other than chloroform THMs compounds are in accordance with values demonstrated by Roccaro et al. (2009). However no correlation was found between raw water tryptophan-like fluorescence and chloroform. Furthermore, tryptophan-like fluorescence was a poorer predictor of chloroform formation compared to peak C emission wavelength ($R^2 = -0.26$ and 0.46 respectively). Differences in chloroform prediction can result from different fluorescence parameters used (fluorescence peak values vs FRI technique) and the size of dataset used for calibration (here just five samples per WTW were analysed).

These findings demonstrate that quantity of OM post-clarification is the most important factor affecting the concentrations of DBPs formed during the disinfection stage. Thus, quantitative fluorescence properties of clarified water can provide invaluable information on OM removal efficiency and the potential DBPs levels in treated waters. In particular peak T2 fluorescence intensity was found to be a good indicator of OM removal ($R^2 = -0.55$), OM

treatability (indicates the low molecular-weight, hydrophilic and microbial OM) and reactivity with chlorine (i.e. formation of CHCl_2Br ; $R^2 = 0.72$). Therefore, a simple fluorescence-TOC removal model based on tryptophan-like fluorescence could be of practical operational use at WTWs. The TOC removal prediction accuracy of this model is poorer than derived for a model incorporating both raw and clarified fluorescence predictors, but could be more practical and feasible at WTWs. A simple fluorescence probe could be installed online to measure tryptophan-like fluorescence of clarified water and therefore be used as a TOC removal and DBPs formation monitor.

Furthermore, several fluorescence properties were demonstrated to exhibit a good correlation with particular DBPs species. Peak C emission wavelength was found to be a good indicator of OM aromaticity, treatability with coagulation, and DBPs formation, particularly chloroform. Likewise, $T_{2\text{int}}/C_{\text{int}}$ fluorescence ratio of raw water demonstrated a proportion of poorly removed tryptophan-like fluorescence to more readily removed peak C fluorescence. Consequently, $T_{2\text{int}}/C_{\text{int}}$ fluorescence ratio discriminates between OM fractions of different treatability and reactivity with chlorine. Moreover, a strong correlation was found between $T_{2\text{int}}/C_{\text{int}}$ and several DBPs species (CHBr_2Cl , CHBr_3 , CHCl_2Br , HAAFP).

In the related study, Roe et al. (2008) attempted to correlate various OM properties derived from isolation techniques with DBPs formation (carried out on the same raw water samples as presented in this study). They showed that hydrophobicity related to chloroform formation which is in accordance with findings in this study. However the authors did not find an OM predictor of dichlorobromomethane which was correlated in this study with peak T fluorescence. These results suggest that fluorescence spectroscopy can yield additional information on OM reactivity compared to other OM characterisation techniques and show the ability of fluorescence peak T at detecting trace amounts of a relevance OM component

(ppb levels of OM). Although peak T fluorescence constitutes a small fraction of the OM pool related to microbial OM, fluorescence spectroscopy enables its accurate detection.

8.3. COMPARISON BETWEEN FLUORESCENCE EEM SPECTROSCOPY AND STANDARD ORGANIC MATTER CHARACTERISATION TECHNIQUES

TOC and UV-Vis absorption are commonly used in drinking water treatment to evaluate the works' performance in OM removal efficiency. Furthermore, several UV-Vis derived parameters have been proposed as surrogate parameters for characterisation of the OM reactivity and treatability, with the UV_{254} and SUVA being the most commonly utilized. TOC measurements provide valuable information on the overall quantity of OM present in water but it does not characterise its reactivity and treatability.

Thus, in this study an application of these standard OM characterisation techniques was compared with the novel approach of fluorescence spectroscopy. OM characterisation with fluorescence spectroscopy was hypothesised to provide additional information on OM quantity, reactivity and treatability compared to TOC and UV-Vis analyses (*Hypothesis H7*).

Fluorescence EEM spectroscopy was found useful in the characterisation of different OM properties and discrimination between different OM fractions. While the UV absorbance technique can only determine the content and properties of the aromatic functional groups in OM, fluorescence can additionally characterise the presence of a low molecular weight, hydrophilic fraction (protein-like fluorescence). This autochthonous OM fraction was corroborated to play an important role in aquatic biochemistry and indicate the presence of microbial pollution in water quality studies. Here, fluorescence intensity of peak T was related to the presence and concentrations of different algae species. Tryptophan-like fluorescence along with peak C emission wavelength (surrogate for degree of hydrophobicity) was found

to provide the best discrimination between different WTWs and water types (raw and clarified). Peak C emission wavelength was found to be equally good to SUVA surrogate parameter for OM aromaticity. From the raw water predictors, solely peak C emission wavelength and SUVA parameter correlated with TOC removal (both $R^2 = 0.36$). The relative contribution of peak T and C fluorophores ($T_{2_{int}}/C_{int}$ fluorescence ratio) was shown to demonstrate the proportion of a labile, low molecular-weight recalcitrant to removal by coagulation OM fraction to a more readily removed, high molecular weight humic-like fraction. Thus, fluorescence spectroscopy was shown to provide an indication of OM origin, physico-chemical properties, reactivity and treatability in conventional drinking water treatment with coagulation, additional to the information obtained from TOC and UV-Vis absorbance analyses.

The OM removal performance between raw and clarified water samples was assessed with the reduction in TOC concentrations. Fluorescence peak C and peak A intensities were shown to be stronger TOC/OM removal predictors compared to UV-Vis absorbance ($R^2 = 0.91, 0.84$ and 0.82 respectively). Thus, in comparison with fluorescence spectroscopy, UV-Vis measurements are prone to underestimation of the OM samples where a low molecular weight fraction is present. These results suggest that fluorescence intensities of humic-like fluorescence provide a better indication of TOC removal compared to absorbance-derived parameters. UV absorbance seems to be a less selective technique and tends to overestimate the OM removal as only the easier to remove by coagulation, hydrophobic fraction is accounted for by absorbance measurements.

UV_{254} and SUVA are commonly used in drinking water treatment for characterisation of OM reactivity with disinfectant towards formation of harmful DBPs. Here, fluorescence spectroscopy was shown to provide DBPs formation prediction of similar accuracy.

Fluorescence was a poorer predictor of chloroform and THMFP but outperformed UV₂₅₄ and SUVA in determination of bromoform, dichlorobromomethane, and dibromochloromethane.

Fluorescence spectroscopy was demonstrated to characterise various qualitative and quantitative OM properties of greatest importance to OM removal efficiency and formation of DBPs. Compared to standard OM characterisation tools routinely measured at WTWs, fluorescence spectroscopy can enable a more comprehensive OM description in terms of its reactivity and treatability. Thus, fluorescence spectroscopy should be considered as an additional and independent OM characterisation tool in drinking water treatment.

8.4. COMPARISON BETWEEN DIFFERENT DATA MINING TECHNIQUES FOR FLUORESCENCE DATA ANALYSIS

Fluorescence excitation-emission spectra contain a substantial amount of information on organic matter characteristics, however special computational and statistical techniques are required to pre-process the data, remove noise and redundant features, like Raman and Rayleigh scatter, and to distinguish patterns of interest from a matrix background. To identify significant features or patterns from multivariate, high-dimensional fluorescence space, a dimensionality reduction, data projection and feature extraction method should be employed.

To facilitate fluorescence data reduction to the most important fluorescence properties and analyse the relationship between fluorescence predictors and TOC removal for 16 WTWs, three different decomposition algorithms (principal components analysis (PCA), parallel factor analysis (PARAFAC), and self-organizing map (SOM)) were evaluated and compared with the standard peak-picking approach in calibration tests (*Hypothesis H5 and H6*). The visual inspection of fluorescence EEMs and peak-picking approach suggested the presence of three main fluorophores in raw and partially-treated water samples: fulvic-, humic-, and

tryptophan-like fluorescence. Components derived from the decomposition algorithms were in accordance with these results. However, the PARAFAC model revealed the presence of more potential fluorophores that could not have been validated for the entire dataset or specific water type (raw and clarified). Therefore, to derive a valid PARAFAC model containing all variations in fluorescence spectra, a solution with fewer components had to be chosen. A lack of good, overall diagnostic for the selection of the number of valid components impedes an interpretation of the PARAFAC model and makes analysis a time-consuming process. Compared to the PARAFAC approach, components derived from the PCA analysis were more difficult to be identified as particular fluorophores on the basis of loadings interpretation. Thus, the PCA algorithm can be successfully used in the initial fluorescence data analysis to provide an insight into data variation and distribution. However, a more advanced analysis requires the prediction of fluorophores, their importance and relative concentration facilitated by the PARAFAC analysis. When the fluorophores composition is uniform between samples and sites, standard peak-picking and the SOM analysis can successfully outperform lengthy PARAFAC modelling. While the peak-picking approach facilitates basic fluorescence data analysis, the SOM model enables advanced interpretation of fluorescence data, samples distribution between sites and water types, and TOC removal-organic matter relationships. The calibration of TOC removal with fluorescence data decomposed with PARAFAC and SOM confirm the similar ability of those models in extraction of the most important fluorescence features.

Although EEM peak-picking analysis is not a robust solution for the analysis of large fluorescence datasets, it is the easiest tool for rapid comparison between a small number of EEMs. Here, analysis of peak fluorescence was applied for analysis of fluorescence changes during the coagulation optimisation (Chapter 6) and through WTWs (Chapter 7).

Fluorescence changes through WTWs were also analysed with the FRI and SOM techniques. In fluorescence analysis the FRI technique is commonly used to quantify fluorescence changes. The fluorescence spectra are integrated over arbitrary defined regions that may contain different fluorophores. Thus, for analysis of particular fluorophores, the peak-picking technique seems to be more selective and accurate. Likewise, the SOM analysis did not provide any additional information on OM properties and simply confirmed the results obtained with the peak-picking approach. As SOM implementation involved tedious data pre-processing, for the purpose of small fluorescence dataset analysis, the pick-picking approach is more effective.

8.5. CHAPTER 8 FIGURES

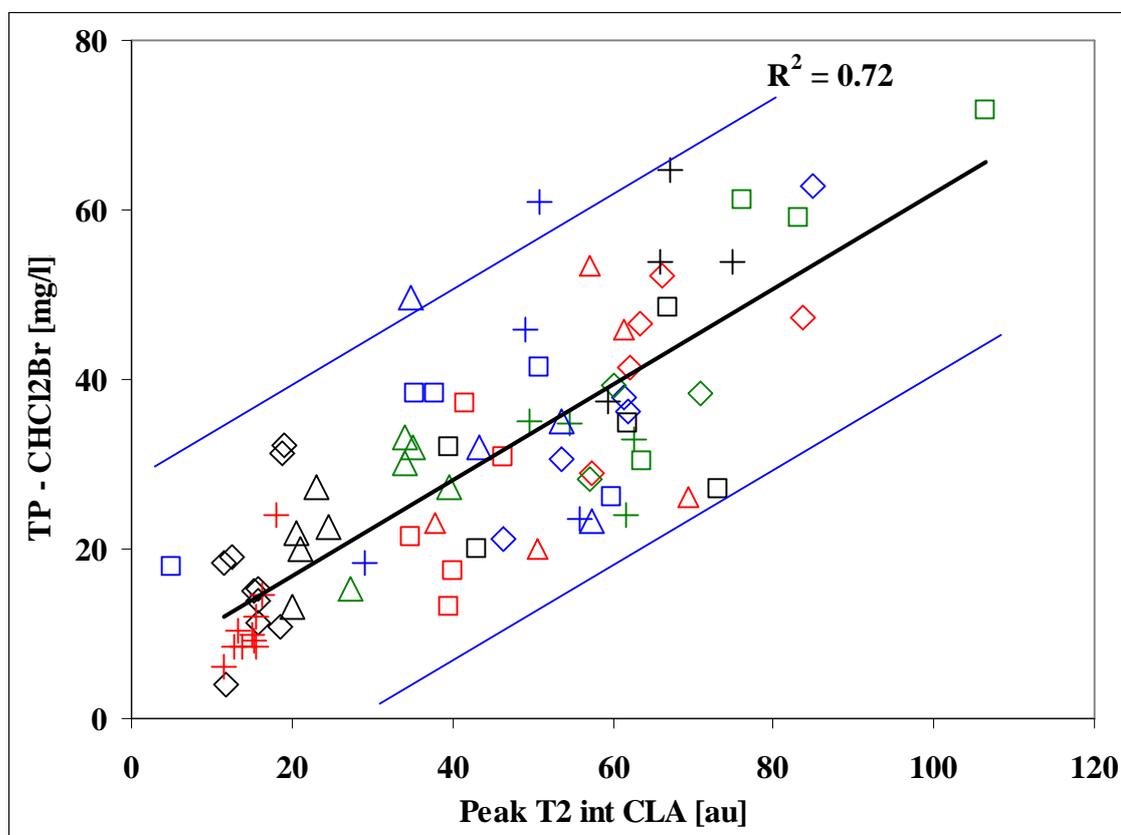


Figure 8.1: The relationship between clarified water peak T2 intensity and formation potential of dichlorobromomethane. For explanation of the labels see Figure 4.2

8.6. CHAPTER 8 TABLES

Table 8.1: OM properties of raw and clarified water and its removal (TOC removal %). HPO – hydrophobic OM, HPI - hydrophilic OM, Int. – intermediate OM character, Micr – microbially-derived OM, Stab. – stable OM properties, Var. – variable OM properties

Group No	Site	Raw				Clarified				TOC
		Character	Concentration	Character	Concentration	Character	Concentration	Character	Concentration	
1	Bamford	Stab.	HPO	Stab.	High	Stab.	HPO	Var.		High
4	Campion Hills	Var.		Var.		Stab.	HPI	Var.		Low
3	Church Wilne	Stab.	HPI, Micr	Var.		Stab.	HPI, Micr	Stab.	Low	Low
2	Cropston	Stab.	HPI	Var.		Stab.	HPI	Stab.	Int.	Low
2	Draycote	Stab.	HPI	Var.		Stab.	HPI	Var.		Low
1	Frankley	Stab.	HPO	Stab.	High	Stab.	HPO	Var.		High
4	Little Eaton	Var.		Var.		Stab.	HPI, Micr	Var.		Low
2	Melbourne	Stab.	HPI	Var.		Stab.	HPI	Var.		Low
5	Mitcheledean	Var.		Var.		Stab.	Int.	Stab.	Int.	Int.
5	Mythe	Var.		Var.		Stab.	Int.	Var.		Int.
3	Ogston	Stab.	HPI	Var.		Stab.	HPI, Micr	Var.		Int.
5	Shelton	Var.	Int.	Stab.	High	Stab.	Int.	Stab.		Int.
5	Strensham	Var.		Var.		Stab.	Int.	Var.		Int.
1	Tittesworth	Stab.	Int.	Stab.	High	Stab.	Int.	Stab.	Int.	High
5	Trimpley	Var.		Stab.	High	Stab.	Int.	Stab.	Int.	Int.
2	Whitacre	Stab.	HPI	Var.		Stab.	HPI	Var.		Low

Table 8.2: Mean disinfection by-products formation potentials ($\mu\text{g/l}$) in raw water

Site	CHCl_3		CHBr_2Cl		CHBr_3		CHCl_2Br		Total THMFP		Total HAAFP	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Bamford	812.8	317.8	0.1	0.0	0.1	0.0	17.6	8.8	830.6	326.4	1013.7	855.9
Campion Hills	205.8	92.6	8.8	4.1	0.8	1.1	43.3	8.9	258.7	97.5	287.5	213.8
Church Wilne	152.0	72.3	8.9	4.1	0.3	0.5	37.8	15.5	199.1	88.4	137.5	105.0
Cropston	372.7	152.2	2.8	0.5	0.1	0.0	35.3	6.2	410.9	157.1	473.7	439.2
Draycote	181.5	89.9	10.7	5.8	0.2	0.3	44.4	20.6	236.9	114.7	273.6	169.0
Frankley	247.2	81.8	0.1	0.1	0.2	0.3	12.0	7.1	259.5	87.4	273.9	255.7
Little Eaton	173.7	99.7	9.4	7.8	0.1	0.0	37.3	17.2	220.4	96.4	166.8	139.3
Melbourne	181.8	52.0	8.0	6.8	0.6	1.1	25.4	14.9	215.8	58.8	225.5	172.8
Mitcheldean	191.8	115.2	3.1	2.6	0.5	0.9	21.0	5.2	216.4	115.9	129.7	95.7
Mythe	261.9	138.1	5.2	5.3	0.8	1.0	33.7	15.0	301.6	141.7	225.9	171.9
Ogston	253.8	56.6	5.6	4.7	0.1	0.0	35.1	10.9	294.6	46.3	244.9	159.3
Shelton	273.4	142.5	2.7	2.6	0.2	0.3	27.5	7.2	303.9	143.7	291.0	241.4
Strensham	243.8	135.0	5.1	4.7	0.3	0.4	32.4	10.5	281.6	132.7	263.8	180.8
Tittesworth	429.6	167.4	1.0	1.0	0.5	0.9	23.9	9.9	455.0	176.7	441.9	292.9
Trimpley	268.2	103.1	3.9	2.2	0.7	0.8	32.5	10.0	305.2	109.7	302.0	248.7
Whitacre	236.5	88.4	12.8	4.0	0.8	1.2	57.1	15.7	307.3	106.3	354.7	216.6

Table 8.3: Disinfection by-products ($\mu\text{g/l}$) formation potential prediction from raw and clarified water fluorescence and absorbance properties. Significant correlations with coefficients > 0.36 highlighted in red

Predictor	CHCl_3	CHBr_2Cl	CHBr_3	CHCl_2Br	Total THMFP	Total HAAFP
T1 int Raw	-0.20	0.58	0.05	0.69	-0.14	-0.33
T2 int Raw	-0.09	0.41	0.11	0.61	-0.03	-0.28
T int Raw	-0.26	0.60	0.08	0.69	-0.19	-0.28
A em Raw	0.06	-0.03	0.16	-0.05	0.06	0.34
A ex Raw	-0.34	0.42	0.17	0.41	-0.29	-0.05
A int Raw	0.27	-0.02	-0.01	0.27	0.29	0.13
C em Raw	0.46	-0.47	-0.09	-0.43	0.41	0.36
C ex Raw	0.17	-0.23	-0.13	-0.23	0.14	0.18
C int Raw	0.31	-0.03	0.01	0.25	0.33	0.26
T2/C int Raw	-0.35	0.48	0.47	0.42	-0.31	-0.49
T1 int Cla	-0.25	0.62	0.01	0.67	-0.18	-0.31
T2 int Cla	-0.27	0.63	0.08	0.72	-0.20	-0.34
T int Cla	-0.29	0.64	0.08	0.72	-0.22	-0.34
A em Cla	-0.08	0.02	0.16	0.12	-0.07	0.09
A ex Cla	-0.36	0.35	0.06	0.47	-0.32	-0.26
A int Cla	-0.19	0.50	0.03	0.64	-0.13	-0.21
C em Cla	0.35	-0.24	-0.29	-0.24	0.32	0.25
C ex Cla	0.07	-0.21	0.02	-0.25	0.05	0.09
C int Cla	-0.22	0.48	0.07	0.63	-0.16	-0.20
T2/C int Cla	-0.15	0.37	-0.21	0.30	-0.12	-0.23
UV ₂₅₄ Raw	0.73	-0.36	-0.09	-0.19	0.72	0.73
UV ₂₅₄ Cla	-0.15	0.43	0.04	0.50	-0.10	-0.22
SUVA Raw	0.65	-0.63	-0.13	-0.55	0.60	0.50
SUVA Cla	0.09	-0.07	-0.01	-0.07	0.08	-0.19
TOC Raw	-0.02	0.05	0.05	0.05	-0.01	0.25
TOC Cla	0.23	-0.23	-0.05	-0.16	0.21	0.19
OM removal A	0.51	-0.70	-0.05	-0.62	0.44	0.46
OM removal T1	0.27	-0.61	-0.02	-0.58	0.21	0.14
OM removal T2	0.46	-0.68	-0.04	-0.60	0.40	0.36
OM removal C	0.54	-0.68	-0.08	-0.61	0.48	0.50

OM removal UV₂₅₄	0.51	-0.69	-0.11	-0.58	0.46	0.46
OM removal SUVA	0.50	-0.61	-0.11	-0.57	0.45	0.54
OM removal TOC	0.50	-0.64	-0.08	-0.58	0.44	0.51

9. CONCLUSIONS

9.1. SUMMARY OF THE FINDINGS

Fluorescence excitation-emission spectroscopy is commonly used in environmental studies for rapid, accurate and comprehensive characterization of OM. Here, fluorescence analysis was used for the first time to characterise OM removal efficiency in drinking water treatment. Fluorescence analysis across the coagulation stage at 16 WTWs enabled qualitative and quantitative characterization of the OM fractions preferentially removed by the coagulation treatment processes. The results obtained in this work were found to prove the testable hypotheses presented in Section 2.5. In particular, fluorescence properties, peak C emission wavelength and peak T fluorescence were shown to describe OM properties in terms of degree of aromaticity and content of low molecular weight, hydrophilic and microbial OM respectively (*Hypothesis H1*). Organic matter removal efficiency was found to highly depend on the site-specific organic matter properties of peak C and peak T fluorescence. It was apparent that the correlation coefficients between fluorescence properties of peak T and peak C and TOC removal were less significant for the site-specific and monthly correlations than for the full dataset. For several sites, a significant contribution of different algae species to peak C and peak T fluorescence was found. Fluorescence properties of peak C and peak T were found to provide good discrimination between WTWs and water types (*Hypothesis H3*).

Furthermore, the reduction in peak C and peak A intensity between raw and clarified water was correlated with OM removal by coagulation measured as TOC reduction (*Hypothesis H2*). For the 16 months of fluorescence data based on raw and clarified water properties from 16 surface WTWs located in the West Midlands region of the UK, the correlation between directly measured TOC removal and fluorescence-derived TOC removal was significantly

high (correlation coefficient value of $R^2=0.90$), indicating a strong linear relationship between the variables. The robustness of a fluorescence OM prediction tool was successfully tested on independent fluorescence data (spot check on Campion Hills, Draycote, Melbourne and Strensham samples collected over summer 2007). Fluorescence provided an improved prediction of OM removal by coagulation compared with UV absorbance measured at 254 nm (*Hypothesis H7*). Thus, peak C fluorescence intensity has been proved to be an efficient surrogate for assessment of TOC removal by coagulation.

Fluorescence spectroscopy was successfully used to characterise OM properties and removal efficiency during coagulation process optimisation. With lowering the coagulation pH, higher fluorescence intensity reductions were observed between raw and clarified waters.

Fluorescence data through selected WTWs clearly demonstrated OM removal between raw water and post-GAC stages. The minimum fluorescence intensity was observed for pre-contact tank stage, while the addition of chlorine impaired the fluorescence signal (fluorescence quenching). TOC removal through the WTW was accompanied by changes in OM composition – compared to raw water, final water exhibited relatively lower concentrations of peak C and peak A fluorophores and higher concentrations of protein-like fluorescence.

The results showed that measurement of fluorescence peak C intensity might be utilized for DBPs formation prediction, as the fluorescence peak C fluorescence intensity correlates with the total amount of organic precursors. Moreover, particular fluorescence properties were correlated with formation potentials of selected DBPs species. Fluorescence peak T2 intensity of clarified water and $T2_{int}/C_{int}$ ratio or raw water were found to be better predictors of dibromochloromethane, dichlorobromomethane, bromoform formation than UV_{254} and SUVA. Clarified water peak T2 intensity was suggested to be of practical operational use at WTWs as

a simple parameter assessing OM properties, removal efficiency, and DBPs formation potential (*Hypothesis H4*).

From the evaluation of several data mining techniques, it was found that extraction of meaningful information on OM properties and variability from large fluorescence datasets can be successfully performed with SOM. However, for rapid assessment of OM fluorescence signature and small fluorescence datasets, standard peak-picking approach combined with visual inspection of EEMs can be a more robust solution (*Hypotheses H5 and H6*).

The simplicity, sensitivity, speed of analysis and low cost, combined with a potential for incorporation into on-line monitoring systems mean that fluorescence spectroscopy offers a robust analytical technique to be used in conjunction with, or in place of, other approaches to OM characterisation and THM formation prediction.

9.2. IMMEDIATE APPLICATIONS FOR WATER COMPANIES

1. The results obtained to date indicate that fluorescence spectroscopy can be effectively utilized at WTWs for characterisation of both OM properties and its removal efficiency.
2. Fluorescence spectroscopy enables quantification of the different OM fractions, including those of key importance to the TOC removal efficiency. The OM fraction (high molecular weight, hydrophobic) readily removed by coagulation is denoted by high peak C emission wavelength. The OM fraction more difficult to remove (low-molecular weight, hydrophilic) is denoted by high peak T intensity and low peak C emission wavelength.
3. The ratio of peak T2 intensity to peak C intensity ($T2_{int}/C_{int}$) was found to be a good indicator of overall water quality, with low values indicating natural upland sources and high values typical of anthropogenically-impacted lowland sources.

4. Humic-like fluorescence (peaks A and C) was found to be more easily removed by coagulation/flocculation/clarification than protein-like fluorescence (peaks T1 and T2).
5. The regression model incorporating fluorescence peak C intensity of raw and clarified water provided a good prediction of OM and OM removal by coagulation for baseline conditions (all sites) and for the optimised coagulation at Campion Hills. Thus, fluorescence analysis provides an efficient framework for comprehensive characterisation of organic matter removal.
6. The fluorescence-derived TOC removal prediction was more robust compared to the corresponding prediction based on UV₂₅₄ measurements of raw and clarified water. Likewise, fluorescence spectroscopy provided a fully comprehensive characterisation of different OM fractions, whereas UV-Vis absorbance spectroscopy highlighted only presence of highly aromatic OM fraction. Thus, the application of fluorescence spectroscopy in drinking water treatment can provide additional monitoring tool of OM properties and removal efficiency.
7. A simple fluorescence-TOC removal prediction based on tryptophan-like fluorescence could be of practical operational use at WTWs when optimising works performance to best remove hydrophilic OM. A simple fluorescence probe could be installed online to measure tryptophan-like fluorescence of clarified water and therefore be used as a TOC removal monitor.
8. Peak T fluorescence intensity demonstrates presence of microbially-derived OM and therefore can be utilized in monitoring of biological activity in raw and treated waters.
9. Fluorescence properties of peak C and peak T were found to provide indication of DBPs formation potentials and therefore can be utilized in assessment of OM

reactivity with chlorine. Thus, peak C and peak T can be used to characterise OM in spatial and temporal terms, and to discriminate between WTW performance.

10. The highest TTHMFP and THAAFP were observed for WTWs from groups 1, 2 and 5, showing that OM character is of lesser importance for formation of DBPs than quantity of OM.

9.3. FUTURE RESEARCH POTENTIAL

Additional research should concentrate in future on:

1. Correlations between fluorescence signature of OM and DBP formation during chlorination. In particular, the impact of chlorine addition on qualitative and quantitative changes in fluorescence EEMs and DBPs formation should be investigated.
2. Application of peak C and T fluorescence for on-line monitoring and early warning of failure of drinking water quality in the distribution system.
3. Characterisation of qualitative and quantitative changes in fluorescence spectra throughout WTWs showing different raw water OM properties.

LIST OF REFERENCES

- Abbt-Braun, G. and Frimmel, F.H. (1999) Basic characterisation of Norwegian NOM samples – similarities and differences. **Environment International**, 25(2-3): 161-180
- Ahmad, S.R. and Reynolds, D.M. (1999) Monitoring of water quality using fluorescence technique: prospect of on-line process control. **Water Research**, 33(9): 2069-2074
- Aiken, G. R., McKnight, D.M., Wershaw, R. L. et al. (1985) **Humic Substances in Soil, Sediment, and Water**. New York: Wiley-Interscience
- Alunkaynak, A. (2007) Forecasting surface water level fluctuations of lake Van by artificial neural networks. **Water Resources Management**, 21(2): 399-408
- Amon, R.M.W. and Benner, R. (1996) Photochemical and microbial consumption of dissolved organic carbon and dissolved oxygen in the Amazon River system. **Geochimica et Cosmochimica Acta**, 60(10): 1783-1792
- Amy, G.L., Chadik, P.A., Chowdhury, Z.K. (1987) Developing models for predicting trihalomethane formation potential and kinetics. **Journal of American Water Works Assoc.**, 79(7): 89-97
- Amy, G.L., Siddiqui, M, Ozekin, K, Zhu, H.W. et al. (1998) Empirical based models for predicting chlorination and ozonation byproducts: haloacetic acids, chloral hydrate and bromate, **EPA Report CX 819579**
- Andersen, C.M. and Bro, R. (2003) Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. **Journal of Chemometrics**, 17: 200-215
- Andrews, J.M. and Lieberman, S.H. (1994) Neural network approach to qualitative identification of fuels and oils from laser induced fluorescence spectra. **Analytica Chimica Acta**, 285(1-2): 237-246

- Antunes, M.C.G. and Esteves da Silva, J.C.G. (2005) Multivariate curve resolution analysis excitation-emission matrices of fluorescence of humic substances. **Analytica Chimica Acta**, 546: 52-59
- Asai, R., Horiguchi, Y. and Yoshida, A. et al. (2001) Detection of phycobilin pigments and their seasonal change in Lake Kasumigaura using a sensitive in situ fluorometric sensor. **Analytical Letters**, 34(14): 2521-2533
- Ates, N., Kaplan, S.S., Sahinkaya, E. et al. (2007a) Occurrence of disinfection by-products in low DOC surface waters in Turkey. **Journal of Hazardous Materials**, 142(1-2): 526-534
- Ates, N., Yetis, U., Kitis, M. (2007b) Effects of bromide ion and natural organic matter fractions on the formation and speciation of chlorination by-products. **Journal of Environmental Engineering**, 133(10): 947-954
- Bahram, M., Bro, R., Stedmon, C. et al. (2006) Handling of Rayleigh and Raman scatter for PARAFAC modelling of fluorescence data using interpolation. **Journal of Chemometrics**, 20: 99-105
- Baker, A. (2001) Fluorescence Excitation-Emission Matrix characterization of some sewage-impacted rivers. **Environmental Science and Technology**, 35(5): 948-953
- Baker, A. (2002) Fluorescence excitation-emission matrix characterization of river waters impacted by a tissue mill effluent. **Environmental Science and Technology**, 36(7): 1377-1382
- Baker, A. (2005) Thermal fluorescence quenching properties of dissolved organic matter. **Water Research**, 39: 4405-4412

- Baker, A. and Spencer, R.G.M. (2004) Characterisation of dissolved organic matter from source to sea using fluorescence and absorbance spectroscopy. **Science of the Total Environment**, 333: 217-232
- Baker, A., Tipping, E., Thacker, S.A. et al. (2008) Relating dissolved organic matter fluorescence and functional properties. **Chemosphere**, 73: 1765-1772
- Banks, J. and Wilson, D. (2002) Low-cost solutions for trihalomethanes compliance. **J. of the Chartered Institution of Water and Environmental Management**, 16(4): 264-269
- Basheer , I.A. and Hajmeer, M. (2000) Artificial neural networks: fundamentals, computing, design, and application. **Journal of Microbiological Methods**, 43(1): 3-31
- Battin, T.J. (1998) Dissolved organic matter and its optical properties in a blackwater tributary of the upper Orinoco river, Venezuela. **Organic Geochemistry**, 28(9-10): 561-569
- Battin, T.J., Kaplan, L.A., Findlay, S. et al., (2008) Biophysical controls on organic carbon fluxes in fluvial networks. **Nature Geosciences**, 1(2): 95-100
- Baytak, D., Aysun, S., Inal, F. et al. (2008) Seasonal variation in drinking water concentrations of disinfection by-products in IZMIR and associated human health risks. **Science of the Total Environment**, 407: 286-296
- Beggs, K., Zachman, B.A., Valenti, C. et al. (2006) Predicting disinfection byproducts using molecular fluorescence. **In Proc. of AWWA WQTC Conf.**, Denver, USA.
- Beggs, K.M.H., Summers, R.S., McKnight, D.M. (2009) Characterizing chlorine oxidation of dissolved organic matter and disinfection by-product formation with fluorescence spectroscopy and parallel factor analysis. **Journal of Geophysical Research-Biogeosciences**, 114, doi:10.1029/2009JG001009

- Bellar, T.A., Lichtenberg, J.J., Kroner, R.C. (1974) The occurrence of organohalides in chlorinated drinking waters. **Journal of American Water Works Assoc.**, 66(11): 703-706
- Belzile, C. and Guo, L. (2006) Optical properties of low molecular weight and colloidal organic matter: Application of the ultrafiltration permeation model to DOM absorption and fluorescence. **Marine Chemistry**, 98: 183-196
- Bengraïne, K. and Marhaba, T.F. (2003) Comparison of spectral fluorescent signatures-based models to characterize DOM in treated water samples. **Journal of Hazardous Materials**, B100(1-3): 117-130
- Beutler, M., Wiltshire, K.H., Meyer, B. et al. (2002) A fluorometric method for the differentiation of algal populations in vivo and in situ. **Journal of Photosynthesis Research**, 72: 39-53
- Bieroza, M., Baker, A., Bridgeman, J. (2009a) Relating freshwater organic matter fluorescence to organic carbon removal efficiency in drinking water treatment. **Science of Total Environment**, 407: 1765-1774
- Bieroza, M., Baker, A., Bridgeman, J. (2009b) Exploratory analysis of excitation-emission matrix fluorescence spectra with self-organizing maps as a basis for determination of organic matter removal efficiency at water treatment works. **Journal of Geophysical Research – Biogeosciences**, doi:10.1029/2009JG000940
- Bieroza, M., Baker, A., Bridgeman, J. (2009c) "Assessing Organics Removal in Water Treatment with Data Mining and Artificial Neural Networks". In: Topping, B.H.V. and Tsompanakis, Y. (eds.), **Proceedings of the First International Conference on Soft Computing Technology in Civil, Structural and Environmental Engineering**. Civil-Comp Press, Stirlingshire, United Kingdom, paper 10, doi:10.4203/ccp.92.10

- Bieroza, M., Baker, A., Bridgeman, J. (2009d) "Fluorescence spectroscopy as a tool for determination of organic matter removal efficiency at water treatment works". In: Boxall, J. and Maksimović, Ć. (eds.), **Computing and Control in the Water Industry 2009: Integrating Water Systems**, Sheffield
- Blaser, P., Heim, A., Luster, J. (1999) Total luminescence spectroscopy of NOM-typing samples and their aluminium complexes. **Environment International**, 25(2-3): 285-293
- Boehme, J., Coble, P., Commy, R. et al. (2004) Examining CDOM fluorescence variability using principal component analysis: seasonal and regional modelling of three-dimensional fluorescence in the Gulf of Mexico. **Marine Chemistry**, 89(1-4): 3-14
- Bond, T., Henriot, O., Goslan, E.H. et al. (2009) Disinfection Byproduct Formation and Fractionation Behavior of Natural Organic Matter Surrogates. **Environmental Science and Technology**, 43: 5982-5989
- Bos, M., Bos, A., van den Linden, W.E. (1993) Data processing by neural networks in quantitative chemical analysis. **The Analyst**, 118(4): 323-328
- Boyle, E.S., Guerriero, N., Thiallet, A. et al. (2009) Optical Properties of Humic Substances and CDOM: Relation to Structure. **Environmental Science and Technology**, 43(12): 4612-4619
- Bro, R. (1997) PARAFAC. Tutorial and applications. **Chemometrics and Intelligent Laboratory Systems**, 38(2): 149-171
- Bro, R. (1998) **Multi-way analysis in the food industry. Models, algorithms, and applications**. PhD dissertation. Department of Dairy and Food Science, Royal Veterinary and Agricultural University, Denmark.
- Brunsdon, C. and Baker, A. (2001) Principal filter analysis for luminescence excitation-emission data. **Geophysical Research Letters**, 29(24): 2156

- Cammack, W.K.L., Kalff, J., Prairie, Y.T. et al. (2004) Fluorescent dissolved organic matter in lakes: Relationship with heterotrophic metabolism. **Limnology and Oceanography**, 49(6): 2034-2045
- Carlson, M. and Hardy, D. (1998) Controlling DBPs with monochloramine. **Journal of American Water Works Assoc.**, 90(2): 95-106
- Chang, E.E., Lin, Y.P., Chiang, P.C. (2001) Effects of bromide on the formation of THMs and HAAs. **Chemosphere**, 43: 1029-1034
- Chang, E.E., Chiang, P.C., Chao, S.H. et al. (2006) Relationship between chlorine consumption and chlorination by-products formation for model compounds. **Chemosphere**, 64(7): 1196-1203
- Chen, Z. and Valentine, R.L. (2007) Formation of N-Nitrosodimethylamine (NDMA) from humic substances in natural water. **Environmental Science and Technology**, 41: 6059-6065
- Chen, W., Westerhoff, P. Leenheer, J.A. et al. (2003) Fluorescence Excitation-Emission Matrix Regional Integration to Quantify Spectra for Dissolved Organic Matter. **Environmental Science and Technology**, 37: 5701-5710
- Chen, Z., and Valentine, R.L. (2008). The influence of the preoxidation of natural organic matter on the formation of N-nitrosodimethylamine. **Environmental Science and Technology**, 42: 5062
- Cheng, W.P., Chi, F.H., Yu, R.F. (2004) Evaluating the efficiency of coagulation in the removal of dissolved organic carbon from reservoir water using fluorescence and ultraviolet photometry. **Environmental Monitoring and Assessment**, 98(1-3): 421-431

- Chow, C.W.K., van Leeuwen, J.A., Drikas, M. et al. (1999) The impact of the character of natural organic matter in conventional treatment with alum. **Water Science and Technology**, 40(9): 97-104
- Chow, C. W. K., Fabris, R., Drikas, M. (2004) A rapid fractionation technique to characterise natural organic matter for the optimisation of water treatment processes. **J. Water Supply Resources and Technology – AQUA**, 53(2): 85–92
- Chowdhury, Z.K., Papadimas, S.P., Olivieri, E.B. (1995) “Use of carbon dioxide for enhanced coagulation: a city of Tempe experience”. In: **Proc. AWWA WQTC Conf.**, New Orleans, LA
- Clark, R.M. (1998) Chlorine demand and TTHM formation kinetics: A second-order model. **Journal of Environmental Engineering**, 1998(1): 16-24
- Clark, R.M., Thurnau, R.C., Sivaganesan, M. et al. (2001) Predicting the formation of chlorinated and brominated by-products. **Journal of Environmental Engineering**, 127(6): 493-501
- Clark, C.D., Jimenez-Morais, J., Jones, G. et al. (2002) A time-resolved fluorescence study of dissolved organic matter in a riverine to marine transition zone. **Marine Chemistry**, 78(2-3): 121-135
- Coble, P.G. (1996) Characterization of marine and terrestrial DOM in seawater using excitation–emission matrix spectroscopy. **Marine Chemistry**, 51: 325–346
- Corine land cover 2000 (CLC2000) 100 m version 9/2007 (2007) European Environmental Agency, <http://www.eea.europa.eu>, Copenhagen.
- Cumberland, S.A. and Baker, A. (2007) The freshwater dissolved organic matter fluorescence-total organic carbon relationship. **Hydrological Processes**, 21(16): 2093-2099

- Daliakopoulos, I.N., Coulibaly, P., Tsanis, I.K. (2005) Groundwater level forecasting using artificial neural networks. **Journal of Hydrology**, 309(1-4): 229-240
- Davies, D.L., and Bouldin, D.W. (1979) Cluster separation measure, **IEEE Trans. on Pattern Anal. and Machine Intel.**, 1(2): 224-227
- de Souza Sierra, M.M., Donard, O.F., Lamotte, M. (1997) Spectral identification and behavior of dissolved organic fluorescent material during estuarine mixing processes. **Marine Chemistry**, 58(1-2): 51-58
- del Castillo, C.E., Coble, P.G., Morell, J.M. et al. (1999) Analysis of the optical properties of the Orinoco River plume by absorption and fluorescence spectroscopy. **Marine Chemistry**, 66: 35-51
- del Vecchio, R. and Blough, N.V. (2002) Photobleaching of chromophoric dissolved organic matter in natural waters: kinetics and modelling. **Marine Chemistry**, 78: 231-253
- del Vecchio, R. and Blough, N.V. (2004) Spatial and seasonal distribution of chromophoric dissolved organic matter and dissolved organic carbon in the Middle Atlantic Bight. **Marine Chemistry**, 89: 169– 187
- Denkhaus, E., Meisen, S., Telgheder, U. et al. (2007) Chemical and physical methods for characterisation of biofilms – Review. **Microchimica Acta**, 158: 1-27
- Despagne, F. and Massart, D.L. (1998) Neural networks in multivariate calibration. **The Analyst**, 123(11): 157R-178R
- Determann, S., Reuter, R., Wagner, P. et al. (1994) Fluorescent matter in the early Atlantic Ocean. 1. Method of measurement and near-surface distribution. **Deep-Sea Research Part I – Oceanographic Research Papers**, 41(4): 659-675

- Determann, S., Lobbes, J.M., Reuter, R. et al. (1998) Ultraviolet fluorescence excitation and emission spectroscopy of marine algae and bacteria, **Marine Chemistry**, 62(1-2): 137-156
- Divya, O. and Mishra, A.K. (2007) Multivariate methods on the excitation emission matrix fluorescence spectroscopy data of diesel-kerosene mixtures: A comparative study. **Analytica Chimica Acta**, 592(1): 82-90
- Dunnick, J.K. and Melnick, R.L. (1993) Assessment of the carcinogenic potential of chlorinated water: experimental studies of chlorine, chloramines, and trihalomethanes. **Journal of the National Cancer Institute**, 85(10): 817-822
- Eaton, A. (1995) Measuring UV-absorbing organics: a standard method. **Journal of American Water Works Assoc.**, 87: 86-90
- Edzwald, J.K. and Tobiasson, J.E. (1999) Enhanced coagulation: US requirements and a broader view. **Water Science and Technology**, 40(9): 63-70
- Edzwald, J.K., Becker, W.C., Wattier, K.L. (1985) Surrogate parameters for monitoring organic matter and THM precursors. **Journal of American Water Works Assoc.**, 77(4): 122-132
- Elkins, K.M. and Nelson, D.J. (2001) Fluorescence and FT-IR spectroscopic studies of Suwannee river fulvic acid complexation with aluminum, terbium and calcium. **Journal of Inorganic Biochemistry**, 87: 81-96
- Elliot, S., Lead, J.R., Baker, A. (2006) Thermal quenching of fluorescence of freshwater, planktonic bacteria. **Analytica Chimica Acta**, 564: 219-225
- Engelen, S., Frosch, S., Jørgensen, B.M. (2009) A fully robust PARAFAC method for analyzing fluorescence data. **Journal of Chemometrics**, 23: 124-131

- Engerholm, B.A. and Amy, G.L. (1983) A predictive model for chloroform formation from humic acid. **Journal of American Water Works Assoc.**, 75(8): 418-423
- Evans, C.D., Chapman, P.J., Clark, J.M. et al. (2006) Alternative explanations for rising dissolved organic carbon export from organic soils. **Global Change Biology**, 12: 2044-2053
- Fabris, R., Chow, C.W.K., Drikas, M. et al. (2008) Comparison of NOM character in selected Australian and Norwegian drinking waters. **Water Research**, 42: 4188-4196
- Fearing, D.A., Banks, J., Guyetand, S. et al. (2004) Combination of ferric and MIEXs for the treatment of a humic rich water. **Water Research**, 38: 2551-2558
- Fellman, J.B., D'Amore, D.V., Hood, E. et al. (2008) Fluorescence characteristics and biodegradability of dissolved organic matter in forest and wetland soils from coastal temperate watersheds in southeast Alaska. **Biogeochemistry**, 88: 169-184
- Francis, R.A., Small, M.J., VanBriesen, J.M. (2009) Multivariate distributions of disinfection by-products in chlorinated drinking water. **Water Research**, 43: 3453-3468
- Freeman, C., Kim, S.Y., Lee, S.H. et al. (2004) Effects of elevated atmospheric CO₂ concentrations on soil microorganisms. **Journal of Microbiology**, 42: 267-277
- Galapate, R.P., Baes, A.U., Ito, K. et al. (1998) Detection of domestic wastes in Kurose River using synchronous fluorescence spectroscopy. **Water Research**, 32(7): 2232-2239
- Gallard, H. and von Gunten, U. (2002) Chlorination of natural organic matter: kinetics of chlorination and of THM formation. **Water Research**, 36(1): 65-74
- Gang, D.D., Segar, R.L., Clevenger, T.S. et al. (2002) Using chlorine demand to predict THM and HAA9 formation. **Journal of American Water Works Assoc.**, 94(10): 76-86
- Gang, D., Clevenger, T.E., Banerji, S.K. (2003) Relationship of chlorine decay and THMs formation to NOM size. **J. of Hazardous Materials**, 96: 1-12

- Gao, H. and Zepp, R.G. (1998) Factors Influencing Photoreactions of Dissolved Organic Matter in a Coastal River of the Southeastern United States. **Environmental Science and Technology**, 32: 2940-2946
- Garcia, J.S., da Silva, G.A., Arruda, M.A. et al. (2007) Application of Kohonen neural network to exploratory analyses of synchrotron radiation x-ray fluorescence measurements of sunflower metalloproteins. **X-ray Spectrometry**, 36(2): 122-129
- Ghosh, K. and Schnitzer, M. (1980) Fluorescence excitation spectra of humic substances. **Canadian Journal of Soil Science**, 60(2): 373-379
- Gjessing, E.T., Egeberg, P.K., Harkedal, J. (1999) Natural organic matter in drinking water - the "NOM typing project", background and basic characteristics of original water samples and NOM isolates. **Environment International**, 25: 145-159
- Golfinopoulos, S.K. and Arhonditsis, G.B. (2002a) Quantitative assessment of trihalomethane formation using simulations of reaction kinetics. **Water Research**, 36: 2856-2868
- Golfinopoulos, S.K. and Arhonditsis, G.B. (2002b) Multiple regression models: A methodology for evaluating trihalomethane concentrations in drinking water from raw water characteristics. **Chemosphere**, 47: 1007-1018
- Gondar, D., Thacker, S.A., Tipping, E. et al. (2008) Functional variability of dissolved organic matter from a productive lake. **Water Research**, 42: 81-90
- Gontarski, C.A., Rodrigues, P.R., Mori, M. Et al. (2000) Simulation of an industrial wastewater treatment plant using artificial neural networks. **Computers and Chemical Engineering**, 24(2-7): 1719-1723
- Goslan, E.H. (2003) **Natural Organic Matter Character and Reactivity: Assessing Seasonal Variation in a Moorland Water**. EngD Thesis, Cranfield University

- Goslan, E.H., Voros, S., Banks, J. et al. (2004) A model for predicting dissolved organic carbon distribution in a reservoir water using fluorescence spectroscopy. **Water Research**, 38(3): 783-791
- Goslan, E.H., Krasner, S.W., Bower, M. et al. (2009) A comparison of disinfection by-products found in chlorinated and chloraminated drinking waters in Scotland. **Water Research**, 43: 4698-4706
- Gregor, J. and Maršálek B. (2005) A simple in vivo fluorescence method for the selective detection and quantification of freshwater cyanobacteria and eukaryotic algae. **Acta Hydrochimica Hydrobiologica**, 33(2): 142–148
- Gregor, I.E., Nokes, C.J., Fenton, E. (1997) Optimising natural organic matter removal from low turbidity waters by controlled pH adjustment of aluminium coagulation. **Water Research**, 31(12): 2949-2958
- Gregor, J., Maršálek B. and Šípková H. (2007) Detection and estimation of potentially toxic cyanobacteria in raw water at the drinking water treatment plant by in vivo fluorescence method. **Water Research**, 41: 228-234
- Gregory, J. (1997) The density of particle aggregates. **Water Science Technology**, 36 (4): 1-13
- Guimet, F., Ferré, J., Boqué, R. et al (2003) Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitation–emission matrix fluorescence spectroscopy. **Analytica Chimica Acta**, 515: 75-85
- Guimet, F., Ferré, J., Boqué, R. (2005) Rapid detection of olive-pomace oil adulteration in extra virgin olive oils from the protected denomination of origin “Siurana” using

excitation-emission fluorescence spectroscopy and three-way methods of analysis.

Analytica Chimica Acta, 544(1-2): 143-152

Häck, M. and Köhne, M. (1996) Estimation of wastewater process parameters using neural networks. **Water Science and Technology**, 33(1): 101-115

Hall, G.J., Clow, K.E., Kenny, J.E. (2005) Estuarial fingerprinting through multidimensional fluorescence and multivariate analysis. **Environmental Science and Technology**, 39(19): 7560-7567

Hallam, N.B., West, J.R., Forster, C.F. et al. (2001) The potential for biofilm growth in water distribution systems. **Water Research**, 34(17): 4063-4071

Hammerstrom, D. (1993) Working with neural networks. **IEEE Spectrum**, 30(7): 46-53

Hautala, K., Peuravouri, J., Pihlaja, K. (2000) Measurement of aquatic humus content by spectroscopic analyses. **Water Research**, 34(1): 246-258

Helms, J.R., Stubbins, A., Ritchie, J.D. et al. (2008) Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. **Limnology and Oceanography**, 53(3): 955-969

Henderson, R., Parsons, S., Jefferson, B. (2008a) The impact of algal properties and pre-oxidation on solid-liquid separation of algae. **Water Research**, 42: 1827-1845

Henderson, R.K., Baker, A., Parsons, S.A. et al. (2008b) Characterisation of algogenic organic matter extracted from cyanobacteria, green algae and diatoms. **Water Research**, 42: 3435-3445

Henderson, R.K., Baker, A., Murphy, K.R. et al. (2009) Fluorescence as a potential monitoring tool for recycled water systems: A review. **Water Research**, 43(4): 863-881

- Henrion, R., Henrion, M., Böhme, M. et al. (1997) Three-way principal components analysis for fluorescence spectroscopic classification of algae species. **Fresenius Journal of Analytical Chemistry**, 357(5): 522-526
- Her, N., Amy, G., Foss, D. et al. (2002) Variations of molecular weight estimation by HP-Size Exclusion Chromatography with UVA versus online DOC detection. **Environmental Science and Technology**, 36: 3393-3399
- Her, N., Amy, G., McKnight, D. et al. (2003) Characterization of DOM as a function of MW by fluorescence EEM and HPLC-SEC using UVA, DOC, and fluorescence detection. **Water Research**, 37: 4295-4303
- Hernes, P.J., Spencer, R.G.M., Dyda, R.Y. et al. (2008) The role of hydrologic regimes on dissolved organic carbon composition in an agricultural watershed. **Geochimica et Cosmochimica Acta**, 72: 5266-5277
- Hopkins, O.S. (2005) Careful consideration necessary when using stepwise regression. **Journal of American Water Works Assoc.**, 97(7): 144
- Hu, J.Y., Wang, Z.S., NG, W.J. et al. (1999) Disinfection by products in water produced by ozonation and chlorination. **Environmental Monitoring and Assessment**, 59: 81-93
- Hua, F. (2000) **The effect of water treatment works on chlorine decay and trihalomethane formation**. PhD Thesis, The University of Birmingham.
- Hua, B., Veum, K., Koirala, A. et al. (2007) Fluorescence fingerprints to monitor total trihalomethanes and N-nitrosodimethylamine formation potentials in water. **Environmental Chemistry Letters**, 5: 73-77
- Hubel, R.E. and Edzwald, J.K. (1987) Removing trihalomethane precursors by coagulation. **Journal of American Water Works Assoc.**, 79 (7): 98-106

- Huber, L. and Frimmel, F.H. (1996) Influence of humic substances on the aquatic adsorption of heavy metals on defined mineral phases. **Environment International**, 22(5): 507-517
- Hudson, N.J., Baker, A., Reynolds, D. (2007) Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters – a review. **River Research and Applications**, 23(6): 631-649
- Hudson, N.J., Baker, A., Ward, D. et al. (2008) Fluorescence spectrometry as a surrogate for the *BOD*₅ test in water quality assessment: an example from South West England. **Science of Total Environment**, 391(1): 149-158
- Hudson et al. (2009) Changes in freshwater organic matter fluorescence intensity with freezing/thawing and dehydration/ rehydration. **Journal of Geophysical Research – Biogeosciences**, doi 10.1029/2008JG000915
- Hur, J., Hwang, S.-J., Shin, J.-K. (2008) Using Synchronous Fluorescence Technique as a Water Quality Monitoring Tool for an Urban River. **Water Air Soil Pollution**, 191: 231-243
- Iriarte-Velasco, U., Alvarez-Uriarte, J.I., Gonzalez-Velasco, J.R. (2006) Monitoring trihalomethanes in water by differential ultraviolet spectroscopy. **Environmental Chemistry Letters**, 4(4): 243-247
- Jackson, P., Hall, T., Young, W. et al. (2008) **A review of different national approaches to the regulation of THMs in drinking water.** DEFRA/DWI.
- Jaffe, R., Boyer, J.N., Lu, X. et al. (2004) Source characterization of dissolved organic matter in a subtropical mangrove-dominated estuary by fluorescence analysis. **Marine Chemistry**, 84(3-4): 195-210

- Jaffe, R., McKnight, D., Maie, N. et al. (2008) Spatial and temporal variations in DOM composition in ecosystems: The importance of long-term monitoring of optical properties. **Journal of Geophysical Research**, 113: 1-15
- Jain, A., and Dubes, R. (1988) **Algorithms for clustering data**. New Jersey - Prentice-Hall.
- Johnstone, D.W. and Miller, C.M. (2009) Fluorescence Excitation–Emission Matrix Regional Transformation and Chlorine Consumption to Predict Trihalomethane and Haloacetic Acid Formation. **Environmental Engineering Science**, 26(7): 1163-1170
- Junk, W.J. and Wantzen, K.M. (2003) “The Flood Pulse Concept: New Aspects, Approaches and Applications - An Update”. In: **Proceedings of the Second International Symposium on the Management of Large Rivers for Fisheries Volume II**, Cambodia
- Kalbitz, K., Geyer, W., Geyer, S. (1999) Spectroscopic properties of dissolved humic substances – a reflection of land use history in a fen area. **Biogeochemistry**, 47: 219-238
- Kalbitz, K., Geyer, W., Gehre, M. (2000) Land use impacts on the isotopic signature (C-13, C-14, C-15) of water-soluble fulvic acids in German fen area. **Soil Science**, 165(9): 728-736
- Kalbitz, K., Schwesig, D., Schmerwitz, J. et al. (2003) Changes in properties of soil-derived dissolved organic matter induced by biodegradation. **Soil Biology and Biochemistry**, 35(8): 1129-1142
- Katsuyama, M. and Ohte, N. (2002) Determining the sources of stormflow from the fluorescence properties of dissolved organic carbon in a forested headwater catchment. **Journal of Hydrology**, 268: 192-202

- Kavanaugh, M.C., Trussell, A.R., Cromer, J. et al. (1980) An empirical kinetic model of trihalomethane formation: Applications to meet the proposed THM standard. **Journal of American Water Works Assoc.**, 72(10): 578-582
- Kelton, N., Molot, L.A., Dillon, P.J. (2007) Spectrofluorometric properties of dissolved organic matter from Central and Southern Ontario streams and the influence of iron and irradiation. **Water Research**, 41: 638-646
- Khoury, A.E., Nicholov, R., Soltes, S. et al. (1992) A Preliminary Assessment of *Pseudomonas aeruginosa* Biofilm Development using Fluorescence Spectroscopy. **International Biodeterioration and Biodegradation**, 30: 187-199
- Kim, H.-C. and Yu, M.-J. (2005) Characterization of natural organic matter in conventional water treatment processes for selection of treatment processes focused on DBPs control. **Water Research**, 39: 4779-4789
- Kitis, M., Karafani, T., Kilduff, J.E. et al. (2001) The reactivity of natural organic matter to disinfection by-products formation and its relation to specific ultraviolet absorbance. **Water Science and Technology**, 43: 9-16
- Kohonen, T. (1998) The self-organizing map. **Neurocomputing**, 21(1): 1-6
- Kohonen, T. (2001) **Self-organizing maps**. 3rd ed., Berlin: Springer.
- Komada, T., Schofield, O.M.E., Remers, C.E. (2002) Fluorescence characteristics of organic matter released from coastal sediments during resuspension. **Marine Chemistry**, 79: 81– 97
- Korshin, G.V., Li, C.-W., Benjamin, M.M. (1997) The decrease of UV absorbance as an indicator of TOX formation. **Water Research**, 31(4): 946-949

- Korshin, G.V., Kumke, M.U., Li, C.-W. et al. (1999) Influence of chlorination on chromophores and fluorophores in humic substances. **Environmental Science and Technology**, 33: 1207-1212
- Korshin, G., Chow, C.W.K., Fabris, R. et al. (2009) Absorbance spectroscopy-based examination of effects of coagulation on the reactivity of fractions of natural organic matter with varying apparent molecular weights. **Water Research**, 43: 1541-1548
- Krasner, S.W., McGuire, M.J., Jacangelo, J.G. et al. (1989) The occurrence of disinfection by-products in US drinking water. **Journal of American Water Works Assoc.**, 91(8): 41-53
- Krasner, S.W., Croue, J.P., Buffle, J. (1996) Three approaches for characterizing NOM. **Journal of American Water Works Assoc.**, 88(6): 66-79
- Laabs, C., Amy, G., Jekel, M. (2004) Organic colloids and their influence on low-pressure membrane filtration. **Water Science and Technology**, 50(12): 311-316
- Lakowicz, J.R. (1999) **Principles of fluorescence spectroscopy**. 2nd Edition, Kluwer Academic/Plenum Publishers
- Lead, J.R., De Momi, A., Goula, G. et al. (2006) Fractionation of freshwater colloids and particles by SPLITT: analysis by electron microscopy and 3D excitation–emission matrix fluorescence. **Analytical Chemistry**, 78: 3609–3615
- Leboulanger, C., Dorigo, U. and Jacquet, S. et al. (2002) Application of a submerge spectrofluorometer for rapid monitoring of freshwater cyanobacterial blooms: a case study. **Aquatic Microbial Ecology**, 30: 83-89
- Lee, K.I., Yim, Y.S., Chung, S.W. et al. (2005) Application of artificial neural networks to the analysis of two-dimensional fluorescence spectra in recombinant E coli fermentation processes. **Journal of Chemical Technology and Biotechnology**, 80(9): 1036-1045

- Leenheer, J. A. (1981) Comprehensive approach to preparative isolation and fractionation of dissolved organic carbon from natural waters and wastewaters. **Environmental Science and Technology**, 15(5): 578–587
- Li, J.K. and Humprey, A.E. (1991) Use of fluorometry for monitoring and control of a bioreactor. **Biotechnology and Bioengineering**, 37: 1043-1049
- Li, C.-W., Benjamin, M.M., Korshin, G.V. (2000) Use of UV spectroscopy to characterize the reaction between NOM and free chlorine. **Environmental Science and Technology**, 34(12): 2370-2575
- Li, Q., Yao, X., Chen, X. et al. (2000) Application of artificial neural networks for the simultaneous determination of a mixture of fluorescent dyes by synchronous fluorescence. **The Analyst**, 125(11): 2049-2053
- Lind, C.B. (1995) “Experiences in TOC removal by polyaluminum hydroxychloride and enhanced coagulation”. In: **Proc. AWWA Annual Conf.**, Anaheim, CA
- Liang, L. and Singer, P.C. (2003) Factors influencing the formation and relative distribution of haloacetic acids and trihalomethanes in drinking water. **Environmental Science and Technology**, 37: 2920-2928
- Lombardi, A.T. and Jardim, W.F. (1999) Fluorescence spectroscopy of high performance liquid chromatography fractionated marine and terrestrial organic materials. **Water Research**, 33(2): 512-520
- Malcolm, R.L. and MacCarthy, P. (1992) Quantitative evaluation of XAD-8 and XAD-8 resins used in tandem for removing organic solutes from water. **Environment International**, 18(6): 597-607

- Marhaba, T., Van, D., Lippincott, R.L. (2000) Rapid identification of Dissolved Organic Matter fractions in water by spectral fluorescent signatures. **Water Research**, 34(14): 3543-3550
- Marhaba, T.F., Bengraïne, K., Pu, Y. et al. (2003) Spectral fluorescence signatures and partial least squares regression: model to predict dissolved organic carbon in water. **Journal of Hazardous Materials**, B97: 83-97
- Marhaba, T.F., Borgaonkar, A.D., Punburananon, K. (2009) Principal component regression model applied to dimensionally reduced spectral fluorescent signature for the determination of organic character and THM formation potential of source water. **Journal of Hazardous Materials**, 169: 998-1004
- Marhuenda-Egea, F.C., Martínez-Sabater, E., Jorda', J. et al. (2007) Dissolved organic matter fractions formed during composting of winery and distillery residues: Evaluation of the process by fluorescence excitation–emission matrix. **Chemosphere**, 68: 301-309
- Marquardt, D. (1963) An algorithm for least-squares estimation of nonlinear parameters, **SIAM Journal on Applied Mathematics**, 11(2): 431-441
- Martens, H. and Naes, T. (1989) **Multivariate calibration**. Chichester: John Wiley & Sons
- Matilainen, A., Lindqvist, N., Korhonen, S. et al. (2002) Removal of NOM in the different stages of the water treatment process. **Environment International**, 28: 457-465
- Maurice, P.A., Cabaniss, S.E., Drummond, J. et al. (2002) Hydrogeochemical controls on the variations in chemical characteristics of natural organic matter at a small freshwater wetland. **Chemical Geology**, 187: 59-77
- McAvoy, T.J., Su, H.T., Wang, N.S. et al. (1992) A comparison of neural networks and partial least-squares for deconvoluting fluorescence spectra. **Biotechnology and Bioengineering**, 40(1): 53–62

- McKnight, D.M., Boyer, E.W., Westerhoff, P.K. et al. (2001) Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. **Limnology and Oceanography**, 46(1): 38-48
- Miano, T.M. and Senesi, N. (1992) Synchronous excitation fluorescence spectroscopy applied to soil humic substances chemistry. **Science of the Total Environment**, 118: 41-51
- Mobed, J.J., Hemmingsen, S.L., Autry, J.,L et al., (1996) Fluorescence Characterization of IHSS Humic Substances: Total Luminescence Spectra with Absorbance Correction. **Environmental Science and Technology**, 30: 3061-3065
- Monteith, D., Stoddard, J.L., Evans, C.D. et al. (2007) Dissolved organic carbon trends resulting from changes in atmospheric deposition chemistry. **Nature**, 450: 537-541
- Mopper, K., and Schultz, C.A. (1993) Fluorescence as a possible tool for studying the nature and water column distribution of DOC components. **Marine Chemistry**, 41(1-3): 229-238
- Morel, E., Santamaria, K., Perrier, M. et al. (2004) Application of multi-wavelength fluorometry for on-line monitoring of an anaerobic digestion process. **Water Research**, 38(14-15): 3287-3296
- Moudgal, C.J., Lipscomb, J.C., Bruce, R.M. (2000) Potential health effects of drinking water disinfection by-products using quantitative structure toxicity relationship. **Toxicology**, 147: 109-131
- Mounier, S., Braucher, R., Benaïm, J.Y. (1999) Differentiation of organic matter's properties of the Rio Negro basin by cross-flow ultra-filtration and UV-spectrofluorescence. **Water Research**, 33(10): 2363-2373

- Murphy, K.R., Stedmon, C.A., Waite, T.D. et al. (2008) Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy. **Marine Chemistry**, 108: 40-58
- Nagao, S., Matsunaga, T., Suzuki, Y. et al. (2003) Characteristics of humic substances in the Kuji River waters as determined by high-performance size exclusion chromatography with fluorescence detection. **Water Research**, 37: 4159-4170
- Najm, I.N., Patania, N.L., Jacangelo, J.G. et al. (1994) Evaluating surrogates for disinfection by-products. **Journal of American Water Works Assoc.**, 86: 98-106
- Nam, S.-N. and Amy, G. (2008) Differentiation of wastewater effluent organic matter (EfOM) from natural organic matter (NOM) using multiple analytical techniques. **Water Science and Technology**, 57(7): 1009-1015
- Newson, M., Baker, A., Mounsey, S. (2001) The potential role of freshwater luminescence measurements in exploring runoff pathways in upland catchments. **Hydrological Processes**, 15: 989-1002
- Nguyen, M.-L., Westerhoff, P., Baker, L. et al. (2005) Characteristics and reactivity of algae-produced dissolved organic carbon. **Journal of Environmental Engineering**, 131(11): 1574-1582
- Nikolaou, A.D., Kostopolou, M.N., Lekkas, T.D. (1999) Organic by-products of drinking water chlorination. **Global Nest**, 1(3): 143-156
- O'Melia, C.R., Becker, W.C., Au, K.-K. (1999) Removal of humic substances by coagulation. **Water Science and Technology**, 40(9):47-54
- Ohno, T., Amirbahman, A., Bro, R. (2008) Parallel factor analysis of excitation-emission matrix fluorescence spectra of water soluble soil organic matter as basis for the

- determination of conditional metal binding parameters. **Environmental Science and Technology**, 42: 186-192
- Park, Y.S., Céréghino, R., Compin, A. et al. (2003) Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. **Ecological Modelling**, 160(3): 265-280
- Parsons, S. and Jefferson, B. (2006) **Introduction to Potable Water Treatment Processes**. Willey-Blackwell
- Parsons, S.A., Jefferson, B., Goslan, E.H. et al. (2004) Natural organic matter – the relationship between character and treatability. **Water Science and Technology – Water Supply**, 4(5-6): 43-48
- Patel-Sorrentino, N., Mounier, S., Benaim J.Y. (2002) Excitation-emission fluorescence matrix to study pH influence on organic matter fluorescence in the Amazon basin rivers. **Water Research**, 36(10): 2571-2581
- Patel-Sorrentino, N., Mounier, S., Benaim, J.Y. (2002) Excitation–emission fluorescence matrix to study pH influence on organic matter fluorescence in the Amazon basin rivers. **Water Research**, 36: 2571-2581
- Patel-Sorrentino, N., Mounier, S., Benaim, L.J.Y. (2004) Effects of UV–visible irradiation on natural organic matter from the Amazon basin. **Science of the Total Environment**, 321: 231-239
- Peiris, B.R., Halle, C., Haberkamp, J. et al. (2008) Assessing nanofiltration fouling in drinking water treatment using fluorescence fingerprinting and LC-OCD analyses. **Water Science and Technology - Water Supply**, 8: 459-465
- Pérez Pavón, J.L., Herrero Martin, S., Garcia Pinto, C. et al. (2008) Determination of trihalomethanes in water samples: A review. **Analytica Chimica Acta**, 629: 6-23

- Persson, T. and Wedborg, M. (2001) Multivariate evaluation of the fluorescence of aquatic organic matter. **Analytica Chimica Acta**, 434: 179-192
- Pinto, A.M., von Sperling, E., Moreira, R.M. (2001) Chlorophyll-A determination via continuous measurement of plankton fluorescence methodology development. **Water Research**, 35(16): 3977-3981
- Pullin, M.J. and Cabaniss, S.E. (1997) Physicochemical variations in DOM-synchronous fluorescence: Implications for mixing studies. **Limnology and Oceanography**, 42(8): 1766-1773
- Randtke, S.J. (1988) Organic contaminant removal by coagulation and related process combinations. **Journal of American Water Works Assoc.**, 80 (5): 40-56
- Reckhow, D.A. and Singer, P.C. (1990) Chlorination by-products in drinking waters: from formation potentials to finished water concentrations. **Journal of American Water Works Assoc.**, 82 (4), 173-180
- Reckhow, D.A., Singer, P.C. & Malcolm, R.L. (1990) Chlorination of humic materials: byproduct formation and chemical interpretations. **Environmental Science and Technology**, 24 (11): 1655–1664
- Reynolds, D.M. (2003) Rapid and direct determination of tryptophan in water using synchronous fluorescence spectroscopy. **Water Research**, 37: 3055-3060
- Reynolds, D.M. and Ahmad, S.R. (1995) The effect of metal ions on the fluorescence of sewage wastewater. **Water Research**, 29(9): 2214-2216
- Reynolds, D.M. and Ahmad, S.R. (1997) Rapid and direct determination of wastewater BOD values using a fluorescence technique. **Water Research**, 31(8): 2012-2018
- Rhee, J.I., Lee, K.I., Kim, C.K. et al. (2005) Classification of two-dimensional fluorescence spectra using self-organizing maps. **Biochemical Engineering J.**, 22(2): 135-144

- Rizzo, L., Belgiorno, V., Meric, S. (2004) Organic THMs precursors removal from surface water with low TOC and high alkalinity by enhanced coagulation. **Water Science and Technology**, 4(5-6): 103-111
- Roccaro, P. and Vagliasindi, F.G. (2009) Differential vs. absolute UV absorbance approaches in studying NOM reactivity in DBPs formation: Comparison and applicability. **Water Research**, 43: 744-750
- Roccaro, P., Chang, H.-S., Vagliasindi, F.G.A. et al. (2008) Differential absorbance study of effects of temperature on chlorine consumption and formation of disinfection by-products in chlorinated water. **Water Research**, 42: 1879-1888
- Roccaro, P., Vagliasindi, G.A., Korshin, G.V. (2009) Changes in NOM Fluorescence Caused by Chlorination and their Associations with Disinfection by-Products Formation. **Environmental Science and Technology**, 43: 724-729
- Roe, J., Sharp, E., Borrill, R. et al. (2007) **Diagnostic techniques to optimise NOM removal**. IWA Conference on Particle Separation, Toulouse.
- Roe, J., Bridgeman, J., Baker, A. (2008) Relating organic matter character to trihalomethanes formation potential: a data mining approach. **Water Science and Technology: Water Supply**, 8(6): 717-723
- Rook, J.J. (1974) Formation of haloforms during chlorination of natural waters. **Water Treatment Examination**, 23(2): 234-243
- Rosario-Ortiz, F.L., Snyder, S.A., Suffet, I.H. (2007) Characterization of dissolved organic matter in drinking water sources impacted by multiple tributaries. **Water Research**, 41: 4115-4128
- Rumelhart, D.E. and McClelland, J.L. (1986) **Parallel distributed processing: Explorations in the microstructure of cognition**. Vol. 1., Cambridge: MA:MIT Press

- Saurina, J., Leal, C., Compano, R. et al. (2000) Determination of triphenyltin in sea-water by excitation–emission matrix fluorescence and multivariate curve resolution. **Analytica Chimica Acta**, 409: 237–245
- Scott, S.M., James, D., Ali, Z. et al. (2003) Total luminescence spectroscopy with pattern recognition for classification of edible oils. **The Analyst**, 128: 966-973
- Senesi, N. (1990) Molecular and quantitative aspects of the chemistry of fulvic acid and its interaction with metal ions and organic chemicals. Part 2. The fluorescence spectroscopy approach. **Analytica Chimica Acta**, 232: 77-106
- Senesi, N., Miano, T.M., Provenzano, M.R. et al. (1989) Spectroscopic and compositional comparative characterization of IHSS reference and standard fulvic and humic acids of various origin. **Science of the Total Environment**, 81-2: 143-156
- Senesi, N., Miano, T.M., Provenzano, M.R. et al. (1991) Characterization, differentiation and classification of humic substances by fluorescence spectroscopy. **Soil Science**, 152: 259–271
- Seredynska-Sobecka, B., Baker, A., Lead, J.R. (2007) Characterisation of colloidal and particulate organic carbon in freshwaters by thermal fluorescence quenching. **Water Research**, 41: 3069 – 3076
- Severn Trent Water (2008) **Water Resources Management Plan 2009**. Volume 1: Appendices – Draft
- Sharp, E.L., Parsons, S.A., Jefferson, B. (2006) Seasonal variations in natural organic matter and its impact on coagulation in water treatment. **Science of the Total Environment**, 363: 183-194

- Sharpless, C.M. and McGown, L.B. (1999) Effects of aluminium-induced aggregation on the fluorescence of humic substances. **Environmental Science and Technology**, 33: 3264-3270
- Simpson, D.R. (2008) Biofilm processes in biologically active carbon water purification – Review. **Water Research**, 42: 2839-2848
- Singer, P.C. (1994) Control of disinfection by-products in drinking water. **Journal of Environmental Engineering-ASCE**, 120(4): 727-744
- Skibsted, E., Lindemann, C., Roca, C. et al. (2001) On-line bioprocess monitoring with a multi-wavelength fluorescence sensor using multivariate calibration. **J. Biotechnology**, 88(1): 47–57
- Smith, C.B., Anderson, J.E., Webb, S.R. (2004) Detection of Bacillus endospores using total luminescence spectroscopy. **Spectrochimica Acta Part A**, 60: 2517-2521
- Smits, J.R.M., Melssen, W.J., Buydens, L.M.C. et al. (1994) Using artificial neural networks for solving chemical problems. Part I. Multi-layer feed-forward networks. **Chemometrics and Intelligent Laboratory Systems**, 22(2): 165-189
- Soh, Y.C., Roddick, F., van Leeuwen, J. (2008) The impact of alum coagulation on the character, biodegradability and disinfection by-product formation potential of reservoir natural organic matter (NOM) fractions. **Water Science and Technology**, 58: 1173-1179
- Sohn, J., Amy, G., Cho, J. et al. (2004) Disinfectant decay and disinfection by-products formation model development: chlorination and ozonation by-products. **Water Research**, 38: 2461-2478

- Song, H., Orr, O., Hong, Y. et al. (2009) Isolation and fractionation of natural organic matter: evaluation of reverse osmosis performance and impact of fractionation parameters. **Environmental Monitoring and Assessment**, 153(1-4): 307-321
- Spencer, R.G.M., Baker, A., Ahad, J.M.E. et al. (2007) Discriminatory classification of natural and anthropogenic waters in two U.K. estuaries. **Science of Total Environment**, 373: 305-323
- Spencer, R.G.M., Stubbins, A., Hernes, P.J. et al. (in press) Photochemical degradation of dissolved organic matter and dissolved lignin phenols from the Congo River. **Journal of Geophysical Research – Biogeosciences**, 114, doi:10.1029/2009JG000968
- Stedmon, C.A. and Markager, S. (2005) Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. **Limnology and Oceanography**, 50(2): 686-697
- Stedmon, C.S., Markager, S., Bro, R. (2003) Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. **Marine Chemistry**, 82(3-4): 239-254
- Stevens, A.A., Slocum, C.J., Seeger, D.R. et al. (1976) Chlorination of organics in drinking water. **Journal of American Water Works Assoc.**, 68(11): 615-620
- Stewart, A.J. and Wetzel, R.G. (1980) Asymmetrical relationships between absorbance, fluorescence, and dissolved organic carbon. **Limnology and Oceanography**, 26: 590–597
- Świetlik, J. and Sikorska, E. (2004) Application of fluorescence spectroscopy in the studies of natural organic matter fractions reactivity with chlorine dioxide and ozone. **Water Research**, 38: 3791-3799

- Thacker, S.A., Tipping, E., Baker, A. et al. (2005) Development and application of functional assays for freshwater dissolved organic matter. **Water Research**, 39: 4559-4573
- Thomas, J.D. (1997) The role of dissolved organic matter, particularly free amino acids and humic substances, in freshwater ecosystems. **Freshwater Biology**, 38: 1–36
- Thompson, B. (1995) Stepwise regression and stepwise discriminant-analysis need not apply here – a guideline. **Educational and Psychological Measurement**, 55(4): 525-534
- Thurman, E. M. (1985) **Organic Geochemistry of Natural Waters**. Dordrecht: Martinus Nijhoff/Dr W. Junk Publishers
- Twardowski, M.S., Boss, E., Sullivan, J.M. et al. (2004) Modeling the spectral shape of absorption by chromophoric dissolved organic matter. **Marine Chemistry**, 89: 69– 88
- Ulsch, A. (1993) “Self-organizing Neural Networks for visualization and classification”. In: O. Opitz, Lausen, B., Klar, R. (eds.) **Information and classification**. Berlin: Springer-Verlag, pp. 307-313
- US EPA (1999) **Research plan for microbial pathogens and disinfection by-products in drinking water**.
- US EPA (2006) **The Occurrence of Disinfection By-Products (DBPs) of Health Concern in Drinking Water: Results of a Nationwide DBP Occurrence Study**. Athens, GA
- Uyak, V. and Toroz, I. (2006) Disinfection by-product precursors reduction by various coagulation techniques in Istanbul water supplies. **J. of Hazardous Materials**, 141(1): 320-328
- Uyak, V., Ozdemir, K., Toroz, I. (2007) Multiple linear regression modeling of disinfection by-products formation in Istanbul drinking water reservoirs. **Science of the Total Environment**, 378: 269-280

- van Leeuwen, J., Daly, R., Holmes, A. (2005) Modeling the treatment of drinking water to maximize dissolved organic matter removal and minimize disinfection by-product formation. **Desalination**, 176(1-3): 81-89
- Velten, S., Hammes, F., Boller, M. et al. (2007) Rapid and direct estimation of active biomass on granular activated carbon through adenosine tri-phosphate (ATP) determination. **Water Research**, 41: 1973-1983
- Volk, C., Bell, K., Ibrahim, E. et al. (2000) Impact of enhanced and optimized coagulation on removal of organic matter and its biodegradable fraction in drinking water. **Water Research**, 34(12): 3247-3257
- Waiser, M.J. and Robarts, R.D. (2004) Photodegradation of DOC in a shallow prairie wetland: evidence from seasonal changes in DOC optical properties and chemical characteristics. **Biogeochemistry**, 69: 263–284
- Wang, G.-S. and Hsieh, S.-T. (2001) Monitoring natural organic matter in water with scanning spectrophotometer. **Environment International**, 26: 205-212
- Weishaar, J.L., Aiken, G.R., Bergamaschi, B.A. et al. (2003) Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. **Environmental Science and Technology**, 37(20): 4702-4708
- Westerhoff, P., Aiken, G., Amy, G. et al. (1999) Relationship between the structure of Natural Organic Matter and its reactivity towards molecular ozone and hydroxyl radicals. **Water Research**, 33(10): 2265-2276
- White, E.M., Baughan, P.P., Zepp, R.G. (2003) Role of the photo-Fenton reaction in the production of hydroxyl radicals and photobleaching of colored dissolved organic matter in a coastal river of the southeastern United States. **Aquatic Sciences**, 65: 402-414

- Wildrig, D.L., Gray, K.A., McAuliffe, K.S. (1996) Removal of algal-derived organic material by preozonation and coagulation: Monitoring changes in organic quality by pyrolysis-GC-MS. **Water Research**, 30(11): 2621-2632
- Wilson, H.F. and Xenopoulos, M.A. (2009) Effects of agricultural land use on the composition of fluvial dissolved organic matter. **Nature Geosciences**, 2: 37-41
- Win, Y.Y., Kumke, M.U., Specht, C.H. et al. (2000) Influence of oxidation of Dissolved Organic Matter (DOM) on subsequent water treatment processes. *Water Research*, 34(7): 2098-2104
- Wolf, G., Almeida, J.S., Pinheiro, C. et al. (2001) Two-dimensional fluorometry coupled with artificial neural networks: A novel method for on-line monitoring of complex biological processes. **Biotechnology and bioengineering**, 72(3): 297-306
- Wolf, G., Almeida, J., Crespo, J.G. et al. (2007) An improved method for two-dimensional fluorescence monitoring of complex bioreactors. **Journal of Biotechnology**, 128: 801-812
- Wu, W.W., Chadik, P.A., Davis, W.M. et al. (2000) "The effect of structural characteristics of humic substances on disinfection by-product formation in chlorination". In: Barrett, S.E., Krasner, S.W. & Amy, G.L. (eds), **Natural Organic Matter and Disinfection By-Products-Characterization and Control in Drinking Water: ACS Symposium Series 761**. American Chemical Society, Washington, DC.
- Wu, F.C., Evans, R.D., Dillon, P.J. (2003a) Separation and characterization of NOM by high-performance liquid chromatography and on-line three-dimensional excitation emission matrix fluorescence detection. **Environmental Science and Technology**, 37(16): 3687-3693

- Wu, W.W., Chadik, P.A., Delfino, J.J. (2003b) The relationship between disinfection by-product formation and structural characteristics of humic substances in chloramination. **Environmental Toxicology and Chemistry**, 22(12): 2845-2852
- Wu, F.C., Kothawala, D.N., Evans, R.D. et al. (2007) Relationship between DOC concentration, molecular size and fluorescence properties of DOM in a stream. **Applications in Geochemistry**, 22: 1659-1667
- Yamashita, Y. and Tanoue, E. (2003) Chemical characterization of protein-like fluorophores in DOM in relation to aromatic amino acids. **Marine Chemistry**, 82: 255– 271
- Yang, X., Shang, C., Lee, W. et al. (2008) Correlations between organic matter properties and DBP formation during chloramination. **Water Research**, 42: 2329-2339
- Yentsch, C.S. and Phinney, D.A. (1985) Spectral fluorescence: an ataxonomic tool for studying the structure of phytoplankton populations. **J. Plankton Research**, 7(5): 617– 632
- Yu, J., Wang, D., Yan, M. et al. (2007) Optimized coagulation of high alkalinity low temperature and particle water: pH adjustment and polyelectrolytes as coagulant aids. **Environmental Monitoring Assessment**, 131: 377-386
- Zhang, H., Qu, J., Liu, H. et al. (2009) Characterization of isolated fractions of dissolved organic matter from sewage treatment plant and the related disinfection by-products formation potential. **Journal of Hazardous Materials**, 164: 1433–1438
- Zika, R. G. (1981) “Marine organic photochemistry”. In: Duursma, E. and Dawson, R. (eds.), **Marine organic chemistry: Evolution, composition, interactions, and chemistry of organic matter in seawater**. Vol. 31, Elsevier Sci. Publ, pp. 299-325
- Zupan, J. and Gasteiger, J. (1991) Neural networks: A new method for solving chemical problems or just a passing phase? **Analytica Chimica Acta**, 248(1): 1-30