



University
of Glasgow

Lenselink, Jeanine N. (2023) *Understanding water chemistry in biofilters using chemometric methods*. PhD thesis.

<https://theses.gla.ac.uk/83488/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk

Understanding water chemistry in biofilters using chemometric methods

Jeanine N. Lenselink

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

School of Engineering
College of Science and Engineering
University of Glasgow



May 2022

Abstract

Biological activated carbon filtration operating in slow flowing mode is a compromise between slow sand filtration and rapid-rate biological activated carbon filtration. It brings together the benefits of both systems: good control of microbial pathogens and dissolved organic matter removal. In this thesis the change in water chemistry in such biofilters was studied as a possible alternative for small-scale decentralised water treatment in Scotland.

Change in chemical water quality was monitored in the influent, effluent, and pore water of lab-scale biofilters of varying lengths. Also, the effect of pore water chemistry on carbon processing by microbial communities from various filter depths was studied by batch experiment. Finally the effect of spatial and temporal water quality variation in Scotland on biofilter performance was studied via a metadata study.

Results showed that dissolved organic matter was removed via multistage adsorption with filter length having a positive impact on the removal efficiency in apparent steady state. Applying the biofilters to other Scottish fresh waters with lower DOM concentrations will increase their performance. Within the filters, different ecological niches were formed. The community at the top processed the easily available low molecular weight acids, the community at the bottom was able to degrade the more recalcitrant humic substances. This indicates that specific microbes inhabit this bottom niche and are able to survive on the limited choice of DOM species.

These research findings contribute to further optimisation of the biofilter design and for the scientific community to further understand biofilter stratification and its impact on DOM removal.

Contents

Abstract	i
Contents	ii
List of Tables	vii
List of Figures	ix
Acknowledgements	xix
Declaration	xxi
1 Introduction	1
2 Literature review	5
2.1 Organic matter	5
2.1.1 Classification	5
2.1.2 Analytical techniques	9
2.1.3 LC-OCD	10
2.1.4 Fluorescence excitation emission matrix spectroscopy	11
2.1.5 BDOC and AOC	13
2.2 Water sources and treatment	14
2.3 Biofiltration	16
2.3.1 Slow and rapid sand filtration	17
2.3.2 BAC filtration	20
2.3.3 Design parameters	29
2.4 Chemometrics	35
2.4.1 Data driven science	35
2.4.2 Data science	36
2.4.3 Data mining	36
2.4.4 Machine learning	37
2.4.5 Deep learning and artificial intelligence	38

2.4.6	Chemometrics	38
2.4.7	Nomenclature and usage	38
2.4.8	Clustering	39
2.4.9	Principle component analysis	40
2.4.10	N-way PCA	42
2.4.11	CANDACOMP/PARAFAC	43
2.4.12	Tucker3	45
2.4.13	PARAFAC and Tucker3 preprocessing	49
2.4.14	Final remarks	50
3	Characterisation of DOM removal and other chemical water quality parameters by laboratory-scale slow-flow BAC biofilters	51
3.1	Introduction	52
3.2	Materials methods	53
3.2.1	Raw water	53
3.2.2	Granular activated carbon	53
3.2.3	Biofilter design	54
3.2.4	Filter bed packing and initiation	56
3.2.5	Experimental run	56
3.2.6	Sampling campaign	57
3.2.7	Chemical analysis	57
3.3	Results and Discussion	62
3.3.1	Water quality and regulatory limits	62
3.3.2	Removal efficiency	64
3.4	Conclusion	78
4	Fluorescence excitation-emission measurements of DOM using a plate reader coupled to PARAFAC analysis	79
4.1	Introduction	80
4.2	Material and methods	82
4.2.1	Plate reader method	82
4.2.2	Gold standard method	84
4.2.3	Data pre-processing	84
4.2.4	PARAFAC	85

4.2.5	Method validation	86
4.2.6	Method application	89
4.3	Results and Discussion	91
4.3.1	PARAFAC component selection	91
4.3.2	Method applicability	97
4.4	Conclusion	105
5	Understanding removal processes of DOM and other chemical parameters by slow-flow BAC filtration using the analysis of pore water	107
5.1	Introduction	109
5.2	Materials methods	111
5.2.1	Biofiltration deconstruction	111
5.2.2	Pore water extraction	111
5.2.3	Analytical techniques	113
5.2.4	Statistical analysis	114
5.3	Results and Discussion	117
5.3.1	Fluorescence components	117
5.3.2	Variable correlation	119
5.3.3	Depth, variables, and time	120
5.3.4	Pore water chemistry for different filter lengths	132
5.3.5	Microbial activity or adsorption	136
5.4	Conclusion	138
6	Differential utilisation of dissolved organic matter compound groups by different biofilter microbial communities	139
6.1	Introduction	140
6.2	Materials Methods	142
6.2.1	Glassware preparation	142
6.2.2	AOC-free <i>Inocula</i> preparation	142
6.2.3	Experimental set-up	143
6.2.4	Cell Abundance Specific growth rate	143
6.2.5	DNA extraction and 16S rRNA gene sequencing	144
6.2.6	16S rRNA gene Sequencing Analyses	144
6.2.7	LC-OCD-UVD-OND sample preparation	145
6.2.8	Statistical analysis	146

6.3	Results and discussion	147
6.3.1	Increase in cell abundance	147
6.3.2	Microbial community structure	148
6.3.3	DOC concentration	149
6.3.4	DOM degradation	150
6.3.5	Differences	153
6.4	Conclusion	155
7	Three-way PCA to study spatial and temporal variation in water quality parameters and its impact on biofilter performance	157
7.1	Introduction	158
7.2	Materials and methods	159
7.2.1	Sampling sites	159
7.2.2	Sampling period	159
7.2.3	Variables	159
7.2.4	Data pre-treatment	159
7.2.5	Substituting missing data	161
7.2.6	Normality distribution and outliers	161
7.2.7	Data pre-processing	161
7.2.8	PCA	162
7.2.9	Tucker3	162
7.2.10	Other statistical analyses	162
7.3	Results and Discussion	163
7.3.1	Spatial and temporal variation within Scotland	163
7.3.2	IN and Pateshill comparison	169
7.3.3	Impact of spatial and temporal variation	172
7.4	Conclusion	174
8	Conclusion	175
8.1	Restatement of research aim and objectives	175
8.2	Summary of key findings of chapter 3 to 7	176
8.2.1	Characterization of DOM and other chemical water quality parameters removal by laboratory-scale slow-flow BAC biofilters	176
8.2.2	Fluorescence excitation–emission measurements of DOM using a plate reader coupled to PARAFAC analysis	176

8.2.3	Understanding removal processes of DOM and other chemical parameters by slow-flow BAC filtration using the analysis of pore water	176
8.2.4	Differential utilisation of dissolved organic matter compound groups by different biofilter microbial communities	177
8.2.5	Three-way PCA to study spatial and temporal variation in water quality parameters and its impact on slow-flow BAC performance	177
8.3	Conclusion and recommendations for future research	178
A	Appendix chapter 3	181
B	Appendix chapter 4	191
B.1	Spectral correction	191
B.2	EEMqual	193
C	Appendix chapter 5	195
C.1	PARAFAC in- and effluent water	197
D	Appendix chapter 6	203
E	Appendix chapter 7	205
	Bibliography	209

List of Tables

2.1	<i>Slow sand filter treatment efficiencies in filter units operated at filtration rates in the range of 0.04 and 0.20 mh⁻¹, temperature above 5°C, and sand bed depths greater than 0.5 m. From Galvis et al. (2002)</i>	19
2.2	<i>Drinking water treatment studies looking at the influence of zone, EBCT and type of filter media on TOC/DOC removal.</i>	33
3.1	<i>Water quality averaged between August to November 2017 provided by Scottish Water</i>	53
3.2	<i>Characteristics Cabot Norit GAC 1240 W used as medium for the biofilters.</i>	54
3.3	<i>Biofilter characteristics and the corresponding operational settings.</i>	54
3.4	<i>Sampling campaign</i>	57
3.5	<i>chemical water quality parameters min, max and mean concentrations in the influent and effluent water of the biofilter together with their regulatory limits. Underscored numbers are measured values above the regulatory limits.</i>	62
4.1	<i>Details of the validation data sets and corresponding PARAFAC models.</i>	90
4.2	<i>Validation parameters of the PR method.</i>	96
4.3	<i>OpenFluor results of the components C1 and C2 present in the water samples analysed by the PR (PR C1 and PR C2) and GS (GS C1 and GS C2) method.</i>	101
4.4	<i>Quantitative specificity of the PR method; signal recovery (%) of QS and TR in the Wiericke and Driebruggen samples analysed by the PR and GS method.</i>	103
5.1	<i>OpenFluor results of the components C1 – C3 present in the pore water samples.</i>	119
5.2	<i>Meaningful interactions for elements g111 and g222 of the Tucker3 [2,2,2]-model of data set Week.</i>	124
5.3	<i>Meaningful interactions for the elements g111 and g122 of the Tucker3 [1,2,2]-model of data set Length.</i>	133
6.1	<i>LC-OCD-OND DOM fractions and corresponding molecular weights and description.</i>	146
6.2	<i>Alpha diversity parameters for the pooled triplicates Sample Number of High quality reads Richness Fisher Alpha Shannon.</i>	149
7.1	<i>Description of the Scottish sampling sites with the variables analysed over a time period.</i>	160

7.2	<i>Meaningful interactions for the elements G111, G112, and G222 of the Tucker3 [2,2,2]-model of the SW data set.</i>	166
A.1	<i>Heatmap sample names</i>	189
D.1	<i>Measured LC-OCD fraction concentration (mean \pm se, n=3) at T = 0 and in the three treatments (TOP, MID, BOT) at T = 23.</i>	203
E.1	<i>data set Waterquality pre-treatment</i>	205

List of Figures

2.1	<i>Path of organic matter from vegetation or animal (waste) to produced drinking water. Rain takes decaying vegetation and animal waste via surface and subsurface flows into the creeks, rivers, and lochs. Drinking water treatment plants use this fresh water and remove the DOC from the water resulting in high quality drinking water that is transported to the houses.</i>	5
2.2	<i>Simplified classification of DOC with respect to its hydrophobicity and acidity.</i>	6
2.3	<i>Chemical formulas of a humic substance molecule (A), carbohydrate molecule (glucose) (B), amino acid molecule structure (C), and LMW molecule (propanone) (D).</i>	8
2.4	<i>Graphical representation of the way DOC, BDOC, and AOC are linked to each other; AOC is part of the BDOC, and BDOC again part of the DOC.</i>	9
2.5	<i>LC-OCD chromatogram with the various DOM fractions and their respective molecular weights. From Simon et al. (2013).</i>	10
2.6	<i>Landscape (A) and contour (B) plots of an EEM of Suwannee River Fulvic Acid. From Ryder et al. (2017).</i>	11
2.7	<i>Example of PARAFAC decomposition of leaf leachate. A data set of EEMs (A) is decomposed into six characteristic fluorescent components (B), followed by recombination to achieve semi quantitative differentiation between the A. Saccharinium and P. Glauca samples (C). Each of the two samples has a different relative composition of the six fluorophores. From (Cuss & Guéguen 2015).</i>	12
2.8	<i>Greenhouse gas emissions footprint by source 2018/19. From Scottish Water (2019).</i>	16
2.9	<i>Schematic representation of a slow sand filter. Water flows into the tank and travels down through a sand filter medium covered with a Schmutzdecke to exit at the water outflow. Microbes live in this Schmutzdecke and facilitate the removal of pathogens and biodegradable organic matter. From Bielefeldt (2011).</i>	17
2.10	<i>Adsorption processes taking place in GAC.</i>	21
2.11	<i>Schematic overview of diffusion of molecules from the mobile phase to the immobile phase within the pores of the GAC. From Liu et al. (2021).</i>	23

2.12	<i>Adsorption of humic substances onto different types of activated carbon with AC/H = hydrogen treated activated carbon, AC/N = ammonized activated carbon, AC0= normal activated carbon. All three filter types show a clear bent in their curve, where the first part is the adsorption of molecules to the large pores and the second less steep part is the adsorption of the molecules in the smaller pores. Difference in adsorption rate is the result of retarded diffusion, the rate limiting step of the total adsorption. From Kołodziej et al. (2014).</i>	24
2.13	<i>TOC removal acclimation for a GAC filter, which becomes a BAC filter once the adsorption capacity is exhausted. Initial removal is dominated by adsorption, while at apparent steady state the removal is dominated by biological removal. Based on Terry & Summers (2018)</i>	26
2.14	<i>Boxplot of TOC removal by inert (n=68) and BAC (n=50) media from various biofiltration studies. Results demonstrate the higher removal of TOC by BAC. From Peterson & Summers (2021).</i>	30
2.15	<i>Simulated TOC removal as a function of EBCT at three temperature ranges for ozonated and nonozonated waters with associated k' values (k' = 0.05 min⁻¹ for 10 °C, k' = 0.09 min⁻¹ for 10 - 20 °C, k' = 0.14 min⁻¹ for 20 °C). From Terry & Summers (2018).</i>	32
2.16	<i>Venn diagram explaining the relationship between data science, data mining, artificial intelligence, machine learning and deep learning. From Kulin et al. (2021).</i>	35
2.17	<i>Data mining method applied to a matrix X (A) which includes multiple samples and multiple variables. A PCA model (B) helps to interpret clustering of the data (blue and green datapoints) in a score plot with help of principal component (1 and 2). From Amigo (2021).</i>	36
2.18	<i>Machine learning procedure showing the three stages of calibration of the model with help of a training data set using a part of the known samples (A), validation of the model with help of a test data set with the other part of known samples (B), and prediction new samples with help of the created model (C). From Amigo (2021).</i>	37
2.19	<i>Non-exhaustive overview of data mining/machine learning tools. Machine learning can be divided into unsupervised (dimension reduction, clustering, and association) and supervised learning (regression and classification) tools. Many methods are part of each of the tools . The methods in red are used in this thesis.</i>	39
2.20	<i>An example of a heatmap. 10 discriminant features (variables) in the column direction with both rows and columns ordered using hierarchical (average linkage) clustering to identify blocks of features of interest. From Gauchotte-Lindsay et al. (2019).</i>	40

2.21	<i>Schematic representation of the concept of PCA. The datapoints in the variable space of 3 variables (A), the centring of the datapoints towards the barycentre (B), the principal component 1 and 2 describing the maximum variations (C), the projected datapoints in the new principle component space (D), the PC1 and PC2 scores that describe a datapoint in the projected principle component space (E), and the loadings describing the relationship between the old variable axis and new principle component axis (F). Based on Amigo (2021).</i>	41
2.22	<i>Representation of a data array with dimensions I, J, and K for samples, variables, and conditions, respectively.</i>	42
2.23	<i>PARAFAC decomposition into multiple trilinear components each having a score vector (<i>a</i>) and two loading vectors (<i>b</i> and <i>c</i>) as well as the error that is left. Together with the loading matrices of A (includes <i>a</i>₁, <i>a</i>₂...<i>a</i>_f), B (includes <i>b</i>₁, <i>b</i>₂...<i>b</i>_f), and C (includes <i>c</i>₁, <i>c</i>₂...<i>c</i>_f).</i>	44
2.24	<i>Tucker3 decomposition. The data array X with I (sample), J (variable), and K (conditions) mode is decomposed in a G core with <i>p</i>, <i>q</i> and <i>r</i> factors for the I, J, and K mode, respectively, together with the three loading matrices (A for sample mode, B for variable mode, and C for condition mode). The remaining unexplained data are part of the error ϵ.</i>	46
2.25	<i>PCA biplot of the Learning to read data set with PC1 explains 96.7% of the data and PC2 21%. The pupils are represented by colour, months by symbols and the tests are presented by the vectors.</i>	47
2.26	<i>Tucker3 model outcome for 'Learning to read' data set. With the pupil's loadings (A), variable's loadings (B), and time loadings (C).</i>	48
2.27	<i>Example of a data array X metricized by frontal slicing for j-scaling (scaling in the column direction).</i>	49
3.1	<i>Biofilter design including the influent jerrycan (1), peristaltic pump that pump the influent water into the filter (2), overflow hole (3), peristaltic pump that keeps the flow rate at 1ml/min (4), and the sample collection bottle (5).</i>	55
3.2	<i>Concentration of aluminium (A), arsenic (B), iron (C), TOC (D). Left y-axis shows the measured concentration of the influent and the 30, 60, 90 cm effluent (mean \pm s.e., <i>n</i> = 3 replicates) over 23 weeks of operation. The right axis together with the dashed line (when present) highlights the regulatory limits.</i>	63

3.3	<i>Removal in % of DOC (A), TOC (B), iron (C), nitrite (D), copper (E), chromium (F), lead (G), chloride (H), sulphate (I) by the 30 (green), 60 (blue) and 90 (purple) cm filte (mean ± s.e., n = 3 replicates) within the biofilter during the first 23 weeks of operation.</i>	66
3.4	<i>Removal in % of phosphate (J), arsenic (K), aluminium (L), antimony (M), cadmium (N), manganese (O), nickel (P), selenium (Q) by the 30 (green), 60 (blue) and 90 (purple) cm filter (mean ± s.e., n = 3 replicates) within the biofilter during the first 23 weeks of operation.</i>	67
3.5	<i>Cluster heatmap of chemical water quality variables removal by the biofilter both rows and columns ordered using hierarchical (complete linkage) clustering. Variables clustered by V1 - V3 and biofilter samples clustered by S1 - S3.</i>	68
3.6	<i>pH (A) and conductivity (B) measured in the influent water and 30, 60, and 90 cm filter effluent (mean ± s.e., n = 3 replicates) of the biofilter over 23 weeks of operation showing the data points of the three filter lengths.</i>	70
3.7	<i>Scatter plot of the conductivity versus chloride concentration measured in the biofilters during the 23 weeks of operation.</i>	71
3.8	<i>SEM-EDX image and spectra of fresh GAC (A) and GAC from the top 10 cm of the biofilter at W23 (B).</i>	72
3.9	<i>DOC removal as function of BV (mean, n = 3 replicates) for different filter lengths (30, 60, and 90 cm) including their (dashed) regression lines over the first 15 week of operation.</i>	74
3.10	<i>Trends of DOC (A) and Biomass (B) removal. DOC removal was measured at week 23 for the three filter lengths (n = 3 replicates). Letters (a, b, c) in (A) denote significant differences with regard to DOC removal between filter lengths assessed by Dunn's post-hoc test on KW H-test. Cell counts were measured at week 5, 9, 12, and 23 for the 90 cm filter as well as the 30 and 60 cm filter at week 23 (n = 3 replicates). Letters (a and b) in (B) denote significant differences between filter lengths at week 23 with regard to cell count assessed by Tukey's post-hoc test on ANOVA.</i>	76
4.1	<i>Raleigh and Raman scatter present in an EEM of water. From Lawaetz & Stedmon (2009).</i>	84
4.2	<i>Schematic overview of PARAFAC pre-processing steps.</i>	85

4.3	<i>PARAFAC component excitation (red) and emission (blue) spectra of QS in water (A) and in matrix (B) analysed by the GS method, QS in water (C) and QS in matrix (D) analysed by the PR method, TR in water (E) and in matrix (F) analysed by the GS method, and TR in water (G) and TR in matrix (H) analysed by the PR method.</i>	92
4.4	<i>Significant correlations ($-.6 > R > .6$) between spectra of TR (A) and QS (B) analysed by the PR method PR and the GS method in water (H₂O) and matrix.</i>	93
4.5	<i>Quantitative specificity: signal recovery of 10, 50, 150 µg/L QS (QS10, QS50, QS150) and 50, 150 µg/L TR (TR50, TR150) (n = 1 replicate) for the GS (A) and PR (B) methods.</i>	94
4.6	<i>Calibration curve of TR (A) and QS (B) measured at day one.</i>	95
4.7	<i>EEMs of PARAFAC component 1 and 2 analysed by the GS method(A and B) and component 1 and 2 analysed by the PR method (C and D) from the water sample data set.</i>	98
4.8	<i>PARAFAC component excitation (red) and emission (blue) spectra of water sample data set analysed by the GS method (C1 and C2 are A and B) and PR method (C1 and C2 are C and D).</i>	99
4.9	<i>Significant correlations ($-.6 > R > .6$) between spectra of C1 and C2 analysed by the PR method (PR) and the GS method (GS).</i>	100
4.10	<i>Quantitative specificity of the PR method; the QS normalized signals of various water samples analysed by the PR and GS method (n = 1 replicate) and the ratio between them.</i>	102
5.1	<i>Schematic representation of the cutting of the biofilter. GAC from the strongly black rimmed cuttings were first sampled for DNA analysis. GAC cuttings of each of the colours brown (0-10 cm), red (10-20 cm), purple (20-30 cm), yellow (50-60 cm) and blue (80-90 cm) were combined for further pore water extraction. Grey cuttings were not used for either analyses.</i>	111
5.2	<i>Schematic overview of the pore extraction steps. (1) Wet GAC is centrifuged after which (2) the dry GAC is collected, and the remaining wet GAC is transferred into pipettes without the bottom part. (3) Pipettes are placed in centrifuge tubes and centrifuged at lower pace. (4) Afterwards, the dry GAC is collected again while the supernatant is collected in a new centrifuge tube and (5) centrifuged at high speed to let the GAC dust in the sample settle. Supernatant was collected for further analysis. . .</i>	112
5.3	<i>Example EEMs of the pore water data set. Sample from week 9 at 0-10 cm from the 90 cm filter (A); sample from week 9 at 50-60 cm from the 90 cm filter (B); sample from week 23 at 0-10cm from the 90 cm filter (C).</i>	117

5.4	<i>Qualitative results of the PARAFAC analysis of the pore water data set. Left: the excitation (red) and emission (blue) spectra of the components C1-C3. Right: their corresponding EEMs.</i>	118
5.5	<i>Significant correlations ($-0.6 > R > 0.6$) of the measured chemical parameter in the biofilter over time and depth for various lengths.</i>	120
5.6	<i>Loading plots of the Tucker3 [2,2,2]-model from data set Week with depths (A), chemical variables (B), months (C).</i>	121
5.7	<i>Core of the Tucker3 'Time' [2,2,2]-model (A) and its unfolded core matrix equivalent represented by the frontal planes (B). g111 and g222 have both a high negative value.</i>	122
5.8	<i>Tucker3 [2,2,2]-model loading plots from data set Week with depth loading plot of component 1 (A1), depth loading plot of component 2 (A2), variables loading plot of component 1 (B1), variables loading plot of component 2 (B2), month loading plot of component 1 (C1), and month loading plot of component 2 (C2).</i>	123
5.9	<i>Concentration Cl^- (A), SO_4^{2-} (B), NO_3^- (C) in the pore water of various depths (mean \pm s.e., $n = 3$ replicates) of the slow-flow BAC filter over several timepoints.</i>	125
5.10	<i>Repeated graphs from chapter 3. In- and effluent concentrations of the group 1 variables Cl^- (A), SO_4^{2-} (B), NO_3^- (C) and Na^+ (D) (mean \pm s.e., $n = 3$ replicates) of the 90 cm filter during 23 weeks of operation.</i>	126
5.11	<i>The fluorescence signal of fluorophore peak H1 (A) and peak H2 (B) in the pore water of various depths (mean \pm s.e., $n = 3$ replicates) of the slow-flow BAC filter over several timepoints.</i>	128
5.12	<i>Fluorescence signal of Peak H1 (stripes) and Peak H2 (unstriped) in the influent (red) and effluent (blue) water of the biofilter at various weeks (mean \pm s.e., $n = 3$ replicates).</i>	129
5.13	<i>Fluorescence signal of peak T (A) and the nitrite concentration (B) at various depths (mean \pm s.e., $n = 3$ replicates) of the biofilter measured at various timepoints.</i>	131
5.14	<i>Tucker3 [1,2,2]-model loading plots from data set Length with depth loading plot of component 1 (A1), variables loading plot of component 1 (B1), variables loading plot of component 2 (B2), month loading plot of component 1 (C1), and month loading plot of component 2 (C2).</i>	134
5.15	<i>Concentrations of fluorophore peak T (A) and nitrite (B) at different depths (mean \pm s.e., $n = 3$ replicates) of the biofilter for various filter lengths.</i>	135
5.16	<i>Significant correlations ($-0.6 > R > 0.6$) of fluorophores and other biomass proxies measured in the biofilter over time and depth for various lengths.</i>	136

6.1	<i>Trends of cell abundances (A) and growth-rates (B). Abundances were measured over the course of the experiment for the three treatments (mean \pm s.e., n = 3 replicates). Stars and ns in B denote significant differences with regard to growth rates between treatments assessed by Tukey's post-hoc test on ANOVA. ns denote no significant difference, Stars denote significant differences between treatments (p<0.05).</i>	147
6.2	<i>Heatmap of the communities in the three treatments at genus level with both rows and columns ordered using hierarchical (average linkage) clustering (A). Venn diagram of the ASVs found in the treatments (B).</i>	149
6.3	<i>Boxplot of the DOC concentration at T = 0 and the three treatments (TOP, MID, BOT) after 23h incubation. Letters a, b and c denote significant differences between treatments regarding DOC concentration. Letters (a,b,c) denote significant differences with regard to DOC concentration in the different treatments assessed by Dunn's post-hoc test on KW H-test. Same letters denote no significant difference, different letters denote significant differences between treatments.</i>	150
6.4	<i>Boxplots of humic substance concentration (A), building block concentration (B), and molecularity of the humic substance pool (C) at T = 0 and in the three treatments (TOP, MID, BOT) at T = 23. Letters (a,b,c) denote significant differences in concentration humic substances (A) and building blocks (B), and molecularity of humic substance pool (C) between the different treatments assessed by Dunn's post-hoc test on KW H-test. Same letters denote no significant difference, different letters denote significant differences between treatments.</i>	151
6.5	<i>Boxplots of the LMW neutrals (A), LMW acids (B), and biopolymers (C) at T = 0 and (TOP, MID, BOT) at T = 23h. Letters a, b and c denote significant differences between treatments regarding concentrations. Letters (a,b,c) denote significant differences in concentration LMW neutrals (A) and acids (B), and biopolymers (C) between the different treatments assessed by Dunn's post-hoc test on KW H-test. Same letters denote no significant difference, different letters denote significant differences between treatments.</i>	153
7.1	<i>PCA biplot of the data set Waterquality with principal component 1 explains 52% of the data and principle component 2 39%. The scores grouped by colour, months by symbols and the loadings are presented by the vectors.</i>	163
7.2	<i>Loading plots of the Tucker3 [2,2,2]-model from the data set Waterquality with sampling sites (A), biochemical variables (B), months (C).</i>	164

7.3	<i>G</i> core of the Tucker3 [2,2,2]-model (A) and its unfolded core matrix equivalent represented by the frontal planes (B). <i>G</i> 111 has a high positive value, <i>G</i> 222 a high negative value, while the other values are of minor importance.	165
7.4	Tucker3 [2,2,2]-model loading plots with sample loading plot of component 1 (A1), sample loading plot of component 2 (A2), variables loading plot of component 1 (B1), variables loading plot of component 2 (B2), time loading plot of component 1 (C1), and time loading plot of component 2 (C2).	167
7.5	Average temperature (A) and precipitation (B) in Scotland throughout the year. Recorded by Metoffice and available at https://www.metoffice.gov.uk	168
7.6	Boxplot of all measurements done by Scottish Water between 2015 and 2019 between or variables DOC (A), <i>dFe</i> (B), <i>dMn</i> (C), <i>E. coli</i> (D), and coliform (E). Datapoints of Pateshill in Oct-Dec2018 (10-12) and Jan-March 2019 (1-3) are pointed out with labels and data from IN labelled in red.	171
8.1	TOC removal as function of bed volumes for a GAC filter that becomes a BAC filter once the adsorption capacity is fully exhausted. Initial removal is dominated by multistage adsorption (macropore filling and micropore filling). Removal will asymptotically reach the steady state (at around 20,000 BV) where removal process is mostly dependent on biological removal. Based on Figure 2.13.	179
A.1	Schematic representation of the biofilter when working in recirculating mode. The water is pumped from the influent jerrycan (1), peristaltic pump brings the influent water into the filter (2), overflow hole runs back the excess of water that is pumped in the filter (3), and after the water is run through the filter it is brought back to the jerrycan.(4)	181
A.2	Schematic representation of the carbon combustion method. From Qualls (2015).	182
A.3	Schematic representation of a typical ion chromatography (IC) analysis process. From Thermo Fisher Scientific (2012a).	183
A.4	Schematic representation of ICP-MS. From Radboud University (n.d.).	184
A.5	Schematic representation of SEM-EDX. From RJL Micro & Analytics (n.d.).	184
A.6	Parameters of water quality at consumers' taps provided by the drinking water quality regulator (2020).	186

A.7	<i>Concentration of nitrate (Top left), phosphate (Top right), copper (Bottom left), chromium (Bottom right). Left y-axis shows the measured concentration of the influent and the h= 30, 60, 90 cm effluent (mean ± s.e., n=3 replicates) over 23 weeks of operation. The right axis together with the dashed line (when present) highlights the regulatory limits.</i>	187
A.8	<i>Concentration of lead (Top left), antimony (Top right), cadmium (Middle left), nickel (Middle right), manganese (Bottom left), and selenium (Bottom right). Left y-axis shows the measured concentration of the influent and the h= 30, 60, 90 cm effluent (mean ± s.e., n=3 replicates) over 23 weeks of operation. The right axis together with the dashed line (when present) highlights the regulatory limits.</i>	188
B.1	<i>Normalized fluorescence emission spectra of 4 standards analysed by the Tecan M200 Pro plate reader (measured) and Chen (1967) (theoretical).</i>	192
B.2	<i>Combined emission correction values.</i>	192
B.3	<i>EEMqual for models 1-5 given by A-E, respectively.</i>	193
B.4	<i>EEMqual of models 6 and 7 given by A-B, respectively.</i>	193
C.1	<i>T² - Q residual plot for the detection of outliers in data set Time. No outliers found</i>	195
C.2	<i>T² - Q residual plot for the detection of outliers in data set Length. No outliers found</i>	195
C.3	<i>EEMqual for the pore water data set. High EEMqual is shown for up to 3 component, while at the 4th component the EEMqual drops. As a result EEMqual indicates 3 components as the correct number.</i>	196
C.4	<i>Scree plot of the Tucker3 model of data set Week with the percentage variance explained by the model when using up to 5,8, and 3 components in the P, Q, and R modes.</i>	196
C.5	<i>EEMQual of the biofilter in- effluent data set.</i>	197
C.6	<i>EEM of the PARAFAC component 1 Peak-T (A), component 2 Peak-H1 (B), and component 3 Peak-H2 (C) present in the in effluent water.</i>	198
C.7	<i>LC-OCD fractions and their contribution to the whole DOC content of the biofilter influent water.</i>	199
C.8	<i>qPCR copies measured at the various depths of the biofilter (mean ± s.e., n = 3 replicates) at different weeks (5, 9, 12, and 23) and lengths (week 23: 30, 60, and 90 cm).</i>	200
C.9	<i>ATP per gram wet GAC measured at the various depths of the biofilter (mean ± s.e., n = 3 replicates) at different weeks (5, 9, 12, and 23) and lengths (week 23: 30, 60, and 90 cm).</i>	201
E.1	<i>Hotelling T² - Q residuals plot for the SW data set; no outliers detected.</i>	206

E.2 *PCA scree plot of the data set Waterquality in which the increase of principle components results in a higher fit of the data.* 206

E.3 *Tucker3 model scree plot of the SW data set with the percentage variance explained by the model when using up to 4, 5, and 12 components in the P, Q, and R modes. . .* 207

Acknowledgements

I would like to start with expressing my gratitude to my supervisor, Dr. Caroline Gauchotte-Lindsay, for her support throughout my PhD and her help to manage having a family during my PhD and the Covid period. I would also like to thank my second supervisor, Prof. Cindy Smith, for her support, ideas and help.

I would like to give a special thanks to Marta Vignola for her continuous advice, support, and encouragement, and for helping me to hopefully publish a first paper together with her.

Additionally, I would like to thank Dominic Quinn, Steve Joyce, Marta Vignola, Stephanie Connolly, William Sloan, Graeme Moore for their collaboration in the biofiltration project. Also, I would like to thank William Peveler from the Chemistry department, Ryan Pereira from the Lyell Center, and Tanya Peshkur from Strathclyde University for their expertise and help with the analytical chemistry aspects of the project. I gratefully acknowledge Scottish Water for funding this project, allowing me the opportunity to conduct this research. I give thanks to Anne McGarrity, Julie Russell and other technicians for their lab assistance and technical knowledge.

To my fellow colleagues - Kelly Stewart, Eri Tsagkari – thank you for the encouragement, listening ear, support, and many laughter. Also, this PhD would not have been possible without the support of many others – to those I have not named individually, please know I am very grateful for your contributions to my achievements and research.

Finally, I am indebted to Cees, Milou and Yannick Willems, Antoinette and Henk Lenselink, Elise Doeschot, Arnoud Willems, and other family and friends for their contribution, love, support, listening ear, patience, and encouragement throughout my studies.

Declaration

I certify that the thesis presented here for the examination for a PhD degree at the University of Glasgow is solely my own work other than where I have indicated that it is the work of others and that the thesis has not been edited by a third party beyond what is permitted by the University's PGR Code of Practice.

The copyright of this thesis rests with the author. Without acknowledgement no quotation from it is permitted. I declare that this thesis does not include work presented for another degree. I declare that this thesis has been produced in accordance with the University of Glasgow's Code of Good Practice in Research.

I acknowledge that if issues are raised regarding good research practice, the examination may be postponed pending the outcome of any investigation of the issues.

Name: Jeanine Nadia Lenselink

Registration Number: xxxxxxxx

List of abbreviations

AOC	Assimilable organic carbon
BAC	Biological activated carbon
BDOC	Biodegradable dissolved organic carbon
DOC	Dissolved organic carbon
DOM	Dissolved organic matter
EBCT	Empty bed contact time
EEM	Excitation emission matrix
GAC	Granular activated carbon
GS	Golden standard
IN	Biofilter influent
KW	Kruskal-Wallis
LMW	Low molecular weight
LOD	Limit of detection
LOQ	Limit of quantitation
NOM	Natural organic matter
PARAFAC	Parallel factor
PCA	Principal component analysis
PR	Plate Reader
QS	Quinine sulphate

RSD Relative standard deviation
SUVA Specific ultraviolet absorbance
TOC Total organic carbon
TR Tryptophan
WTW Water treatment works

Chapter 1

Introduction

Scottish Water provides drinking water to the majority of the population of Scotland. Roughly 80% of Scottish Water's drinking Water Treatment Works (WTW) serve only 20% of the population. Overall, Scotland has more than 20,000 water supplies that serve a population of fewer than 50 people, while 10,000 supplies serve only one household (Scottish Water 2018). Installing, maintaining, and managing such large number of small rural water supplies is economically and environmentally costly. Having a simple household solution would be ideal for small communities or single households, but a reliable product is not on the market to date. Either it is too expensive or requires too much maintenance. One of the challenges of these point-of-use technologies are the challenging and variable chemical water quality present in the river and lochs (Scottish Water 2015). To add to the problem, long term UK data shows that Dissolved Organic Carbon (DOC) has increased with this trend expected to continue (Scottish Water 2012). The innovations in point-of-use water treatment systems can help to improve the affordability and quality of drinking water for rural communities and those hardest to reach.

For single stage filtration systems treating surface water, the primary objective is the control of microbial pathogens. Secondly, the control of chemical water quality is important as high concentrations contribute to colour, taste, and smell of the water as well as the possible regrowth of pathogens in the storage or transport system (Matilainen et al. 2011, Volk et al. 1997). Slow flow Biological Activated Carbon (BAC) filtration is a compromise between slow sand filtration and rapid-rate BAC filtration. It brings together the benefits of both systems for off-grid drinking water treatment: good control of microbial pathogens and Dissolved Organic Matter (DOM) removal from the start of operation. The small scale and low flow-rate of the system results in a lower output of treated water compared to the large scale slow sand or rapid-rate BAC filtration. However for single households this should not be an issue, as they only use a fraction of the water normally provided by large scale systems.

Studies concerning filters that use slow flow rates for drinking water treatment are very limited (Moona et al. 2021), and publications using slow-flow to enhance BAC filtration efficiency are not available. In this thesis a large laboratory scale experiment was conducted in which the chemical aspects of the

slow flow BAC filters were studied during the first six months of operation. These results contributed to the further understanding and development of a small-scale water filtration system that can provide drinking water in remote areas in Scotland. This multidisciplinary project is funded by Scottish Water. My contribution to the overall project is looking at the water chemistry passing these filters.

The aim of this thesis is to evaluate slow flow BAC filters as possible application for small scale water treatment systems by looking at the chemical parameters of the water and the influence of filter length with help of chemometric methods. To achieve the aims of this project, the following objectives were identified:

1. To characterise DOM and other chemical water quality parameter removal by pilot filters during the first months of operation and see how filter lengths impacts this process.
2. To develop a simple cost-effective fluorescence by excitation-emission spectroscopy method combined to Parallel Factor Analysis using a plate reader for the analysis of DOM.
3. To further understand removal processes of DOM and other chemical parameters by the biofilters through depth-resolved pore water analysis.
4. To examine the degradation of DOM by filter microbial communities at the different depths of a biofilter.
5. To assess how water source and seasonal changes can impact the biofilters.

Chapter 2 introduces DOM, the various analytical techniques that can be used to measure DOM, as well as the techniques used to remove DOM during drinking water treatment. Chapter 2 also presents a review of literature surrounding biofiltration. Finally, the chapter introduces the concept of chemometrics and highlights the methods used in the thesis.

Chapter 3 discusses objective 1 and presents the study of the change in concentration of DOM and other chemical water quality parameters after filtration by a pilot scale filters. Removal of these chemical water quality parameters over time and impact of filter length are described and evaluated.

Chapter 4 goes into objective 2 and presents the validation of fluorescence excitation-emission matrix spectroscopy combined to parallel factor analysis using a plate reader for the characterization and semi-quantitation of DOM in freshwater samples. A comparison is drawn with the results obtained via a golden standard method.

Chapter 5 focuses on objective 3 and presents the study of the concentration of various chemical water quality parameters within the pores of the BAC filter medium. Concentrations are measured through depth and over time.

Chapter 6 discusses objective 4 and presents the study of the degradation of DOM by filter microbial communities at the different depths of a biofilter. Samples from 3 depths of the filter are collected and their growth, DOM degradation, and community composition studied.

Chapter 7 focuses on objective 5 and presents the data study of spatial and temporal variation in Scottish fresh waters and the impact on the biofilter processes described in chapter 3, 5, and 6.

Finally, Chapter 8 summarises the major findings of this research, and highlights recommendations for further research.

Chapter 2

Literature review

2.1 Organic matter

2.1.1 Classification

Natural Organic Matter (NOM) is a complex mixture of organic compounds. NOM is a breakdown product of plants and other vegetation and is widely abundant in nature. Hydrological flow path through enriched organic and mineral soils take the NOM into the creeks, rivers and lochs (Aitkenhead-Peterson et al. 2007, Muller & Tankéré-Muller 2012, Soulsby et al. 2001), see Figure 2.1. When dissolved in water NOM is referred to as Dissolved Organic Matter (DOM).

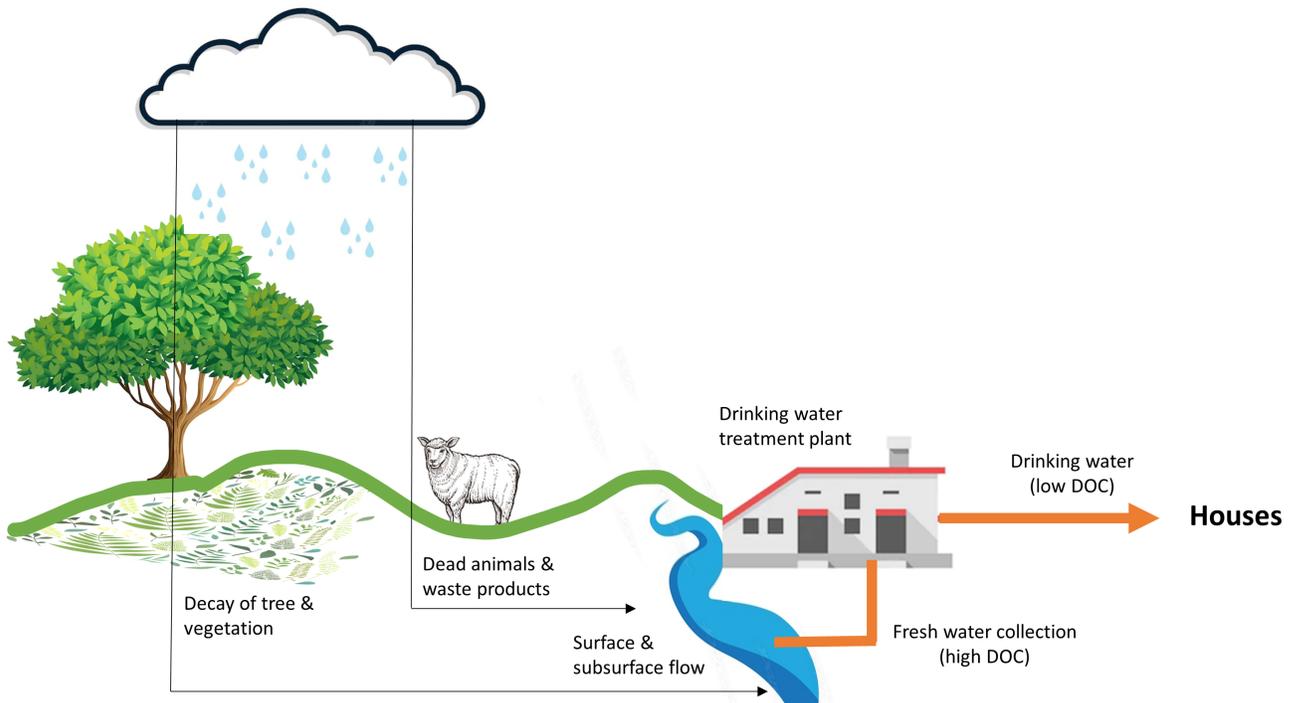


Figure 2.1: *Path of organic matter from vegetation or animal (waste) to produced drinking water. Rain takes decaying vegetation and animal waste via surface and subsurface flows into the creeks, rivers, and lochs. Drinking water treatment plants use this fresh water and remove the DOC from the water resulting in high quality drinking water that is transported to the houses.*

Hereafter, DOM is used instead of NOM, as most organics will be dissolved in waters. DOM (predominantly humic and fulvic acid) can give the characteristic yellow-brown colour to the natural waters because of peaty soils. The presence of DOM is a concern during drinking water treatment because of its influence on water quality. DOM is made up from aromatic and aliphatic hydrocarbon structures that have attached amide, carboxyl, hydroxyl, ketone, and various minor functional groups, which can form aggregates in natural waters increasing the complexity of the DOM (Leenheer & Croue 2003). Removal of DOM improves water quality by removing compounds that create taste, odour and smell (Matilainen et al. 2011). Moreover, it removes compounds that become carcinogenic when getting in contact with disinfectants (Kitis et al. 2002). Also it decreases the microbial growth potential in distribution systems (Matilainen et al. 2011, Peterson & Summers 2021, Terry & Summers 2018). Although millions of individual DOM molecules exist, DOM can be divided in classes depending on specific characteristics.

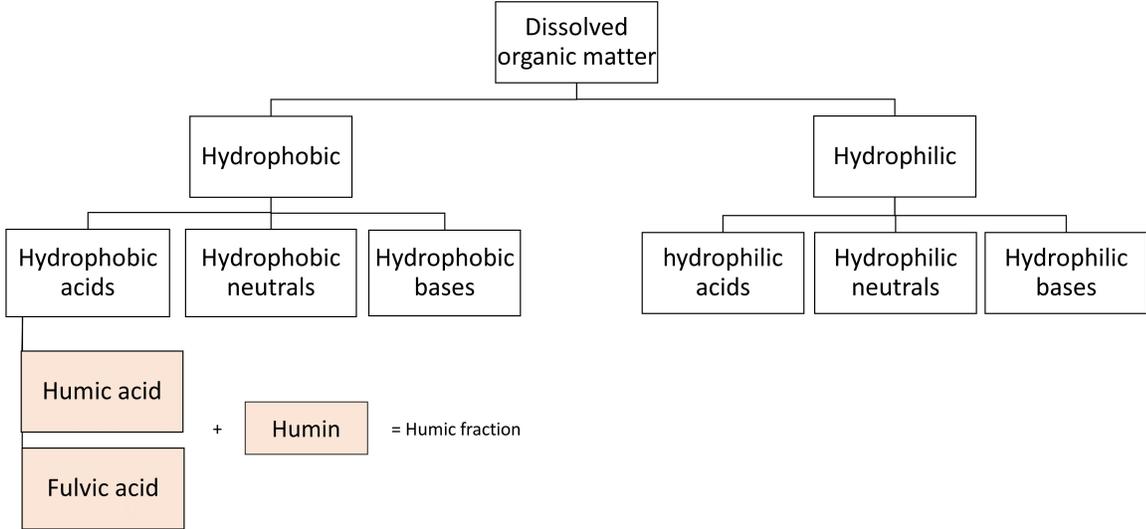


Figure 2.2: Simplified classification of DOC with respect to its hydrophobicity and acidity.

Leenheer & Croue (2003) wrote a frequently cited paper on DOM classification. Figure 2.2 shows a simplified version of this classification and highlights the separation of hydrophobic and hydrophilic acids, bases, neutrals. Two main components that are part of the hydrophobic acids are the highly soluble fulvic acid (soluble at all pH) and slightly less soluble humic acid (soluble at pH > 2) (Critenden et al. 2005, Sillanpää et al. 2014). These humic fractions are the main component of DOM in water supplies, often accounting for up to 50% (American Water Works Association 1999, Sillanpää et al. 2014, VanLoon & Duffy 2005). It is believed that relatively small primary molecular structures of 100-2000 Da form macromolecular structures also known as aggregates (Leenheer & Croue 2003,

Gerke 2018). They are high in aromaticity, low in nitrogen content, contain conjugated double bonds and are anionic polyelectrolytes (American Water Works Association 1999, Crittenden et al. 2005, Matilainen et al. 2010). Carboxyl and phenolic groups cause their negative charge and they contain components like aromatic carbonyl and methoxyl, which can have active surface sites (American Water Works Association 1999, Croue et al. 1999). Hydrophobic neutral examples are hydrocarbons, aliphatic carbon chains > 5 C with alcohols, amides, esters, ketones, and aldehyde groups, > 9 C aliphatic carboxylic acids and aliphatic amines, aromatic carboxylic acids aromatic amines of three rings and greater (Leenheer & Croue 2003, Leenheer & Huffman 1979).

Hydrophilic compounds contain aliphatic carbon chains and nitrogenous compounds and are largely composed of Low Molecular Weight (LMW) organics (Huber et al. 2011, Sillanpää et al. 2014, Matilainen et al. 2011, Świetlik et al. 2004). Carbohydrates and sugars are examples of the hydrophilic neutrals, while proteins and amino acids are basic compounds (Bhatnagar & Sillanpää 2017, Croue et al. 1999, Owen et al. 1995). Although hydrophilic compounds are often classified as LMW organics, previous studies sometimes also refer to them as components with higher MW than the humic fractions (Krasner et al. 1996). These high molecular weight biopolymers are formed from polysaccharides, proteins and amino sugars (Huber et al. 2011). Extracellular polymeric substances are examples of HMW biopolymers, with a MW of approximately 10 kDa. A study by Volk et al. (1997) showed for headwaters of the White Clay Creek in southeastern Pennsylvania DOM was comprised of 72 - 78% humic substances, 9-16% carbohydrates, 2 - 3% amino acids and 15 - 22% OM with a molecular mass > 100 kDa. Example of molecular formulas of the different DOM classes are given in Figure 2.3.

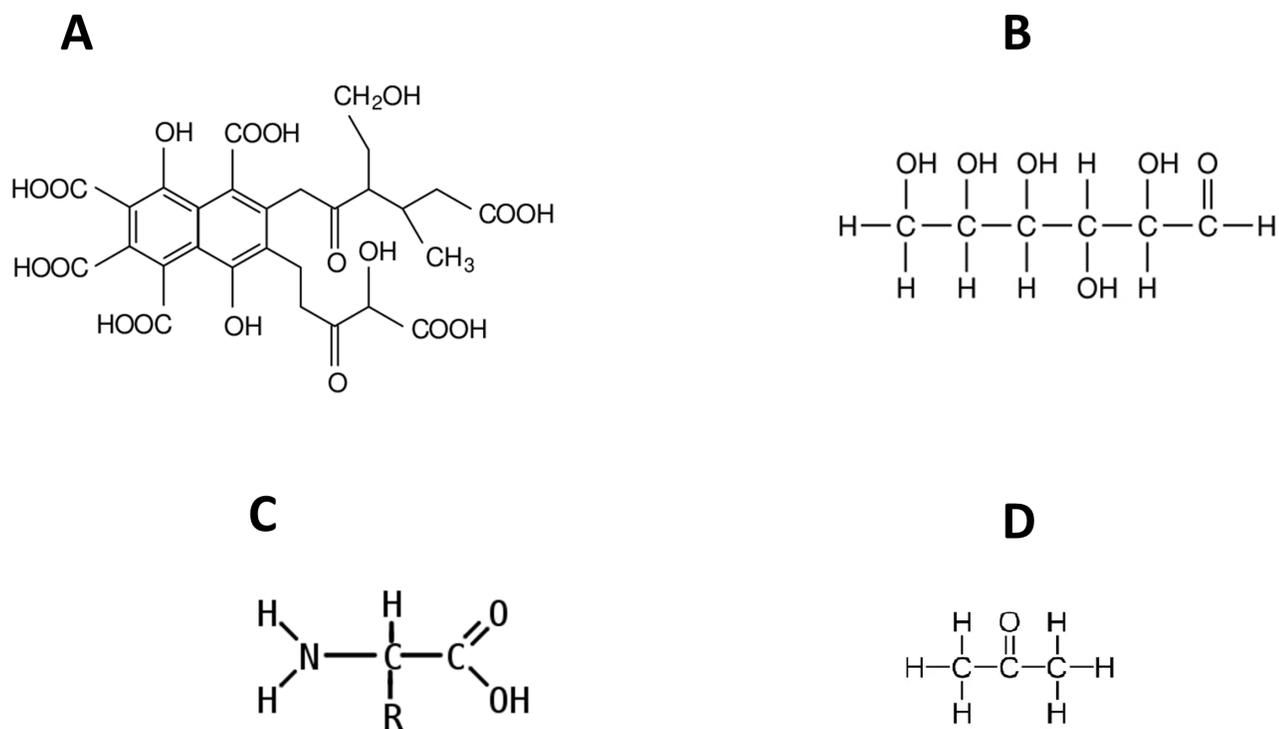


Figure 2.3: Chemical formulas of a humic substance molecule (A), carbohydrate molecule (glucose) (B), amino acid molecule structure (C), and LMW molecule (propanone) (D).

DOM found in drinking water sources includes a biodegradable fraction and a non-biodegradable fraction. This difference can also be used for classification. Biodegradable organic matter is of interest for drinking water utilities as it provides a carbon source for microbial regrowth in the distribution system. The collective measure of Biodegradable Dissolved Organic Carbon (BDOC) and Assimilable Organic Carbon (AOC) are the most commonly used parameters to measure biodegradable organic matter under aerobic conditions (Escobar & Randall 2001, Hammes 2008). BDOC measures change in organic carbon content sample due to microbial metabolism. BDOC represents the portion of DOC that is biodegraded by heterotrophic microorganisms for energy production resulting in the production of CO₂ or other organic substances and the growth of biomass, while AOC is the portion of BDOC that is most readily used by bacteria and converted to cell mass (Escobar & Randall 2001, Juhna & Melin 1985, Terry & Summers 2018) (Figure 2.4). AOC and BDOC can be seen as complementary measures that are both important indicators for water biostability (Escobar & Randall 2001).

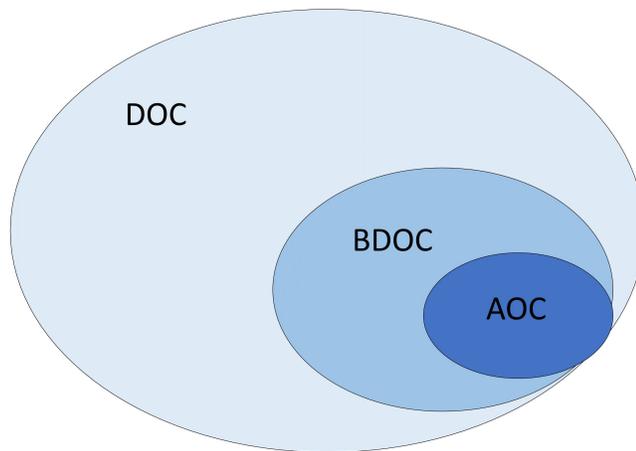


Figure 2.4: *Graphical representation of the way DOC, BDOC, and AOC are linked to each other; AOC is part of the BDOC, and BDOC again part of the DOC.*

2.1.2 Analytical techniques

Depending on the classification DOM can be measured by various techniques. Matilainen et al. (2011) published a comprehensive literature review on the methods used in the characterization of DOM in relation to drinking water treatment. Analytical methods can be grouped as (1) general parameters, (2) biological test, (3) isolation and concentration, (4) fractionation, and (5) Element composition and structures. For the scope of this thesis only the general parameters that are commonly used during drinking water monitoring, fractionation by Liquid Chromatography Organic Carbon Detection, also known as LC-OCD, fluorescence spectroscopy and biological tests will be further discussed. These methods are not used interchangeably but complement each other, as they all measure a different characteristic of the DOM pool.

In practice, NOM/DOM is monitored by the measurement of TOC, DOC, colorimetric tests, absorption of UV-light at 254 nm (254) or Specific UV Absorbance (SUVA). TOC and DOC are the most convenient parameters for analysing. TOC is the sum of all organic carbons in the water including the particulate and DOC is the fraction of TOC after filtration, mostly through a 0.45 µm filter, when existing inorganic carbon is removed by acidification. TOC and DOC are both quantitative measures where all available (dissolved) organic carbon in a water sample is combusted, converted to CO₂, and its CO₂ concentration measured. It gives a bulk number of the total amount of organic carbon in the sample but gives no information on OM character. Similarly, water sample colour only gives an indication of the concentration of humic and fulvic substances, as these substances give a distinct

colour in fresh water as a result of breakdown product of leaf litter. This measure, however, does not give information on other fractions of OM (Uyguner-Demirel & Bekbolet 2011). Specific UV absorbance at 254 nm, often used as proxy for DOC, only gives an indication of the aromaticity of the organic matter (Sillanpää et al. 2014). SUVA is the UV absorbance of a given sample at 254 nm divided by the DOC concentration of the sample. This ratio describes the nature of DOM in the water in terms of hydrophobicity and hydrophilicity (Matilainen et al. 2011); SUVA larger than 4 indicates mainly hydrophobic and especially aromatic material, whilst a SUVA smaller than 3 illustrates mainly hydrophilic material.

2.1.3 LC-OCD

A technique that is gaining popularity, is analysis by size-exclusion chromatography in combination with organic carbon detection, UV detection and organic nitrogen detection known as LC-OCD-UVD-OND (Huber et al. 2011) and is an established method to separate the DOM pool to identify five different DOM classes (Figure 2.5).

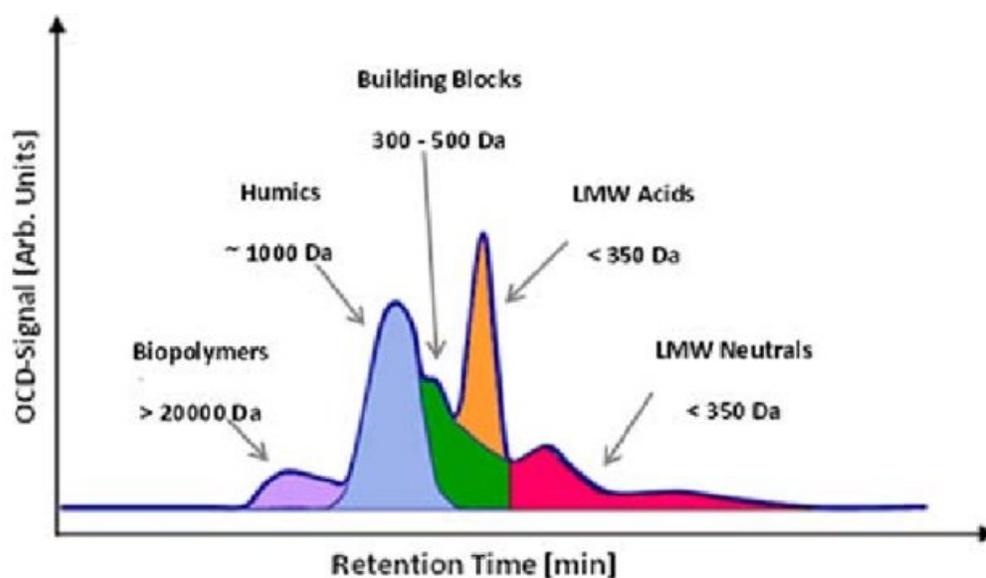


Figure 2.5: LC-OCD chromatogram with the various DOM fractions and their respective molecular weights. From Simon et al. (2013).

These classes include (1) Biopolymers that are likely hydrophobic, HMW (10,000 g/mol or higher), largely non-UV-absorbing Extracellular polymeric substances with saturated structures, polysaccharides and some contribution of proteins or amino sugars (Huber et al. 2011); (2) Humic Substances;

with higher molecular weight at 1000 g/mol with UV-absorbing aromatic molecular aggregates of relatively small molecules, stabilised by the hydrophobic effect and hydrogen bonds (Gerke 2018); (3) Building blocks that are UV-absorbing humics of lower-molecular-weight (300 – 500 g/mol) that have been shown to include microbial breakdown products of humic substances (Huber et al. 2011, Velten et al. 2011a); (4) Low-Molecular-Weight neutrals; 350 g/mol) characterised as non-UV-absorbing, weakly or uncharged hydrophilic or amphiphilic compounds that can include alcohols, aldehydes, ketones, and amino acids hydrophilic; and (5) LMW acids; 350 g/mol).

2.1.4 Fluorescence excitation emission matrix spectroscopy

Fluorescence spectroscopy is a spectrochemical method in which the analyte molecules are excited by light at a certain wavelength and the emitted radiation is measured at a different wavelength. Fluorescence spectroscopy has become popular in the water industry due to its potential application as a monitoring technique. An advantage of fluorescence techniques over traditional methods, such as UV254 and colorimetric tests, is the better sensitivity and selectivity (Bridgeman et al. 2011). The 3-D fluorescence Excitation–Emission Matrix (EEM) spectrophotometric technique gives a 3D contour map of excitation wavelength vs. emission wavelength vs. fluorescence intensity and visualises a range of different fluorophores covering the excitation and emission wavelengths range from 250 nm to 500 nm (Figure 2.6).

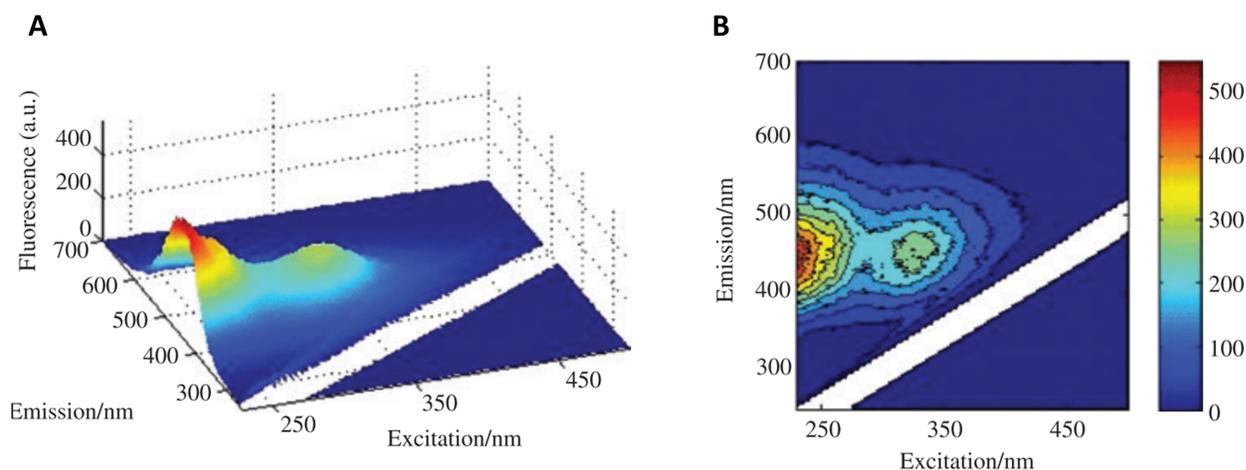


Figure 2.6: *Landscape (A) and contour (B) plots of an EEM of Suwannee River Fulvic Acid. From Ryder et al. (2017).*

The combination of the intensity of each component within a sample makes up for the total fluorescence signal at any excitation-emission pair in the EEM. Spectroscopic EEM measurements can be performed directly on complex DOM mixture. Often the signals of different fluorophores overlap, presenting a challenge for the interpretation of the 3D plot. A variety of advanced data analysis techniques exist for the understanding of fluorescence EEM including peak picking (Coble 1996), fluorescence regional integration (FRI) (Chen et al. 2003), Principal Component Analysis (PCA) (Persson & Wedborg 2001), PARAllel FACTor (PARAFAC) analysis (Murphy et al. 2013b, Stedmon & Bro 2008) and Self-Organizing Maps (Bieroza et al. 2011). In the chemical sciences, PARAFAC analysis introduced by Harshman (1970) is used to decompose trilinear multi-way data arrays and facilitates the identification and quantitation of independent underlying signals, termed ‘components’ (Murphy et al. 2013b). For each component (analyte), the pure excitation and emission spectra are found as well as their relative concentration in the sample. Cuss & Guéguen (2015) included a clear figure of the PARAFAC process of their samples as shown in Figure 2.7.

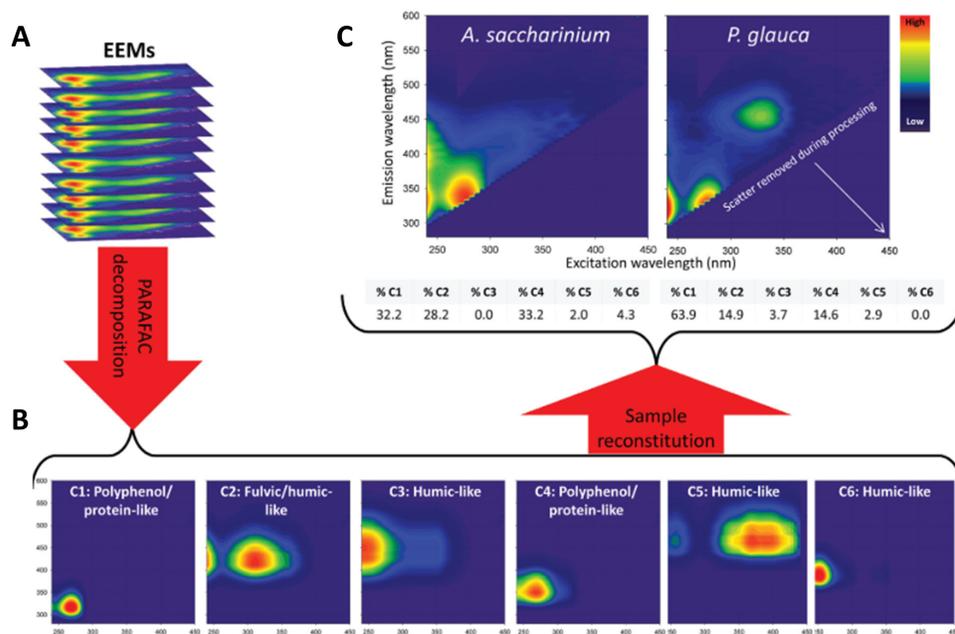


Figure 2.7: Example of PARAFAC decomposition of leaf leachate. A data set of EEMs (A) is decomposed into six characteristic fluorescent components (B), followed by reconstitution to achieve semi quantitative differentiation between the *A. Saccharinum* and *P. Glauca* samples (C). Each of the two samples has a different relative composition of the six fluorophores. From (Cuss & Guéguen 2015).

Top left (A) the EEMs of various samples are shown stacked on top of each other creating a data array with fluorescence signal intensity as content. After PARAFAC decomposition using six components these six components are visualised at the bottom (B). Then, for two samples (*A. saccharinium* and *P. Glauca*) the relative presence of each component is given in percentage (C). For example, in the *A. saccharinium* sample the signal of the first fluorophore (here called C1) makes up for 32.2% of the total fluorescence signal, while in *P. Glauca* C1 makes up for 63.9% of the total fluorescence signal.

In raw water, EEMs typically contain two major fluorescence peaks described as humic-like and protein-like fluorescence maxima (Wu et al. 2003, Bagthoth et al. 2011). Humic-like fluorescence can be divided in humic-like and fulvic-like fluorescence (Chen et al. 2003), while protein-like fluorescence includes tryptophan-like and tyrosine-like fluorescence (Hudson et al. 2008). Tryptophan (e.g., traditionally defines peak T) has its excitation/emission wavelengths at 270–290 / 340–360 nm, while tyrosine (e.g., traditionally defines peak B) has its peak maxima at 270–280 / 300–320 nm (Hudson et al. 2008, Sillanpää et al. 2014, Yang et al. 2015). These components have been associated with an autochthonous source (Coble et al. 2014) and are believed to be compounds containing at least one aromatic ring (Barsotti et al. 2016), including amino acids, DNA, lignin, and polycyclic aromatic hydrocarbons (PAHs). (Carstea et al. 2016). Detection indicates possible microbial activity (Elliott et al. 2006, Henderson et al. 2009), more specifically, microbial metabolic activity (Fox et al. 2017).

Humic-like components show variations within different studies, commonly ranging between Ex/Em maxima 250–400 / 372–495 nm. Compounds contributing to these region contain two or more aromatic rings (Barsotti et al. 2016), are from an allochthonous origin of terrestrial input (Coble et al. 2014) and include among other things humic substances, lignin, PAHs, aromatic ketones, quinones and flavonoids (Carstea et al. 2016, 2020, Coble et al. 2014).

2.1.5 BDOC and AOC

BDOC and AOC can be quantified by biological tests. The BDOC concentration is determined as the difference between the initial maximum DOC concentration and the minimum DOC concentration observed after an incubation period (Escobar & Randall 2001). Hammes & Egli (2005) developed a method for AOC detection using flow-cytometric enumeration and a natural microbial consortium as inoculum. In the method a water sample is inoculated by the natural microbes and incubated at 30 °C. Samples are taken between 0 and 23 h and cell enumeration during the linear growth phase determined by flow-cytometry. The growth rate for the natural microbial consortium is finally determined and converted to AOC with help of a conversion factor. So, while biodegradable organic matter is best

measured by the collective measure of BDOC and AOC under aerobic conditions (Servais 1994, Terry & Summers 2018), the bulk measurement DOC is still commonly reported when studying the removal of DOM in drinking water systems.

2.2 Water sources and treatment

The composition of DOM is dependent on water source. Scottish Water uses a variety of drinking water sources varying in their water composition and quality. There are lochs and reservoirs (183 sources), rivers and burns (171 sources), and boreholes and springs (87 sources) (Drinking Water Quality Regulator 2020a). Water of these sources is treated by one of the 233 Water Treatment Works (WTW) ranging from large supplies providing water to whole cities to very small works that serve only small communities consisting of a few properties. The water signature determines what type of treatment technique is required. Water from springs and boreholes is generally of a higher quality and only requires simple filtering followed by disinfection. Upland water sources such as streams, lochs, and reservoirs, need additional treatment processes. More extensive treatment is required for water from lowland sources such as rivers, as it can contain pollutants.

In most cases, the first step in water treatment is the removal of leaves, weeds, and sticks by mesh screens. Second, particulate matter needs to be removed. Various treatment options are available for this second step. Often, coagulation using alum (aluminium sulphate) is used to help bind impurities and colloids together, followed by flocculation in which polymers are added to further increase the size of the particles formed during coagulation. The aggregates created from the coagulation and flocculation processes then form themselves into a suspended mass called a sludge blanket, and this blanket is removed by sedimentation, and water is then allowed to settle in tanks. Third, rapid gravity filtration is used to remove remaining impurities through direct filtration using sand or a mixture of coal and sand. Granular Activated Carbon (GAC) is typically used for the removal of trace contaminants, like taste and odour compounds (American Water Works Association 1999, Crittenden et al. 2005, Velten et al. 2007). Finally, disinfection with chlorine, chloramination or UV is essential to ensure that water-borne diseases are eliminated, and that the drinking water that is supplied meets The Public Water Supplies (Scotland) Regulations 2014.

The quality of supplied water is regulated by the Drinking Water Quality Regulator for Scotland, who make sure that drinking water supplies meet the requirements (Drinking Water Quality Regulator 2020a, Scottish Statutory Instruments 2015). Microbial water quality is monitored by its *Coliform*, *E.*

coli, and *Clostridium perfringens* content. The chemical water quality on the other hand include pH, turbidity, metals (aluminium, iron, lead, manganese, and nickel), nitrite, odour, taste, radon, benzo-3,4-pyrene, and more. TOC and DOC are not regulated via a threshold limit. As long as there is no abnormal change, the quality is guaranteed (Scottish Statutory Instruments 2015). However, organic substances (predominantly humic and fulvic acid) can give the characteristic yellow-brown colour, odour and taste.

Water quality is measured at different points in a water chain: after treatment at the WTW, in service reservoirs (storage points), and at consumer's tap. In 2020, 67,373 tests were done at the WTWs of which only 30 (0.04%) failed, while 137,681 tests taken at consumers taps of which 99.95% complied with the standards (Drinking Water Quality Regulator 2020a), demonstrating the high quality water Scottish Water produces. A full list of all monitored parameters are presented yearly by the Drinking Water Quality Regulator (Drinking Water Quality Regulator 2020b). The failed samples mostly struggled with *coliform*, manganese, and iron (Drinking Water Quality Regulator 2020b). *Coliforms* are a group of bacteria widely found in the environment, however, they should not be present in the water supply (Eden 2014). Manganese and iron are widely abundant elements in nature and serve as micronutrient (Grose et al. 1998). Elevated levels impair drinking water quality, as both metals cause discolouration of water, though, it is unlikely to be harmful to health from short-term or long-term exposure (Frisbie et al. 2012, Drinking Water Quality Regulator 2020a). In Scotland, undesirable DOC, manganese (> 0.05 mg/L) and iron concentrations (>2 mg/L), occur predominantly in upland freshwaters (northern and western parts of Scotland) (Heal 2001). These upland waters are very soft, exhibiting low mineralization, but contain high concentrations of DOC, iron, and manganese. The organic nature of catchment soils and acidic pH favour manganese and iron mobilisation (Grose et al. 1998, Heal 2001).

Overall, high quality water is produced in Scotland, but the installation, maintenance, and management of the systems are costly, especially in small rural areas. Moreover, these traditional systems leave a large carbon footprint. Scottish Water had an operational carbon footprint for drinking water treatment in 2018/19 of 62,560 tCO₂e (tonnes of carbon dioxide equivalent) (Scottish Water 2019). Grid electricity together with process emissions contributed the most to the carbon footprint, see Figure 2.8.

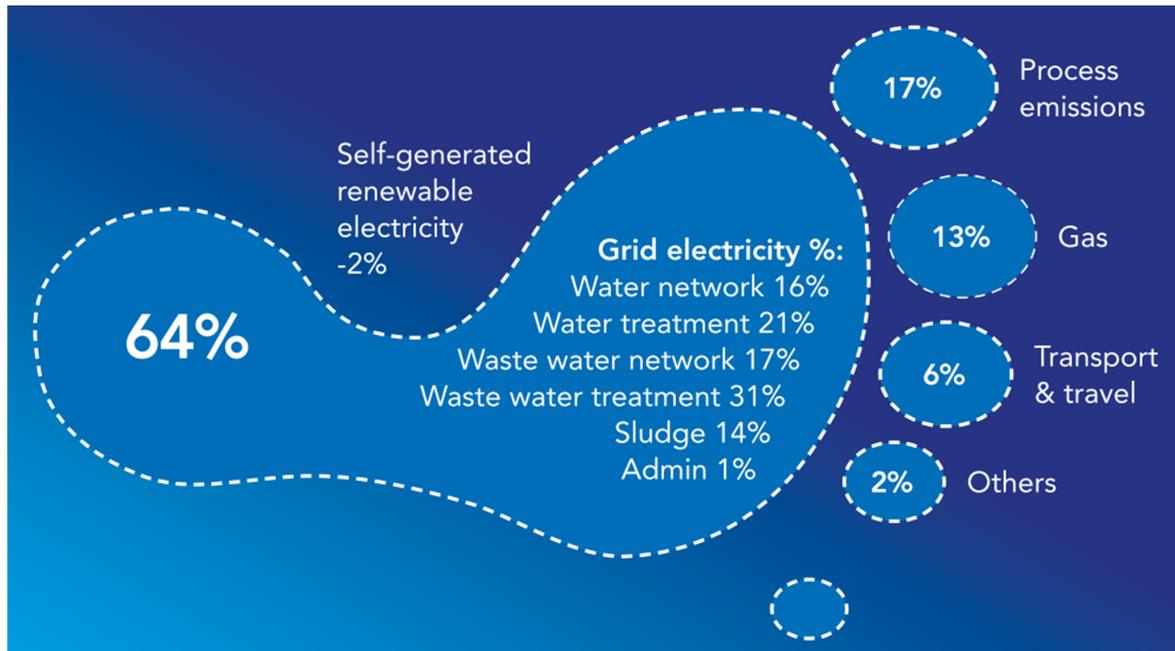


Figure 2.8: Greenhouse gas emissions footprint by source 2018/19. From Scottish Water (2019).

2.3 Biofiltration

An economical and environmental friendly water treatment technique is biological activated filter treatment (Korotta-Gamage & Sathasivan 2017). Biofiltration can be a stand alone process or part of a treatment train, for example, placed after sedimentation. These systems can consist of the previously mentioned GAC material, sand, or anthracite with microbial communities living on the granules. Biodegradable organic matter is the primary substrate for aerobic heterotrophic growth in these biofilters (Peterson & Summers 2021), thereby removal part of the DOC concentration. Although biofilters might be a promising alternative for water treatment, the filter process is complex and still poorly understood (Wu et al. 2018). Biofilters include bank filtration systems (Jeyakumar et al. 2017), slow sand filters (Guchi 2015), and rapid-rate biofilters (Simpson 2008, Terry & Summers 2018). This review is limited to slow sand and rapid-rate BAC filters, as the slow flow BAC filter, central in the project, is a compromise between these two techniques.

2.3.1 Slow and rapid sand filtration

Slow sand filtration are single stage filtration systems that have demonstrated excellent removal of pathogens, while still being able to remove DOM, and other unwanted compounds. These biologically driven systems have low maintenance and a low carbon footprint. Slow sand filtration dates back to 1829 in Paisley, Scotland, where John Gibb supplied water to the city from the slow sand filtration at his bleachery (Baker 1948). Slow sand filtration has been recognised as a simple, reliable and efficient treatment technology and a most effective unit treatment process in improving water quality (Galvis et al. 2002). It can be a good alternative to coagulation, sedimentation, and rapid gravity filtration. A schematic representation of a slow sand filter design is given in Figure 2.9.

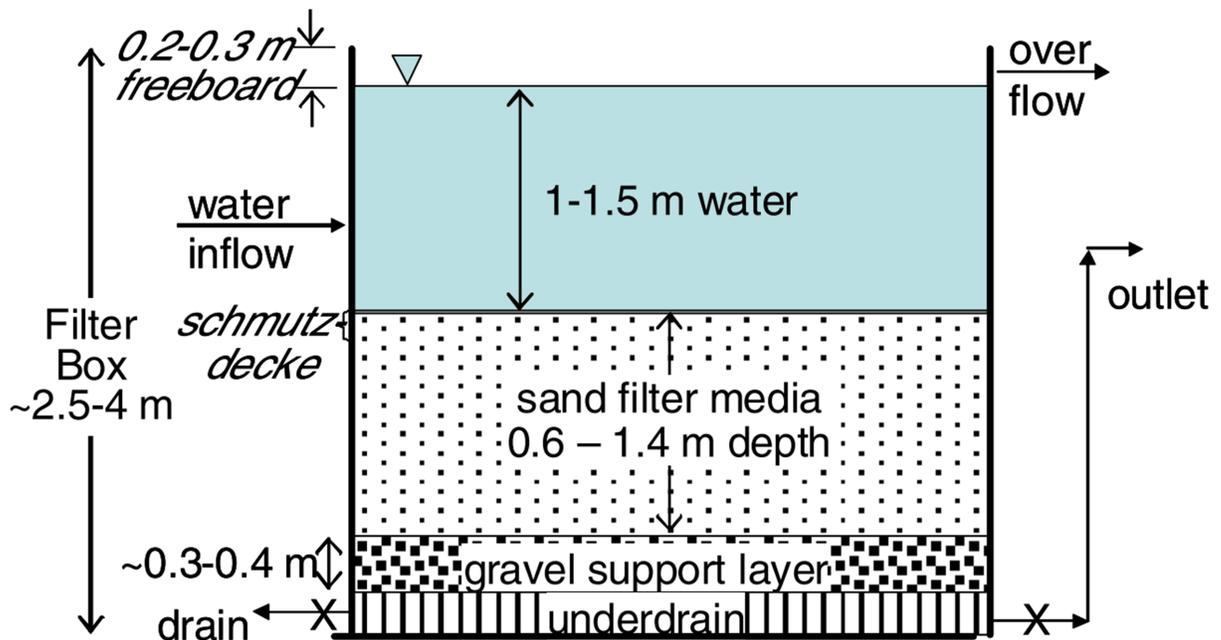


Figure 2.9: Schematic representation of a slow sand filter. Water flows into the tank and travels down through a sand filter medium covered with a Schmutzdecke to exit at the water outflow. Microbes live in this Schmutzdecke and facilitate the removal of pathogens and biodegradable organic matter. From Bielefeldt (2011).

The principle of slow sand filtration is simple. Only a supernatant water layer, sand bed, gravel and an outlet hose are required for the system to work. The supernatant water layer provides a head of water that drives the water through the filter bed. This filter bed is usually sand, because of its low costs and durability. The gravel at the bottom provides a support for the filter sand bed and prevent the sand from clogging the underdrain piping. The water passes the filter bed through gravity with a retention time

of several hours. Inert particles, organic material, and microorganisms such as bacteria and viruses are being removed by physical filtration and biological degradation in the sand bed. Most of these processes happen at the top of the filter, where deposits of particulate and algal matter, combined with dense biomass, form a layer called a Schmutzdecke (Guchi 2015). Periodically, the Schmutzdecke needs to be skimmed of and the sand needs washing.

Slow sand filtration is an appropriate technology for drinking water treatment in rural areas, as it improves the physical, chemical, and microbiological quality of water in a single treatment process without the addition of chemicals, and can produce an effluent low in turbidity and free of bacteria, parasites and viruses (Guchi 2015). However, slow sand filtration does not always remove all harmful substances to the extent required by relevant drinking water quality standards, much depending on the nature, composition, and concentration of the components in the influent waters (Galvis et al. 2002). When using disinfection the necessity of full bacterial and suspended material removal is lower, and over the years rapid sand filtration has become a good alternative as they can process a lot of water with a smaller footprint (Yamamura 2014). Table 2.1 gives the slow sand treatment efficiencies as collected by (Galvis et al. 2002) which have been achieved in filter units operated at filtration rates in the range of 0.04 and 0.20 m h^{-1} , temperature above 5°C, and sand bed depths greater than 0.5 m. From the treatment efficiencies it is clear that slow sand filtration works very well for the removal of microbial contamination. Also, the BDOC fraction (mean 60%) is reduced significantly.

TOC removal ranges from <15-25% and UV-254 from 5-35%, meaning that a part of the organic carbon removed includes aromatic fractions. Still, some TOC remains which might result in water with noticeable colour, smell, and taste. Indeed, slow sand filtration is not considered an effective tool for removal of colour in the water. In combination with coagulation and disinfection, the necessity of full bacterial and suspended matter removal is lower, and over the years rapid sand filtration has become a good alternative as it can process a lot of water at a narrower site (Yamamura 2014).

Table 2.1: *Slow sand filter treatment efficiencies in filter units operated at filtration rates in the range of 0.04 and 0.20 mh⁻¹, temperature above 5°C, and sand bed depths greater than 0.5 m. From Galvis et al. (2002)*

Parameter	Removal	Comments
Enteric bacteria	90-99.9%	Reduces by low temperature, increased hydraulic rates, shallow sand beds and decreased contaminant level
Enteric viruses	99-99.99%	
Giardia cysts	99-99.99%	
Turbidity	<1 NTU	The level of turbidity and the nature and distribution of particles affect treatment
Pesticides	0-100%	Affected by the rate of biodegradation
DOC	5-40%	Mean around 16%. Removal appears to be site specific and varies with raw water and OM character
UV-absorbance (254 nm)	5-35%	
True colour	25-40%	30% average, with colour associated with organic material and humic acids.
UV-absorbance (400 nm)	15-80%	Mean 34%, but upland water sources 42% and lowland water sources 26%
TOC;COD	<15-25%	
AOC	14-40%	Mean 26%
BDOC	46-75%	Mean 60%
Iron, manganese	30-90%	Fe levels>1 mg/L reduce the filter runs

2.3.2 BAC filtration

Rapid-rate BAC filtration is an alternative environmentally friendly method for the removal of bioavailable organic compounds that also needs low maintenance. The main objective of a rapid-rate BAC filters is biodegradable organic matter removal, rather than the control of microbial pathogens. In textbooks, rapid-rate BAC filtration is often not identified as a specific water treatment process, it is rather discussed in connection with biological activity in GAC filters (Juhna & Melin 1985). Although rapid-rate BAC filtration can have various media, adsorbates as well as non-adsorbates, many use adsorptive GAC.

Using GAC, the process could be described in three stages: physical adsorption, concurrent adsorption/biological degradation and biological degradation on its own (Aktaş & Çeçen 2007a, Simpson 2008). While biodegradable organic matter in non-adsorptive media filters only takes place after microbial communities have colonized the filter, adsorption by the GAC results in removal of DOM even before biodegradation has started (Brown et al. 2020, Simpson 2008). Terry & Summers (2018) performed a comprehensive literature data analysis in which the various operational conditions of rapid-rate BAC filtration experiments were compared as well as their treatment efficiencies. They found that the biofilters (n = 117) operating in an Empty Bed Contact Time (EBCT) range of 2-38 min removed 12% (median) of the influent TOC. Rapid-rate BAC filtration is therefore considered an adequate method for the removal of biodegradable DOM.

Granular Activated Carbon

Activated carbon is the general term used to describe carbon-based materials which contain internal pore structures. Various carbonaceous rich materials can be used to produce activated carbon such as coal, coconut shell, wood, and lignite. Activated carbon has a large porosity, high surface area, pore structure consisting of macro-, meso-, and micropores as well as a wide variety of functional groups present on the activated carbon surface. The functional groups carboxyl, carbonyl, phenols, lactones, quinones are responsible for the uptake of pollutants. The functional groups as well as the size of the pores are derived from the activation process, precursors, thermal treatment and post chemical treatment (Bhatnagar et al. 2013). How the activated carbon is modified will determine the performance of the activated carbon for specific contaminant removal. Bhatnagar et al. (2013) wrote a comprehensive review about the various activated carbon modification methods. For instance acid treatment of carbon increases the acidic property, removes the mineral elements, and improves the hydrophilic nature of surface. Base (alkaline) treatment of activated carbon produces positive surface charge which contributes to the adsorption of negatively charged species in higher amounts.

In drinking water biofilter, basic modified GAC is often used which more easily adsorbs the negatively charged DOM.

Adsorption

The adsorption of molecules onto the GAC is a combination of (1) pore filling and (2) interaction of the adsorbates with the carbon medium (MacDonald & Evans 2002, Terzyk 2004), see Figure 2.10. Pore filling is only possible if the molecules fit into the GAC pore space (Gauden et al. 2014, Kowalczyk et al. 2010) and volume (Deng et al. 2017). Pores larger than 50 nm in diameter are classified as macropores, pore sizes between 2 - 50 nm are classified as mesopores, and pores smaller than 2 nm are called micropores. Small pollutants are more likely to adsorb in micropores, DOM in mesopores and bacteria in macropores (Moreno-Castilla 2004).

Not only pore filling, but also GAC-adsorbate interaction is an important part of the adsorption process. It consists of non-electrostatic and electrostatic interactions, of which hydrophobic attraction is part of the non-electrostatic interactions (Figure 2.10) (Karanfil & Dastgheib 2004, Kołodziej et al. 2014, Matsui et al. 2015, Moreno-Castilla 2004, Nam et al. 2014, Pendleton et al. 2002, Rivera-Utrilla et al. 2001). DOM like humic acids, fulvic acids and other hydrophobic compounds, will have more frequent hydrophobic attraction (Leenheer & Croue 2003). On the other hand, hydrophilic compounds of DOM like sugars, proteins, amino sugars and amino acids will tend to be more hydrophilic and water soluble (Matilainen et al. 2010, Sillanpää et al. 2014) and therefore less attracted to hydrophobic GAC. H-bonds can be formed between molecules and functional groups on the GAC. In basic GAC the nitrogen-containing groups can enhance the interaction between carbon and acid molecules via dipole–dipole, H-bonding, or covalent bonding (Bhatnagar et al. 2013).

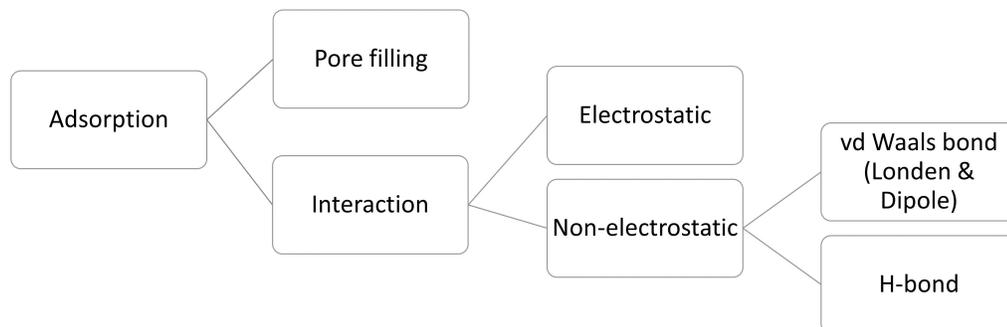


Figure 2.10: *Adsorption processes taking place in GAC.*

Electrostatic forces attract negatively charged compounds to the slightly positively charged carbon surface (Bjelopavlic et al. 1999, Kołodziej et al. 2014, Li et al. 2003, Liang et al. 2018, Nam et al. 2014, Terzyk et al. 2012). Kołodziej et al. (2014) studied the adsorption of humic acids by GAC. They found that a more alkaline carbon surface increased the uptake of humic acids, as the GAC was more positively charged and the DOM negatively charged at pH 7 (Bhatnagar & Sillanpää 2017, Kołodziej et al. 2014, Matilainen et al. 2010, Simpson 2008, Yapsaklı et al. 2009). In this case, a combination of dispersion forces and electrostatic interactions enhance attraction (Kołodziej et al. 2014, Newcombe et al. 1997). Kołodziej et al. (2014) also found that for negatively charged GAC the humic acid adsorption was much lower, as the electrostatic interactions repelled the GAC and humic acids.

According to Velten et al. (2011), the degree of adsorption of NOM fractions on GAC increases with decreasing molecular size (humic < building blocks < LMW organics), while the biopolymers do not adsorb. Greater portions of the internal GAC surface area can be accessed if NOM size is smaller (Bhatnagar & Sillanpää 2017). This depends of course on the size of the GAC pores, of which micropores are the most suitable for NOM adsorption (Ando et al. 2010, Bjelopavlic et al. 1999). During adsorption of organic matter, the adsorption capacity decreases rapidly over time (Crittenden et al. 2005, Gibert et al. 2013, Simpson 2008, Velten et al. 2011a) GAC filtration is therefore not considered as an effective treatment technology for the removal of these larger fouling components, but rather for NOM with larger percentage of LMW NOM (Velten et al. 2011a).

Adsorption kinetics

Under unsaturated conditions GAC has a bimodal pore size distribution (De Smedt & Wierenga 1979, Gauden et al. 2007, Nguyen & Do 2000, 1999), meaning that the pores sizes can be divided into two groups, the larger pores, and the smaller pores. The larger pores are the areas where water can run through more easily, while the smaller pores are like a sink or a source for adsorbates (De Smedt & Wierenga 1979). In many studies the adsorption isotherm of molecules on GAC is best described by the Langmuir isotherm (Demiral & Gündüzoğlu 2010, Faria et al. 2004, Foo & Hameed 2010, Ouakouak & Youcef 2017). The Langmuir adsorption isotherm assumes that that one active site can bind one molecule and adsorption is reversible. It is known that adsorption kinetics is dependent on various factors such as mass transfer, diffusion control, chemical reactions, and particle diffusion. The kinetic of the adsorption process most often follows the pseudo-second-order or intraparticle diffusion kinetics model (Álvarez-Torrellas et al. 2016, Demiral & Gündüzoğlu 2010, Kołodziej et al. 2014, Pohlman 1940). A schematic overview of the diffusion steps is given in Figure 2.11.

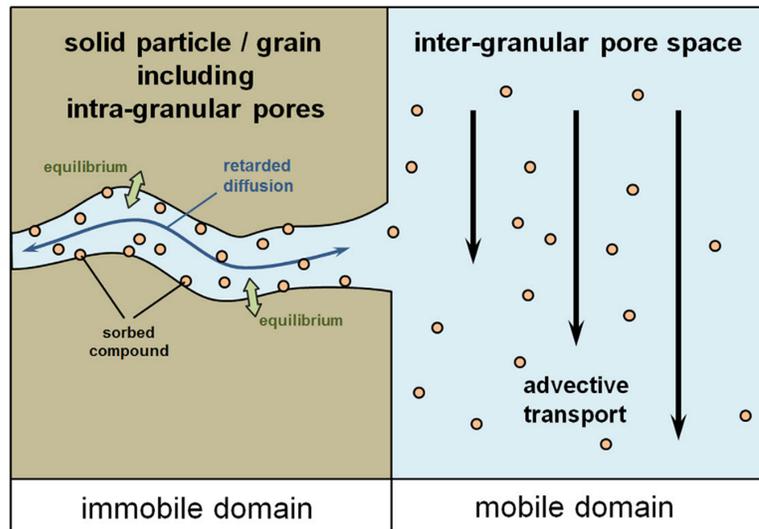


Figure 2.11: Schematic overview of diffusion of molecules from the mobile phase to the immobile phase within the pores of the GAC. From Liu et al. (2021).

The water in the larger pores is considered as mobile phase and the water in the smaller pores as immobile phase. Adsorption kinetics is thought of as adsorption of molecules in the mobile phase to larger pores walls (A) followed by the diffusion of such adsorbed molecules into small micropores through the pore mouth barriers (B) (Kołodziej et al. 2014, Nguyen & Do 2000, Pohlman 1940). This diffusion (also called retarded pore diffusion) continuously removes molecules from the large pores as long as the concentration in the mobile water/large pore adsorption sides is higher than that of the immobile phase ‘pushing’ the molecules into the micropores. This diffusion is much slower than the adsorption of molecules in the macropores and is therefore rate limiting step. The opposite also holds, when molecules are leaching from the filter the concentration in the small pores is high and molecules will travel via the large pores to the water that leaves the filter.

Figure 2.12 shows the difference in adsorption between the larger and smaller pores. Molecules adsorb via an equilibrium to the wall of the larger pores. At first the concentration around the micropore mouth increases and retarded pore diffusion brings these molecules into the micropore where the immobile phase is and the concentration still low. In the micropore the molecule adsorbs onto the wall via an equilibrium. When the concentration of the molecules in the immobile phase of micropores increases to a concentration that exceeds the concentration of the mobile phase in the large pore, the molecules diffuse back again. When all adsorption sites at the mobile phase are occupied, the removal

will continue with the rate of the diffusion of molecules into the immobile phase. As a result, there is not enough time to adsorb all molecules that passes a filter and therefore an initial breakthrough (but not full) will take place. Adsorption via the micropores continues with this slow removal continuing for a long period of time, also known as tailing (Pohlman 1940). If a greater part of the GAC contains small pores, there is more immobile water resulting in a quicker initial breakthrough and longer tailing (slow removal). The lower the micropore content of the GAC, the more mobile water there is, the longer it takes until initial breakthrough and the shorter the tailing (Pohlman 1940).

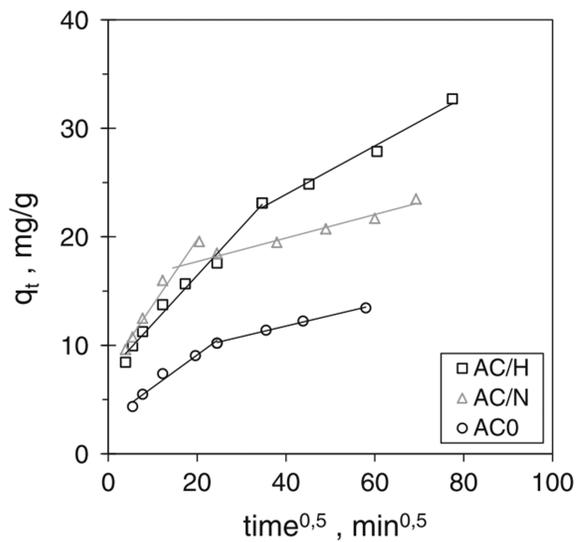


Figure 2.12: Adsorption of humic substances onto different types of activated carbon with AC/H = hydrogen treated activated carbon, AC/N = ammonized activated carbon, AC0= normal activated carbon. All three filter types show a clear bent in their curve, where the first part is the adsorption of molecules to the large pores and the second less steep part is the adsorption of the molecules in the smaller pores. Difference in adsorption rate is the result of retarded diffusion, the rate limiting step of the total adsorption. From Kołodziej et al. (2014).

Biodegradation

The shift in adsorption processes to biological processes takes place when the GAC is getting saturated and simultaneously bacteria start to colonize the GAC material. The biofilm is described to be in an acclimation phase (Servais 1994, Simpson 2008). In this process activated carbon is enriched with DOM, nutrients and high oxygen concentrations (Ghosh et al. 1999, Simpson 2008, Stewart et al. 1990). Microorganisms naturally present in the water attach to the GAC media surface. The rough porous surfaces of the granular particles serve as immobilization sites (Scholz & Martin 1997, Ser-

vais 1994). There they multiply and form a colony as they feed on the rich supply of adsorbed and entrapped organic matter, waterborne nutrients and other microorganisms (Korotta-Gamage & Sathasivan 2017, Simpson 2008). This colony forms a matrix to help keeping the structure together and develops into a mature biofilm. The environment within the biofilter can support the growth of a wide variety of naturally occurring microorganisms (Simpson 2008).

The biofilm that grows on the filter material is capable of biodegrading a significant fraction of nutrients entrapped in the GAC pores, DOC adsorbed to the GAC surfaces and other contaminants, minerals and microorganisms in the raw water (Zhang & Huck 1996). These biological processes start to remove biodegradable organic matter (Servais 1994) and these processes are even able to remove recalcitrant biodegradable organic matter (Korotta-Gamage & Sathasivan 2017). While the microbial community is growing, the second stage of concurrent adsorption and biological degradation process takes place. The saturation of the GAC continues, and finally the DOC removal is completely replaced by biological degradation, reaching a steady state (Simpson 2008). The common understanding is that in steady state the removal of DOC during biofiltration is controlled by the biomass concentration (Carlson & Amy 1998). However, the relationship between biomass and DOC removal is not linear (Urfer et al. 1997, Wang et al. 1995), and an increased biomass does not automatically mean a similar increase in DOC removal.

A graphical representation of these processes is given in Figure 2.13 on the following page. A biofilter made from inert media such as sand requires sufficient operational time until the biofilter is starting to remove TOC. When using GAC the initial TOC removal is high and decreases over time as a result of exhaustion. Concurrently, biodegradation starts to play a role and after sufficient time, the filter solely relies upon microbial breakdown of TOC. At this point the filter is referred to as a BAC filter. The transition from GAC filter to BAC filter is time dependent (Velten et al. 2011a), and much still remains unclear about when the filter can actually be called a true BAC filter. It is believed that a steady state biofilm and DOC effluent concentration is reached after 90 days in slow sand filter systems (Boon et al. 2011, Feng et al. 2010, Velten et al. 2011a). For a BAC filter, determining steady state is much more complicated. This problem is clearly set out by Peterson & Summers (2021), who observed in their metadata study of 89 wastewater treatment biofilter studies, the median TOC removal for biofilters with less than 10,000 Bed Volumes (BV; the number of times that water has passed the filter bed) treated was nearly four-fold higher than for those with greater than 40,000 BV treated, 70% and 18%, respectively. Under steady state conditions the average TOC removal should be independent of BV throughput and TOC removal is expected to be similar between the 10,000 and 40,000 BV (Peterson & Summers 2021), which was clearly not the case. When they reclassified the data set for adsorptive

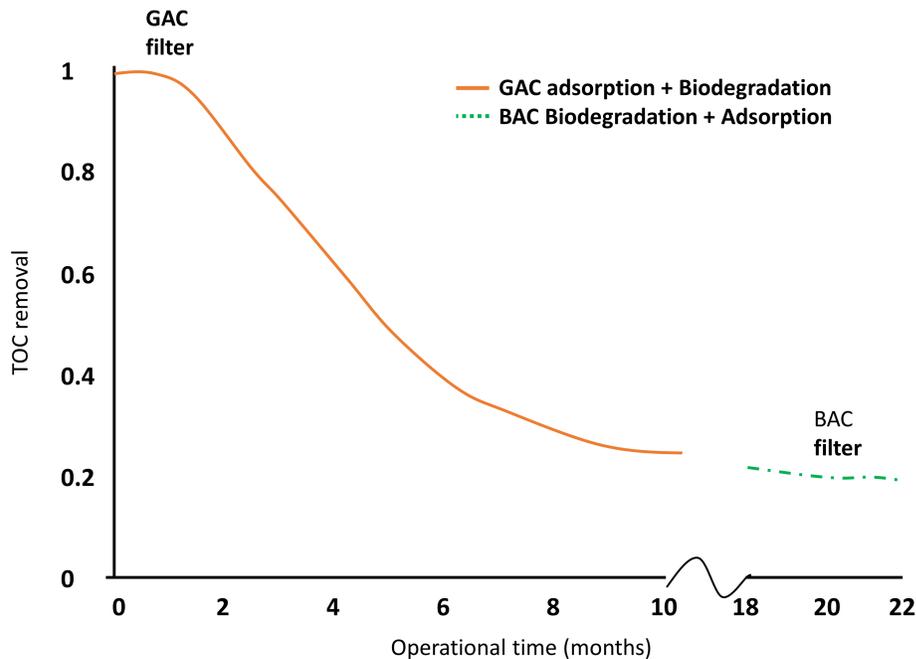


Figure 2.13: *TOC removal acclimation for a GAC filter, which becomes a BAC filter once the adsorption capacity is exhausted. Initial removal is dominated by adsorption, while at apparent steady state the removal is dominated by biological removal. Based on Terry & Summers (2018)*

biofilters media filters in steady state as “BAC” when throughput exceeded 20,000 BV or when relative throughput was 1.0 or higher and as “GAC” when neither threshold was achieved, they found that of 89 filters, 40 were classified as “GAC”. Peterson & Summers (2021) demonstrated that a significant part of the filters were wrongly described as steady state BAC filters. The reported TOC removal in the original studies was, due to the low BV, probably influenced by adsorptive processes.

Why are BAC filters often wrongly classified? According to Peel & Benedek (1980), after a rapid fall in adsorption efficiency to a lower level the removal actually continues, virtually unchanged, over an extended period of time, as a result of slow adsorption. The adsorption phase can last for months or longer depending on the characteristics of the adsorbate and adsorbent (Peterson & Summers 2021). For example, slow adsorption of large molecular weight DOM into GAC micropores can take months (Peel & Benedek 1983). In various water treatment operations these types of breakthrough curves (such as Figure 2.12) have been observed and have been wrongly assumed to be caused by biological removal following an initial period of high adsorptive uptake (Peel & Benedek 1983), while actually

all removal could be ascribed to some form of adsorption. Moreover, the sorption process can continue on the surface of the biofilm covering the filter media (Klimenko et al. 2002, Quintelas et al. 2010), which is referred to as biosorption. The apparent steady state can therefore be easily mistaken for steady state behaviour.

Colonisation, substrate utilization and stratification

Alongside (slow) adsorption, microbial communities start to inhabit the filters (Simpson 2008). Most of the microorganisms present in the biofilter live in a biofilm. Biofilms consists of microbial cells either immobilized at the surface of the GAC (substratum) or embedded in an extracellular polymeric substances (Simpson, 2008). The GAC pores (1 - 100 nm) are often too small for the bacteria (> 200 nm) to enter. Biofilms in biofilters contains bacteria, both aerobic and anaerobic, as well as protozoa which are grazing on the bacteria (Juhna & Melin 1985). The type and structure of bacteria is dependent on water chemistry and composition of DOC. As previously mentioned, the colonisation of microorganisms of the BAC filter requires a carbon source. These carbon molecules can be extracted from the water that passes through the filter, or from the GAC on which they adsorb (Herzberg et al. 2003). Understanding of the way microorganisms access these substrates is limited. Different theories are proposed that describe this multistep biological/physical process. Various studies use the theory of the bioregeneration of the GAC material (Aktaş & Çeçen 2006, 2007b, De Jonge et al. 1996, El Gamal et al. 2018, Hanaki et al. 1997, Herzberg et al. 2003, Nath & Bhakhar 2011, Sirotkin et al. 2001). Bioregeneration of activated carbon involves the use of microbial communities to regenerate the capacity and surface of the carbon (El Gamal et al. 2018). Two methods of bioregeneration as pointed out by El Gamal are (1) bioregeneration through concentration gradient and (2) due to exoenzymatic reaction. (1) A concentration gradient causes desorption of organics into the bulk solvent and when microbes metabolize adsorbed contaminants, the equilibrium is shifted resulting in further desorption. (2) The exoenzymatic hypothesis was introduced by Perotti-Rotman (1974) and explains that substrates are biodegraded after desorption from the GAC. Bacteria are too large to enter the micro- and mesopores of the GAC where the substrates are adsorbed. Exocellular enzymes, however, can enter these pores and react with the substrates. The reaction products show weak adorability and diffuse out of the pores, where they are biodegraded by the bacteria.

However, Xiaojian et al. (1991) evaluated this theory by looking at the properties of enzymes and the distribution of exoenzymes in the carbon pores. They concluded that the theory of Perotti-Rotman was incorrect, and substrates can only be removed by either biodegradation or adsorption. Velten et al.

(2011) concluded that NOM adsorption is irreversible. In their entire study period, concentrations of all NOM fractions decreased along the length of the filter bed. According to them this observation suggests that displacement of one adsorbed NOM fraction by another did not occur and that humics, building blocks, and LMW compounds adsorbed irreversibly. This could indicate that adsorbed carbons are not available to the microbes.

The highest concentration of biomass is present at the top of the filter and decreases with increasing filter depth (Boon et al. 2011, Chen et al. 2016, Liao et al. 2013, Moll et al. 1999, Persson & Wedborg 2001, Urfer & Huck 2001, Velten et al. 2011, Wang et al. 1995, Zhang et al. 2016). The high biomass content in the top is a direct result of the highest concentration of the bioavailable DOM present in the top of the filter. Indeed, (Velten et al. 2011) demonstrated that the DOM concentration of their O₃/BAC filter from start-up with virgin GAC media was the highest concentration in the top and decreased through filter depth, similar to the biomass. Also Liao et al. (2013) showed in their BAC filter a decline in AOC and DOC concentration with depth. Chen et al. (2016) showed in their pilot-scale biofilter treating raw water that DOC, biopolymers, and to a lesser extent humic substances concentration decreased through filter depth, while there was no clear trend for lower MW building blocks and acids. Zhang et al. (2016) monitored dissolved organic nitrogen and other physiochemical parameters through filter depth and demonstrated a decrease in the pH, dissolved oxygen, and nitrite, while DOC, and ammonium did not change, and nitrate increased. These results suggest that the greatest level of DOM removal occurs at the top of the filter, resulting in the highest biomass content in the top layer. However, Chen et al. (2016) demonstrated that the percent removal of FEEM protein-like materials and LC-OCD biopolymers increased with increasing bed depth. Especially biopolymers demonstrated a high removal. The percent removal of humic-like materials was consistently low (<13%) but also increased for increasing media depth, similar to low MW neutrals. LC-OCD B and LMW acids did not have any distinct removal trend.

In the study conducted by Boon et al. (2011) the removal efficiency was taken as the percent removal per biomass which demonstrated that the highest DOC removal efficiency was found in the bottom of the filter. This was especially the case for humic substances and building blocks. It appears that the bacteria able to degrade more recalcitrant carbon are living deeper in the filter (Boon et al. 2011, Juhna & Melin 1985). Indeed, the study conducted by Velten et al. (2011a) that demonstrated a DOC gradient through filter depth in the maturing biofilter showed that at the start of the biofilter acclimatisation bacteria in the bottom layers are deprived of easily available DOC. The adsorption in the top of the filter creates a nutrient-rich micro-environment on the surface of the GAC granules that is favourable to biological growth (Herzberg et al. 2003, Li & DiGiano 1983, Urfer et al. 1997).

2.3.3 Design parameters

Although a biofilter is working via a natural process, the design of the filter and the operational settings influence its performance. There are various factors influencing biofilter performance which are comprehensively set out in various studies (Basu et al. 2016, Korotta-Gamage & Sathasivan 2017, Peterson & Summers 2021, Terry & Summers 2018). Physiochemical parameters of the influent water that most significantly affect BAC performance include DOM concentration and characteristics, available nutrients (Lauderdale et al. 2012, Nemani et al. 2018), temperature (Emelko et al. 2006, Moll et al. 1999, Moona et al. 2018), turbidity, and pH (Korotta-Gamage & Sathasivan 2017). Operational settings that have an influence on DOM removal include the type of medium (Emelko et al. 2006, Lechevallier et al. 1992, Persson et al. 2006, Yapsakli & Çeçen 2010), water pre-treatment (Lauderdale et al. 2012, Lechevallier et al. 1992, Yapsakli & Çeçen 2010), renewal of material (Moona et al. 2018), backwashing strategies (Emelko et al. 2006, Liao et al. 2014), as well as EBCT (Arnold et al. 2018, Basu et al. 2016, Chowdhury et al. 2010, Emelko et al. 2006, Korotta-Gamage & Sathasivan 2017, Lechevallier et al. 1993, Terry & Summers 2018, Thiel et al. 2006). The filter media and EBCT will be discussed in more details because these two parameters were controlled in the biofiltration experiment. Temperature control, pre-treatment, and backwashing were outside the scope of this study.

Media selection

Various granular materials can be used in biofilters such as sand, GAC, anthracite, and expanded ceramics. A survey of 38 North American utilities found biofilter media configurations were predominantly anthracite/ sand (37%) and GAC/sand (37%), while GAC alone (23%) and sand alone (3%) designs were less abundant (Terry & Summers 2018).

Often, GAC is used for biofilter media as it offers several advantages over inert media including better TOC removal. It has to be kept in mind that performance of biofilters that use GAC as media can be substantially higher due to residual TOC adsorption capacity when prior throughput and influent TOC are low (Basu et al. 2016). Another benefit of GAC is that it supports more dense microbial populations than sand or anthracite (i.e. 4 to 8 times more biomass per gram of media), due to a combination of factors including porosity, surface area, surface roughness, surface charge and adsorption capacity (Basu et al. 2016). The higher microbial density is expected to translate into higher BDOC removal within a biofilter. The metadata analysis by Peterson & Summers (2021) on 118 wastewater treatment studies (of which the BAC filters had processed over 20,000 BV) indeed found a slightly higher TOC

removal by the GAC compared to the inert media; 23% by GAC and 15% by inert media (Figure 2.14). The largest differences between GAC and inert material in TOC removal were on approximately 24 percentage points higher in BAC filters; these studies used long EBCT, 60 min and 40 min, respectively (Farre et al. 2011, Reaume et al. 2015).

Also for drinking water treatment, a selection of paired studies that evaluated both inert and BAC media (with 20,000 + BV) under the same operational conditions report on average a higher TOC removal with BAC media than inert media similar to that shown in Figure 2.14, with some exceptions (Bourgin et al. 2018, Chowdhury et al. 2010). Overall, there is evidence that BAC media can be expected to outperform inert media. However, when very large differences are noticeable this may indicate non-steady state behaviour in the BAC which should be further evaluated.

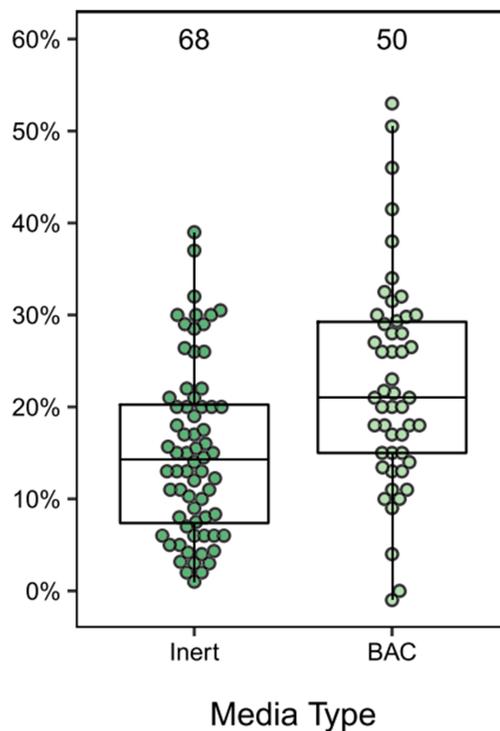


Figure 2.14: Boxplot of TOC removal by inert ($n=68$) and BAC ($n=50$) media from various biofiltration studies. Results demonstrate the higher removal of TOC by BAC. From Peterson & Summers (2021).

Empty bed contact time

One of the crucial parameters that impacts DOM removal efficiency is the EBCT (Basu et al. 2016, Korotta-Gamage & Sathasivan 2017, Peterson & Summers 2021, Terry & Summers 2018). It is the measure of the water residence time in the biofilter which is calculated by the filter media depth divided by the flow rate. The influence of EBCT on DOM removal has previously been studied by sampling a biofilter at different depths (Boon et al. 2011, Chen et al. 2016, Chowdhury et al. 2010, Emelko et al. 2006), using different loading rates (Moona et al. 2021) or different filter lengths. This parameter should be optimized as long as the chosen settings do not conflict with the hydraulic efficiency or particle removal goals. Moreover, there is also a practical limit or threshold as increasing EBCT either results in longer filter bed depths at a set high loading rate, or more filter area (larger filters) at a set filter depth.

TOC removal is independent of the rate the water goes through the filter (filtration rate) as long as the EBCT remains constant (Servais 1994). Increasing the BV increases the EBCT and thereby increasing the DOC removal (Korotta-Gamage & Sathasivan 2017). Similarly, increasing the filtration rate, decreases the EBCT and thereby decreasing the DOC removal (Carlson & Amy 1998). A decreased DOC removal as a result of an increased filtration rate may also be due to bad filter acclimatization (Basu et al. 2016). In other words, the biomass is unable to assimilate the dissolved carbon substrates as efficiently at this higher loading rate. There is not time for the biomass to reacclimate to the new biofilter conditions (Carlson & Amy 1998).

In wastewater treatment studies, the EBCT had a significant impact on the DOM removal. High DOM removal rates of up to 90% are reported in studies on wastewater treatment (Gerrity et al. 2011, Hu et al. 2005, Ka et al. 2020, Pipe-Martin 2010, Rattier et al. 2012a). Hu et al. (2005) demonstrated a 30% increase in DOC removal by the pilot-scale zeolite biofilters when the EBCT was increased from 5 to 30 min, beyond which no improvement was found. Rattier et al. (2012a) used an EBCT of 1 hour with flow rate of 30 L/h to clean raw water with a DOC concentration of on average 7 mg/L by BAC filter. They witnessed a removal of approximately 40%. Pipe-Martin (2010) investigated EBCT varying from 50 to 150 minutes and demonstrated an increase of approximately 60%. According to Peterson & Summers (2021) extending EBCT above 20 or 30 min can have a significant benefit in wastewater treatment.

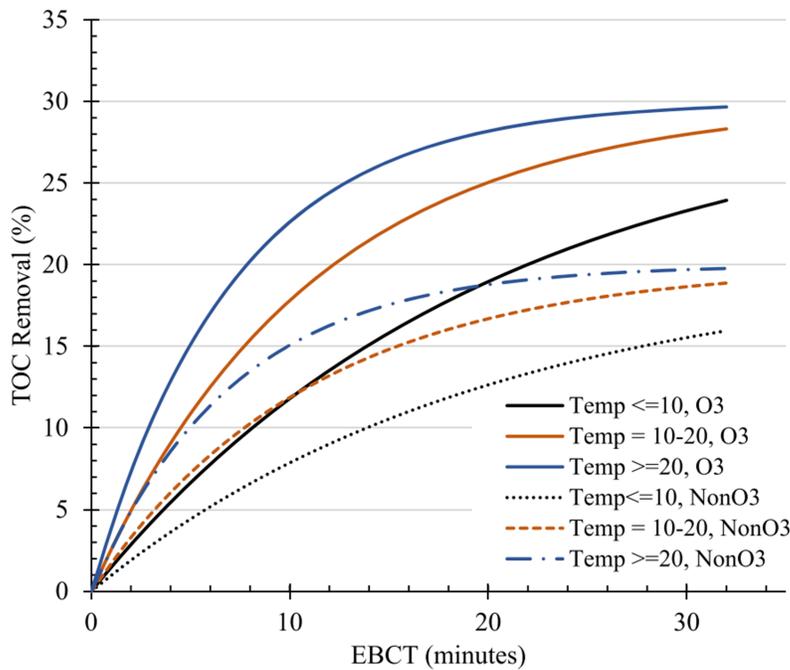


Figure 2.15: Simulated TOC removal as a function of EBCT at three temperature ranges for ozonated and nonozonated waters with associated k' values ($k' = 0.05 \text{ min}^{-1}$ for $10 \text{ }^\circ\text{C}$, $k' = 0.09 \text{ min}^{-1}$ for $10 - 20 \text{ }^\circ\text{C}$, $k' = 0.14 \text{ min}^{-1}$ for $20 \text{ }^\circ\text{C}$). From Terry & Summers (2018).

For drinking water treatment, the metadata study of Terry & Summers (2018) summarized that for all temperature and oxidation conditions ($n = 117$), biofilters operating in an EBCT range of 2 - 38 min removed 12% (median) of the influent TOC. They demonstrated that a higher EBCT increases the TOC removal, for the GAC as well as the inert media (Figure 2.15). A longer EBCT gives a longer contact time between organic substrates and biomass and promotes diffusion through the biofilm together with enhanced utilization of the substrate, which leads to a higher DOM removal (Carlson & Amy 2001, Lechevallier et al. 1992, Rattier et al. 2012, Wu & Xie 2005, Zhang et al. 2017). Moreover, extracellular hydrolysis of large organic substrates is demonstrated to happen in the bulk water and increased residence time increases the removal rate of these non-diffusible organics (Aktaş & Çeçen 2006, Rohold & H. 1993). The required higher EBCT for elevated DOM removal suggests that a considerable portion of the BDOC is relatively recalcitrant, needing longer EBCT for removal (Peterson & Summers 2021). Incremental DOM removal decreases at longer EBCT as the more easily biodegradable substrate is preferentially removed in the upper filter leaving only more recalcitrant compounds in the lower filter (Zearley & Summers 2012).

Table 2.2: Drinking water treatment studies looking at the influence of ozone, EBCT and type of filter media on TOC/DOC removal.

Ozone	EBCT	Media	% Removal TOC/DOC	Study
No	8	GAC (100.000 BV) Anthracite	11-14 1-3	Thiel et al., 2006
	16	GAC (100.000 BV) Anthracite	15-20 2-7	
Yes	9.2	GAC (24.000 BV)	29	Wang et al., 1995
		Anthracite/sand Sand	16 20	
Yes	60	GAC (45.000 BV)	46	Farre et al., 2011
		Sand	23	
Yes	40	GAC (1400 BV)	40	Reaume et al., 2015
		Sand	20	
No	15-30	GAC (0-30.000 BV)	18	Ho et al., 2011
		Sand	5	
Yes	2-20	GAC (8500-17000 BV)	19	Arnold et al., 2018
		Anthracite	9	
Yes	7.5 12	GAC (16 y.o)	24.2	Hooper et al., 2020
		Anthracite	19.8	
Yes	14 10	GAC (20.000 BV)	18	Bourgin et al., 2018
		Sand	20	
Yes	30	GAC (unknown)	45	Hubele et al., 1985 (From Terry&Summers 2018)
		S	20	
Yes	17-36	GAC (several y.o.)	25	Emelko et al., 2006
		A/S	23	
			20	
			14	
No	7	GAC (exhausted)	9	Chowdhury et al., 2009

Table 2.2 on the previous page shows a selection of drinking water treatment studies using various media that looked at the influence of EBCT on DOC removal. In most cases the increased EBCT showed an increased DOC removal, but the influence of the EBCT varied among studies. Lechevallier et al. (1992) demonstrated that, although the reduction of the AOC fraction can be achieved with 5 - 10 min EBCT, DOC removal increased from 30% to 50% when the EBCT lengthened from 5 min to 20 min in a GAC filter. Also Moona et al. (2021) who used a non-absorptive media demonstrated that long contact times (>30min) can be advantageous for operating biological filters. Li et al. (2006) reported that the optimum EBCT for a BAC filter that processes ozonated water is 15 min. Not all studies found an increased DOC removal for an increased EBCT. Hozalski et al. (1995) did not find any difference in DOC removal when increasing the EBCT from 4 - 20 minutes. Also, Kalkan et al. (2011) reported a similar increase in TOC removal for their thermally-activated carbon (7%, EBCT 9 to 18 min) but no increase removal for a chemically-activated carbon.

Overall, biofiltration is still considered a black box. Most studies manipulate the operational conditions to find optimal performance of TOC removal without fully understanding why the filter behaves as it does. Various steps can be made to enhance the understanding of the biofilter processes, so that further optimization can be achieved. First of all, only a very few number of studies combine the collected chemical and microbial data of the biofilter. This is interesting, as a biofilter in steady state thrives on the microbial processes. Analysing the chemistry in combination with the biology of the biofilter can contribute to unravelling this black box. Second, DOM analysis is mostly done by TOC/DOC analysis, while the use of other techniques remains limited. By combining more advanced DOM analysis method such as fluorescence spectroscopy, LC-OCD, and AOC / BDOC, a better understanding of the DOM removal processes can be found. Third, most studies only look at the in- and effluent concentrations. As a result the processes within the biofilter which might hold the key to finding aspects of the removal process that can be enhanced. Diving into the biofilter and understanding what happens inside the biofilter by looking at the ecological mechanisms that are taking place can therefore be a next step in the further understanding and possible enhancement of biofilter performance.

2.4 Chemometrics

2.4.1 Data driven science

Science and the way research is undertaken are changing rapidly. The increase of data generation is present in all scientific disciplines including environmental chemistry and engineering. A interpretation of high dimensional data is complex. Moreover, data-intensive science represents a paradigm shift. It asks different kinds of questions and performs science with a different process compared to theory- or hypothesis-driven science (McCue & McCoy 2017). However, data-driven research is not entirely hypothesis-free. It often starts with a broad hypothesis and generates large volumes of quantitative data relevant to that hypothesis, while often being unbiased by prior knowledge and letting the data speak for itself. These data are then explored to generate more specific and mechanistic hypotheses through eliminative induction (finding truth by using evidence to eliminate false competitors) or abductive reasoning (making a probable conclusion from what you know) (Pietsch 2015). In this way data-driven science does not follow the assumption that the scientist has all the scientific insight to generate the best hypotheses; instead, it assumes that the scientist has a basic understanding of the subject and they uses data to generate specific hypotheses (Pietsch 2015). Various disciplines can be used in data-driven research. Figure 2.16 shows a Venn diagram, which illustrates the relation between data science, data mining, artificial intelligence, machine learning and deep learning, explained in more detail in the following section.

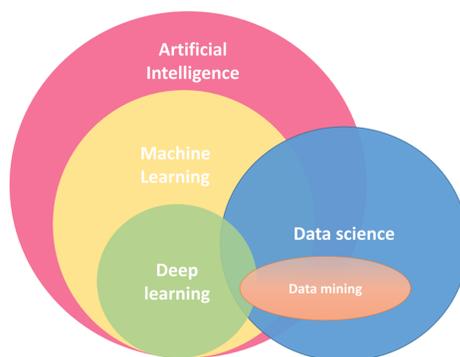


Figure 2.16: Venn diagram explaining the relationship between data science, data mining, artificial intelligence, machine learning and deep learning. From Kulin et al. (2021).

2.4.2 Data science

Data science is the scientific discipline that studies everything related to data, from data collection, data storage, data analysis, data pre-processing, data visualization, data interpretation, data-based decision making, determining how to create meaningful output from data and how to communicate relevant data (Kulin et al. 2021). Data science makes use of data mining, machine learning, deep learning, artificial intelligence techniques, and other approaches such as: trial and error algorithms, statistics, causal inference, etc (Kulin et al. 2021).

2.4.3 Data mining

Data mining can be defined as a set of methods used to extract usable information from large data sets (Amigo, 2021). Data mining implies that the usable information is already in the data, however, the complexity of the data makes it impossible to find this useful information without powerful mathematical tools. Basically, data mining aims at patterns that already exist in the data, but which are hidden due to a large number of samples and variables, the noise of the data, or the difficulty in linking more than two variables at the same time in a univariate fashion (one variable at a time) (Amigo 2021). Data mining methods can be divided into clustering or dimension reduction methods (Figure 2.17). It is important to remember that data mining does NOT include any procedure of learning (discussed in the next section), and therefore the methods used for data mining are of an unsupervised nature and are also called unsupervised machine learning tools (Bro & Amigo 2020).

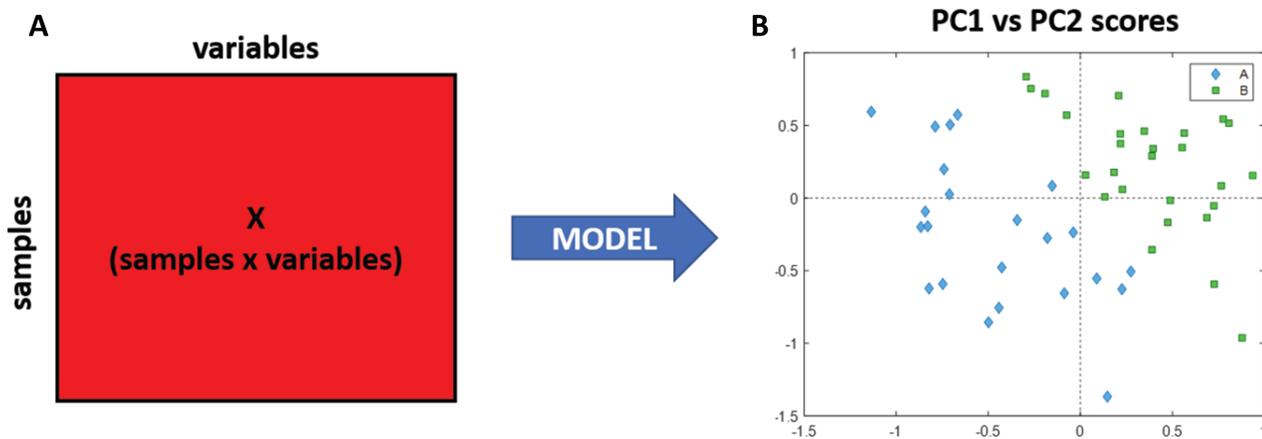


Figure 2.17: Data mining method applied to a matrix X (A) which includes multiple samples and multiple variables. A PCA model (B) helps to interpret clustering of the data (blue and green datapoints) in a score plot with help of principal component (1 and 2). From Amigo (2021).

2.4.4 Machine learning

Machine learning is a subset of artificial intelligence (Figure 2.18). It aims to develop algorithms that can learn from historical/known data and make informed decisions based on what is learned (Amigo 2021, Kulin et al. 2021). The learning procedure makes the algorithm reliable enough to predict any property in new data that have not been used for learning. Because machine learning includes this learning, the methods used for machine learning are of a supervised nature (Bro & Amigo 2020). Learning is often described as training, and as a result machine learning algorithms are composed of two steps: training (A, calibrating), testing the model (B), and applying the model created in B to a new data set (C) (Figure 2.18). Examples of supervised machine learning methods are regression models, classification models.

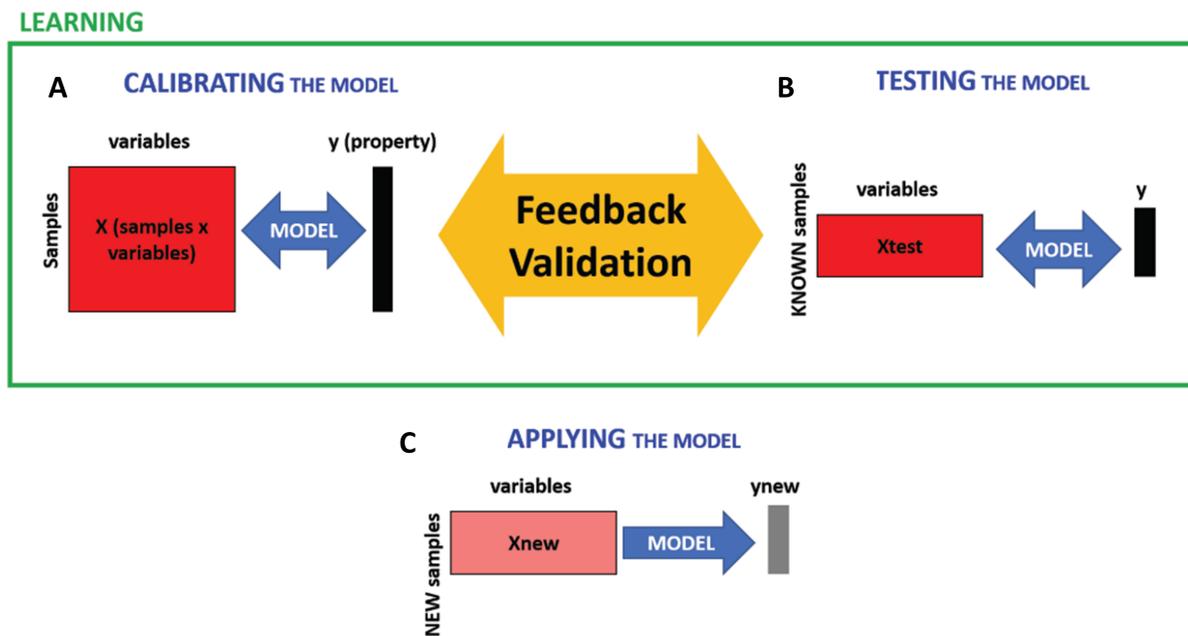


Figure 2.18: Machine learning procedure showing the three stages of calibration of the model with help of a training data set using a part of the known samples (A), validation of the model with help of a test data set with the other part of known samples (B), and prediction new samples with help of the created model (C). From Amigo (2021).

2.4.5 Deep learning and artificial intelligence

Deep learning is again a subset of machine learning including methods that permit software to train itself to perform tasks, like image recognition, by exposing multi-layered neural networks to vast amounts of data (Bro & Amigo 2020). Deep learning, machine learning and data mining are all part of the artificial intelligence, which is defined as any technique that enables computers to mimic human intelligence, using logic, if-then rules, decision making (Bro & Amigo 2020).

2.4.6 Chemometrics

The term chemometrics was coined in 1972 by Svante Wold and Bruce Kowalsky. The most accepted definition of chemometrics is the chemical discipline that uses mathematical, statistical, and other derived methods employing formal logic to (a) design or select optimal measurement procedures and experiments and (b) provide maximum relevant chemical information by analysing chemical data (Amigo 2021). Another definition could be that chemometrics is the application of artificial intelligence (therefore including data mining, machine learning, artificial neural networks, and deep learning) to data coming from Chemical Systems (Amigo 2021). It means that since the start of the paragraph it was all about chemometrics. But what makes chemometrics different from the statistical analysis in other disciplines? Although all disciplines can use the same methodologies, understanding what the data mean and how to interpret it requires a person that understands chemistry.

2.4.7 Nomenclature and usage

An overview of the possible methods that can be used when applying machine learning to data are given in Figure 2.19. The definitions in red are discussed in more detail in the next section as they will be used in the upcoming chapters.

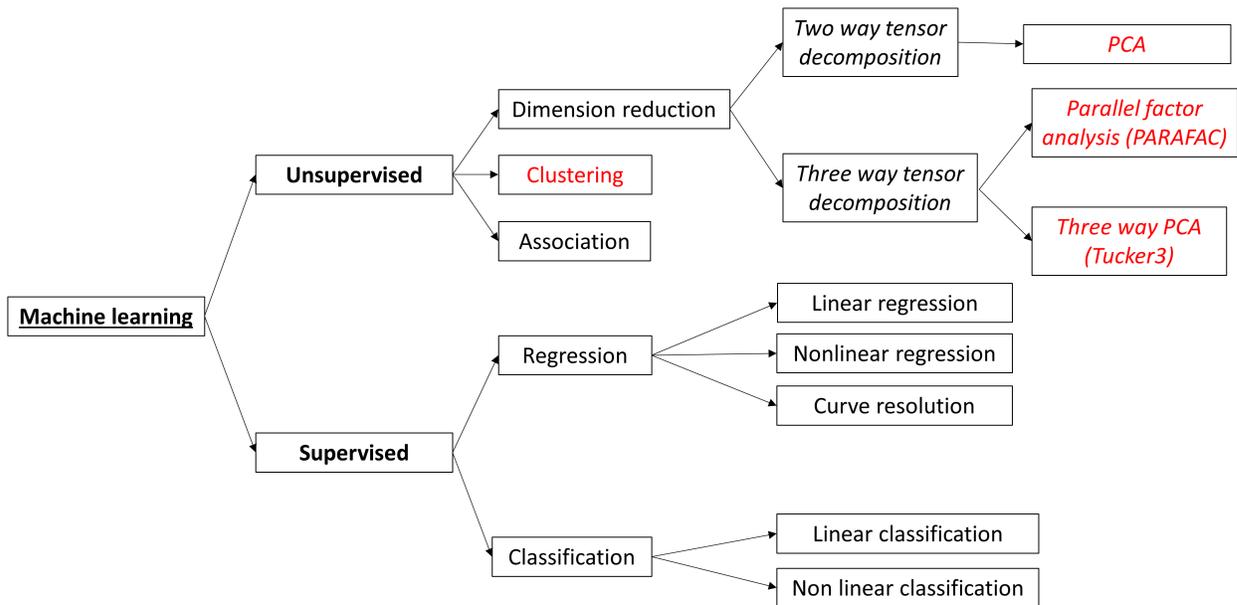


Figure 2.19: *Non-exhaustive overview of data mining/machine learning tools. Machine learning can be divided into unsupervised (dimension reduction, clustering, and association) and supervised learning (regression and classification) tools. Many methods are part of each of the tools . The methods in red are used in this thesis.*

2.4.8 Clustering

Cluster analysis or clustering is a possible exploratory data analysis strategy in which a set of objects is grouped in such a way that the objects in the same group (called a cluster) are more similar to each other than to those in other groups (clusters). Cluster analysis is the general task which can be achieved by various algorithms that differ greatly in their way of forming cluster. Possible types of clustering are connectivity-based clustering (Hierarchical clustering), centroids-based clustering, distribution-based clustering, and density-based clustering (Bro & Amigo 2020). In this thesis the hierarchical clustering heatmap will be used. Heatmaps allow for simultaneously visualise clusters of samples and variables (see Figure 2.20 for an example of a heatmap).

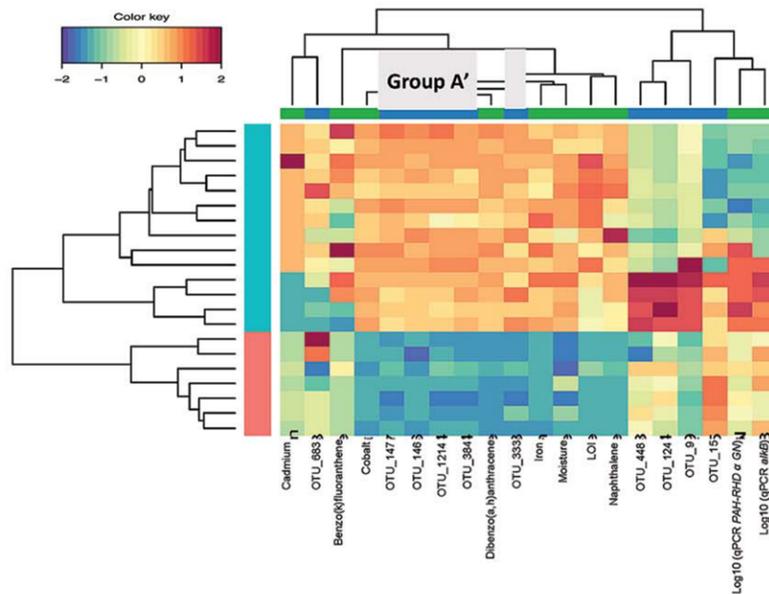


Figure 2.20: An example of a heatmap. 10 discriminant features (variables) in the column direction with both rows and columns ordered using hierarchical (average linkage) clustering to identify blocks of features of interest. From Gauchotte-Lindsay et al. (2019).

On the left side of the heatmap the hierarchical clustering is done for the samples and on top of the heatmap it was done for the variables. The columns/rows of the data matrix are re-ordered according to the hierarchical clustering by average linkage (specifying the distance between two clusters as the average distance between objects from the first cluster and objects from the second cluster). In this way, blocks of ‘high’ and ‘low’ values are grouped in the data matrix. The data matrix is visualized by a colour scheme which contributes to finding the variables that appear to be characteristic for each sample cluster.

2.4.9 Principle component analysis

When a couple of variables is measured over a set of samples a multivariate data set can be constructed, which can also be called a matrix, second-order tensor, or two-way data. Correlations and differences (variance) between samples reflected by a group of measured variables can be studied by pattern recognition models (Amigo 2021). The most widely used method is the PCA model. PCA is used in exploratory data analysis and a schematic overview of the PCA concept is given in Figure 2.21.

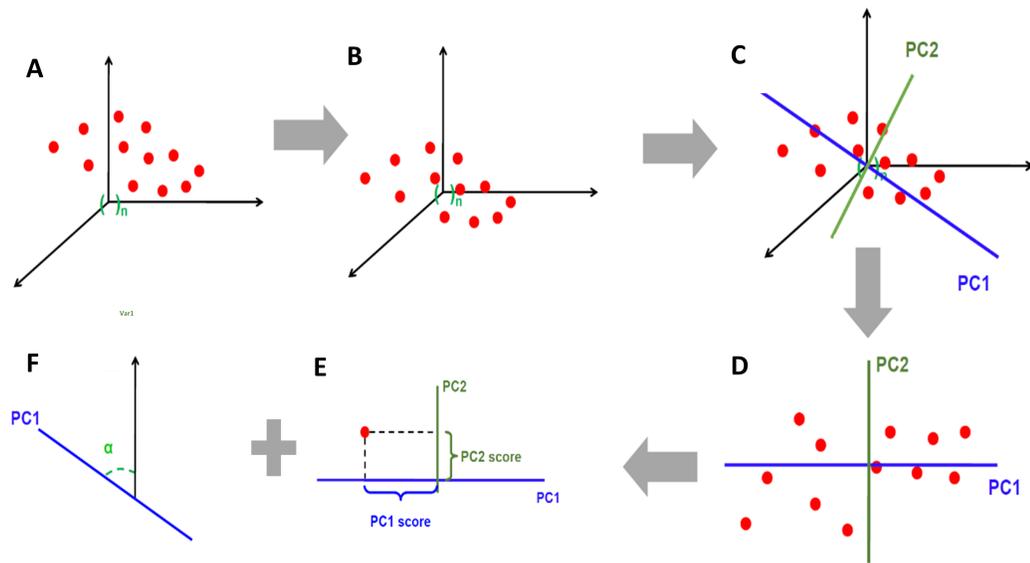


Figure 2.21: Schematic representation of the concept of PCA. The datapoints in the variable space of 3 variables (A), the centring of the datapoints towards the barycentre (B), the principal component 1 and 2 describing the maximum variations (C), the projected datapoints in the new principle component space (D), the PC1 and PC2 scores that describe a datapoint in the projected principle component space (E), and the loadings describing the relationship between the old variable axis and new principle component axis (F). Based on Amigo (2021).

In PCA, the data set is presented in a new “space”, with the number of dimensions equal to the number of variables (n) (Figure 2.21 A). In other words, each new axis corresponds to each variable. In this new space each m sample is now a “point” in this new space, therefore there will be as many points as samples. In most cases, the barycentre of that cloud of points is placed to the centre of the variable space by normalization (centring), (Figure 2.21 B). First, the direction of the maximum variability in the data is explained by PC1, which must pass by the centre of the variable space. Second, the direction of the maximum variability that PC1 did not explain is explained by PC2 which must pass the centre of the variable space and must be orthogonal to PC1 (Figure 2.21 C). Now, PC1 and PC2 are projected in a new space called the principal components space (Figure 2.21 D). How the samples points are related with the new axis is described by the scores, which is the distance of the projected samples to the centre of the new axes (Figure 2.21 E). Loadings describe the relationship between the old axes (variables) and the new axis (PCs) (Figure 2.21 F). In simple words, PCA decomposes the data matrix into a set of so-called scores and so-called loadings.

2.4.10 N-way PCA

When data are collected for multiple samples analysing a range of variables under various conditions data will be present in three dimensions instead of two as is in a matrix. This three dimensional data are also known as a data array, data cube, third order tensor, or three way data. When dealing with even more dimensions, data are referred to as N-th order tensor, where N is the number of dimensions. In this thesis three dimensions are used, and therefore this type of data is referred to as data array. A representation of a data array is given in Figure 2.22. By X the three-mode data array is denoted of order (I,J, and K). Such an array originated from the characterization of i objects by j variables under k different conditions.

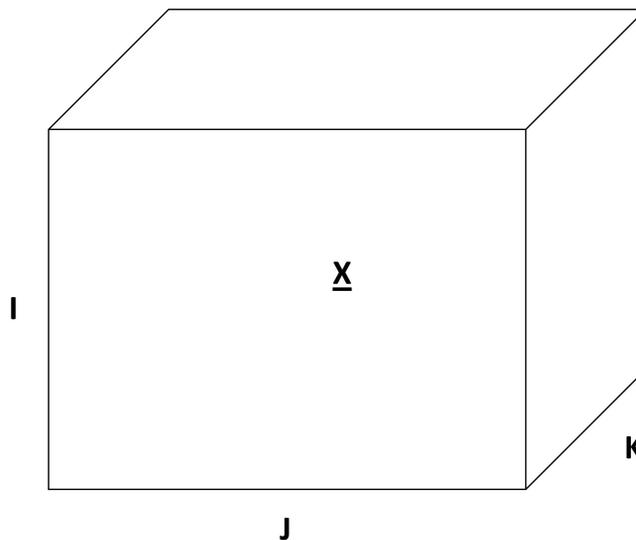


Figure 2.22: Representation of a data array with dimensions I , J , and K for samples, variables, and conditions, respectively.

Exploratory analysis such as PCA cannot be applied directly to data arrays. The data array can be unfolded into a data matrix, but also the original structure can be used. For this, multimode data analysis is required, more specific, N-way principal component analysis (also referred to as multiway PCA) (Henrion 1993). Examples of high order PCA are canonical decomposition/Parallel Factor analysis (CANDECOMP/PARAFAC, and hereafter called PARAFAC analysis) and Tucker3 analysis. PARAFAC and Tucker3 analysis are both multilinear decomposition methods, which decompose a tensor into sets of scores and loadings, similar to PCA, that describes the data in a more condensed form than the original data array (Bro 1997). PARAFAC analysis can be considered a constrained version of Tucker3 analysis, and Tucker3 analysis a constrained version of two-way PCA. A data

set that can be modelled adequately by PARAFAC analysis can thus also be modelled by Tucker3 or two-way PCA. In other words, a two-way PCA model always fits data better than a Tucker3 model, which again will fit better than a PARAFAC model. This also means that when a PARAFAC model is adequate, Tucker3 and two-way PCA models have the possible disadvantage of modelling noise or model the systematic variation which is unwanted. As Occam's razor stated in the fourteenth century, 'always use the simplest possible model', which is also known as the law or principle of parsimony. In this sense, PCA model can be considered the most complex and flexible model, while PARAFAC analysis is the most simple and restricted model (Bro 1997). Generally, scientists find two-way PCA easier to use than the multilinear methods, however, to apply PCA the tensor requires unfolding to a matrix where the variables in the unfolded modes get mixed up and as a result the effect of one variable is not associated with one but many elements of a loading vector. Using a more structured approach, like PARAFAC analysis, will result in a simpler model which also has consequences for the fit; this is often poorer. The reason for using multi-way methods is therefore not to obtain better fit, but rather more adequate, robust, and interpretable models (Bro 1997).

2.4.11 CANDACOMP/PARAFAC

Already in 1927 Hitchcock proposed the idea of the expressing a tensor as the sum of a finite number of rank-one tensors (Hitchcock 1927); and in 1944 Cattell suggested ideas for parallel proportional analysis and the idea of multiple axes for analysis (circumstances, objects, and features) (Cattell 1944). It took until 1970 until the concept finally became popular after its third introduction to the psychometrics community, in the form of CANDECOMP (canonical decomposition) by Carroll & Chang (1970) and PARAFAC analysis by (Harshman 1970). In this thesis, the term PARAFAC analysis will be used.

A decomposition of a data array is made into trilinear components (factors) as shown in Figure 2.23 (Bro 1997, Kolda & Bader 2009, Kroonenberg 2009). Each component consists of one score vector (a) and two loading vectors (b and c), however, often all three vectors are called loading vectors. The PARAFAC model output is given by three loading matrices, A, B, and C with elements a_{if} , b_{jf} , and c_{kf} . The PARAFAC model is found to minimize the sum of squares of the residuals, e_{ijk} .

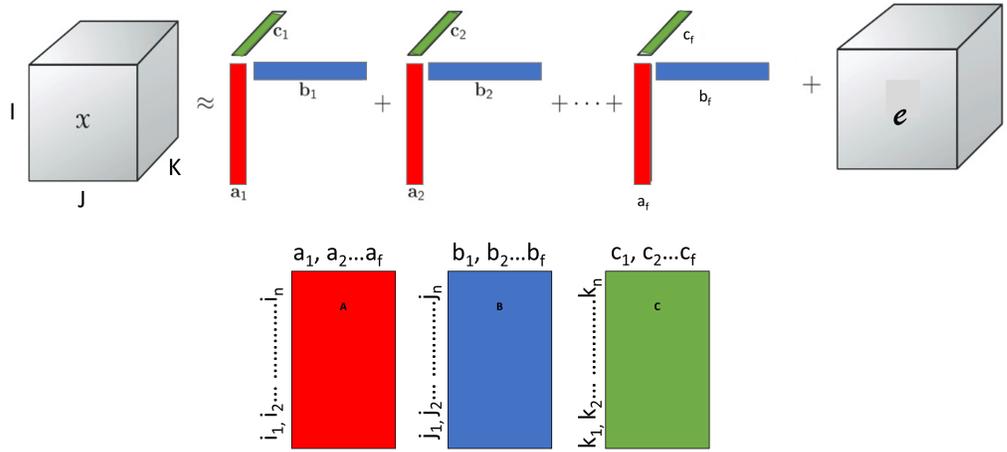


Figure 2.23: *PARAFAC decomposition into multiple trilinear components each having a score vector (a) and two loading vectors (b and c) as well as the error that is left. Together with the loading matrices of A (includes $a_1, a_2 \dots a_f$), B (includes $b_1, b_2 \dots b_f$), and C (includes $c_1, c_2 \dots c_f$).*

The basic PARAFAC model for a three-way array X with elements x_{ijk} consisting of the scores of I subjects on J variables under K conditions is a direct extension of standard two-mode PCA, as shown in Equation 2.1, where F is the number of factors (components):

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk} \quad (2.1)$$

While in PCA adding numbers of principle components will increase the complexity of the model and create a better fit, in PARAFAC analysis an exact number of components/factors need to be chosen that creates the only possible model. This one possible combination of components makes the model unique (Kolda & Bader 2009). When too few components are included, a part of the data X will be included in the error. Using too many components will add part of the error to the components meaning that error will be interpreted as meaningful data. The choice for number of components is therefore crucial in PARAFAC analysis and the most difficult task. There are many aspects that can be discussed regarding PARAFAC (Bro 1997, Bro & Kiers 2003, Kolda & Bader 2009, Kroonenberg 2009, 1983), however, for the scope of this thesis they are not further discussed. Despite the more challenging concept of PARAFAC analysis compared to PCA, various open source packages are available to perform PARAFAC analysis including informative examples and example datasets (Giordani et al.

2015, Murphy et al. 2013b, Pucher et al. 2019).

2.4.12 Tucker3

The Tucker decomposition was first introduced by Tucker (1963). Tucker's 1966 article is the most comprehensive of the early literature and is generally the one most cited. The Tucker decomposition is also a form of higher-order PCA. In PARAFAC analysis all directions (I, J, and K) have similar number of components/factors, however, this can vary in Tucker3. Therefore decomposing a data array looks a bit different. It decomposes a data tensor into a core tensor (G) multiplied (or transformed) by a matrix along each direction (A, B, and C), see Figure 2.24 (Kolda & Bader 2009). Also here, the part of the data that is not included in the model becomes part of the error (ϵ).

The basic Tucker3 model for a three-way array X with elements x_{ijk} consisting of the loadings of I sample on J variables under K conditions is another direct extension of standard two-mode PCA, as shown in Equation 2.2, where P, Q, and R are the number of factors (components): x_{ijk} is, in this case, the value of the measurement referring to the i^{th} , j^{th} and k^{th} :

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} + e_{ijk} \quad (2.2)$$

The values P, Q and R are the number of components selected to describe the first, the second and the third mode, respectively, of the data array. P, Q and R are not necessarily the same for each mode, as in PARAFAC analysis. The elements a_{ip} , b_{jq} and c_{kr} belong, respectively, to the component matrix A ($I \times P$), describing the sampling sites (objects), B ($J \times Q$), describing the chemical water quality variables, and C ($K \times R$), describing the sampling times (conditions) (Figure 2.24). Each of these matrices can be interpreted as a loading matrix in the normal PCA. G_{pqr} denotes the elements (p, q, r) of the $P \times Q \times R$ core array G, and e_{ijk} denotes the error term for element x_{ijk} and is an element of the $I \times J \times K$ array E. In other words, the final result is given by three sets of loadings together with a core array that describes the relationship among them.

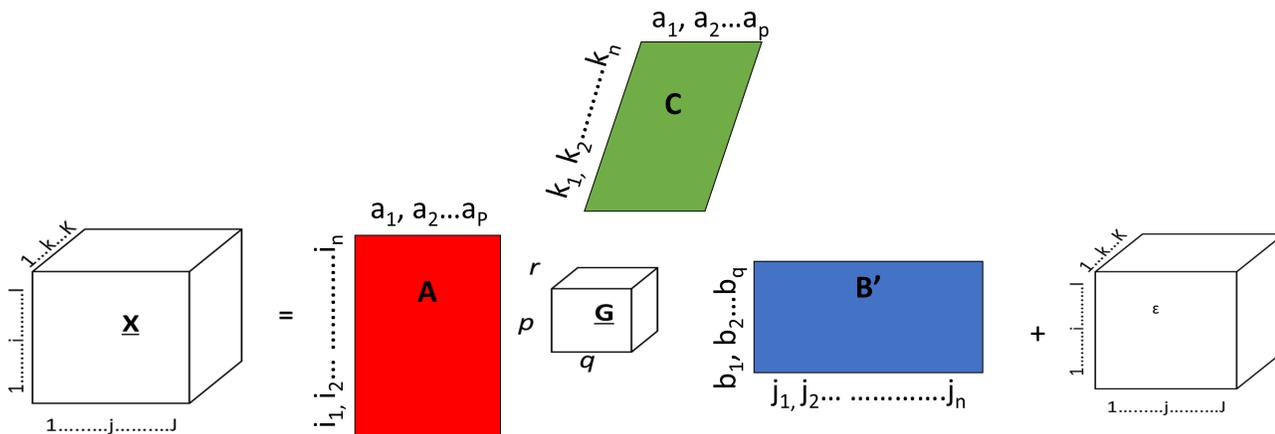


Figure 2.24: Tucker3 decomposition. The data array X with I (sample), J (variable), and K (conditions) mode is decomposed in a G core with p , q and r factors for the I , J , and K mode, respectively, together with the three loading matrices (A for sample mode, B for variable mode, and C for condition mode). The remaining unexplained data are part of the error ϵ .

The number of components in each direction are chosen as low as possible while still accounting for a significant amount of variance, similar to 2-PCA. There is no single possible outcome for Tucker3 analysis and increasing the number of components in each mode will increase the fit and consequently making the model more complex. The Tucker3 model is therefore not unique, similar to PCA, but unlike PARAFAC analysis. Core array G must be a superdiagonal matrix, having non-zero elements only on its main diagonal (Bro 1997). The squared elements of G reflect the strength of the interactions amongst the three modes and are central to the interpretation of the model (Pardo et al. 2013). Tucker3 model interpretation must consider simultaneously the sign and magnitude of the non-null elements of G . In the case of a cubic core array, a series of orthogonal rotations can be made on the three spaces (site, variables and month), looking for the common orientation for which the core array is as much as possible body-diagonal (Brouwer & Kroonenberg 1991). Also for Tucker3 models various open source tools are available (Andersson 1998, Giordani et al. 2015).

The difference between PCA and Tucker3 analysis can be illustrated by the example data set ‘Learning to read’ provided by Bus (1982). The data set describes the process of learning to read of seven pupils ($I = 7$). Five tests ($J = 5$) were used to evaluate the learning process with tests given the letters L, P, Q, S, and R. The pupils were tested weekly from week 3 to week 47 except for eight holidays weeks, hence $K = 37$. The aim of the study was to investigate the learning process and whether the performances of the pupils were equal over time (Giordani et al. 2015). Applying PCA to this ‘Learning to read’ data set with the measured variables in the columns and the tracking of pupils over time in the

rows, results in a plot with severe overlap of scores with no clear pattern (Figure 2.25). Answering the research question is not (easily) possible.

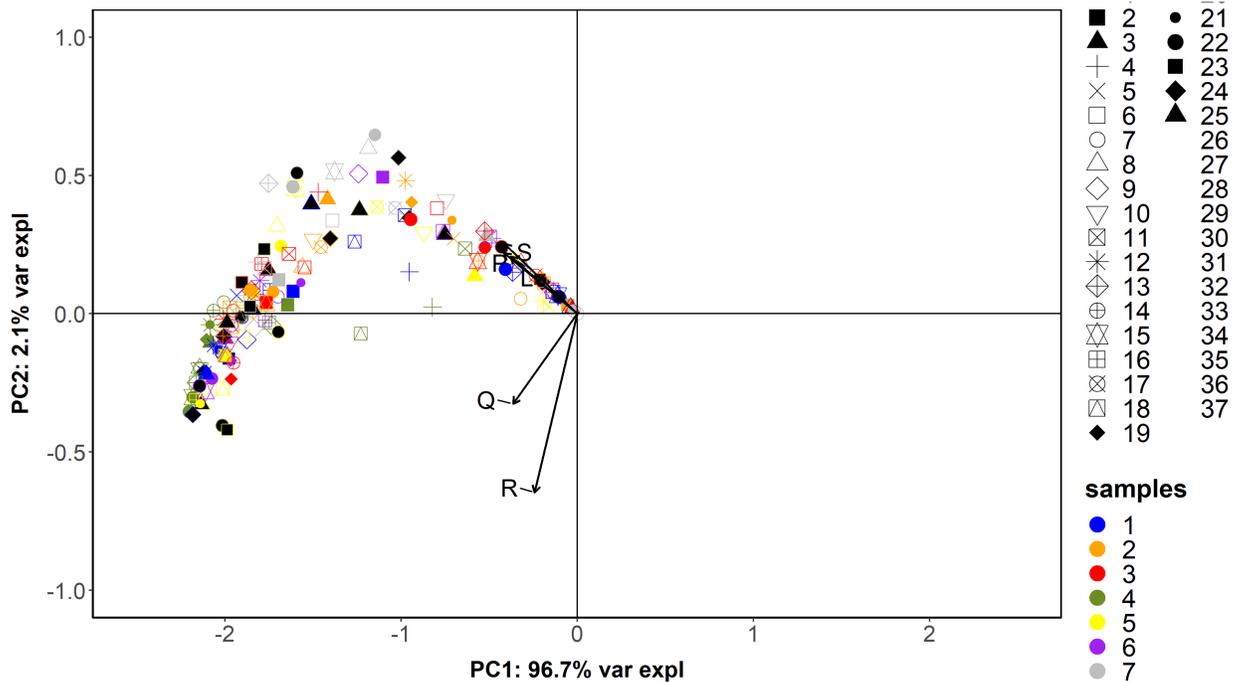


Figure 2.25: PCA biplot of the Learning to read data set with PC1 explains 96.7% of the data and PC2 21%. The pupils are represented by colour, months by symbols and the tests are presented by the vectors.

Giordani et al. (2015) therefore applied three-way PCA to this ‘Learning to read’ data set with the pupils, variables, and time in the three different modes creating a data array. There are various ways to present the outcome of a three-way PCA model, with the one chosen by Giordani et al. (2015) given in Figure 2.26. A full explanation of the model can be found in (Giordani et al. 2015). In short: the most appropriate model was formed by 2 components in mode A (pupils), 1 component in mode B (tests) and 2 components in mode C (time). For the best visual representation of the results the loadings of mode A and B were given by their loading values, while mode C was graphically presented by the loading values over time (Figure 2.26).

A. Pupils' loadings

n.1	1.06	-0.42
n.2	0.96	-0.30
n.3	0.99	-0.38
n.4	1.28	1.00
n.5	1.16	0.19
n.6	1.09	-0.01
n.7	0.89	-0.42
n.1	1.06	-0.42

B: Variable's loadings

L	0.91
P	1.00
Q	0.87
S	0.99
R	0.58

C: Time's loadings

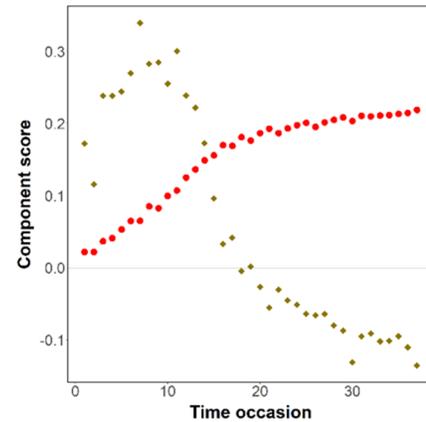


Figure 2.26: Tucker3 model outcome for 'Learning to read' data set. With the pupil's loadings (A), variable's loadings (B), and time loadings (C).

The component scores C1 and C2 demonstrated a clear pattern over time, representing performance level (C1) and learning rate (C2). The first and second C-mode components were related, respectively, to the first and second A-mode components. The pupils whose component scores A1 and A2 were high, were those who had a performance level (C1 with respect to A1) and a learning rate (C2 with respect to A2) above average. For example, pupil 4 was the best student: his (her) scores were the highest (scores 1.28 and 1.00). From Figure 2.26 it becomes clear that the information of the three-way PCA model makes for an orderly interpretation of the data and thereby outperforms the PCA model.

Also various studies have successfully applied three-way PCA to answer environmental questions. (Dong et al. 2010, Engle et al. 2014, Giussani et al. 2008, Leardi et al. 2000, Pardo et al. 2013). Leardi et al. (2000) used three-way PCA to study the environmental information by 4 years of monitoring of the various quality parameters of different waters of the Venice lagoon. Engle et al. (2014) analysed groundwater quality parameters collected for three years from 14 monitoring wells to identify changes and processes affecting interactions.

2.4.13 PARAFAC and Tucker3 preprocessing

Tucker3 and PARAFAC analysis are very much alike. A PARAFAC model is a constrained version of Tucker3 analysis with the number of components $P = Q = R$ and having all the superdiagonal elements of G equal to 1. For both methods, the data sets require pre-processing which homogenises scales and units without altering the differences among the sample sites and among the sampling times. For PCA, this problem is solved by autoscaling in the variable (column) direction of the matrix. For PARAFAC and Tucker3 analysis, this problem is commonly solved by performing a j -scaling when dealing with a data array (Henrion 1993, Kroonenberg 1983). The three-way array X (with I , K , and J modes) is matricized to a two-way matrix X_b having $I \times K$ modes in the row direction and the J mode in the column direction (Figure 2.27). On this matrix autoscaling was performed in the variable (column) direction. As a result, the global variance of each variable was set to one, and the differences among the objects and the conditions are preserved. J -scaling calculates averages over two modes. This operation removes some offsets but at the same time may introduce new offsets, thereby introducing artificial variation that the model also has to handle (Leardi et al. 2000). More suitable pre-treatments are currently under study (Bro & Kiers 2003).

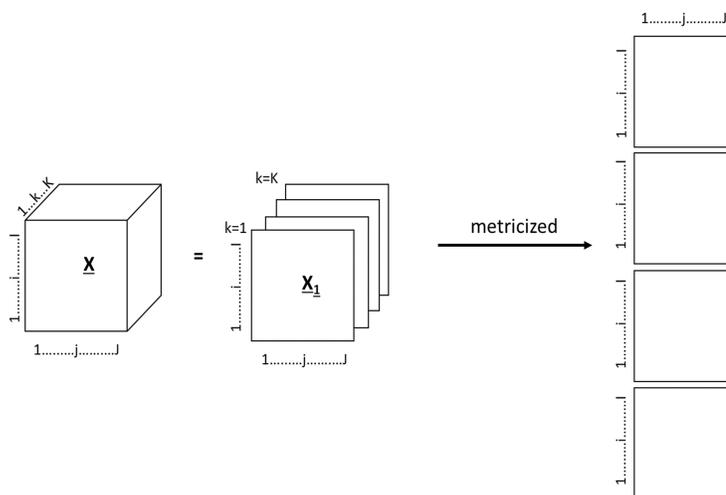


Figure 2.27: Example of a data array X matricized by frontal slicing for j -scaling (scaling in the column direction).

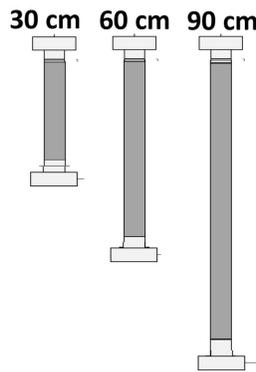
2.4.14 Final remarks

Amigo (2021) wrote a nice and comprehensive manuscript about the definitions of machine learning, data mining, artificial intelligence and big data analysis, defining their application ranges and its application in analytical chemistry, also called chemometrics. According to him, there has been quite a lot of confusion in the literature about the terms chemometrics, machine learning, data mining, deep learning and artificial intelligence the last few years. But looking back in time, all these methods are actually part of what people half a century ago would call statistics, or a century ago just mathematics.

He also warns against the vast amplitude of the words “data science” and the little importance that sometimes is given to the word data. He states that ‘Only by understanding the scientific problem and the complexity of the data, will we be able to choose a proper data analysis methodology, if it is needed.’ In the light of this he warns that a good outcome might not necessarily be a perfect model; the Garbage In = Garbage Out principle is relevant when creating models. Therefore, the quality and information that data can provide should be checked, as well as the purpose of the data and how they were obtained. The consequences of bad data are (1) wrong hypothesis: trying to find what is not in the data, (2) wrong design of the experiment: hoping that the measurements contain variation that is not contemplated in the design of the experiment, (3) overuse of resources: using a measurement instrument just because it is available in the laboratory. Finally, people tend to use more sophisticated models, just because they look fancy. However, using more complex models on a simple, small data set could lead to a model that is not correctly applicable to new data set. Therefore, the simpler model, the better.

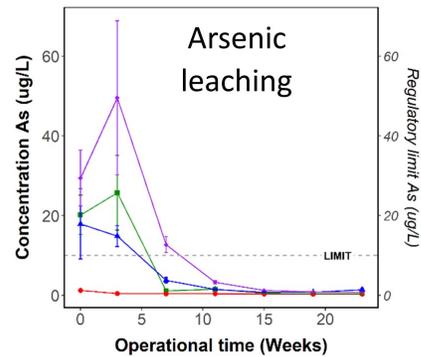
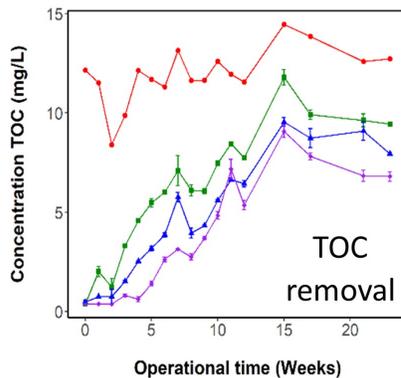
Chapter 3

Characterisation of DOM removal and other chemical water quality parameters by laboratory-scale slow-flow BAC biofilters



Monitoring 20 chemical water quality parameters in pilot scale biofilters from start up to 23 weeks runtime

removal TOC + iron
leaching arsenic + aluminium



Longer filter = highest removal in apparent steady state as result of mainly slow adsorption

Length

- INFLUENT
- EFFLUENT 30cm filter
- ▲ EFFLUENT 60cm filter
- ▼ EFFLUENT 90cm filter

CONTRIBUTORS

The data used in this chapter are part of a larger biofiltration project in which multiple biofilters were constructed and run for a 6-month period by Dr Marta Vignola (MV) and postgraduate researchers Dominic Quinn (DQ) and Steve Joyce (SJ). DQ has designed and constructed the biofilters. Filter operation, pH and conductivity monitoring were done as a team. Water sample analysis for microbial water quality parameters was done by DQ. Filter modelling was done by SJ, while *Legionella* and cell counts were measured by MV. I carried out all chemical water quality parameters monitoring. When data were produced by others, it will be clearly indicated in the text.

3.1 Introduction

For single stage filtration systems treating surface water, the primary objective is the control of microbial pathogens. Although the removal of DOM is a non-acute issue, high concentration contributes to colour, taste, smell of the water as well as the possible regrowth of pathogens in the storage or transport system (Matilainen et al. 2011, Volk et al. 1997).

Slow sand filters are an example of such single stage filtration system that have demonstrated excellent removal of pathogens, while still being able to remove DOM, iron and manganese somewhat (Guchi 2015). rapid-rate BAC filtration is an alternative environmentally friendly, smaller scale method for the removal of bioavailable organic compounds that also needs low maintenance and benefits from the adsorption of GAC. The main objective of a rapid-rate BAC filtration is biodegradable organic matter removal, rather than the control of microbial pathogens. Combining the two methods into a slow flow BAC filter for the off-grid treatment of drinking water could bring together the benefits of both systems: control of microbial pathogens and increased DOM removal from the start of operation.

The objective of this chapter is therefore to (1) characterize DOM removal and other chemical water quality parameters and (2) investigate the impact of filter length for a slow-flow BAC. To this aim, multiple pilot-scale biofilters of various sizes were built and used for a water treatment experiment over 6 months. Filter performance was quantified by measuring twenty chemical water quality parameters in the influent and effluent water. The chemical composition of the effluent water was compared to water quality regulations. Also, the removal of DOM, regulated metals and anions were closely monitored from virgin to steady state. Only filter length varied in the experiments, other operational parameters and influent water were constant.

3.2 Materials methods

3.2.1 Raw water

Untreated reservoir water was collected from Pateshillwater treatment works. This water source was chosen for its proximity to the laboratory. Water quality parameters are presented in Table 3.1. Approximately 500 L of raw water was collected every two weeks. The water was prefiltered on-site using a submersible pump and a 10 µm polypropylene cartridge filter. The raw water was then stored at room temperature until usage.

Table 3.1: *Water quality averaged between August to November 2017 provided by Scottish Water*

Parameter	Unit	Value
Colour	<i>mg/l Pt/Co</i>	168
Hydrogen ion	<i>pH value</i>	7-8
TOC	<i>mg-C/l</i>	18
Ammonium	<i>mg/l</i>	<0.02
Turbidity	<i>NTU</i>	1.5
Aluminium	<i>µg/l</i>	264
Iron	<i>µg/l</i>	647
Manganese	<i>µg/l</i>	43
Presumptive E. coli	<i>CFU in 100ml</i>	7
Presumptive coliforms	<i>CFU in 100ml</i>	22
conductivity	<i>µS/cm</i>	118
temperature	<i>°C</i>	22

3.2.2 Granular activated carbon

Cabot Norit GAC 1240 W was used as a filter bed medium. Its properties can be found in Table 3.2. Prewashing of the GAC was done in batches of 260 g of GAC by rinsing it twice with 260 ml of 18.2 M Ohm deionised carbon-free (hereafter referred to as Milli-Q) water. Another 260 ml of Milli-Q water was added, followed by shaking of the bottle. Trapped air was released by opening the lid. The GAC in Milli-Q water was then sonicated at full power for 10 minutes in a sonicating bath. The GAC was left to settle in the Milli-Q water for 48 hours. This procedure was repeated until sufficient GAC was washed to fill all filters.

Table 3.2: *Characteristics Cabot Norit GAC 1240 W used as medium for the biofilters.*

Characteristic	Unit	Value
Particle size > 12 mesh (1.70 mm)	<i>mass-%</i>	Min. 10
Particle size < 40 mesh (0.425 mm)	<i>mass-%</i>	Max. 5
Moisture (as packed)	<i>mass-%</i>	Max. 5
Iodine number	-	975
Methylene blue adsorption	<i>g/100g</i>	20
Total surface area (B.E.T.)	<i>m²/g</i>	1100
Apparent density	<i>kg/m³</i>	500
Density backwashed and drained	<i>kg/m³</i>	445
Ball-pan hardness	-	97
Effective Size D10	<i>mm</i>	0.6 - 0.7
Uniformity coefficient	-	1.7
Ash content	<i>mass-%</i>	12
Water soluble Ash	<i>mass-%</i>	0.1
pH	-	Alkaline
Dechlorination halving value	<i>cm</i>	2.5

3.2.3 Biofilter design

Biofilters with 30 and 60 cm GAC filter beds were constructed in triplicate, while 90 cm GAC filters bed were constructed in four separate triplicates, resulting in 18 individual filters. The layout of the experimental biofilters and dimensions of key components are sketched in Figure 3.1 on the next page and the biofilter characteristics and operational settings are given in Table 3.3.

Table 3.3: *Biofilter characteristics and the corresponding operational settings.*

Characteristic	Unit	30cm filter	60cm filter	90cm filter
GAC depth	cm	30	60	90
Column inner diameter	<i>cm</i>	2.6	2.6	2.6
Filter volume	<i>cm³</i>	159	319	478
Packed bed density	<i>kg/cm²</i>	4.5 ± 0.1	4.5 ± 0.1	4.5 ± 0.1
Flowrate	<i>ml/min</i>	1.06	1.06	1.06
Empty bed contact time	<i>h</i>	2.5	5.0	7.5
Run period	<i>h</i>	3864	3864	3864
Bed volume	-	1546	773	515

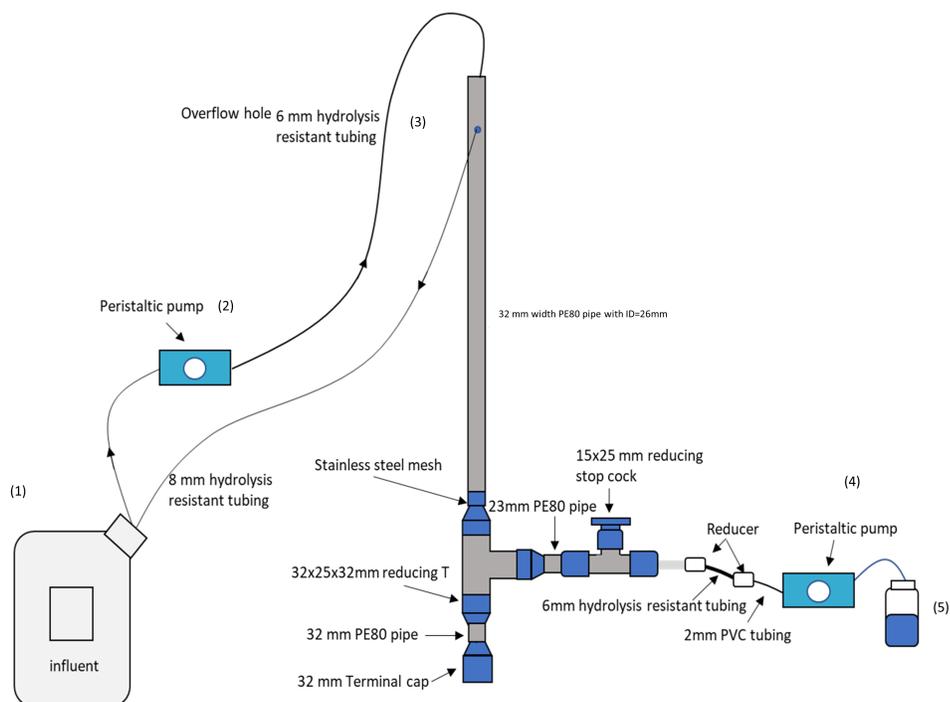


Figure 3.1: *Biofilter design including the influent jerrycan (1), peristaltic pump that pump the influent water into the filter (2), overflow hole (3), peristaltic pump that keeps the flow rate at 1ml/min (4), and the sample collection bottle (5).*

For the 30, 60 and 90 cm filters PE80 pipes with an internal diameter of 2.6 cm and length of 60, 90 and 120 cm, respectively, were used. This pipe diameter was associated with the plumbing material used to construct the filters. For future up-scaling, multiple pipes can be bundled resulting in a higher volume of water to be treated. Stainless steel meshes at the ends of the pipes with an aperture of 75 μm were used to hold the GAC in place. Peristaltic pumps (Watson Marlow 300 series with three attached pump heads, 313 OEM) were used to supply water to each filter triplicate. An overflow hole was drilled 50 mm below the top-end of each filter to allow for excessive water supply to flow back to the jerrycan creating a constant water head for all filters. Water exiting the filter initially passed through a needle valve that was set to create a 1 ml/min flow rate for each filter. However, soon after the start of the experiment, the needle valve was replaced by peristaltic pump system which permitted much greater control of the flow rate. This peristaltic pump (Watson Marlow 300 series peristaltic pump with a 5 channel microcassette pump head) was set to 7 rpm, providing a flow rate of 1 ml/min. One pump was used per filter triplicate.

3.2.4 Filter bed packing and initiation

The GAC was added to each filter in small batches by pouring the GAC Milli-Q suspension through a funnel into the neck of the biofilter. The GAC was pushed down with a self-designed packing tool, consisting of a circular plastic disk with a 28 mm diameter attached to a 1 meter stainless steel pole, until resistance was felt. The pipes were filled until 300 mm below the top so that each filter could have a 232 ml head of water above the GAC packing.

Figure 3.1 shows the set-up of the biofilter while filtering reservoir water. Before doing so, the system needed to be flushed. For this Milli-Q water was circulated through the filters to clean the GAC. This flushing is not shown in Figure 3.1, but can be found in Appendix A, Figure A.1 (pg. 181). In short, water flowed from an autoclaved 1L Nalgene bottle (1) to the top of the filter via a peristaltic pump (2) working at 70 RPM. Overflow at the entrance of the filter (3) was circulated back to the bottle. The flow rate was regulated by another peristaltic pump (4) at the exit of the filter at 1 ml/min. The effluent water was collected in the bottle (1) and circulated back to the top. The filters were run in this self-cycling way for two weeks, with the Milli-Q water being changed after 3, 5 and 7 days. No measurements were done during this self-circuiting period. After two weeks of circulation Milli-Q water, GAC had to be added the filters where the GAC had compacted.

3.2.5 Experimental run

After two weeks of circulation, raw water was introduced to the filters. The filters were run continuously in a downstream mode. One jerrycan 30 L supplied three replicate filters of each filter set. Containers with raw water were replaced with new ones before they ran dry to maintain continuous water flow. For each of the 18 filters effluent water was collected separately in 1 L Nalgene bottles. The 30 and 60 cm filters as well as one of the 90 cm filter triplicates were operated, and chemical water quality parameters monitored for 23 weeks and were afterwards deconstructed to examine microbial communities within the filter as well as the water chemistry of the pore water. The three remaining 90 cm filter triplicates were each run for a different period of time, being 5, 9, and 12 weeks and each triplicate was deconstructed at the end of their operation also to analyse the microbiology and chemistry of the inside of the filter (See Chapter 5 Section 5.2, pg. 111 for deconstruction and pore water chemistry analysis).

3.2.6 Sampling campaign

Room temperature, pH, and Conductivity, TOC, DOC, anions (nitrate (NO_3^-), nitrite (NO_2^-), phosphate (PO_4^{3-}), sulphate (SO_4^{2-}), chloride (Cl^-), and cations (ammonium (NH_4^+)) and metals (Aluminium (Al), Chromium (Cr), Manganese (Mn), Iron (Fe), Nickel (Ni), Copper (Cu), Arsenic (As), Selenium (Se), Cadmium (Cd), Antimony (Sb) and Lead (Pb) of the raw water (hereafter called influent) and effluent water was measured. Sampling campaign is found in Table 3.4.

Table 3.4: *Sampling campaign*

Parameter	Week
Temperature, Conductivity, pH	0, 2-5, 7-9, 11, 12, 15-19, 21, 23
TOC/DOC, Anions	0-12, 15, 17, 19, 21, 23
Metals (90 cm)	0-4, 6-12, 15, 17, 19, 21, 23
Metals (30 and 60 cm)	0, 3, 7, 11, 15, 19, 23

Prior to every sampling, 100 mL reusable borosilicate glass media bottles with polypropylene cap (Fisherbrand™) were washed with common detergent and thereafter rinsed thrice with deionised water, followed by an acetone rinse, and glassware subsequently heated in a Muffle oven to 550 °C for at least 6 hours.

The influent water of the 30, 60 and 90 cm filters was collected separately at the top of each filter length just before the water enters the replicate columns. Effluent water was collected approximately 15 hours after the influent water allowing sufficient time for the newly changed influent to pass through the filters. For the 30, 60, and 90 cm filter 6, 3, 2 BV passed the filter, respectively. All bottles were filled to the top to avoid headspace and stored at 4 °Celsius.

3.2.7 Chemical analysis

TOC/DOC analysis

TOC and DOC were analysed using combustion (TOC-L, Shimadzu, Japan) (Shimadzu 2012). See Appendix A, Figure A.2 (pg. 182) for a schematic representation of the technique. Sample pre-treatment was done within 24 hours after sampling and analysis was done the same day of sample preparation. For TOC measurements, 10 ml disposable borosilicate vials (Fisherbrand™), were filled with 8 - 10 mL sample after the collection bottle was gently shaken. For DOC measurements, similar

vials were used to prepare 8-10 mL sample after filtration through 0.2 μm polyEther sulfone membrane filters (Whatman Puradisc).

For all weeks, the instrument was run using a pre-set method. In short, non-purgeable organic carbon was used as the measure of TOC / DOC. Samples were loaded in the autosampler. In the machine 50 μl water samples were sparged to which a small amount 1.5% of 1 M hydrochloric acid was added, in this way the inorganic carbon in the sample was converted to carbon dioxide during 1.5 min of sparging using 80 mL nitrogen carrier gas. This carbon dioxide was removed, and the remaining total carbon was measured with help of 680 $^{\circ}\text{C}$ combustion catalytic oxidation achieving total combustion inside TC combustion tubes filled with a platinum catalyst. Between the samples, the system was flushed twice. The integration time was between 0 and 4 minutes and 50 seconds. Every sample was analysed twice when standard deviation was 0.1 and coefficient of variation 2%. When two samples exceeded these values a third analysis was done with the best two outcomes used.

For each sample, the machine analysed using 2-3 technical replicates depending on the relative standard deviation of the results. Also, 3 vials with 100 mg-C/L potassium hydrogen phthalate (TC stock) solution were loaded in the autosampler from which the machine made its own dilutions towards 2, 4, 6, 8, 10 mg-C/L, measured their signal and calculated the calibration curve. With help of this calibration curve, the measured signals of the samples were automatically converted to concentrations. Every 5-10 sample a blank was measured as well as a positive control (5 mg-C/L) at the end of the analysis. The average Limit Of Quantitation (LOQ) and Limit Of Quantitation (LOD) throughout the experimental run was 0.59 ± 0.2 and 0.20 ± 0.067 mg-C/L, respectively. Data were extracted as .txt file for further data analysis.

Anions and cations analysis

NO_3^- , NO_2^- , PO_4^{3-} , SO_4^{2-} , Cl^- , and NH_4^+ were monitored by ion chromatography (DionexTM IC900 and IC1100 for anions and cations, respectively) (Thermo Fisher Scientific 2012*a,b*). See Appendix A, Figure A.3 (pg. 183) for a schematic representation of the technique. For sample preparation, the same day as the sampling 0.5 ml AS-DV Autosampler PolyVials vials (DionexTM), were filled with 0.5 mL sample volume after filtration through a 0.2 mm Chromacol nylon syringe filter (Thermo ScientificTM) and syringe. Tubes were capped with the compatible caps (DionexTM). Samples were analysed by ion chromatography the same day of sample preparation. For all weeks, the machine run using a pre-set method. The method was in loop mode having a delay volume of 125 μl , a delivery speed of 4 mL/min, and a flush factor of 10. In short, an eluent brings with help of a pump 50 μl sample via

a guard column, which removes contaminants, to the separation column (AS23 column for the IC900 and CS16 column for the IC1100) where the ion chromatography takes place. The AS23 column separates oxyhalides and common inorganic anions in drinking water, groundwater, wastewater, and other diverse sample matrices. The CS16 is optimized for the determination of unequal concentration ratios of adjacent eluting cations such as sodium and ammonium in diverse sample matrices.

The separated ions pass the suppressor that selectively enhances detection of the sample ions while suppressing the conductivity of the eluent. The conductivity cell measures the electrical conductance of the sample ions and produces a signal based on a chemical or physical property of the analyte. The runtime for each sample was 30 minutes. Data were analysed in Dionex™ Chromeleon (version 7.1.3.2425) and data extracted as .csv files for further data analysis.

For each sample, the machine analysed 1 technical replicate to minimize sample load. Also, 0.05 (or 0.1), 0.25, 0.5, 0.75, and 1 mg/L dilution series for an anion and cation standard were made separately using the similar vials as the samples. Anion standard (Dionex™ Combined Seven Anion Standard II) included $\text{Cl}^- = 100 \text{ mg/L}$; $\text{NO}_2^- = 100 \text{ mg/L}$; $\text{NO}_3^- = 100 \text{ mg/L}$; $\text{PO}_4^{3-} = 200 \text{ mg/L}$; $\text{SO}_4^{2-} = 100 \text{ mg/L}$. Cation standard (Dionex™ Combined Six Cation Standard II) included $\text{NH}_4^+ = 250 \text{ mg/L}$.

Calibration standards and samples were loaded in the autosampler, signals were measured, and the calibration curve calculated automatically. With help of this calibration curve, the measured signals of the samples were converted to concentrations. Every 5-10 sample a blank was measured together as well as a positive control (0.5 mg/L) at the end of the analysis. The LOQ / LOD for the cation and anion standards monitored throughout the experimental run was: $\text{NH}_4^+ = 0.2 \pm 0.2 / 0.067 \pm 0.067$, $\text{NO}_2^- = 0.08 \pm 0.06 / 0.027 \pm 0.02$, $\text{NO}_3^- = 0.11 \pm 0.06 / 0.037 \pm 0.02$, $\text{Cl}^- = 0.8 \pm 0.7 / 0.3 \pm 0.023$, $\text{SO}_4^{2-} = 0.4 \pm 0.3 / 0.13 \pm 0.1$, and $\text{PO}_4^{3-} = 0.3 \pm 0.1 / 0.1 \pm 0.03 \text{ mg/L}$. In- and effluent concentrations of ammonium and nitrite were below the LOQ during the course of the experiment and their data not further reported.

Metal analysis

Dissolved metals were analysed by Inductively Coupled Plasma Mass Spectrometry, also known as ICP-MS (Radboud University n.d.). See Appendix A, Figure A.4 (pg. 184) for a schematic representation of the technique. All dissolved metals which are relevant for EU regulatory requirements (Appendix A, Figure A.6, pg. 186) except B, Na and Hg, were included in the measurements.

For sample preparation, 15 ml plastic centrifuge tubes (Fisher Scientific™) were filled with 10 - 15 mL samples. Samples were filtered through the 0.2 mm Chromacol nylon syringe filter (Thermo Scientific™). Two drops of pure nitric acid were added to the sample for conservation.

The samples were analysed by the chemistry department for ICP-MS at Strathclyde University within 6 months using three technical replicates. In short, ICP-MS is a technique to determine low-concentrations (range: ppb = parts per billion = µg/l). The sample solution is introduced into the device by a peristaltic pump. There it becomes vaporized in a spray chamber. The resulting aerosol is injected into an argon-plasma that has a temperature of 5726 - 7726 °C. Inside the plasma torch, solution is removed from the sample. Moreover, the sample gets atomization and ionization occurs. Only a small amount part of the ions produced in the plasma further penetrates to the mass-spectrometer part where it is being analysed. The LOQ / LOD was determined as: Al = 0.02 / 0.007, Cr = 0.07 / 0.023, Mn = 0.02 / 0.007, Fe = 3 / 1, Ni = 0.006 / 0.002, Cu = 0.006 / 0.002, As = 0.03 / 0.01, Se = 0.6 / 0.2, Cd = 0.008 / 0.003, Sb = 0.01 / 0.003, Pb = 0.02 / 0.007 µg/L, respectively.

SEM

GAC samples were analysed by Scanning Electron Microscopy with Energy Dispersive X-ray spectroscopy, also known as SEM-EDX (Thermo Fisher Scientific n.d.). A schematic overview of the technique is given in Appendix A, Figure A.5 (pg. 184). For sample preparation, virgin GAC as well as GAC collected from the top 10 cm of the 90 cm filter at week 23 were dried at room temperature and transferred to sterile 50 mL containers for analysis at a later time. SEM-EDX makes high resolution images of surface topography, with excellent depth of field. It uses a highly-focused, scanning (primary) electron beam. These primary electrons enter a surface with an energy and produce many low energy secondary electrons. The image of the sample surface is constructed by measuring secondary electron intensity as a function of the position of the scanning primary electron beam. Moreover, backscattered electrons and X-rays are also generated by the primary electron bombardment. Atomic number(s) of the element(s) within the sample are correlated to the intensity of backscattered electrons. The analysis of characteristic X-rays emitted from the sample gives the more quantitative elemental information.

Statistical analysis

All statistical analyses were carried out using R (version 4.0.2). The correlations were determined by Pearson correlation using an alpha significance level of 0.05. For filter length comparison, significant differences were ascertained by Kruskal-Wallis (KW) H test using a alpha significance level of 0.05,

as the data were not normally and homogeneously distributed as well as sample size small, followed by the Dunn's post-hoc test (equal variances, unequal group sizes) to examine the pairwise comparison with p-values adjusted by the Bonferroni method. The results were expressed by the mean for each group together with its variance (expressed by the standard error).

Biomass in apparent steady state was quantified for the three filter lengths by measuring the ATP on the GAC measured by MV and significant difference was determined by ANOVA with data being normal and homogeneously distributed (equal variances, unequal group sizes), followed by the Tukey post-hoc test to discriminate between filter lengths with an 95% confidence level.

Non-parametric multiplicative simple imputation was used to replace values below detection limit in the compositional data set following Martin-Fernandez et al. (2003). A cluster heatmap with rows and columns ordered using hierarchical (complete linkage) clustering was used to visualize clusters of samples and variables, and to simultaneously help find the variables that appear to be characteristic for each sample cluster.

3.3 Results and Discussion

3.3.1 Water quality and regulatory limits

Table 3.5: *chemical water quality parameters min, max and mean concentrations in the influent and effluent water of the biofilter together with their regulatory limits. Under-scored numbers are measured values above the regulatory limits.*

Parameters	Influent (µg/L)			Effluent (µg/L)			Limit (µg/L)
	Min	Max	Mean ± sd	Min	Max	Mean ± sd	
Iron	<u>268</u>	<u>459</u>	<u>384 ± 54</u>	45	<u>340</u>	<u>250 ± 90</u>	200
Manganese	1.2	3.7	2.1 ± 0.7	1	8	2 ± 1	20
Copper	0	4	2 ± 1	0.0	4.0	1.1 ± 1.0	2000
Lead	0.1	0.7	0.3 ± 0.1	0.0	0.7	0.2 ± 0.2	10
Aluminium	42	100	80 ± 20	100	<u>900</u>	<u>200 ± 200</u>	200
Nitrate	0.3	0.8	0.6 ± 0.2	0.07	0.68	0.11 ± 0.09	50000
Nitrite	<LOQ	<LOQ	-	<LOQ	<LOQ	-	500
Chloride	18	27	20 ± 3	0	30	19 ± 7	250000
Sulphate	0	6	3 ± 2	0	7	2 ± 2	250000
pH	7.3	8.1	7.9 ± 0.2	7.0	9.2	7.7 ± 0.4	6.5 - 9.5
Conductivity	111	131	118 ± 4	98	134	116 ± 8	2500 µS/cm
TOC	10	15	13 ± 1	0	13	6 ± 4	-
Antimony	0.09	0.46	0.17 ± 0.08	0.1	3.3	0.5 ± 0.7	5
Arsenic	0.2	2.5	0.5 ± 0.5	0	<u>85</u>	10 ± 20	10
Nickel	0.50	1.4	0.8 ± 0.2	1	12	4 ± 2	20
Cadmium	<LOQ	0.11	0.02 ± 0.03	0.0	0.6	0.1 ± 0.1	5
Chromium	0.3	1.5	0.5 ± 0.2	0.0	0.7	0.3 ± 0.2	50
Selenium	0	8	3 ± 2	0	<u>16</u>	4 ± 3	10
Phosphate	0.180	0.180	0.180 ± 0.001	0.2	0.9	0.3 ± 0.2	-

Twenty chemical water quality parameters were monitored in the influent and effluent of biofilters operating from start up to 23 weeks of operation. Table 3.5 presents an overview of the minimum, maximum, and mean concentrations of these parameters in both waters. Sixteen parameters did not show concentrations above the regulatory drinking water limits, neither in the influent or the effluent. The trend of iron, TOC, aluminium, and arsenic was evaluated in more detail (Figure 3.2).

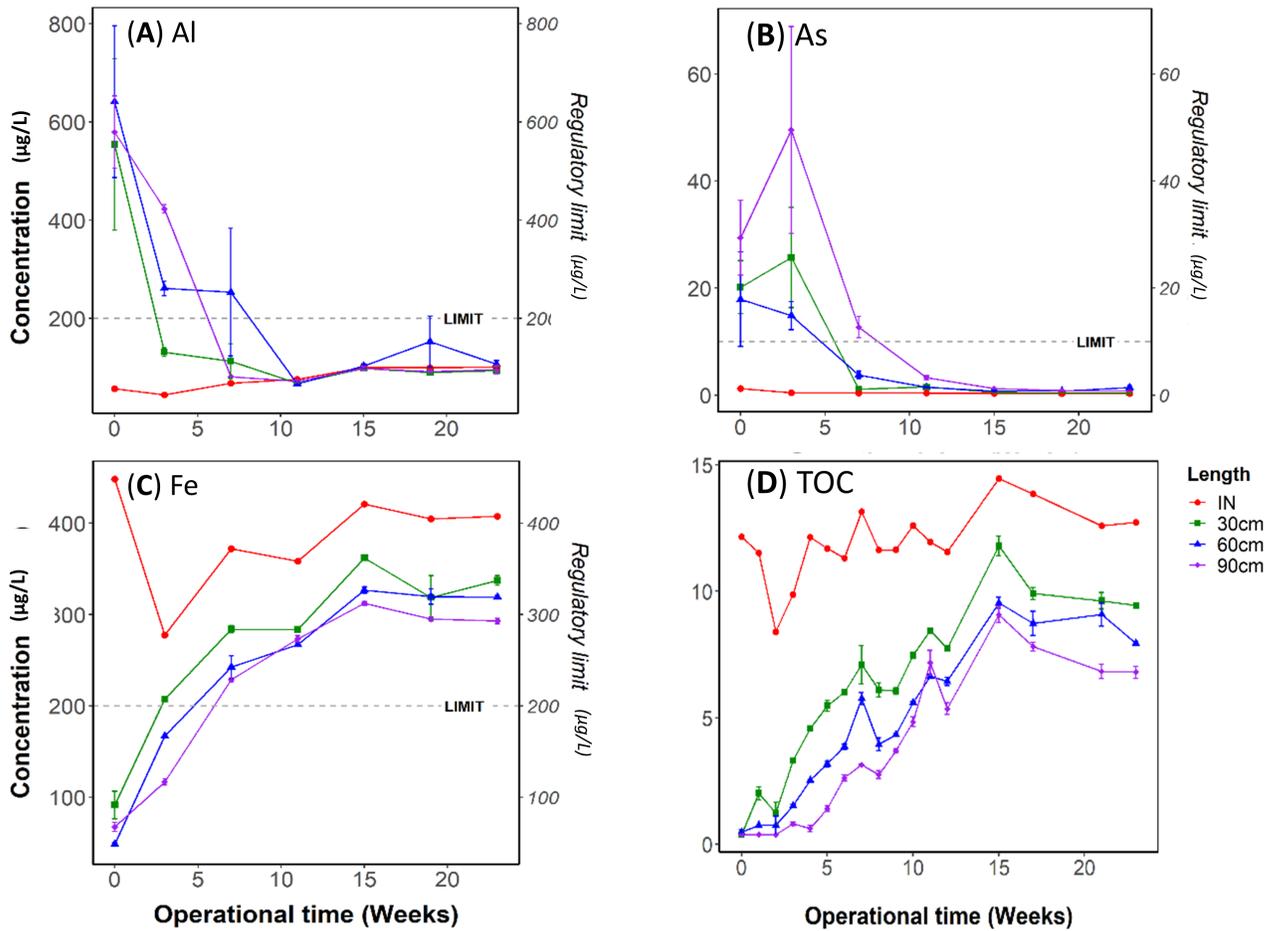


Figure 3.2: Concentration of aluminium (A), arsenic (B), iron (C), TOC (D). Left y-axis shows the measured concentration of the influent and the 30, 60, 90 cm effluent (mean \pm s.e., $n = 3$ replicates) over 23 weeks of operation. The right axis together with the dashed line (when present) highlights the regulatory limits.

Iron concentration in the influent was higher than the regulatory drinking water limit, while aluminium and arsenic showed exceeding levels in the effluent. The trends of the other parameters can be found in Appendix A, Figure A.7 and A.8 (pg. 187 and 188). Up to the first ten weeks of the experiment aluminium and arsenic concentrations in the effluent were higher than the regulatory limits, while not in the influent (Figure 3.2 A and B). Aluminium and arsenic concentrations in the influent water were 80 ± 20 and 0.50 ± 0.50 µg/L, respectively, while the effluent water at the start of the experimental run was 600 ± 100 and 20 ± 7 µg/L. These effluent concentrations were approximately 3 and 2 times higher than their regulatory limit, respectively. After approximately 10 weeks these increased levels in the effluent diminished. This suggests that the GAC material initially added these components to the water. Throughout the experiment, iron influent concentration varied from 270 to 460 µg/L,

averaging 380 µg/L with a standard deviation of 50 µg/L. This exceeded the regulatory limit of 200 mg/L. Measurements show that filtration removed the iron significantly during the first weeks of operation, irrespective of filter length (Figure 3.2 C). The effluent iron concentration, however, gradually increased already shortly after the start of the experiment. The 30, 60 and 90 cm filter surpassed the regulatory limit after approximately week 3, 5 and 7, respectively. This suggests that iron could be a problematic chemical water quality parameter when biofilters are solely being used for drinking water treatment in rural Scotland.

TOC was removed significantly during the first weeks of operation for all three filter lengths (Figure 3.2 D). TOC is not mentioned in regulations but is still relevant in drinking water context because it may colour water or add odours and taste, and hence influence people's perception of clean drinking water (Scottish Water 2015). Moreover, it contributes to regrowth of bacteria in storage and distribution systems (Peterson & Summers 2021, Terry & Summers 2018, van der Kooij & van der Wielen 2011). However, it does not cause health risks (Drinking Water Quality Regulator 2020a). TOC concentration in the influent water ranged from 10 to 15 mg/L, averaging 13 mg/L with a standard deviation of 1 mg/L. Similar to iron, the TOC effluent concentration seems to gradually increase over time after initial removal, possibly as a result of GAC saturation. These observations show that the long-term performance of the filters to remove TOC is not yet fit for purpose. However, as long as the bioavailable fractions are being removed from the water it is safe to store for later consumption irrespective of DOM content (Hijnen et al. 2014, van der Kooij & van der Wielen 2011).

3.3.2 Removal efficiency

For a more detailed evaluation of the biofilter performance, the removal efficiency (Equation 3.1) of the various chemical water quality parameters was monitored over time (Figure 3.3 and 3.4). Some parameters had removal efficiencies above zero, while other were below zero. This negative removal meant that the concentration in the effluent was higher than the influent. In such case, the filter was leaching.

Results show that the filters removed DOC, TOC, iron, nitrate, copper, chromium, lead, chloride, and sulphate (Figure 3.3 A-I). DOC, TOC, iron, and nitrate were removed over the whole operational period. The 30 and 90 cm filters behaved significantly differently for these chemicals, resulting in a later onset of breakthrough and a higher removal in steady state for the 90 cm filter (Dunn's post-hoc test on KW; TOC $H(151) = 18.50$, DOC $H(151) = 17.26$, (iron) $H(63) = 16.78$, (nitrate) $H(151) =$

22.15, $p_{\text{adj}} < .001$). Copper, chromium, and lead were also removed by the filter, however, their removal values occasionally dropped below zero, meaning the effluent concentration was higher than that of the influent. These metals demonstrated low concentrations in the influent and effluent (See Appendix A, Figure A.8, pg. 188), and the magnitude of the measurements could be within the error resulting in the occasional higher effluent than the influent. Chloride and sulphate were only removed at the start of the experiment after which the chloride removal dropped to zero and the sulphate removal became negative. Phosphate, arsenic, aluminium, antimony, and cadmium showed negative percentage removal during the first weeks of the experiment (Figure 3.4 J-N). The negative percentage removal was a result of the division of the higher effluent concentration by the influent concentration.

$$Removal(\%) = \frac{Concentration_{influent} - Concentration_{effluent}}{Concentration_{influent}} \times 100 \quad (3.1)$$

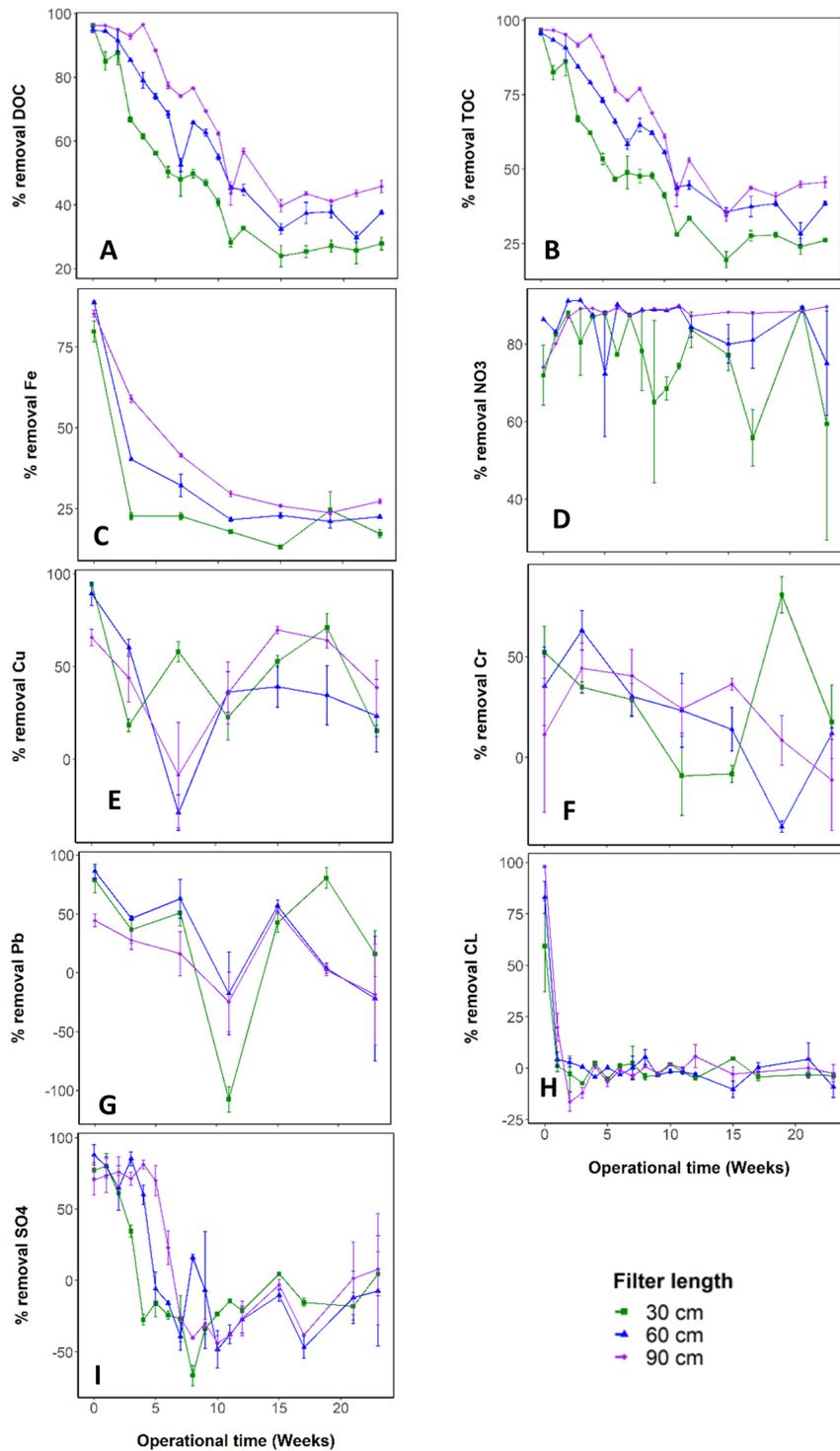


Figure 3.3: Removal in % of DOC (A), TOC (B), iron (C), nitrite (D), copper (E), chromium (F), lead (G), chloride (H), sulphate (I) by the 30 (green), 60 (blue) and 90 (purple) cm filter (mean \pm s.e., $n = 3$ replicates) within the biofilter during the first 23 weeks of operation.

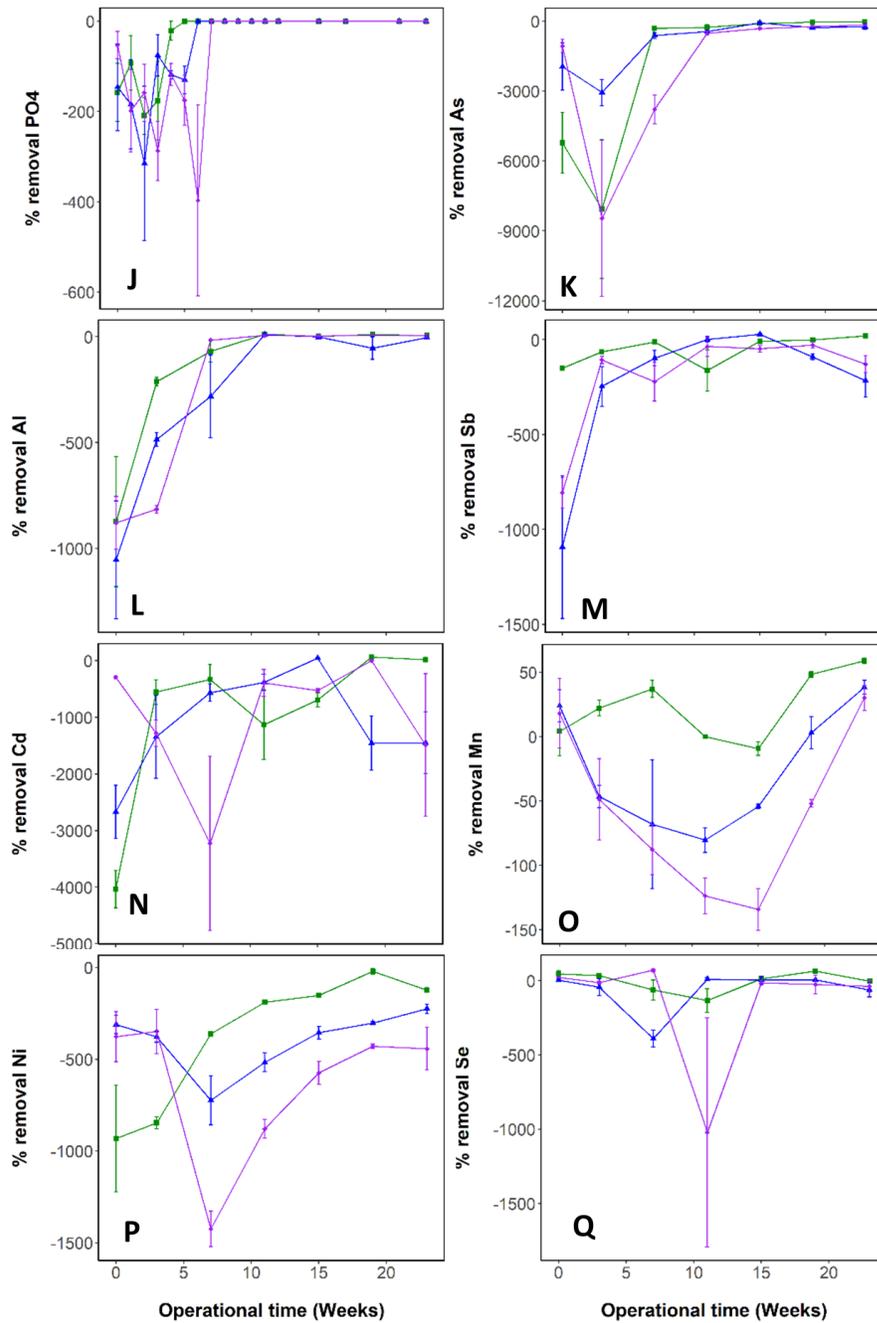


Figure 3.4: Removal in % of phosphate (**J**), arsenic (**K**), aluminium (**L**), antimony (**M**), cadmium (**N**), manganese (**O**), nickel (**P**), selenium (**Q**) by the 30 (green), 60 (blue) and 90 (purple) cm filter (mean \pm s.e., $n = 3$ replicates) within the biofilter during the first 23 weeks of operation.

Heatmap

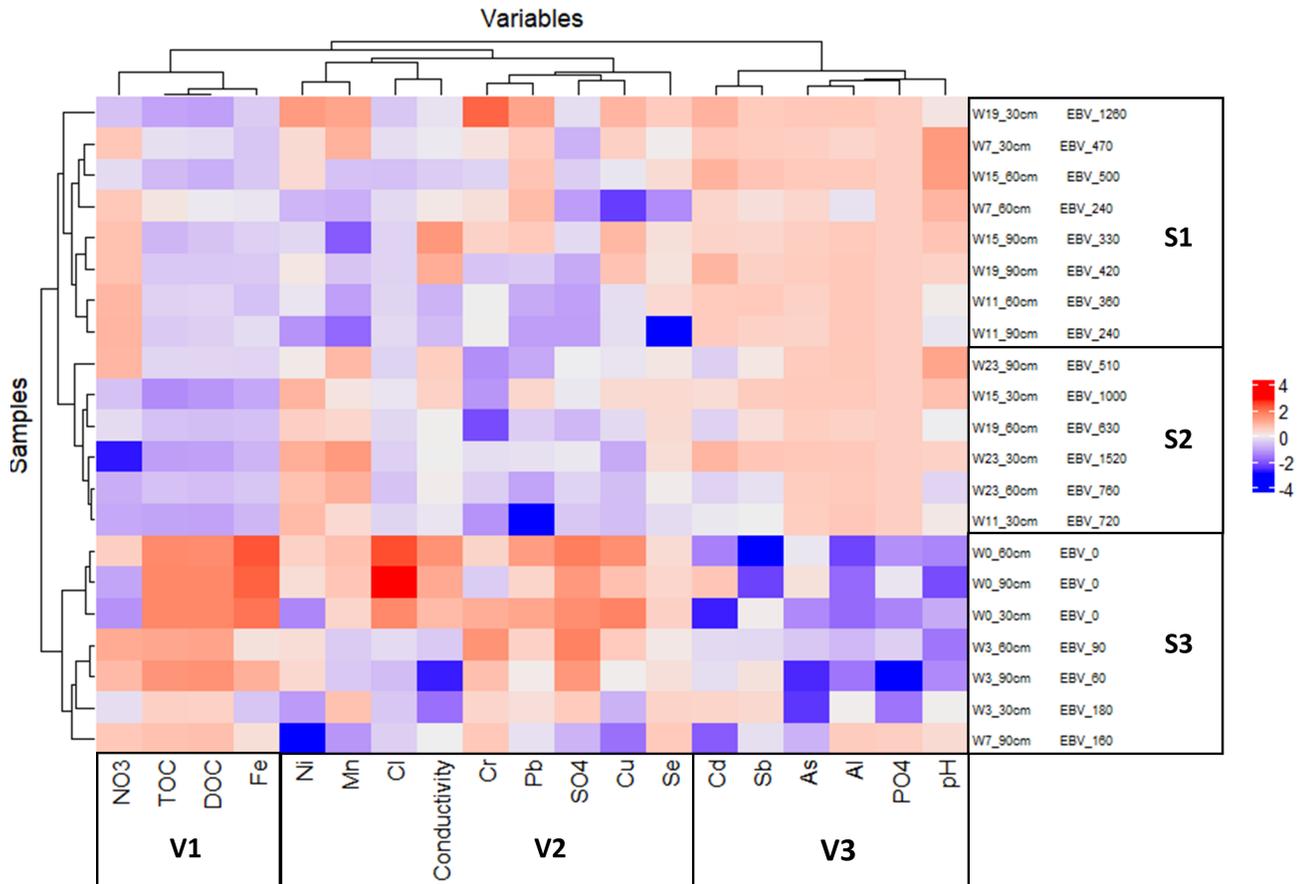


Figure 3.5: Cluster heatmap of chemical water quality variables removal by the biofilter both rows and columns ordered using hierarchical (complete linkage) clustering. Variables clustered by V1 - V3 and biofilter samples clustered by S1 - S3.

The multivariate information about the removal of the chemical water quality parameters was also visualized by a heatmap (Figure 3.5) which combined and visualized the removal (influent concentration-effluent concentration) between two dimensions (chemical water quality variables and samples) in order to identify patterns. The characteristics of the samples and corresponding names are given in Appendix A, Table A.1 (pg. 189). To facilitate clustering of samples having different EBCTs, operation time was expressed as throughput in BV (Equation 3.2).

$$BV = \frac{\text{Volume of treated water}}{EBCT} \quad (3.2)$$

The samples were divided in three sections: (S1) clusters the samples with high BV (> 500), (S2) clusters the samples with medium BV (200 - 500), (S3) clusters low BV (< 200). The variables on the other hand clustered into three sections (V1 - V3). V1 included variables that were removed by the filter, while V3 included variables that leached from the filter during the first week of operation. V2 are the remaining variables. In the heatmap, dark blue represents a negative removal (leaching), while dark red denotes a high removal. The dendrograms of the variables (on top of the heatmap) and samples (on the left side of the heatmap) illustrate the arrangement of the clusters. In other words, to what extent samples or variables behave similarly.

Variable cluster V2

Cluster V2 includes nickel (Ni), manganese (Mn), chloride (Cl^-), conductivity, chromium (Cr), Lead (Pb), sulphate (SO_4^{2-}), copper (Cu) and selenium (Se). All these variables did not demonstrate a clear removal or leaching in the first weeks. For example, nickel leaching was constantly present but reduced over time, with the 90 cm excreting significantly more nickel than the 30 cm filter (Dunn's post-hoc test on KW; $H(63) = 10.29$, $p.\text{adj} = 0.06$) (Figure 3.4 O). Manganese leached from the 60 and 90 cm filters, while the 30 cm filter showed some removal. As result significance difference was present between the 30 and 60 cm filters as well as 30 and 90 cm filters (Dunn's post-hoc test on KW; $H(63) = 17.82$, $p.\text{adj} < 0.001$). Selenium concentration difference between in- and effluent was negligible (Figure 3.4 Q), while the other metals (chromium, lead, and copper) were removed from samples with low BV while at higher BV occasional leaching was picked up (Figure 3.3 E - G). This negative removal could be a result of leaching but could also be a result of small variations and low concentrations of the variable between in- and effluent. At very low concentration small variations can result in relatively high negative removal. Chloride was removed during the first week and afterwards showed minor leaching (Figure 3.3 H), sulphate showed significant removal during the first weeks of the experiment, which lasted longer than that of chloride (Figure 3.3 I). Removal reversed to leaching between week 7 and 10 with higher effluent concentrations compared to influent. After week 10 in- and effluent concentrations are statistically equal. Interestingly, nitrate (NO_3^-), another anion, was part of cluster V1 and demonstrated almost full removal throughout the experimental run.

The question is why nitrate keeps on being removed while chloride and sulphate reverse into leaching. Electrosorption studies have demonstrated a preferential initial adsorption of chloride on the GAC over nitrate (Chen et al. 2015, Mubita et al. 2019). In both studies adsorption of nitrate became domi-

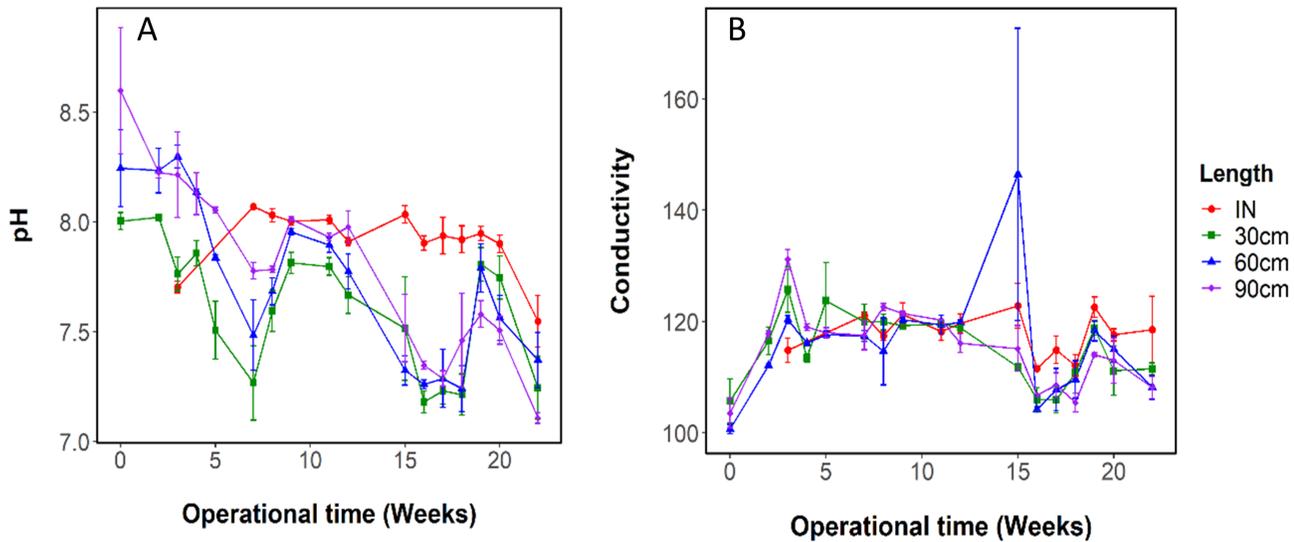


Figure 3.6: *pH (A) and conductivity (B) measured in the influent water and 30, 60, and 90 cm filter effluent (mean \pm s.e., $n = 3$ replicates) of the biofilter over 23 weeks of operation showing the data points of the three filter lengths.*

nant over that of chloride, which even caused the chloride to desorb again. With selectivity coefficient $K_{Cl/NO_3^-} > 1$, the competitiveness of nitrate is stronger than chloride, and as a result desorption of chloride occurs. Chen et al. (2015) did not observe obvious selectivity between chloride and sulphate. This would suggest that they behave similarly. These findings could explain the significant removal of chloride and sulphate in the present study during the first weeks of operation followed by the (minor) leaching from the filter.

The conductivity clustered most with chloride in the heatmap. Figure 3.6 (B) shows the conductivity in the influent and effluent during the experiment. Although the influent concentration was not registered in the first two weeks, a concentration higher than the effluent can be expected, because of the minimal variation of conductivity over time. Indeed, conductivity was, similar to chloride, lower in the effluent in the first weeks, while it increased quickly. When plotting the change in chloride against the change in conductivity (Figure 3.7), it shows a moderate correlation (Pearson correlation, $r(61) = .57$, $p = .001$). At the start of the experiment negatively charged molecules are retained by the filter causing the conductivity in the effluent to decrease. When the chloride is leaching from the filter the conductivity increases to above the value of the influent. The conductivity was less influenced by the sulphate leaching, probably as a result of much lower concentrations of sulphate (average 2 mg/L) compared to chloride (average 19 mg/L).

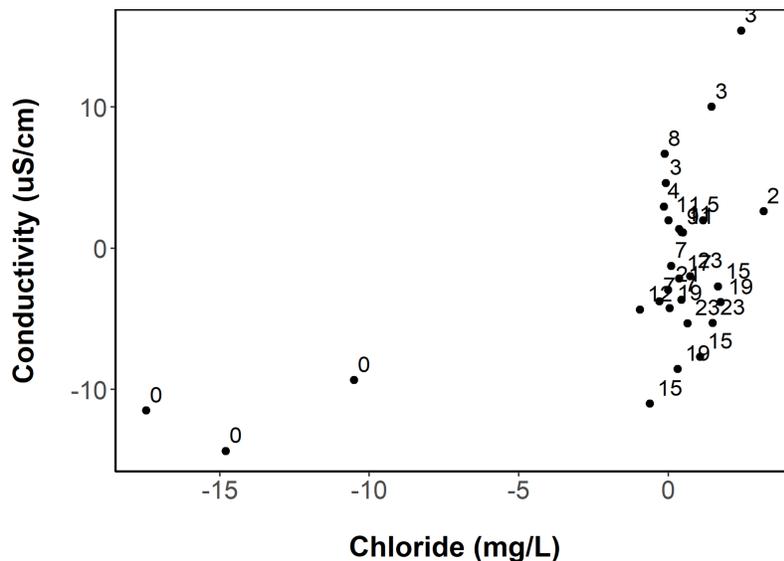


Figure 3.7: Scatter plot of the conductivity versus chloride concentration measured in the biofilters during the 23 weeks of operation.

Variable cluster V3

Cluster V3 includes the chemicals aluminium (Al), arsenic (As), phosphate (PO_4^{3-}), antimony (Sb), and cadmium (Cd) together with pH. This third variable cluster consisted of chemical water quality variables that demonstrated overshoot at the first week of operation when the pH was elevated. Their removal was negative in sample cluster S1 (blue), which disappeared in sample cluster S2 and S3. Their leaching was also clearly noticeable in Figure 3.4 J-N. The discard happened randomly in the three filters with no clear influence of filter length. The leaching of the metals is the result of desorption of metals from the pre-existing GAC surface–metal complexes (Al-Attas et al. 2018). The used Norit W1240 GAC is made from steam activated coal that contains ash (12% total and 0.1% soluble). Indeed, leaching metals were already present in the virgin GAC. SEM-EDX imaging recorded the presence of aluminium and iron (Figure 3.8).

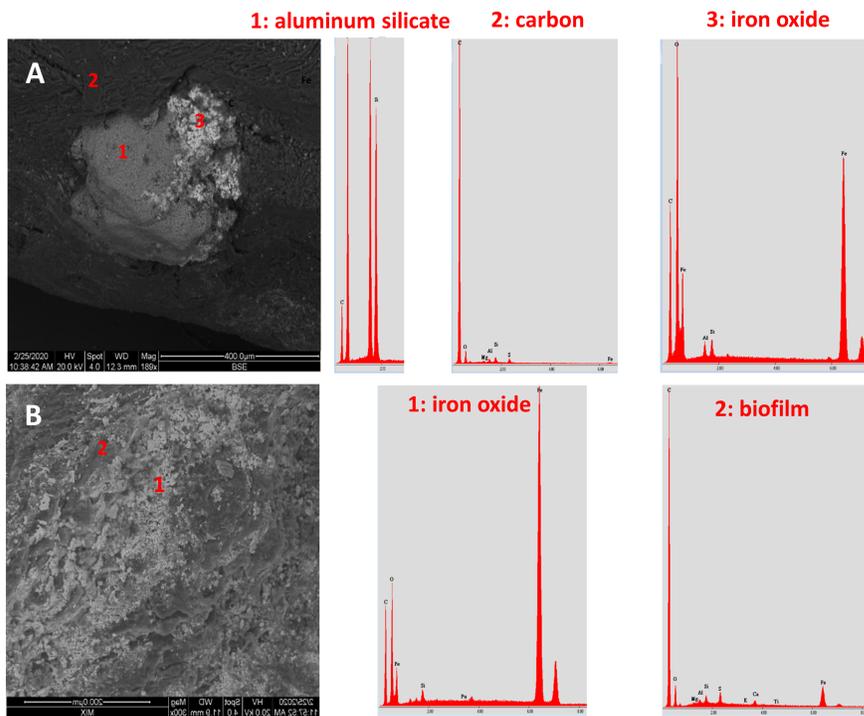


Figure 3.8: SEM-EDX image and spectra of fresh GAC (A) and GAC from the top 10 cm of the biofilter at W23 (B).

Leaching of metals from the GAC filter is a common phenomenon during the start-up of a GAC system (Grieco 2021), and the pH variation plays an important role in this overshoot (Chen et al. 2003). At the start, the surface functional groups on the virgin alkaline GAC are protonated when getting in contact with water, resulting in a rise in pH (Boehm 1994). Protonation is followed by the charge-neutralizing of the surface with negatively charged molecules present in the water such as anions and NOM (Farmer et al. 2015). This decreases the pH over time (Grieco 2021). In the present study, a rise in pH was indeed noticeable during the first week because of filtration as well as a decrease in conductivity (Figure 3.6). After week 5, the pH of the effluent decreased below the influent pH for the remaining weeks and the conductivity increased. When the pH in the filter is elevated, metals are able to leach from the carbon (Grieco 2021). In this present study, leaching did not depend on filter lengths as the overshoot took place until approximately W7-W10 for all filter lengths with high variation between all measurements.

Leaching of metals and phosphate could have been prevented by sufficient flushing of the GAC. Grieco (2021) advises 30 BV to reduce arsenic effluent concentration below 5 $\mu\text{g/L}$ and 50 BV for the pH to stabilize. Although approximately 50 BV was reached in the present study, recirculated DI water

was used which was only replaced 3 times. During recirculation an equilibrium must have built up, slowing down the flushing of the GAC. Monitoring for the rinse water could have identified this.

Variable cluster V1

Cluster V1 includes TOC, DOC, iron, and nitrate (NO_3^-) removal. This cluster represents the chemical water quality variables that were constantly being removed during filtration (Figure 3.3 A-D). It is positively associated with sample cluster S3. This means that at the start of the experiment, where BV were low, their removal was high. Cluster S1 and S2 (higher BV) contain more blue coloured blocks, indicating that the removal of these chemical parameters decreased for higher BV.

Adsorption

In the experiment, an almost hundred percent removal was visible for DOC, TOC, and nitrate in the early weeks of the experiment (Figure 3.2 and A.8). With biodegradation not being installed yet, this removal is the result of adsorption of the chemicals to the GAC. In water with near neutral pH, the alkaline GAC is positively charged and facilitates the adsorption of negatively charged NOM and anions by electrostatic interactions (Faria et al. 2004, Grieco 2021, Kołodziej et al. 2014, Paredes et al. 2016). With the filter still being unsaturated, the removal of these chemicals is driven by the adsorption during the first weeks of operation (Simpson 2008). Electrostatic interactions, pore filling, hydrophobic interaction, hydrogen bonds and p - p interaction all contribute to this adsorption process (Bhatnagar et al. 2013). The metals, however, are positively charged and the previously described processes are not expected to apply to them. In water, un-complexed metals are not commonly found (Adusei-Gyamfi et al. 2019), and are often complexed to organic or inorganic molecules. At near neutral pH the order of adsorption to humic substances has been reported as $\text{Fe} = \text{Pb} = \text{Al} = \text{Cr} = \text{Cu} > \text{Cd} > \text{Ni} > \text{Mn}$, with sorption efficiency increasing with rise in pH (Hart & Davies 1981, Kerndorff & Schnitzer 1980). In metal-NOM complexation metal binds to already existing ionized sites on the NOM molecule or by displacing a proton from its position (Adusei-Gyamfi et al. 2019). When DOM is removed by GAC via adsorption, it consequently removes the complexed metal, as GAC adsorption is believed not to dissociate the NOM-metal complexes (Al-Attas et al. 2018). In the results, this is clearly visible for DOC/TOC and iron. TOC, DOC, and iron showed a similar trend (Figure 3.3). Indeed, iron correlated strongly with DOC (Pearson correlation; $r(61) = .879$, $p < .001$) and is removed when TOC/DOC is removed.

GAC saturation

DOC behaved similarly to TOC (Pearson correlation; $r(149) = .995$, $p < .001$), and only DOC will be further discussed. All three filters showed almost 100% DOC removal at the start of the experimental run (Figure 3.3 A). However, the removal efficiency declined over time. This was most likely the result of GAC saturation, because no biodegradation is taking place yet at the start of the operation run (Korotta-Gamage & Sathasivan 2017). The 90 cm filter was expected to treat at least three times the amount of water in at least three times the number of weeks before saturation compared to 30 cm, as there was three times the amount of GAC (Biswas & Mishra 2015, Lee et al. 1983). Interestingly all filters plateaued at approximately 15 weeks after treating the same amount of water. Instead of needing the same amount of BV until equilibrium, the 90 cm filter needed much less BV, compared to the other two filters. This becomes clearer in Figure 3.9, where DOC removal efficiency is plotted against the BV. The 90 cm filter started breakthrough at 291 BV and reached apparent steady state at approximately 307 BV (slope = -0.20), while the 60 and 30 cm filter needed started breakthrough at 379 and 470 BV, respectively, and reached steady state at 462 (slope = -0.14) and 924 BV (slope = -0.08), respectively (Linear regression; $F(36,2) = 31.31$, $p < 0.05$).

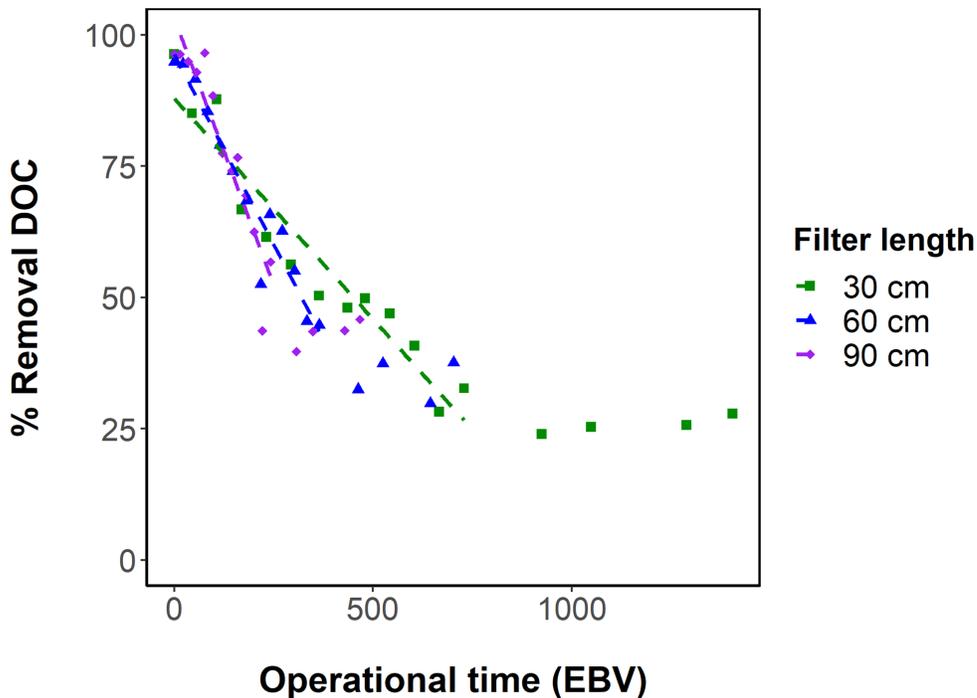


Figure 3.9: *DOC removal as function of BV (mean, n = 3 replicates) for different filter lengths (30, 60, and 90 cm) including their (dashed) regression lines over the first 15 week of operation.*

Apparent steady state

From week 15 the filter appears to be in an apparent steady state (3.3 A B), with significant difference between the filter lengths at week 23 (Dunn's post-hoc test on KW; $H(36) = 26.6$, $p_{\text{adj}} < 0.001$). The 30, 60 and 90 cm filters removed $26 \pm 3\%$, $34 \pm 3\%$, $43 \pm 2\%$, respectively. The scientific consensus for a BAC filter is that it is fully functioning through biodegradation when DOM removal is in steady state (Terry Summers, 2018a).

First, we can evaluate the removal efficiency at the assumed steady state. These removals ranged from approximately 26 - 43% depending on filter length and are much higher than the reported DOM removal in rapid-rate BAC filtration (10 - 15%) and slow sand filtration (16%) (Guchi 2015, Terry & Summers 2018). There are drinking water treatment studies that have reported TOC or DOC removal above the commonly accepted 10 - 16%: 6 - 30% at 6-30 min EBCT (Arnold et al. 2018), 17-37% at 18 min EBCT (Kalkan et al., 2011), 23 - 51% at 18 min EBCT (Farre et al. 2011), 22-63% at 18 - 120 min EBCT (Reungoat et al. 2011), 70% at 60 min EBCT (Pipe-Martin 2010), 60% at 14.4 hour EBCT (Ka et al. 2020), and 37% at EBCT of 1 hour (Rattier et al. 2012a). All these studies use adsorptive media. There seems to be a link between the adsorptive character and the elevated removal efficiencies reported by these studies.

Secondly, we can look at the BV needed until reaching the steady state caused by microbial activity solely. The apparent steady state of the present study was achieved between 300 and 950 BV depending on filter length. It is a much lower BV than the advised 20,000 BV indicated by Peterson & Summers (2021). It means that not enough BV have passed through the filters in the current study to have exhausted the GAC. As a result, GAC adsorption is expected to still play a large role.

Third, we can look at the microbial activity in the filter. During steady state, it is expected for DOM removal to be a function of biomass concentration (Carlson & Amy 1998), and therefore a higher DOM removal in steady state suggests a higher microbial activity. During saturation of the filters, microbial communities start to inhabit the GAC while utilizing biodegradable organics (Korotta-Gamage & Sathasivan 2017, Scholz & Martin 1997, Simpson 2008). Although this biodegradation contributes to the removal of DOM, the formation of biomass decreases the alkalinity of the GAC, making it less easy for negatively charged molecules such as the DOM and anions to adsorb (Mahajan et al. 1980, Rattier et al. 2012a). In the present study, it is not clear when biodegradation start playing a role in the removal of DOM, but already in week 5 biomass (cells) was present on the GAC (Figure 3.10).

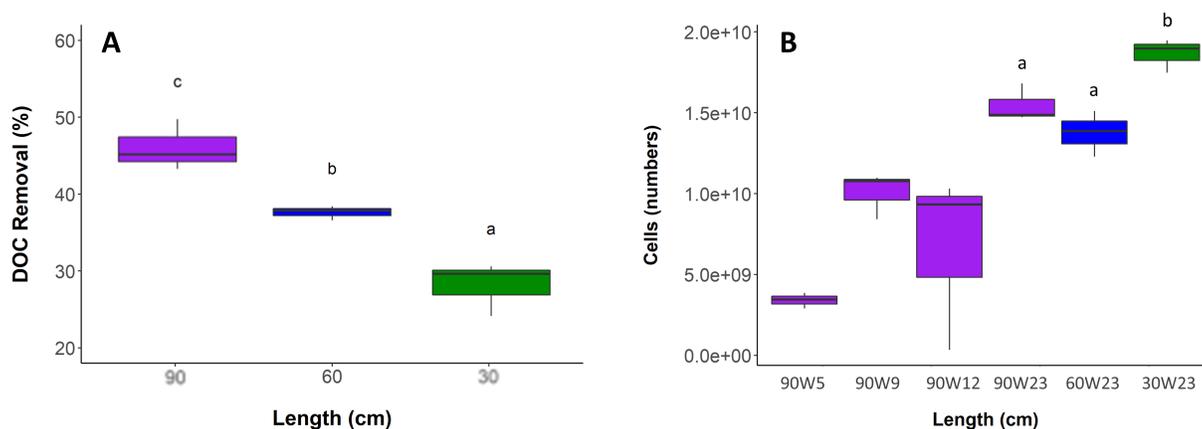


Figure 3.10: Trends of DOC (A) and Biomass (B) removal. DOC removal was measured at week 23 for the three filter lengths ($n = 3$ replicates). Letters (a, b, c) in (A) denote significant differences with regard to DOC removal between filter lengths assessed by Dunn's post-hoc test on KW H-test. Cell counts were measured at week 5, 9, 12, and 23 for the 90 cm filter as well as the 30 and 60 cm filter at week 23 ($n = 3$ replicates). Letters (a and b) in (B) denote significant differences between filter lengths at week 23 with regard to cell count assessed by Tukey's post-hoc test on ANOVA.

At week 23 the 30, 60 and 90 cm filters removed $26 \pm 3\%$, $34 \pm 3\%$, $43 \pm 2\%$. The total biomass in the filters, on the other hand, was $2 \cdot 10^{10} \pm 6 \cdot 10^8$, $1 \cdot 10^{10} \pm 8 \cdot 10^8$ and $2 \cdot 10^{10} \pm 7 \cdot 10^8$ cells for the 30, 60 and 90 cm filters respectively (Figure 3.10). Interestingly, while the shortest filter showed the lowest DOC removal efficiency, it had the highest biomass content. This suggests that adsorption still plays an important role even when the removal efficiency appears to be in steady state.

All these factors point out that in the present study the apparent steady state for DOM removal is not a result of solely biodegradation, but that adsorption is still playing a significant role. Adsorption kinetics of carbon is modelled as a series of two consecutive processes: nonselective adsorption of molecules in larger pores followed by the movement of such adsorbed molecules into the small micropores through the pore mouth barriers (Kołodziej et al. 2014, Nguyen & Do 2000, Pohlman 1940). After a rapid fall in adsorption efficiency to a lower level when the nonselective adsorption sites are occupied, adsorption actually continues, apparently unchanged, over an extended period of time, also known as tailing (slow adsorption) (Pohlman 1940). This adsorption phase can last for months or longer depending on the characteristics of the adsorbate and adsorbent (Peterson & Summers 2021). Because this dual-rate kinetic model produces a similarly shaped breakthrough curve, various water treatment operations have been wrongly classified as biological removal following an initial period of high adsorptive uptake (Peel & Benedek 1983), while actually all removal could be ascribed to

some form of adsorption. This apparent steady state can therefore be easily mistaken for steady state behaviour caused by microbial activity. To conclude, in the present study the apparent steady state is highly likely causing the slow adsorption of the DOC into the micropores. The longest filter benefits the most from these slow adsorption processes resulting in the highest removal efficiency for longer periods of time during this apparent state. How much of the witnessed DOM removal can be ascribed to biodegradation or biosorption remains unknown. Analytical techniques that better characterize and quantify DOM fractions (such as AOC, BDOC, fluorescence or LC-OCD) could contribute to quantification of the DOM fraction being biodegraded and/or adsorbed. Therefore in chapter 4 a fluorescence spectroscopy method is designed to characterize and semi-quantify DOM which can help to characterize and monitor the behaviour of less biodegradable humic substances and the more readily available proteins. Among other methods, this method will then be applied in chapter 5 to analyse the pore water of the biofilter for further understanding of the removal processes of DOM within the biofilter.

Other V1 parameters

Looking at the other chemical water quality variables of cluster V1, nitrate was also fully removed at the start of the operation (Figure 3.3 D). Removal of nitrate by GAC has been shown previously (Demiral & Gündüzoğlu 2010, Mubita et al. 2019). Over time the concentrations of the 30 and 60 cm filters rose above the LOQ resulting in a lower removal efficiency. However, removal efficiencies remained high during the whole operational run. On average removal efficiency was $75 \pm 20\%$, $84 \pm 10\%$, $87 \pm 5\%$ for the 30, 60 and 90 cm filters, respectively, with the 30 and 60 cm as well as the 30 and 90 cm filters showing statistical difference (Dunn's post-hoc test on KW; $H(151) = 22.51$, $p.\text{adj} < 0.001$).

As described before, iron was removed in a similar way as DOC with a large removal at the start of the experiment and a decline over time until steady state. In steady state, iron removal efficiency was $18 \pm 7\%$, $22 \pm 2\%$, $26 \pm 2\%$ for the 30, 60 and 90 cm filter. Only the 30 and 90 cm filters behaved statistically differently from each other (Dunn's post-hoc test on KW; $H(27) = 14.10$, $p.\text{adj} < 0.001$).

Overall, results of the present study show that longer filters have higher removal rates because of its high EBCT and extended slow-adsorption. Microbial data provided by DQ demonstrated that filter length did not influence the pathogen removal and that all microbial processes are taking place within the first 30 cm of the filter. Therefore, from a microbial point of view short biofilters with low EBCT can be used for the treatment of the water. However, from a chemical point of view the longer filter has the longest slow adsorption and the highest removal of DOM and iron for the longest period of time.

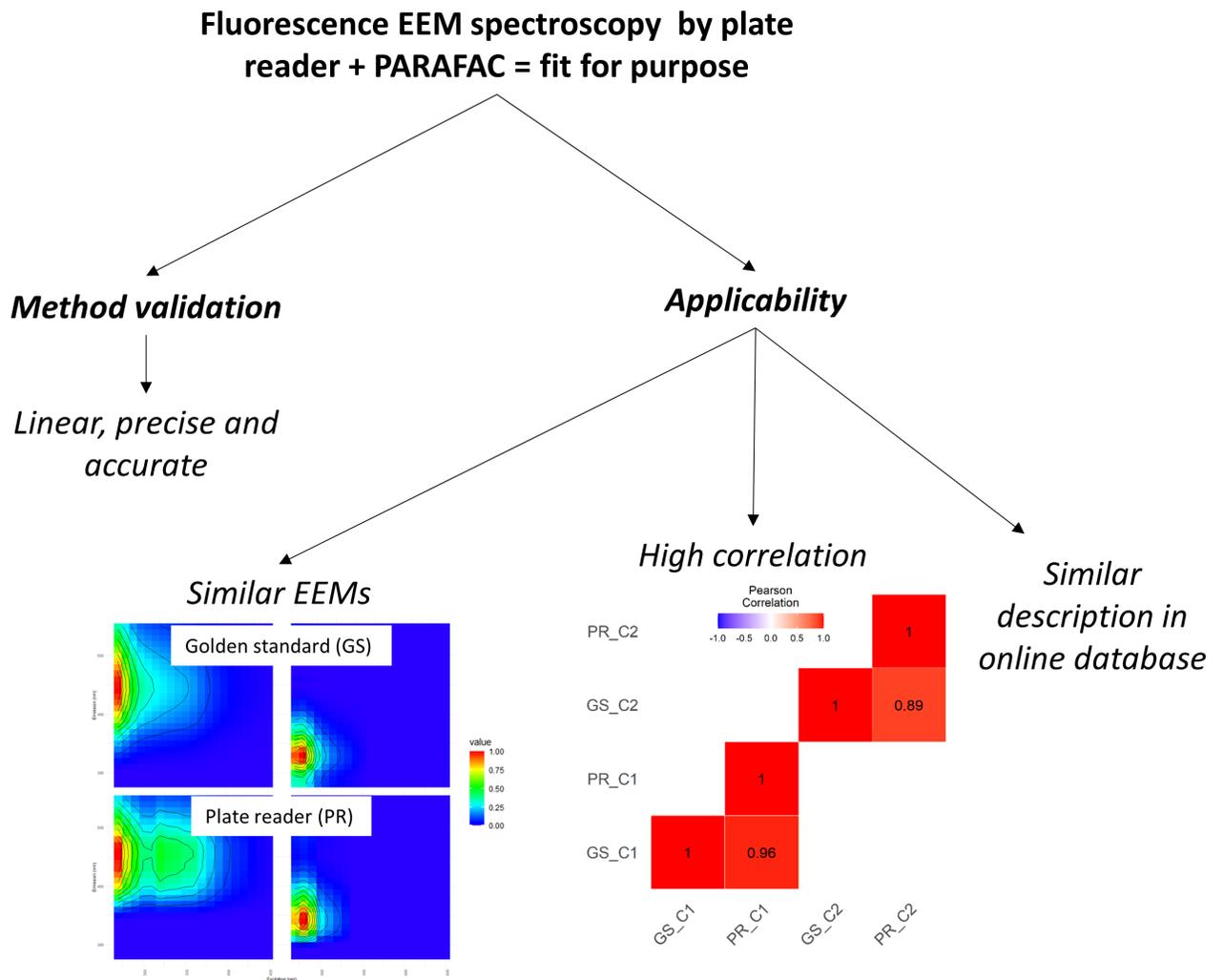
3.4 Conclusion

The objective of this chapter was to (1) quantify DOM removal and other chemical water quality parameters and (2) explore the impact of filter length for a biofilter during the first six months of operation. Results from this study suggest that:

- Slow-flow BAC filtration with long filter length resulted in a significantly higher DOC removal in apparent steady state compared to the smaller filters.
- In the apparent steady state not only biodegradation, but also other processes such as slow-adsorption by GAC or biosorption by the biofilm play a role in the removal of DOC which might explain the increased DOM removal by slow-flow BAC filtration compared to other BAC filter studies.
- Longer filters have a higher capacity to remove DOM, because of GAC material is available for slow-adsorption.
- The slow-flow BAC filter did not meet the criteria for providing clean drinking water. It was not able to filter iron until a concentration below the EU standards. Moreover, the DOC concentration increased to such high concentration that colour was visible. Also, arsenic leached from the filter during the first weeks.
- The choice for filter length is dependent on the treatment goal for the Scottish off-grid system. If one of the important aspects is the significant reduction of TOC and iron, solutions for the removal of these chemical parameters need to be found before the system can be deployed.
- Most of the other anion, cations and metals are present in (very) low concentrations and are not a problem for the quality of the treated water. It is, however, recommended to flush the GAC thoroughly and monitor the flush water to avoid pH and conductivity fluctuations at the start of the operation resulting in the leaching of harmful metals such as arsenic.
- To what extent microbial utilization plays a role in the removal of DOM is not clear from this chapter. Therefore in chapter 4 a fluorescence spectroscopy method is designed to characterize and semi-quantify DOM which can help to characterize and monitor the behaviour of less biodegradable humic substances and the more readily available proteins. This method will then be applied in chapter 5 to analyse the pore water of the biofilter for further understanding of the removal processes of DOM within the biofilter.

Chapter 4

Fluorescence excitation-emission measurements of DOM using a plate reader coupled to PARAFAC analysis



4.1 Introduction

To produce safe and high-quality drinking water, the effective removal of DOM is essential. The results of chapter 3 have demonstrated that slow-flow BAC filtration is able to remove DOM. In biofilters, DOM is one of the chemical parameters that plays an important role in the ecological mechanisms happening in the biofilter. It is used as substrate by the microorganisms that will inhabit the filter and can steer microbial community composition (Wu et al. 2018). Results of chapter 3 did not unveil to what extent microbial breakdown of biodegradable DOC played a role during filtration. For this, a method that characterises the DOM better is required. One of these methods is EEM fluorescence spectroscopy is a proven popular technique to characterize and monitor dissolved organic matter (DOM) in natural as well in engineered systems (Carstea et al. 2020). Although this technique does not provide absolute concentrations, it can characterize the DOM to a certain extent and give relative fluorescence signals (Ryder et al. 2017). PARAFAC has become one of the most commonly used techniques to analyse the fluorescence EEM data (Sanchez et al. 2013), with a repository of published organic fluorescence spectra available online (Murphy et al. 2014a). Although PARAFAC is not considered a simple method, various tools are freely available that can assist with the PARAFAC model building (Bro & Vidal 2011, Murphy et al. 2013*a,b*).

Top-end fluorescence spectrometers quickly and accurately produce EEMs over a wide range of excitation and emission wavelengths using small intervals (hereafter called resolution). This, together with the fact that no sample pre-treatment and ample sample volume are required, makes fluorescence excitation–emission measurements a popular analytical method. However, these top-end fluorescence spectrometers are not always present in laboratories. A possible alternative could be the use of a plate reader for fluorescence EEM measurements. Plate readers are multifunctional and possibly present in laboratories where fluorescence spectrometers are missing. Plate readers that can make fluorescence scans are able to produce a non-automated EEM, while newer versions can even do so fully automated. Possible challenges when fluorescence EEM are produced by a plate reader are the microplates and the resolution of the machine. The microplates that are commonly used in plate readers are made from material that causes autofluorescence. Although special microplates are available that counteract the autofluorescence (Greiner bio-one 2020), interference of autofluorescence signal during PARAFAC modelling cannot be ruled out. Also, when using a non-automated plate reader, a number of manual steps have to be taken and consequently a lower number of excitation and emission wavelengths will be used. This will impact the resolution of the produced EEM.

The use of the plate reader to analyse DOM by fluorescence EEM spectroscopy and subsequent analysis of data by PARAFAC has the potential to make the approach more widely applicable and more affordable; notably for less well-resourced laboratories where fluorimeters are not commonly present in laboratories. The objective of this chapter is therefore to develop a fit for purpose analytical method using a non-automated plate reader to characterize and semi-quantify DOM with help of PARAFAC. This innovative method using a plate reader has not been reported to date. The method was validated by evaluating its specificity, linearity, accuracy, and precision with help of two reference standards that cover the DOM EEM range and by applying the method to a small data set of water samples analysed by the method itself and a method commonly accepted in the field of expertise (the golden standard).

4.2 Material and methods

4.2.1 Plate reader method

Plate reader + Microplates

The Tecan M200 Pro plate reader (Tecan, Switzerland) was used for method development with the following settings: excitation scan modus, excitation bandwidth 5 nm (250-315 nm) and 10 nm (316-450), emission bandwidth 20 nm, excitation scan number 21, gain 60, number of flashes 25, integration time 20 μ s, lag time 0 μ s, settle time 0 ms, z-position 20000 μ m. For each EEM, the analysis was done in 3 sections being excitation/emission (1) 250/280, 250-260/290, 250-270/300, 250-280/310, 250-290/320, 250-300/330, 250-310/340, 250-320/350, 250-320/360, 250-370/370; (2) 250-380/380, 250-390/390, 250-400/400, 250-410/410, 250-420/420, 250-430/430, 250-440/440, 250-450/450, 250-450/460, 250-450/470; (3) 250-450/480, 250-450/490, 250-450/500, 250-450/510, 250-450/520, 250-450/530, 250-450/540, 250-450/550. Samples were analysed in a 96 wells microplate (UVStar®, black chimney flat bottom μ Clear®, Greiner, Austria). This microplate made from cycloolefin copolymers was chosen because of its low costs, low level of autofluorescence, along with very good transparency in lower UV wavelengths (down to 230 nm) (Greiner bio-one 2020). The black chimney shaped (tube shaped) wells with flat transparent bottom prevented signal crossing from other wells.

EEM Absorbance data

For each sample, as well as blank (Milli-Q water), 330 μ l was transferred to one of the wells of the microplate with help of a pipette. This provides a pathlength of approximately 1 cm which is the standard pathlength of a spectroscopy cuvette. The fluorescence intensity of the sample and the blank was measured for the EEM ranging from excitation wavelength 250-450 nm and emission wavelength 280-550 nm with 10 nm interval for both wavelengths. Afterwards, the absorbance of both sample and blank were measured from 250-550 nm with 10 nm interval. During analysis, a maximum of two rows of the microplate was analysed at the time to minimise the risk of evaporation. The plate reader did not order the data for each sample as EEM. A self-build MATLAB tool (MATLAB R2016B) was used to extract the EEM together with the excitation and emission wavelengths for every sample/blank (Appendix B, Section B.2, pg. 193). The absorbance with its corresponding wavelengths was also extracted from the raw datafile. Data were stored according to the PARAFAC data analysis method developed by Pucher et al. (2019), including a metadata file with sample information.

Spectral correction factor data

Raw fluorescence data are often biased due to variations in the efficiency at which the various wavelengths of light are transmitted through the monochromators and the imperfections in the optical components or their alignment (Murphy et al. 2013*b*). This can create systematic biases in the data set with inaccurate excitation or emission spectra and must be corrected through spectral correction. In spectral correction, every combination of excitation and emission wavelength within the EEM is multiplied by a correction factor from a correction matrix (the correlation matrix is the excitation correction vector \times emission correction vector) that is specific to the instrument in use. The emission spectral correction vector was based on the method of Lakowicz (2006), and is described in Appendix B, Section B.1, pg. 191. The excitation wavelength of the current set-up did not need further correction. The excitation correction vector ranging from 250-470 nm was therefore set to 1. Both the excitation and emission correction files are stored according to Pucher et al. (2019).

Raman calibration data

Fluorescence measurements at different timepoints or using different instrument settings can cause fluorescence signal variation. For data comparison an intensity calibration is required. A simple method developed by Lawaetz & Stedmon (2009) uses the integrated area of a water Raman peak for the calibration of the fluorescence intensity. This water Raman band is a result of non-elastic scatter and runs through the EEM (Figure 4.1). The wavelength-dependent Raman cross-section of water is directly proportional to the integral of the Raman peak (A_{rp}) (Figure 4.1) and a fixed property of water (Faris & Copeland 1997). In most cases, the area under the peak at excitation wavelength 275 or 350 nm is used (Lawaetz & Stedmon 2009, Murphy et al. 2010, 2013*b*). For the Plate Reader method (hereafter called PR method), the A_{rp} at 420 nm excitation wavelength of a blank is used to calibrate the results, because background noise due to autofluorescence of the system interferes with the A_{rp} at excitations 275 and 350 nm. For the calculation of the A_{rp} , the fluorescence emission intensities, ranging from 460 to 520 with 10 nm interval at 420 nm excitation wavelength, were extracted from the first measured DI blank of the day and stored in a separate file with its corresponding wavelengths. Raman calibration files are stored according to (Pucher et al. 2019).

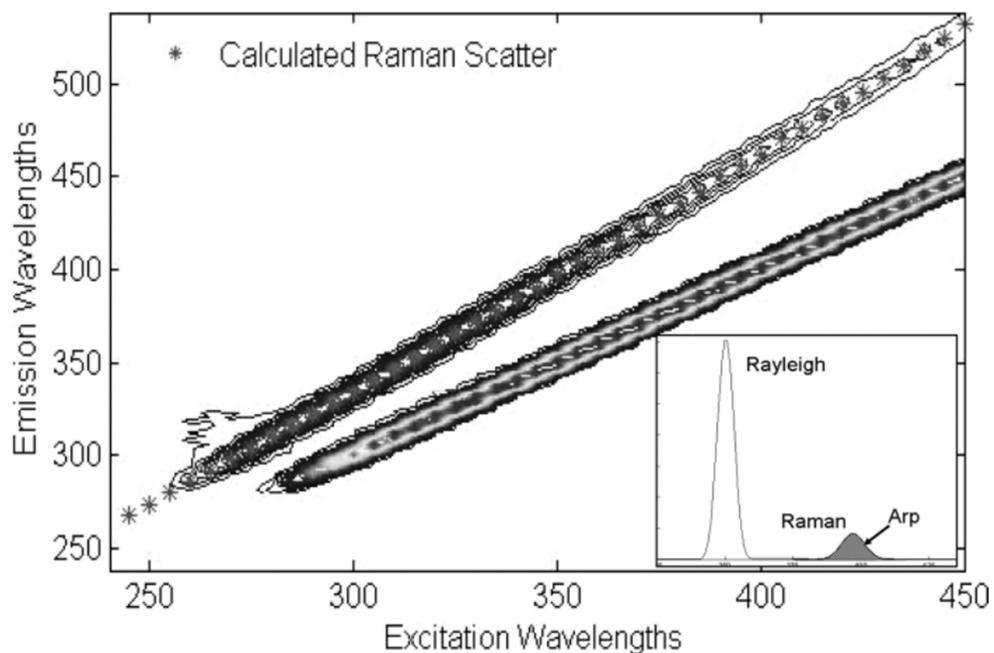


Figure 4.1: *Raleigh and Raman scatter present in an EEM of water. From Lawaetz & Stedmon (2009).*

4.2.2 Gold standard method

The Gold Standard method (hereafter called GS method) used the Duetta (Horiba, Japan) equipped with a 3.5 mL quartz cuvette using the following settings: Excitation range 250-470 nm with 10 nm steps increment, emission range 280-550, automatic correction of inner filter effect, excitation/emission band pass set at 5 nm, integration time 1s and detector accumulations 1. For each sample, the cuvette was loaded with 3 mL sample and placed in the machine. The absorbance and fluorescence were measured consecutively, and the inner filter effect was automatically corrected. Afterwards, a blank was measured, and its signal subtracted from the signal of the sample. Samples were Raman calibrated as described previously using excitation wavelength 350 nm. Data preprocessing and PARAFAC carried out in a similar way as the PR method.

4.2.3 Data pre-processing

To build a PARAFAC model the raw EEM needed pre-processing to correct systematic biases in the data set. For data preprocessing, the staRdom package for R was used (Pucher et al. 2019), following the steps given in Figure 4.2.

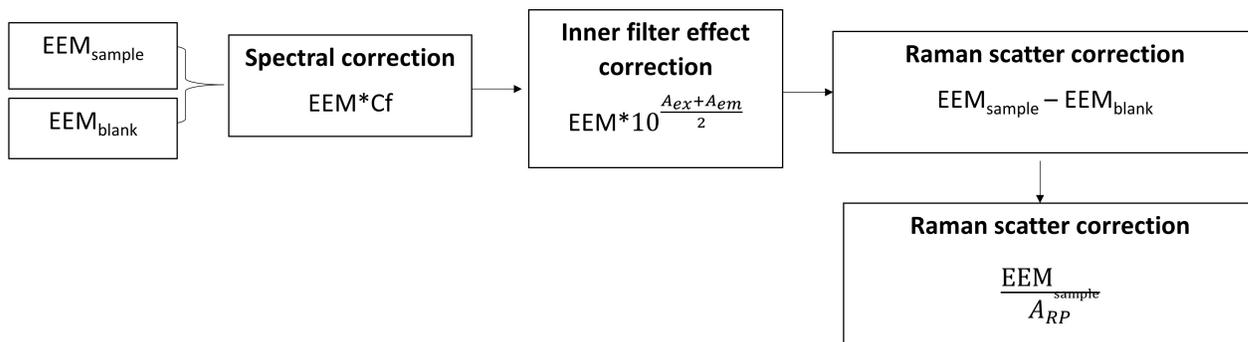


Figure 4.2: Schematic overview of PARAFAC pre-processing steps.

Step (1) Sample EEMs were spectrally corrected by element-wise multiplication with the correction matrix (excitation correction vector x emission correction vector) as described previously. (2) Sample EEMs were subsequently corrected for large direct scattering of the incident light at the same wavelength as the excitation by deleting the values of these excitation-emission wavelength pairs (Stedmon & Bro 2008). (3) Sample and blank EEMs were corrected for the inner filter effect which can result in an underestimation of emitted fluorescence. During inner filter effect the dissolved analytes attenuate either the excitation, the emission light, or both as it passes through the sample (Ryder et al. 2017). This effect was corrected with help of the inner filter correction equation (Murphy et al. 2010). (4) Sample EEMs were corrected for non-elastic scatter by subtracting the blank spectrum from the sample spectrum. Although the necessity of Raman scatter removal by blank subtraction has been questioned in previous studies (Murphy et al. 2010, Zepp et al. 2004), this correction step is necessary in the PR method. In fact, in addition to the removal of Raman scatter (Lawaetz & Stedmon 2009, Murphy et al. 2010), this correction step removes the autofluorescence signal caused by the plate reader and/or microplate. (5) Raman calibration normalized the sample EEMs by dividing the sample EEM by the area under the Raman peak of the blank.

4.2.4 PARAFAC

A PARAFAC model was built with help of the staRdom package in R (version 4.1.1) (Pucher et al. 2019), described in short here. Exploratory phase: initial PARAFAC models were created for a first set of components (1-6 factors). For all investigated components, sample outliers and problematic excitations and emissions were inspected and removed where needed. Three measures are used to decide which possible models described the data set best (including all fluorophores without including noise), being (1) fit-values, (2) Core Consistency Diagnosis (CORCONDIA) and (3) split-half

validation. Fit-values describe the variance that is explained by the model. Core consistency diagnosis evaluates the degree of trilinearity of the PARAFAC loadings by comparison of the least squares Tucker3 core (Tucker 1966) calculated for these and a superdiagonal core of ones according to the assumption that the model can be represented as a constrained Tucker3 model with the core corresponding to a superdiagonal of ones (Bro & Kiers 2003, Sanchez et al. 2013). The core consistency, which is expressed as a percentage, implies the degree of fitting of the Tucker3 core with respect to the assumption of the model. Core consistency of a one component model is always 100% and decreases as the number of components in the model increases. Core consistency exhibits a significant reduction when an additional component is added after the appropriate number of constituents has been reached. Split-half analysis proposed by Harshman & Sarbo (1984) uses multiple split-half tests, where various models are created and compared after dividing the data set in half in different ways (Murphy et al. 2013b, Peleato et al. 2016).

An overall measure of model quality was defined by EEMqual (Bro & Vidal 2011). This measure is the combination of three tests (fit-values, core consistency diagnosis and split-half validation) and is expressed by a single value (Equation 4.1). All three tests have a value between 0 and 100, where 0 is bad and 100 perfect. When all three measures are close to hundred, the EEMqual was close to hundred, and the model is good. In general, the most suitable model is the model with the highest number of components which still has a high EEMqual (Equation 4.1). In the current method, models with increasing number of components were tested and the one with the highest number that is still providing a good model (high EEMqual) was chosen.

$$EEMqual = Fit \times Coreconsistencediagnosis \times Splithalf \quad (4.1)$$

4.2.5 Method validation

The PR method was validated for fitness-for-purpose. The following validation parameters : selectivity, linearity, limit of detection and quantitation, accuracy (recovery) and precision (repeatability and intermediate precision) were investigated by measuring two fluorophores, Quinine Sulphate (QS) and TRyptophan (TR), on two days within 2 weeks. QS excitation/emission range is <250-400 / 400-600 and TR excitation/emission range is 260-300 / 300-450 nm. These two analytes thereby cover the section of the EEM similar to humic-like (excitation/emission range = 250-400 / 400-550 nm) and protein-like (excitation/emission range = 250-300 / 300-400 nm) fluorescence components present in fresh water.

Validation standards

200 µg/L stock solutions were prepared: tryptophan (Sigma Aldrich, UK) was prepared in the absence of added buffer at pH 7 in (1) DI water (Elix© Pure Water Systems, Merck) and (2) river water and (3) 50%-50% mix of DI and river water; QS (Sigma Aldrich, UK) in 0.1 H₂SO₄ was examined at a concentration of 15 mg/L and at pH 1.5 in (3) DI water and (4) 50% - 50% mix of DI and river water. Pure DI water and river water were used as blanks; samples were stored at 4 °Celsius in the dark before analysis within 48 h.

Validation parameters

selectivity Selectivity is the extent to which a method can determine particular analyte(s) in a mixture(s) or matrix(ces) without interferences from other components of similar behaviour (FDA 2020). Three concentrations (10, 50, 150 µg/L) of each analyte dissolved in pure water as well as in matrix (1 : 1 river water : DI water) were created and measured in triplicate by the PR method as well as the GS method. For each method a separate PARAFAC model was created. The spectral shape/peak position (qualitative) as well as the signal intensity (quantitative) was compared between solvent (water and matrix) and among methods (PR ad GS methods). Spectral shape and peak position were compared by visual plot inspection as well as a Pearson correlation in which the combined excitation and emission spectra of each analyte were compared among solvents and methods with data being normality distributed.

Signal intensity was compared among the two methods by looking at their signal recoveries of the standards in matrix where signal in water was used as reference (Equation 4.2).

$$Recovery = \frac{(\text{Concentration in matrix})}{(\text{Concentration in water})} \times 100 \quad (4.2)$$

Linearity Linearity is the ability of the method to produce results that are directly proportional to analyte concentration within a given range (FDA 2020). It is evaluated by the slope, intercept, standard error of slope, standard error of intercept, correlation coefficient of a six-point calibration curve. For this, six calibration standards (10, 30, 50, 100, 150 and 200 µg/L) were measured in triplicate, with subsequent regression analysis on the data.

Limit of detection and quantitation Limit Of Detection (LOD) is the lowest amount of analyte in a sample which can be detected but not necessarily quantitated as an exact value, while the LOQ is the level above which quantitative results may be determined with acceptable accuracy and precision (FDA 2020). The LOQ were calculated with help of Equation 4.3.

$$LOQ = \frac{(\text{standard error of intercept})}{\text{slope}} \times 10 \quad (4.3)$$

Accuracy and precision The Quality Controls were used to determine the accuracy of the method (recovery) and its precision (repeatability and intermediate precision). Accuracy is the nearness of a result or the mean of a set of measurements to the true value and is expressed in recovery (Equation 4.4) (FDA 2020). Precision is the agreement between a set of replicate measurements without assumption of knowledge of the true value and was expressed by the repeatability within 1 day (Equation 4.5) as well as the intermediate precision over 2 days (Equation 4.6) (FDA 2020). For this, three quality controls (10, 50 and 150 µg/L) were measured in sixfold for both fluorophores on two different days.

$$Recovery = \frac{(\text{Measured concentration})}{(\text{Used concentration})} \times 100 \quad (4.4)$$

$$Repeatability = \frac{\text{standard deviation}}{\text{mean}} \times 100 \text{ (within 1 day)} \quad (4.5)$$

$$Intermediate\ precision = \text{mean} \times 100 \text{ (over 2 day)} \quad (4.6)$$

4.2.6 Method application

Water samples

Six water samples were collected from various water sources in The Netherlands and one in Scotland: *Amsterdam channel (NL)*, *lake 's-Gravenbroek (NL)*, *lake 's-Gravenkoop (NL)*, *river Wiericke (NL)*, *fosse Driebruggen (NL)*, and *Kelvin Glasgow (GB)*. Two additional samples were created by diluting river Wiericke and fosse Driebruggen 1 : 1 with DI water.

Application

The PR and GS methods were applied to a small data set of water samples ($s=8$) to verify the environmental applicability of the PR method. This was done by looking at the qualitative and quantitative specificity of the PR method. First the correct number of PARAFAC components was determined. Second, for the qualitative specificity, the spectral shapes of the components of the models of both methods were compared. This was done by a Pearson correlation. Next, both excitation/emission spectra were cross checked with components previously reported in the literature using the OpenFluor database (Murphy et al. 2014). For each component the number of correlated components in the database (Tucker's Congruence Coefficient $> 0.95\%$) were given together with their reported characteristics of the three best matching components. Tucker's Congruence Coefficient is an index of the similarity between the compared factors (components). The hits with the components in the database were compared among the two methods. Third, for the quantitative specificity, signal intensities of the components were compared among the methods. Also, signal intensities of the components were compared among diluted and undiluted samples. This last step was necessary to check the influence of absorbance in samples with high fluorescence intensities. An overview of the various parameters analysed, and corresponding models produced is given in Table 4.1.

Table 4.1: *Details of the validation data sets and corresponding PARAFAC models.*

Validation parameter	Matrix	Conc.	Reps.	GS method	model name	Nr. samples per model	Comments
Selectivity	DI & River	10, 50, 150	3	Yes	(1)PRmethod&DI	9	250 & 260 nm excitation removed for GS method
					(2)PRmethod&river	9	
					(3)GSmethod&DI	3	
					(4)GSmethod&river	3	
Regression (linearity&LOQ)	DI	10, 30, 50, 100, 150, 200	3	No			
Accuracy (Recovery)	DI	10, 50, 150	6	No	(5)	108	
Precision (Repeatability& intermediate precision)	DI	10, 50, 150	6	No			
Method applicability	-	water samples	1	Yes	(6)	8	250 & 260 nm excitation removed for GS method
					(7)	8	

4.3 Results and Discussion

4.3.1 PARAFAC component selection

The aim of this study was to validate the use of a plate reader for the characterization and semi-quantitation of DOM with help of PARAFAC. First various validation parameters (selectivity, linearity, accuracy, and precision) were evaluated using two reference standards (QS and TR) dissolved in water and a river water matrix, followed by the demonstration of the method's applicability to environmental samples.

For each validation data set from Table 4.1, a separate PARAFAC model was built. The first step was to find the correct number of components per data set to subsequently build the individual PARAFAC models. The correct number of components that describes all present fluorophores while excluding the background noise was selected with help of the results of EEMqual. EEMqual was the combined measure of the fit-values, CORCONDIA and split-half validation. Results are given in Appendix B, section B.2, pg. 193, and show that 2 components were needed to build for models 1 to 7.

Selectivity

Selectivity was studied to demonstrate that fluorescence components were at similar position in the EEM's of the PR method and GS method in water and in matrix, as well as with similar signal intensity. Correct position is hereafter called qualitative selectivity, while the right signal intensity is hereafter called quantitative selectivity. The excitation/emission spectra of the analytes in ultra-pure and matrix (river water) of the GS method and PR method are shown in Figure 4.3, respectively. During PARAFAC model building, the GS method showed significant background scatter at excitations 250 and 260 nm in most low concentration samples. These excitation wavelengths needed to be excluded and were consequently missing in their excitation spectra (Figure 4.3). For the PR method, no abnormalities were found in the spectra during model building.

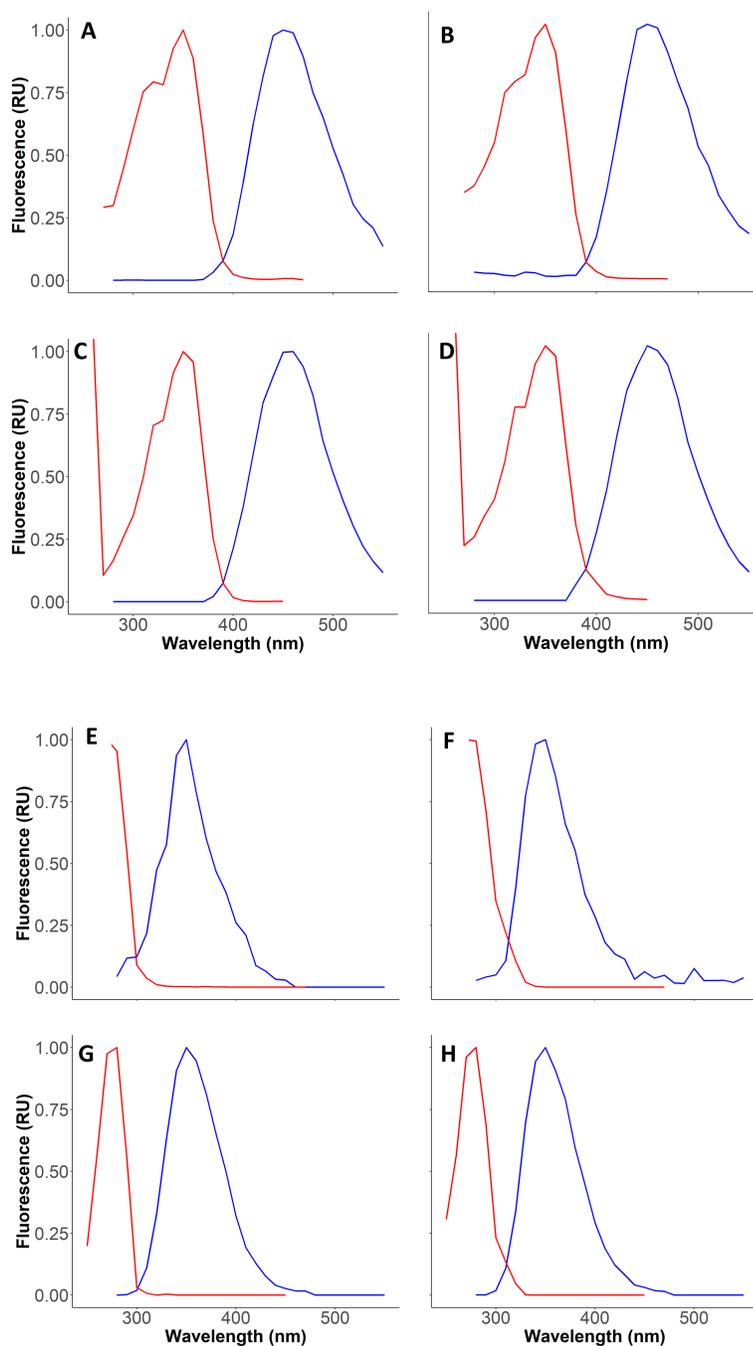


Figure 4.3: *PARAFAC* component excitation (red) and emission (blue) spectra of *QS* in water (**A**) and in matrix (**B**) analysed by the *GS* method, *QS* in water (**C**) and *QS* in matrix (**D**) analysed by the *PR* method, *TR* in water (**E**) and in matrix (**F**) analysed by the *GS* method, and *TR* in water (**G**) and *TR* in matrix (**H**) analysed by the *PR* method.

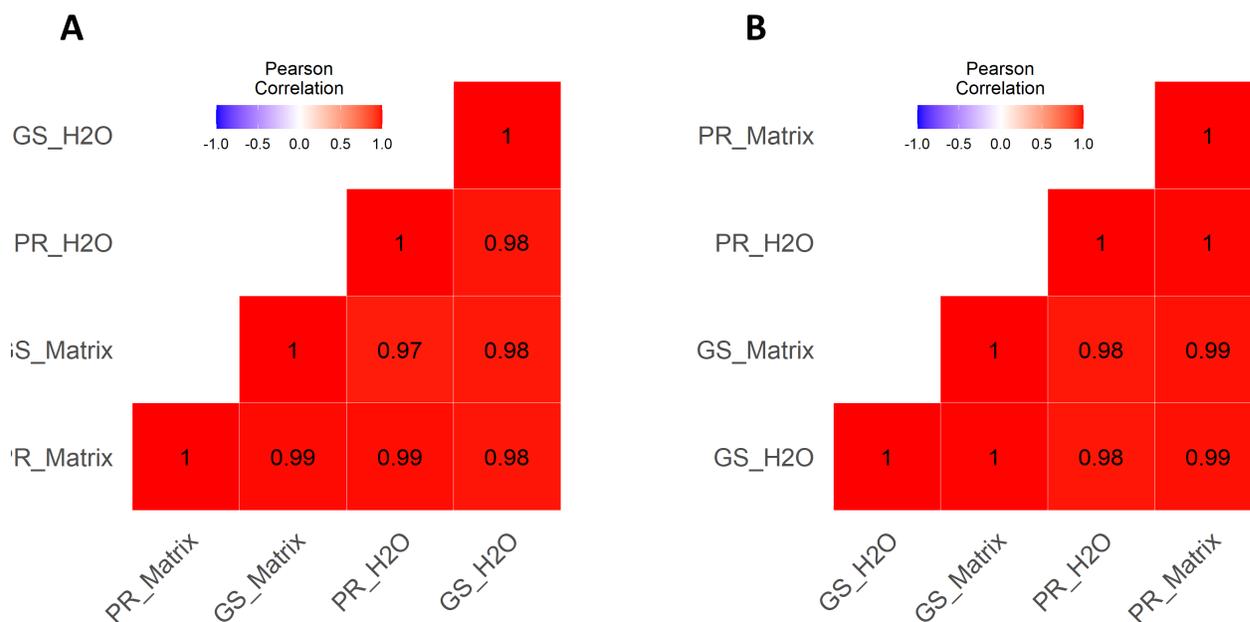


Figure 4.4: Significant correlations ($-.6 > R > .6$) between spectra of TR (A) and QS (B) analysed by the PR method PR and the GS method in water (H_2O) and matrix.

To judge the qualitative selectivity of the PR method for each analyte, the excitation/emission spectra were compared between methods and solvent. Visual comparison of the PR and GS method showed similar spectra position and shape (QS: Figure 4.3 A vs C, TR: Figure 4.3 E vs G). Only the shape of excitation wavelengths 250 and 260 could not be compared, because of the removal of these wavelengths from the GS model. Visual comparison of the water and matrix samples of the PR method also showed similar spectra position and shape (QS: Figure 4.3 C vs D, TR: Figure 4.3 G vs H). Overall, the components demonstrated excellent Pearson correlation ($R(49) > .97$, $p < .001$) among the methods and solvents (Figure 4.4). Therefore, the PR method demonstrated a good qualitative selectivity.

To judge the accuracy of the signal intensities of the measured analytes, the signal recovery (Equation 4.2) in river water matrix was compared among the PR and GS methods. For both methods, the matrix interfered in some way with the fluorescence signal intensity. Apart from PR method TR50 and TR150, all other samples experienced matrix effect resulting in negative recovery (Figure 4.5). The recovery of the GS method ranged from -40 to -8.6%, the PR method ranged from -17 to 36%. The 10 $\mu\text{g/L}$ TR was excluded because of an error in sample preparation. Overall, the PR method recoveries were within acceptable limits of $\pm 20\%$, apart from the 50 $\mu\text{g/L}$ tryptophan sample (TR50) analysed by the PR method (FDA 2020).

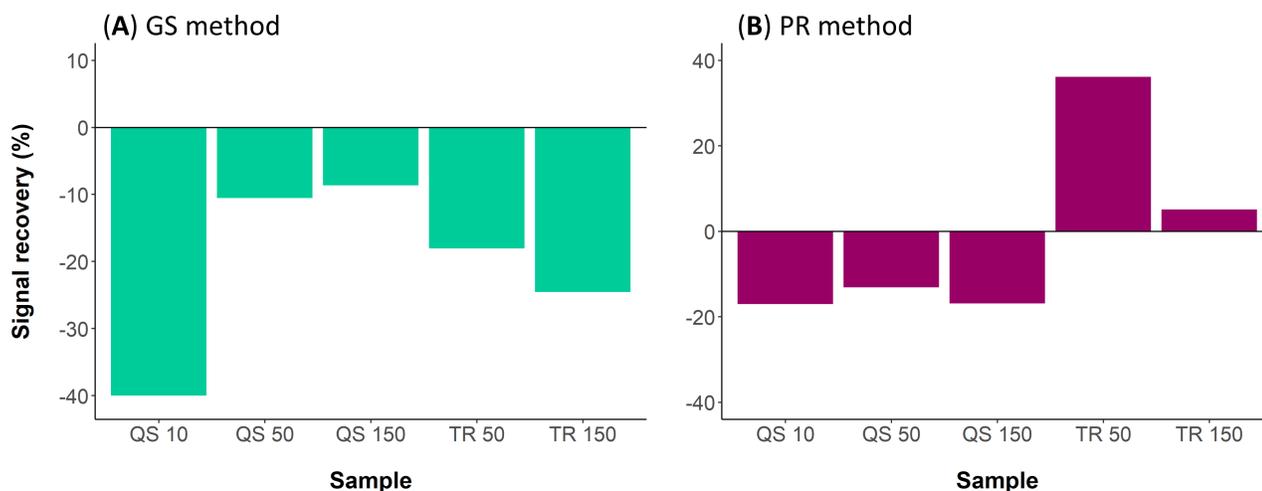


Figure 4.5: *Quantitative specificity: signal recovery of 10, 50, 150 µg/L QS (QS10, QS50, QS150) and 50, 150 µg/L TR (TR50, TR150) (n = 1 replicate) for the GS (A) and PR (B) methods.*

Signal quenching of QS can be a result of the inner filter effect. This effect is due to dissolved analytes attenuating either the excitation light, the emission light, or both as they pass through the sample (Ryder et al. 2017). While other methods are available for this correction (Larsson et al. 2007), the equation proposed in Lakowicz (2006), and used in the present study, is the commonly accepted correction method. With the optical density of the samples in the study under the absorption limit ($A < 1.5$) (Kothawala et al. 2013, Ryder et al. 2017), the mathematical inner filter effect correction approach is assumed to be correct. Another possible quenching mechanism is energy transfer (Lakowicz 2006, Mounier et al. 2017). Energy transfer is a dynamic quenching process in which non-radiatively energy (without absorption or emission of photons) transfers from donor (in the excited state) to the acceptor. The donor and acceptor then are coupled by a dipole–dipole interaction. This process only occurs when the emission spectrum of a fluorophore donor overlaps with the absorption spectrum of the acceptor, which does not need to be fluorescent. In the present study, both TR and QS emission spectrum (donor) overlapped with the absorption spectrum of humic-like fluorescence acceptor present in the river water matrix. So, the insufficient recovery can be a result of these principles. On the other hand, the positive extreme recovery of the 50 µg/L tryptophan sample (TR50) by the PR method can be the result of overcorrection of the inner filter effect. Both problems could be resolved by sample dilution. However this dilution step is not always possible and also adds a sample handling disadvantage, which is a source of error (Ryder et al. 2017). Moreover, when two fluorophores are present in a water sample and one of them is present in high concentration and the other is present with minimal signal intensity,

dilution can positively affect the highly present fluorophore signal, while removing the signal of the other fluorophore. It is therefore best to analyse both diluted and undiluted samples, when sufficient sample volume is available.

Linearity and detection limits

Second, the linearity and detection limits were investigated by using quantitative data from model 5 (Table 4.1). A 6-point calibration curve (10, 30, 50, 100, 150 and 200 $\mu\text{g/L}$) was created on two independent days using 3 replicates per calibration point. The calibration curve of QS and TR measured at day 1 are given in Figure 4.6. The method showed a good linearity on both days. Regression analysis indicated that the regression coefficients for TR as well as QS were above 0.99 and the LOQ around 10 $\mu\text{g/L}$ for both analytes (Table 4.2).

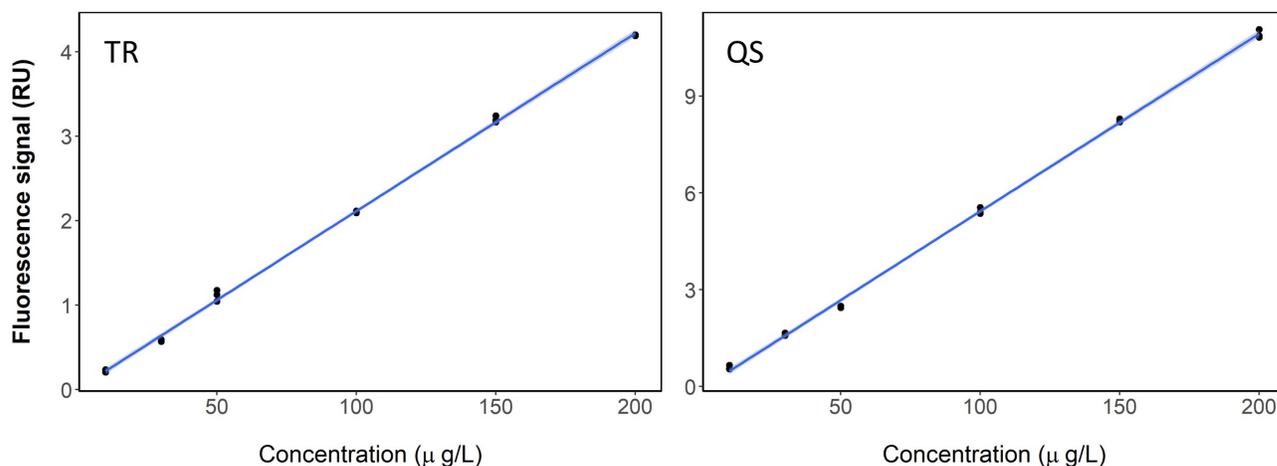


Figure 4.6: Calibration curve of TR (A) and QS (B) measured at day one.

Accuracy and precision

Third, the accuracy and precision were investigated by using quantitative data from model 5 (Table 4.1). The accuracy was evaluated by looking at analyte recovery (Table 4.2). For most samples the recovery of the PR method was good (recovery between 90-110%). Only the recovery of the low concentration samples (10 $\mu\text{g/L}$) varied. Most low concentration samples showed recovery above 100%. Again, this suggests that low signals might be overestimated by the PR method. All TR samples measured at day 2 showed a relatively low recovery compared to their day 1 counterpart. In

general, the accuracy was within acceptable limits. The precision was evaluated by looking at measurement repeatability in one day (repeatability) and over two days (intermediate precision) (Table 4.2). Measurement repeatability and intermediate precision illustrated a high precision under the various circumstances (% Relative Standard Deviation (RSD) < 10). Only the intermediate precision of the 10 µg/L TR samples (18% RSD) deviated.

Table 4.2: Validation parameters of the PR method.

			Tryptophan		Quinine sulphate		
			Day1	Day2	Day1	Day2	
Linearity and detection limit	Corr. Coeff.		0.999	0.998	0.998	1.000	
	Slope /sensitivity		0.02	0.03	0.06	0.06	
	SE ^a of slope		1.66·10 ⁻⁴	3.05·10 ⁻⁴	4.40·10 ⁻⁴	3.47·10 ⁻⁴	
	Intercept		8.00·10 ⁻³⁷	2.00·10 ⁻³	-0.08	-0.07	
	SEa of intercept		0.027	0.03	0.05	0.04	
	LOQ ^b		8.88	11.45	8.99	6.17	
Precision	Repeatability (%RSD) ^c	10	6.13	3.09	4.18	1.09	
		50	1.86	2.65	0.93	2.18	
		150	0.98	0.47	3.36	0.90	
	Intermediate precision (%RSD) ^c	10		17.64		5.48	
		50		5.13		1.81	
		150		6.85		3.64	
Accuracy	Recovery range (%) ^d	10	105–125	79-87	115–129	108-111	
		50	99-104	89-96	99–102	97-103	
		150	101-104	89-90	98-108	95-98	
	Mean recovery (%) ^d ± SD ^e	10		116± 7	84±3	119± 9	109±2
		50		101±2	91±2	101±1	99±2
		150		102±1	90±0	102±3	97±1

^a SE = standard error

^b Limit of quantitation (LOQ) = 10 × (SE of intercept/slope)

^c % of Relative Standard Deviation (%RSD) = (SD/mean) × 100

^d Recovery (%) = (measured concentration/used concentration) × 100

^e SD = standard deviation

4.3.2 Method applicability

The PR method as well as the GS method were applied to a set of environmental samples to judge the qualitative and quantitative specificity of the PR method. For both methods independent models were created (model 6 and 7 for the PR and GS method, respectively, Table 4.1). For qualitative specificity, the component of the PR method was compared to that of the GS method by visual inspection of their excitation/emission spectra and Pearson correlation, as well as comparison of their nomenclature in the OpenFluor database. For quantitative specificity, signal intensities of the components were compared among the methods as well as the comparison between diluted and undiluted samples. This last step was necessary to check the influence of absorbance in samples with high fluorescence intensities.

Qualitative specificity

The components given by both methods are illustrated by the EEM (Figure 4.7) and corresponding spectra (Figure 4.8). The excitation spectrum of component 1 (C1) differed slightly among the two methods, while the emission spectra look similar (Figure 4.8 A and C). C1 of the GS method had one maximum (270 nm), the PR method on the other hand had two peaks (270 and 320 nm). With replicates showing similar results, it remains unclear what caused this second peak at 320 nm.

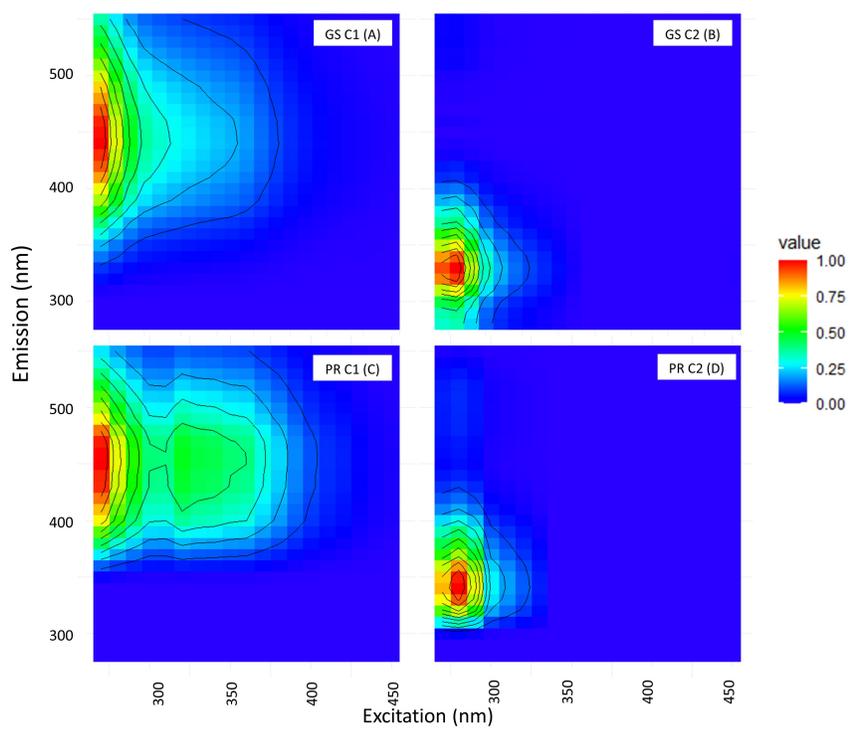


Figure 4.7: *EEMs of PARAFAC component 1 and 2 analysed by the GS method(A and B) and component 1 and 2 analysed by the PR method (C and D) from the water sample data set.*

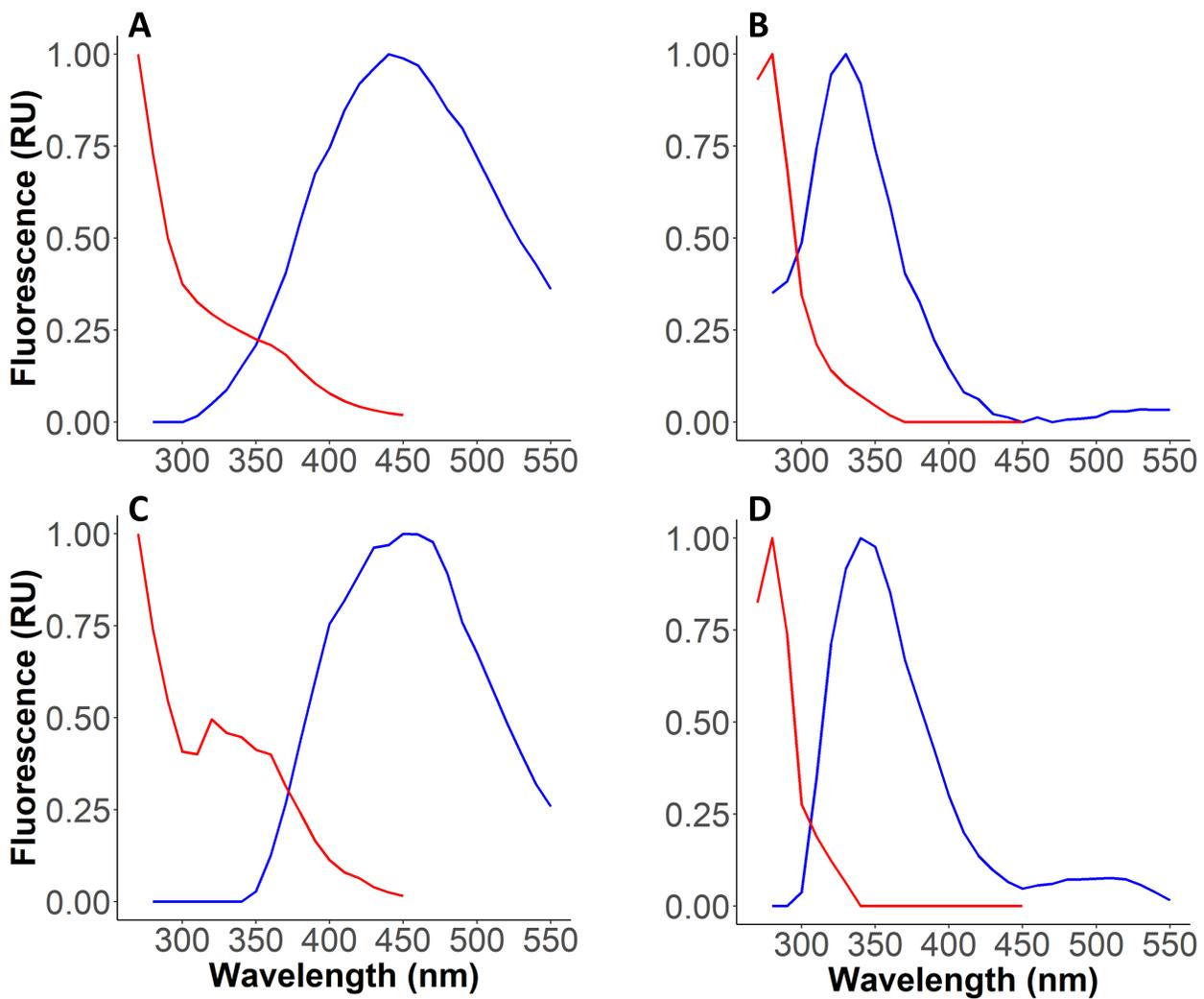


Figure 4.8: PARAFAC component excitation (red) and emission (blue) spectra of water sample data set analysed by the GS method (C1 and C2 are **A** and **B**) and PR method (C1 and C2 are **C** and **D**).

For component 2 (C2), the excitation and emission spectra of both methods looked similar (Figure 4.8 B and D). The peak position of component C1 and C2 found by the GS method were at excitation/emission < 270 / 440 nm (Figure 4.8 A) and 280 / 330 (Figure 4.8 B), respectively. The peak position of component C1 and C2 found by the PR method were at excitation/emission 270(320) / 450 nm (Figure 4.8 C) and 280 / 340 (Figure 4.8 D), respectively. Between the two methods both components C1 and C2 were very similar (C1: $r(45) = .96$, $p < .001$ and C2: $r(45) = .89$, $p < .001$) (Figure 4.9).

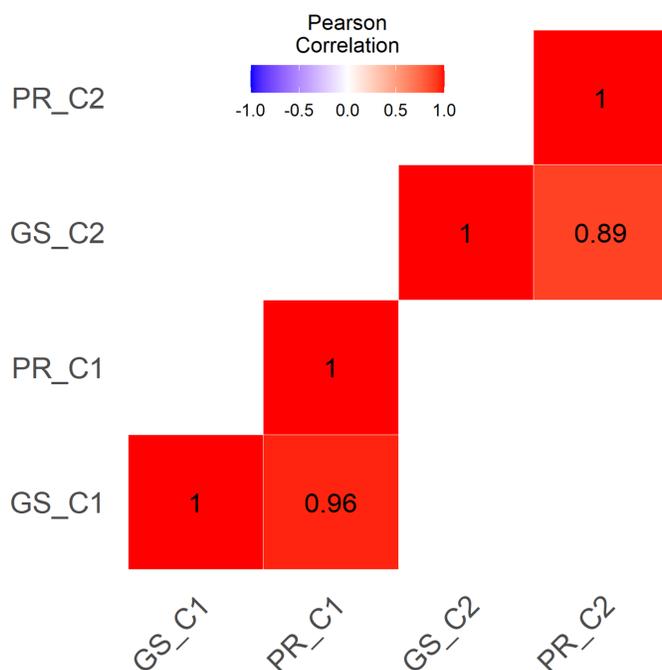


Figure 4.9: Significant correlations ($-.6 > R > .6$) between spectra of C1 and C2 analysed by the PR method (PR) and the GS method (GS).

Cross checking the component spectra of both methods with OpenFluor database (Murphy et al. 2014) yielded similar component descriptions. Table 4.3 includes the characteristics of the components C1 and C2 analysed by the PR and GS method. For both methods, C1 matched with the references that described the component as (terrestrial) humic-like fluorescence. For C2, the results of both methods matched with references describing this component as protein-like/tryptophan-like fluorescence. This demonstrates that despite the small discrepancies in spectral shape and peak position, components were correctly classified.

Table 4.3: *OpenFluor* results of the components C1 and C2 present in the water samples analysed by the PR (PR C1 and PR C2) and GS (GS C1 and GS C2) method.

Comp.	Ex/Em	OpenFluor matches (>0.95)	Components in previous studies matched in OpenFluor	Tucker's Congruence Coefficient	Ex/Em Description
PR C1	260(320)/450	19	C1, Chen et al. (2017), river, porewater, bottom water	Ex= 0.987 Em= 0.999	265(310)/440 Terrestrial Humic-like
			C2, Brogi et al. (2019), Ocean	Ex= 0.990 Em= 0.989	250,320/455 Humic-like
			C1, Derrien et al. (2019) Soil	Ex= 0.984 Em= 0.994	220/446 nm humic-like
PR C2	280/340	31	C3, Garcia et al. (2021) Ocean	Ex= 0.994 Em= 0.993	280/342 protein-like, aliphatic compounds likely derived from recent microbial production
			C3, DeFrancesco & Guenguen (2021) Ocean	Ex= 0.996 Em= 0.984	tryptophan-like 280/335
			C3, Osburn (2016) Coastal waters	Ex= 0.991 Em= 0.987	protein (tryptophan) 275/350
GS C1	<270/440	14	C3, Eder et al. (2022) Stream	Ex= 0.975 Em= 0.998	<245/414 humic-like, terrestrial,
			C3, Murphy et al. (2014) Stream/estuary	Ex= 0.982 Em= 0.986	<245/430 -
			C3, Kida et al., Antarctic lakes and streams	Ex= 0.996 Em= 0.972	250/440 terrestrial humic-like
GS C2	280/330	46	C5, Cawley et al. (2012), estuary, ocean	Ex= 0.996 Em= 0.997	270/<340 Peak T1
			C4, Yamashita et al. (2021), Stream	Ex= 0.995 Em= 0.991	280/330 Protein-like
			C5, Zhou et al. (2019), Stream/ocean	Ex= 0.996 Em= 0.987	275/328 Protein, tryptophan & tyrosine mix

¹ Peak T, <260/435–460, Humic like; Peak C, 320–360/420–460, Humic like; Peak B, 275/310, Tyrosine-like or protein-like (Coble 1996).

Humic-like fluorescence signals have been subclassified with peaks in low excitation ranges being called Coble's peak A or fulvic-like, while the peak in the higher excitation range is called Coble's peak C or humic-like (Chen et al. 2003, Coble 1996, Ishii & Boyer 2012). Peak A humic substances are smaller with less aromaticity compared to the humic substances responsible for the fluorescence signal of peak C (Barsotti et al. 2016, Ishii & Boyer 2012, Wu et al. 2003). When this subclassification is applied to C1 of the PR method, small and larger humic-substance components (Coble's peak A and C) can be detected. For the GS method, C1 only shows 1 peak in the smaller excitation wavelength range (Peak A). In the larger humic-substance region, the C1 of the GS method has some signal, but no peak (Figure 4.7). The PR and GS method differ in this subclassification and therefore this approach is debatable when using the PR method.

Quantitative specificity

First, quantitative specificity was studied by comparing signal intensities of the GS and PR method, as shown in Figure 4.10. For all water samples, the signal intensity of the fluorophores was normalized by dividing the measured intensity by the intensity of the QS standard. These QS normalized signals of the GS method were higher than the PR method (paired T-test, $t(7,1) = -4.63$, $p < 0.05$); between 12-15% times larger. Comparability of quantitative results between different instruments remain an issue to date (Goletz et al. 2011). With the precision of the PR method at $2.3 \pm 1.7\%$, the difference between the methods of 12-15% is much larger. However, 12-15% is still acceptable.

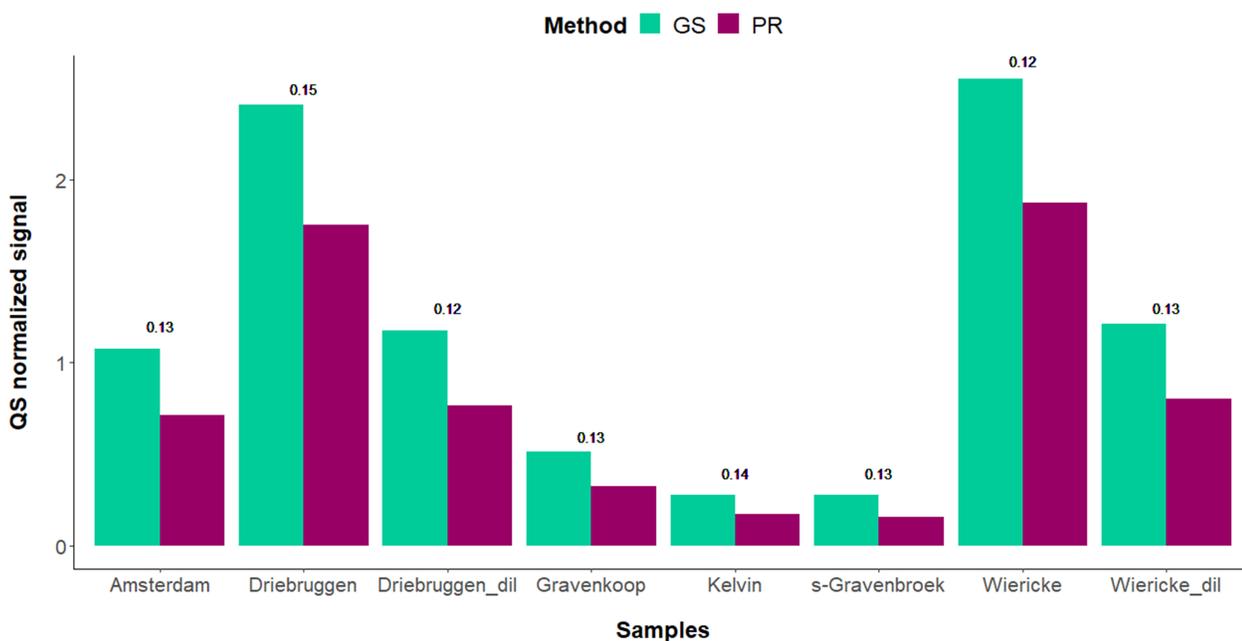


Figure 4.10: *Quantitative specificity of the PR method; the QS normalized signals of various water samples analysed by the PR and GS method ($n = 1$ replicate) and the ratio between them.*

Second, quantitative specificity was studied by looking at the diluted and undiluted samples (Table 4.4). Signal recovery of the fluorophores in diluted samples with respect to the undiluted samples lied between 86-87% for the PR method. The recovery of the GS method, on the other hand, was between 95-98% and demonstrated a better signal recovery of the diluted samples. However, recovery of both methods was within the 15% deviation, and therefore recoveries were both acceptable.

Table 4.4: *Quantitative specificity of the PR method; signal recovery (%) of QS and TR in the Wiericke and Driebruggen samples analysed by the PR and GS method.*

Analyte	Method	Water	Original signal	Signal	Diluted signal corrected	Recovery (%)
QS	PR	Wiericke	1.87	0.80	1.60	86
		Driebruggen	1.75	0.76	1.52	87
	GS	Wiericke	2.55	1.21	2.42	95
		Driebruggen	2.41	1.18	2.36	98
TR	PR	Wiericke	17.62	7.55	15.10	86
		Driebruggen	16.51	7.19	14.38	87
	GS	Wiericke	2.04	0.97	1.94	95
		Driebruggen	1.93	0.94	1.88	97

Some considerations

Some methodological constrains have affected the results of the present study. First, the Raman peak needed to be measured at a different excitation wavelength (420 nm) compared to the commonly accepted excitation wavelength (270 or 350 nm). While qualitatively the PARAFAC components compared well with other studies via the OpenFluor database, direct comparison of the quantitative results was not possible in this way. Using QS normalized signals did improve the comparability between the GS and PR method. The PR method systematically underrated the signal intensity compared to the GS method. When data of two different methods need to be compared, the difference between the two methods can be estimated by measuring a standard and correct data accordingly. Alternatively, the Raman peak at 350 and 420 could be measured by a GS method and the correction factor between 350 and 420 nm can be used to adjust the PR method results. However, even with careful calibration, variance between similar instruments can be very significant (6–10% in intensity) (Goletz et al. 2011). Direct quantitative comparison is questionable but can be useful for, for example, metadata analysis of various published studies.

Second, analysis of the microplate could only be done 2 rows at the time because non-interfering lids that prevent evaporation were not manufactured to date. This caused a longer total analysis time as well as extra manual steps that need to be taken. Although high end equipment makes all corrections automatically creating a simple and quick analysis, the plate reader works with a 96-well microplate. This would allow for high throughput, when cycloolefin lids are purchasable. Cycloolefin lids have low autofluorescence and prevent evaporation of the sample. This would make it possible to analyse 96 samples at the time. Overall, the PR method demonstrated validation parameters within acceptable limits when samples are used with a concentration above 10 µg/L. Where possible, samples with high

absorbance should be analysed diluted and undiluted to check the influence of the absorbance on the signal. Using 96-well plates at once can result in a high throughput with the benefit of using replicate measurements. This could help to detect and eliminate measurement variation.

Despite these limitations, the results provide evidence that even low resolution tools, such as a plate reader, can be used for the general classification of DOM fluorophores. This is important because high end equipment is not commonly present in all laboratories. Plate readers on the other hand are widely used in research, quality control and manufacturing processes in the pharmaceutical and biotechnological industry and academic organizations.

4.4 Conclusion

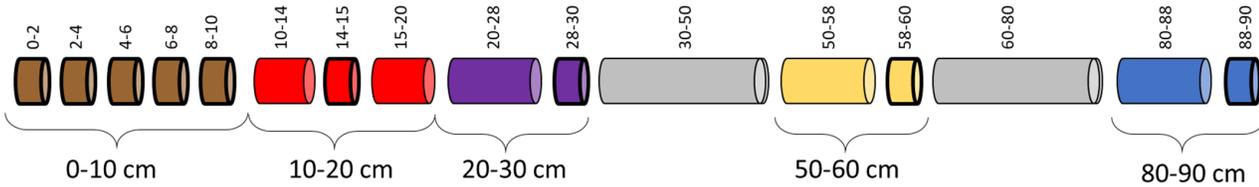
The objective of this chapter was to develop fit for purpose and cost-effective analytical method using a non-automated plate reader to characterize and semi-quantify DOM with help of PARAFAC. The use of the plate reader to analyse DOM by fluorescence EEM spectroscopy and subsequent analysis of data by PARAFAC has not been reported to date. Results from this study suggest that:

- In line with the hypothesis, the proposed fluorescence-spectrophotometric technique is an appropriate method for the identification and semi-quantification of fluorescence components within the DOM spectral range and therefore fit for purpose.
- PR method demonstrated validation parameters within acceptable limits when samples are used with a concentration above 10 µg/L. While small differences between the two methods is apparent, the application of fluorescence EEM spectroscopy in combination with PARAFAC is able to generally characterize DOM in a similar way as high-end equipment.
- Where possible, samples should be diluted to overcome possible matrix effect. It is also recommended to use replicates during analysis, especially when low concentrations are assumed. This could help to detect and eliminate measurement variation.
- This innovative method can be applied to pore water samples described in chapter 5. The volume of these samples is too small for DOC analysis and fluorescence EEM spectroscopy by plate reader can be used as a proxy.

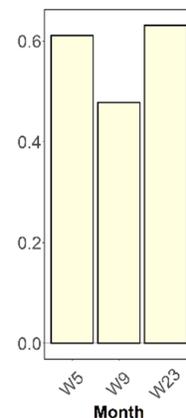
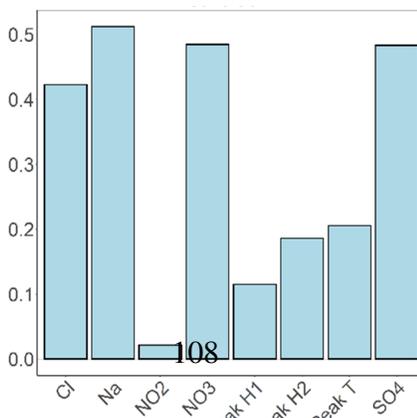
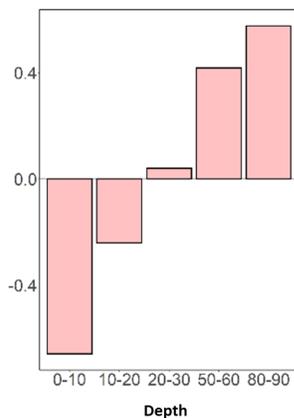
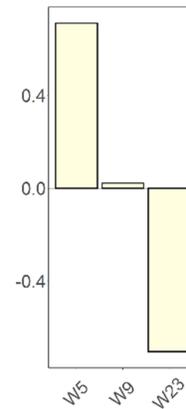
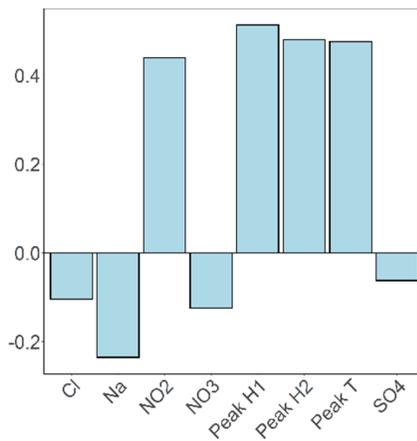
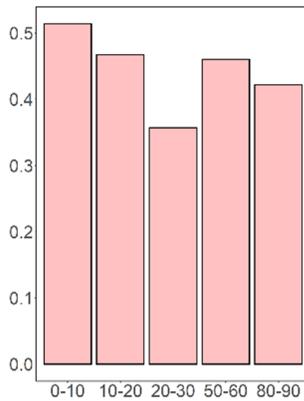
Chapter 5

Understanding removal processes of DOM and other chemical parameters by slow-flow BAC filtration using the analysis of pore water

Biofilter GAC cuttings



Measuring 8 water quality parameters through depths over time by three-way PCA



Variables

Contribution

The data used in this chapter are part of a larger biofiltration project in which multiple biofilters were constructed and run for a 6-month period by Dr Marta Vignola (MV) and postgraduate researchers Dominic Quinn (DQ) and Steve Joyce (SJ). Filter deconstruction and sampling were done as a team. DNA sequencing on the GAC samples was done by DQ. ATP analysis on the GAC samples was done by MV. I carried out pore water extraction and chemical water quality parameter analysis, with the help of SJ at week 12 during my absence. When data were produced by others, it will be clearly indicated in the text.

5.1 Introduction

The BAC filter is still considered a black box (Zhu et al. 2010). Very diverse communities have been found to populate biofilters and significant steps have been made to increase the understanding of their composition (Hu et al. 2020, Lautenschlager et al. 2014, Oh et al. 2018, Palomo et al. 2016), and their role in the treatment process (Bai et al. 2013, Pinto et al. 2012, Ye et al. 2001). However, the true challenge is still identifying the ecological mechanisms responsible for their colonization and assembly and how this might affect their functions and the quality of the final drinking water (Chen et al. 2021, Li et al. 2021).

Results from chapter 3 have demonstrated that slow-flow BAC filtration removes organic compounds and various anions and cations and that EBCT is an important design parameter, which increased the time until the start of breakthrough as well as increased the removal efficiency in apparent steady state. However, it remains unclear how much of this removal of DOM was the result of slow-adsorption and/or biodegradation/biosorption. Biodegradation takes place within the filter and is expected to influence the chemistry of the water.

To unravel the processes that are taking place within the BAC filter detailed information is needed about the change in concentration of water quality parameters in the pores of the filter over time. Water quality parameters such as TOC, DOC, BDOC, AOC, DON, protein-like and humic-like fluorophores, and LC-OCD fractions of DOM within biofilters have been monitored in the past (Boon et al. 2011, Chen et al. 2016, Liao et al. 2013, Persson et al. 2006, Tränckner et al. 2008, Urfer & Huck 2001, Velten et al. 2011a, Wang et al. 1995, Zhang et al. 2016). However, these studies used water samples drawn from various depths of the filter. This technique carries the risk of measuring

the short circuiting water instead of the water that enters the pores. Short circuiting is a condition that occurs when water flows along a nearly direct preferential pathway from the inlet to the outlet of the biofilter, often resulting in shorter contact times with the GAC in comparison with the presumed retention times. This water has a different chemistry compared to the water that enters the pores, where microbial processes are taking place.

When parameters are measured at regular time intervals at various depths the data set has information in three dimensions (modes), being sampling sites, variables, and time. These data are structured in a data array, with biofilter depths, variables and time all having their own mode. This structure is more complex to interpret and requires data reduction methods, such as PCA, to disclose hidden information. For data arrays this is done by three-way PCA, which allows for a good insight into the data structure, showing correlation between variables, and helping in the description of environmental problems.

The objective of this chapter is therefore to further understand removal processes of DOM and other chemical parameters by slow-flow BAC filtration using the analysis of pore water. To do so, the same multiple pilot-scale slow-flow BAC filters of various sizes constructed in chapter 3 were used for this. Changes in chemical water quality composition were studied by measuring seven parameters in the pore water. Fluorescence signal as proxy of DOC was one of them measured with help of the method developed in chapter 4. Data were interpreted by chemometric methods such as Tucker3. Only filter length varied in the experiments, other operational parameters and influent water were constant.

5.2 Materials methods

5.2.1 Biofiltration deconstruction

All details concerning the set-up and experimental operation of the 18 biofilters are given in Chapter 3, Section 3.2.1 till Section 3.2.5 (pg. 53). Three individual 90 cm filter triplicates were deconstructed at week 5, 9, and 12, while the remaining 90 cm filter triplicate as well as the 30 and 60 cm filter triplicates were deconstructed at week 23. Deconstruction was done by interrupting the influent water to the filter system and draining the system for 15 minutes. The biofilter was disconnected and cut in sections of various sizes (see Figure 5.1, so that information about stratification of microbial communities and biochemistry could be obtained. While DNA sampling only required little GAC, the pore water analysis required much more GAC. For this reason, multiple cuttings needed to be combined to be able to extract enough pore water.

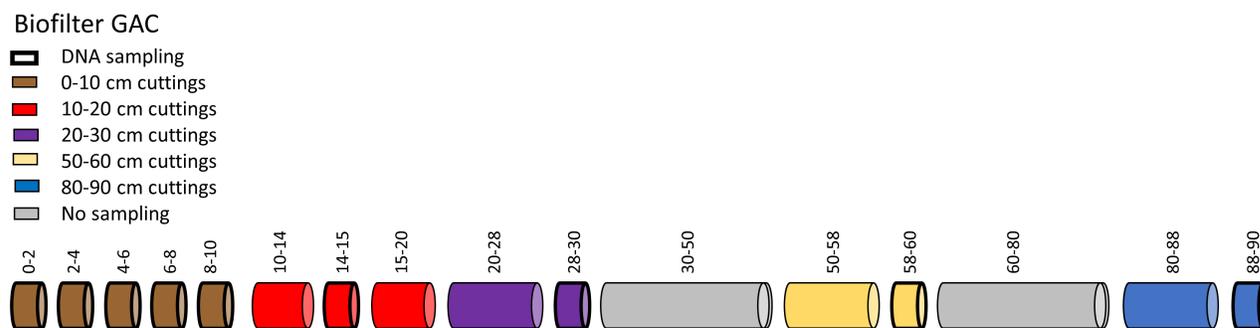


Figure 5.1: Schematic representation of the cutting of the biofilter. GAC from the strongly black rimmed cuttings were first sampled for DNA analysis. GAC cuttings of each of the colours brown (0-10 cm), red (10-20 cm), purple (20-30 cm), yellow (50-60 cm) and blue (80-90 cm) were combined for further pore water extraction. Grey cuttings were not used for either analyses.

5.2.2 Pore water extraction

A schematic overview of the pore water extraction is given in Figure 5.2. (1) 50 mL centrifuge tubes were filled with wet GAC material and (2) centrifuged (Thermo scientific Heraeus Multifuge X1R centrifuge) 5000 rpm for 5 minutes. (3A) The dry GAC material at the top of the centrifuged tubes was collected in separate labelled containers. (3B) The remaining wet GAC was transferred into a 15 mL syringe and the syringe placed in a clean labelled 50 mL centrifuge tube with closed lid. (4)

This tube was centrifuged for 5 minutes at 1000 rpm and (5A) extracted water collected in a 15 ml centrifuge tube. (5B) The remaining dry GAC in the syringe was added to its corresponding collection pot. (6) When all pore water of the sample was collected, the 15 mL tube was centrifuged at 5000 rpm for 5 minutes to sink the suspended solids. (7) The supernatant was transferred to an 8 ml amber glass vial (Fisherbrand™) using a Pasteur pipette. Multiple samples were centrifuged simultaneously, and tubes were reused after a rinse with MQ water and being dried. Samples were stored at 4 degrees in the dark until analysis by ion chromatography (within 24 hours) or fluorescence EEM spectroscopy (within 4 days).

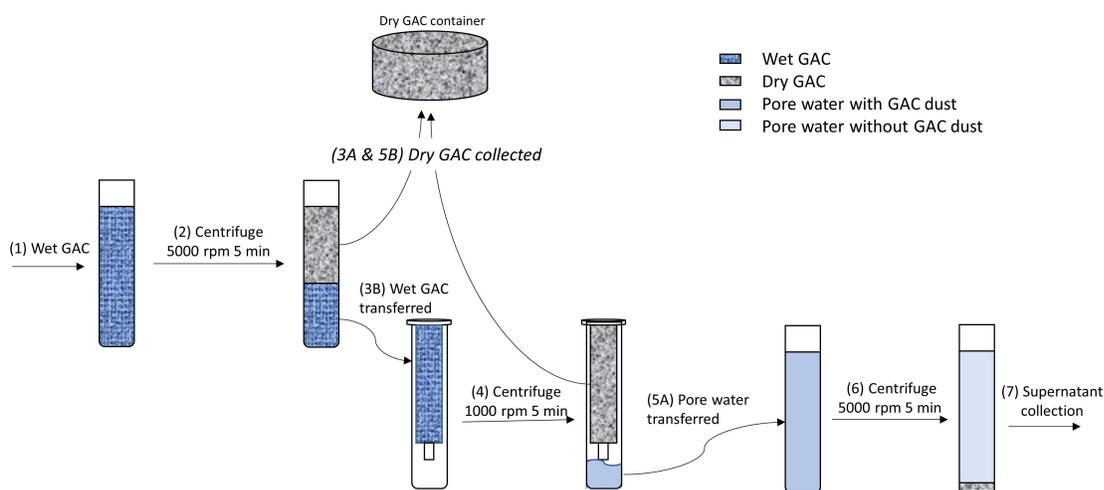


Figure 5.2: Schematic overview of the pore extraction steps. (1) Wet GAC is centrifuged after which (2) the dry GAC is collected, and the remaining wet GAC is transferred into pipettes without the bottom part. (3) Pipettes are placed in centrifuge tubes and centrifuged at lower pace. (4) Afterwards, the dry GAC is collected again while the supernatant is collected in a new centrifuge tube and (5) centrifuged at high speed to let the GAC dust in the sample settle. Supernatant was collected for further analysis.

5.2.3 Analytical techniques

Ion chromatography

Five anions (nitrate (NO_3^-), nitrite (NO_2^-), phosphate (PO_4^{3-}), sulphate (SO_4^{2-}), and chloride (Cl^-)) and two cations (ammonium (NH_4^+) and sodium (Na^+)) were analysed by ion chromatography (see Chapter 3, Section 3.2.6, pg. 57 for the analytical procedure). In short, 0.5 mL sample was filtered through a 0.2 μm nylon filter and analysed by ion chromatography using one technical replicate. The use of a single replicate was chosen to overcome the large sample load and required machine usage. The LOQ corresponds to on average 0.12, 0.06, 0.44, 0.40, 0.16 mg/L for NO_3^- , NO_2^- , Cl^- , SO_4^{2-} , and Na^+ , respectively. These anions and cations were chosen because of their role in the microbial processes (NO_3^- , NO_2^- , and SO_4^{2-}), as well as their non-microbial counterpart to understand physiochemical processes (Na^+ and Cl^-). NH_4^+ (at pH 7-8) and PO_4^{3-} had concentrations below their LOQ, and data were not used.

Fluorescence

The remaining sample volume from the pore water extraction was collected in 8 ml disposable amber glass vials. Because the remaining volume was very small, TOC/DOC analysis was not possible. Fluorescence EEM spectroscopy by plate reader on the other hand only required 330 μL and was therefore used as a proxy for TOC. Unfiltered samples were analysed by fluorescence EEM spectroscopy using the M200Pro plate reader (Tecan, Switzerland) (as developed in chapter 4) using three technical replicates when sufficient sample volume was left. Results of chapter 4 advise to analyse diluted and undiluted samples to cover for possible influence of matrix effect. However, the water volume collected during pore extraction did not give enough volume to do so. Therefore only the undiluted samples were analysed.

Chapter 4, Section 4.2.6 (pg. 89) explains the PARAFAC data preprocessing in detail. In short: for all samples the EEM and absorbance were measured, together with at least 1 blank sample EEM and absorbance per day as well as 1 Raman scan (Ex. = 420 nm, Em. = 460-550 nm) of the blank per day. EEM samples were Raleigh Scatter corrected, Raman corrected, Inner Filter Effect corrected and normalized to Raman units (see figure 4.2). The LOQ of the technique was 0.2 Raman unit. However, using the LOQ would exclude substantial number of samples. The technical and biological replicates demonstrated high repeatability and as a result the limit of detection of 0.06 Raman unit was used as a cut off for sample inclusion. Therefore, 137 samples were used to build the PARAFAC model. Chapter 4, Section 4.2.4 (pg. 85) explains the PARAFAC modelling. In short, visual spectral inspec-

tion, residual examination, and the EEMqual (which looks at model fit, CORCONDIA, and split-half validation) were used to determine the correct number of components. The model was built followed by the extraction of the qualitative and quantitative data.

5.2.4 Statistical analysis

All data analysis was carried out using R (version 4.0.2). All results were gathered in a table and data entries below the LOQ were substituted by 0.65 times their LOQ.

Normality and outliers

The normal distribution of each variable was tested. Shapiro–Wilk test was applied to check the distribution pattern of the variables. For both data sets, not all variables demonstrated a normal distribution ($p < 0.05$), which is a common phenomenon in environmental data (Leardi et al. 2000). This was confirmed by frequency histograms demonstrating skewness. A logarithmic transformation was therefore applied to the variables of both data sets to transform the data into a (pseudo-) normal distribution.

Outliers were evaluated by the Hotelling $T^2 - Q$ residual plot generated from an initial PCA model (Goueguel 2020, Bro & Amigo 2020). Q-residuals were calculated as the distance between the original position of an object to the principle component space (squared orthogonal Euclidean distance) and showed how well the object is fitted by the PCA model. The distance from the projection of the object to the principle component space and the origin was used to get the Hotelling T^2 distance (score distance). The T^2 distance visualised extreme objects. Two critical limits were used in the Hotelling T^2 vs Q residual plot. (1) The hotelling T^2 cut-off for outlying the squared score distances is the 0.975 quantile of the Chi-Square distribution with k degrees of freedom (Equation 5.1) (Goueguel 2020). (2) The cut-off values for the Q-residual distances were obtained using the Wilson-Hilferty approximation for a Chi-Squared distribution (Equation 5.2) (Goueguel 2020). Outliers were inspected for both data sets, but none were found based on these cut-off values (See Appendix C, Figure C.1 and Figure C.2, pg. 195).

$$SD > \sqrt{(X_{k,0.975}^2)} \quad (5.1)$$

$$OD > [median(OD + Z_{0.975} \times MAD(OD^{2/3}))]^{3/2} \quad (5.2)$$

Three-way modelling by Tucker3

Three-way PCA, more precise Tucker3, was applied to two individual data sets. To understand the change in chemical variables through depth over time data of 5 depths (0-10, 10-20, 20-30, 50-60, 80-90 cm), 8 variables (nitrate, nitrite, sulphate, chloride, sodium, peak T, peak H1, and peak H2), and 3 timepoints (week 5, 9, and 23) were combined to a data array (data set Time). Week 12 was not included in the data array, because fluorophore data of that week were missing as a result of wrong sample storage and subsequently missing analysis. To understand the similarity between the three filter lengths, the change in chemical variables through depth for the various filter lengths was analysed at week 23 having 3 depths (0-10, 10-20, 20-30 cm), 8 variables (nitrate, nitrite, sulphate, chloride, sodium, peak T, peak H1, and peak H2), and 3 filter lengths (the 30, 60, and 90 cm filter). From the three filter lengths only the top 30 cm were used, and all data combined to a data array (data set Length).

Data pre-processing

Each data set required pre-processing which homogenises scales and units without altering the differences among the sample sites and among the sampling times. A j-scaling was used. The three-way array X (with I, J, and K modes) was matricized to a two-way matrix X_b having I x K modes in the row direction and the J mode in the column direction. On this matrix variables were log transformed to make the data semi-normally distributed, because data were not normally distributed. Afterwards, autoscaling was performed in the variable (column) direction. As a result, the global variance of each variable was set to one, and the differences among the objects and the conditions are preserved. J-scaling calculates averages over two modes.

Tucker3

For the Tucker3 analysis the ThreeWay toolbox was used in R (version 4.0.2) (Giordani et al. 2015). A description of the Tucker3 approach is given in chapter 2 section 2.4.12 (pg. 45). The T3 code provided an interactive Tucker3 analysis with the following main steps: (1) provide the number of entities in the P, Q, and R mode. (2) determine the type of centring and normalization (already done, so not executed). (3) PCAs of super matrices with slices of the 3way array next to each other, thus 3 super matrices were analysed by PCA, and for each component matrix and illi was found. The results are used to create a generalized scree plot, which can be used to find the number of components for each mode. (4) specific convergence criterion is set to $1 \cdot 10^{-6}$ by default. The statistical validity of the obtained component matrices was assessed by a bootstrap procedure for computing the confidence

intervals for the solution (Giordani et al. 2015). The analysis was based on 500 bootstrap samples and matching via optimal transformation towards full solutions.

Other statistical analyses

The correlations were determined by Pearson correlation using an alpha significance level of 0.05. For filter depths significant differences were ascertained by KW H-test as the data were not normally and homogeneously distributed as well as sample size small using an alpha significance level of 0.05, followed by the Dunn's post-hoc test (equal variances, unequal group sizes) to examine the pairwise comparison with p-values adjusted by the Bonferroni method. The results were expressed by the group mean together with its variance (expressed by the standard error).

5.3 Results and Discussion

5.3.1 Fluorescence components

Chemical water composition was monitored through filter depth over time to understand the change happening through the filter. Apart from anion and cations, the fluorescence signal was monitored in various filter depths over time. To be able to study the change in presence of the fluorophores, the fluorophores were extracted from the fluorescence EEM samples ($n = 137$) by PARAFAC analysis. Examples of pre-processed EEMs are given in Figure 5.3. This figure shows that there is a difference in signal intensity and ratio between eligible fluorophores when comparing different depths (A compared to B) and weeks (A and C), and filter length.

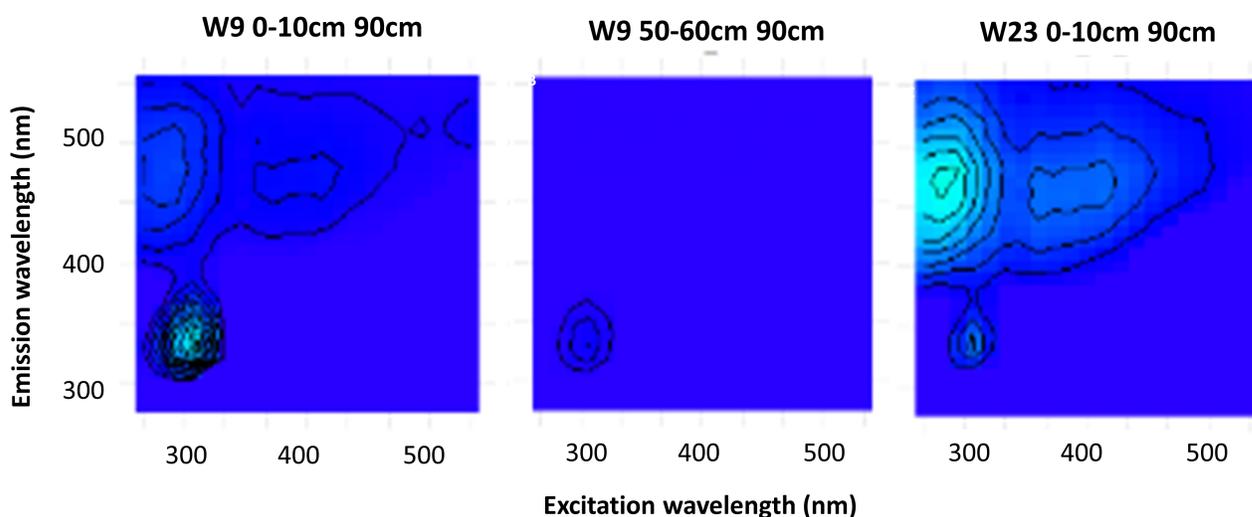


Figure 5.3: Example EEMs of the pore water data set. Sample from week 9 at 0-10 cm from the 90 cm filter (A); sample from week 9 at 50-60 cm from the 90 cm filter (B); sample from week 23 at 0-10cm from the 90 cm filter (C).

The appropriate number of components was selected that explained all meaningful fluorophores while excluding meaningless scatter with help of the EEMqual results (See Appendix C, Figure C.3, pg. 196). Three components were needed to build a PARAFAC model, because at component four the EEMqual dropped significantly. With this information a three components PARAFAC model was created, and qualitative and quantitative data produced. The qualitative data (shape and position of the components) are presented in Figure 5.4.

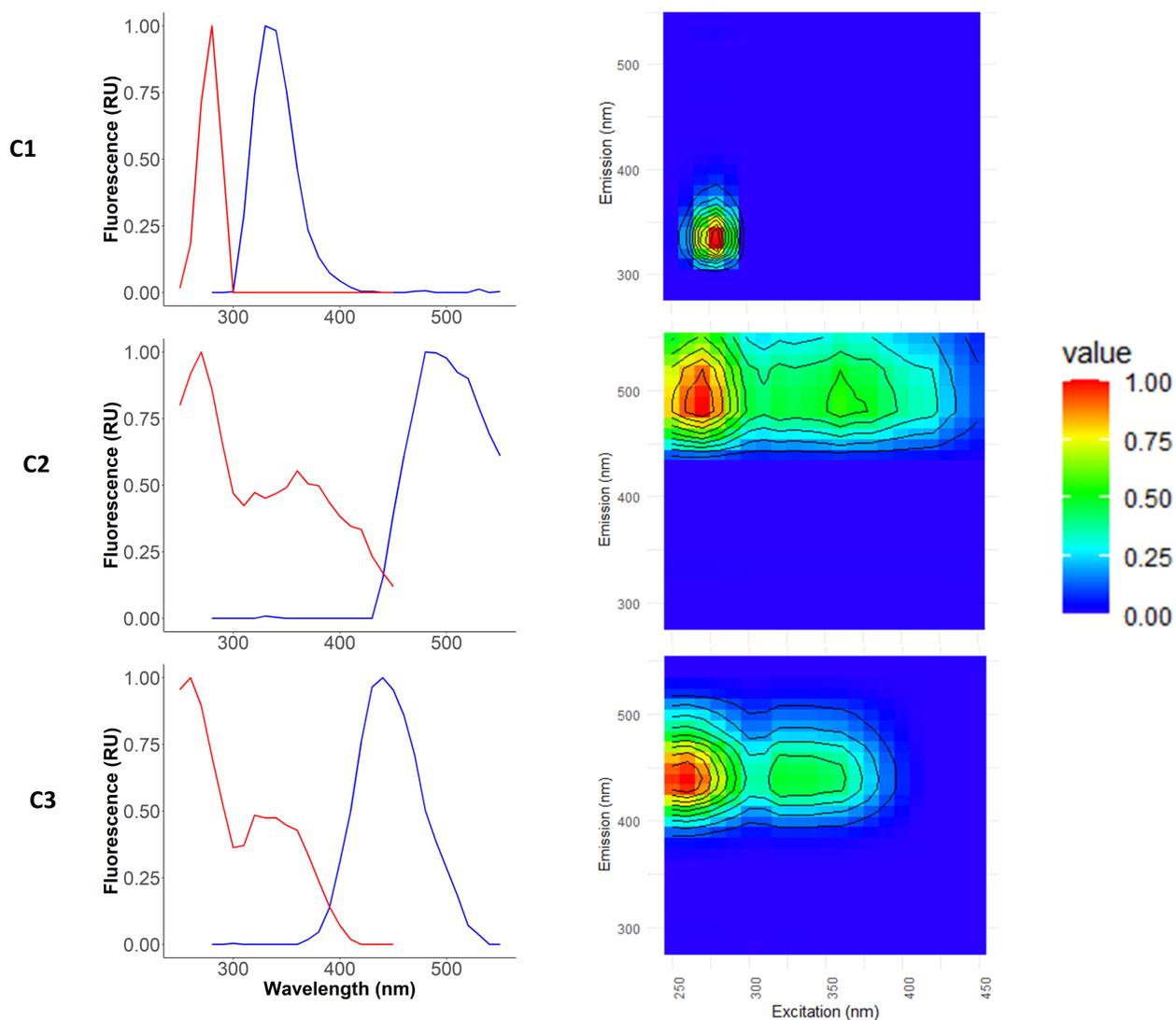


Figure 5.4: *Qualitative results of the PARAFAC analysis of the pore water data set. Left: the excitation (red) and emission (blue) spectra of the components C1-C3. Right: their corresponding EEMs.*

In Figure 5.4 component 1 showed one peak (280 nm) for the excitation and one peak (330 nm) in the emission spectrum C1. The peak position of component C2 and C3 were at excitation/emission 270(360)/480 nm and 260(320)/440 nm, respectively. Component C2 and C3 had a moderate correlation (Pearson correlation, $r(47) = .38$, $p = .007$), while the correlation of C1 with both components C2 and C3 was absent.

Cross checking the component spectra with OpenFluor database (Murphy et al., 2014) yielded component description given by other studies. Table 5.1 includes the characteristics of the components C1, C2, and C3. C1 matched with references describing this component as protein-like / tryptophan-like fluorescence (hereafter called peak T). Both C2 and C3 matched with the references that described the component as (terrestrial) humic-like fluorescence (hereafter called peak H1 and peak H2).

Table 5.1: *OpenFluor results of the components C1 – C3 present in the pore water samples.*

Comp.	Ex/Em	OpenFluor matches (>0.95)	Components in previous studies matched in OpenFluor	Tucker's Congruence Coefficient	Ex/Em Description
C1	280/335	4	C1, Kida et al. antarctic lakes & streams	Ex= 0.952 Em= 0.979	Autochthonous, protein-like
			C3, Brianchi et al., coastal & open waters	Ex= 0.961 Em= 0.969	Amino-acid-like DOM
			H1, Huarong et al. river water	Ex= 0.969 Em=0.954	Protein-like
C2	270(360)/490	>25	C2, Murphy et al., Stream, river, estuary & ocean	Ex= 0.998 Em=0.996	Terrestrial humic-like material
			C2, Graeber., stream	Ex= 0.992 Em=0.998	
			C1, Eder et al., stream	Ex= 0.994 Em=0.995	
C3	260(320)/440	16	C1, Lambert et al., rivers	Ex= 0.990 Em=0.967	terrestrial humic-like component
			C1, Stedmon., seawater	Ex= 0.979 Em=0.972	terrestrial humic-like component
			C2, Osburn et al., estuary	Ex= 0.974 Em=0.978	aromatic, conjugated macromolecular substances of terrestrial origin

5.3.2 Variable correlation

After fluorescence analysis, the monitored chemical parameters now included: fluorophore peaks H1 H2, and T, as well as nitrate, nitrite, chloride, sulphate, and sodium. The first impression of the relationships between these variables was given by the Pearson correlation among the variables. An overview of all meaningful correlations is given in Figure 5.5. Three variable groupings can be found in the matrix: (1) peak H1 and peak H2, (2) peak T and H1, H2 and NO_2^- , (3) Na^+ , Cl^- , and SO_4^{2-} .

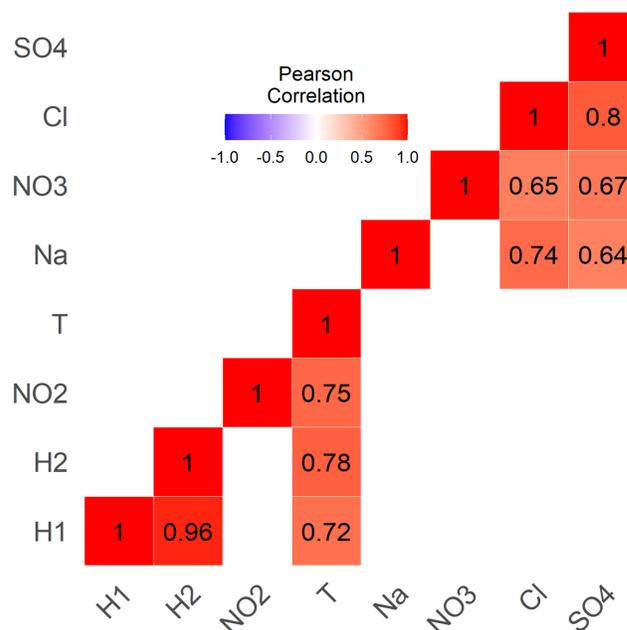


Figure 5.5: Significant correlations ($-.6 > R > .6$) of the measured chemical parameter in the biofilter over time and depth for various lengths.

5.3.3 Depth, variables, and time

Tucker3 model

To interpret the chemical variables over time for various sampling sites simultaneously, Tucker3 was applied to data set Time. For Tucker3, the data set was arranged in a three-way data array with modes: 5 filter depths (0-10, 10-20, 20-30, 50-60, 80-90cm), 8 measured chemical variables (H1, H2, Pr, NO_3^- , NO_2^- , SO_4^{2-} , Cl, Na^+) and over 3 time points (week 5, 9, and 23) from the 90 cm filter. Therefore, the data set was arranged as a three-dimensional array with modes: 5 (depths) \times 8 (variables) \times 3 (time).

To find the appropriate number of components for each mode, initial Tucker3 models were created with up to 5, 8, and 3 components in respectively the p, q, and r mode. The scree plot was used to visualise how many additional percent of variance each model explained when adding extra components (See Appendix C, Figure C.4, pg. 196). Generally, the optimal complexity of the Tucker3 model is the one that requires the smallest number of components, but still describes relatively high fraction of data variance. Figure C.4 scree plot of the Tucker3 model of data set Week with the percentage variance explained by the model when using up to 5, 8, and 3 components in the P, Q, and R modes.

For data set Week, the complexity that shows as much variance but is still interpretable is the model with 2 components in each mode, also known as a [2,2,2]-model, accounting for 70% variance. The statistical validity of the bootstrap computing confidence intervals for the current fit was between 66-87%. Such a medium variance with environmental data is not unusual as a result of the very high noise related to the great variability of environmental and sampling conditions (Leardi et al. 2000).

It has two components in mode P (depth), two components in mode Q (the chemical variables) and two components in mode R (weeks). Results of the Tucker3 model are reported in Figure 5.6 with loading plots of Depths (P), Variables (Q) and Weeks (R). Each of these diagrams may be inspected separately, e.g. in Figure 5.6 B a clear discrimination between the chloride and nitrite is noticeable.

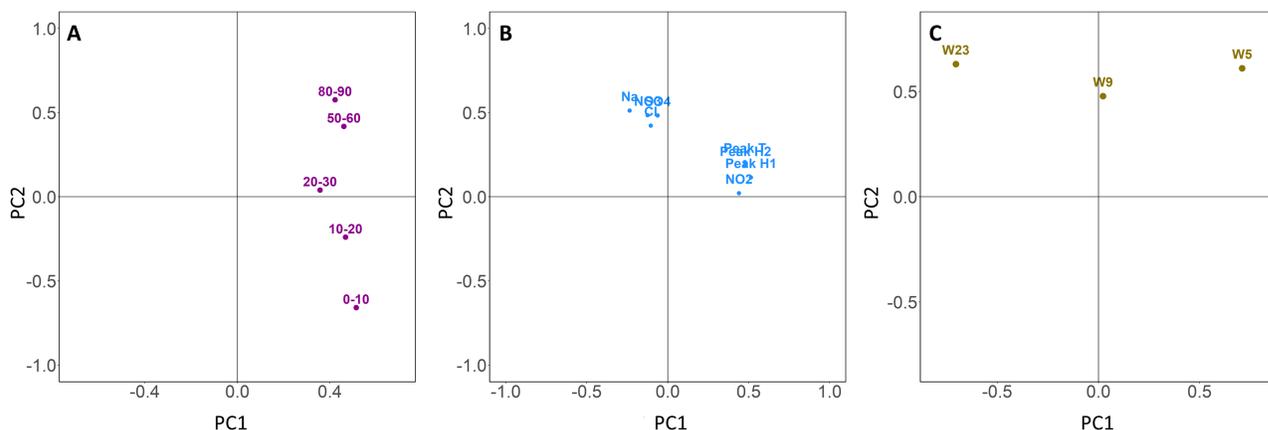


Figure 5.6: Loading plots of the Tucker3 [2,2,2]-model from data set Week with depths (A), chemical variables (B), months (C).

More information, however, can be expected by the joint interpretation of all loading plots. For this, the information in G core is required. The G core of the [2,2,2]-model is shown in Figure 5.7 A together with its unfolded two frontal planes (time slices) in Figure 5.7 B. In total the core matrix of the [2,2,2]-model has eight elements (g111 up to g222). The two parts of this unfolded matrix correspond to the first and second components of ‘Time’, while the rows relate to the components of ‘Depth’

and the columns within each block refer to the components of ‘Variables’. From the core matrix it appeared that a diagonalization has been obtained, because the two major parts of the variance are covered by the body diagonal elements, g111 and g222 (Henrion 1993). The remaining elements have minor importance.

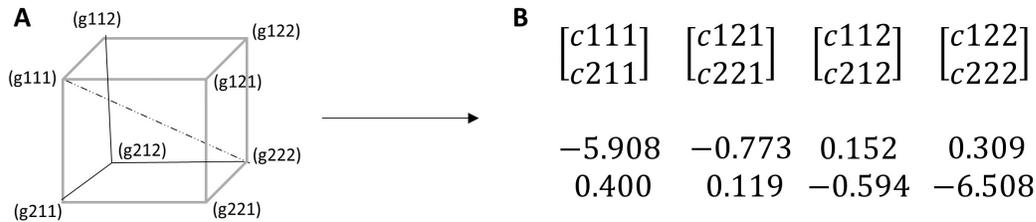


Figure 5.7: Core of the Tucker3 'Time' [2,2,2]-model (A) and its unfolded core matrix equivalent represented by the frontal planes (B). g111 and g222 have both a high negative value.

With most of the variance covered by g111 and g222, these elements together describe almost the full G core of Figure 5.7 A, which simplified lead to a trilinear model with three terms, see Equation 5.3.

$$x_{ijk} = a_{i1}b_{j1}c_{k1}g_{111} + a_{i2}b_{j2}c_{k2}g_{222} + e_{ijk} \quad (5.3)$$

Loading plots can be interpreted simultaneously with respect of their respective signs and magnitudes. This means that only large magnitude loadings with the right combination of signs can be interpreted together. For a better visual interpretation, the loading of the two components for each mode in Figure 5.6 are plotted in individual plots in Figure 5.8. Figure 5.8 A1 and A2 present the loading plots of the two components of the depth's mode, while Figure 5.8 B1 and B2 present the loading plots of the two components of the variable's mode, and Figure 5.8 C1 and C2 present the loading plots of the two components of the month's mode.

G111 and g222 interactions together describe almost the full G core, which can be simplified to a trilinear model with two terms, see, Equation 5.3. Loading plots can be interpreted simultaneously with respect of their respective signs and magnitudes. An example of the interpretation of the Tucker3 model is as follows: Term 1 of Equation 5.3 (explaining 16% of information) has a negative core element, g111, and the interaction will be important for samples with high positive B1 in combination with negative C1, as well as negative B1 together with positive C1 loadings compensating the negative sign of g111. This means that positive large A1 (all depths), positive large B1 (peak T, peak H1, peak

H2, and NO_2^-), and large negative C1 (W23) can be interpreted together. Term 2 of Equation 5.3 (explaining 54% of information) also has a negative core element, g222, and also here the interaction will be important for samples with high negative A2 loadings. More of these interactions are present and an overview is given in Table 5.2.

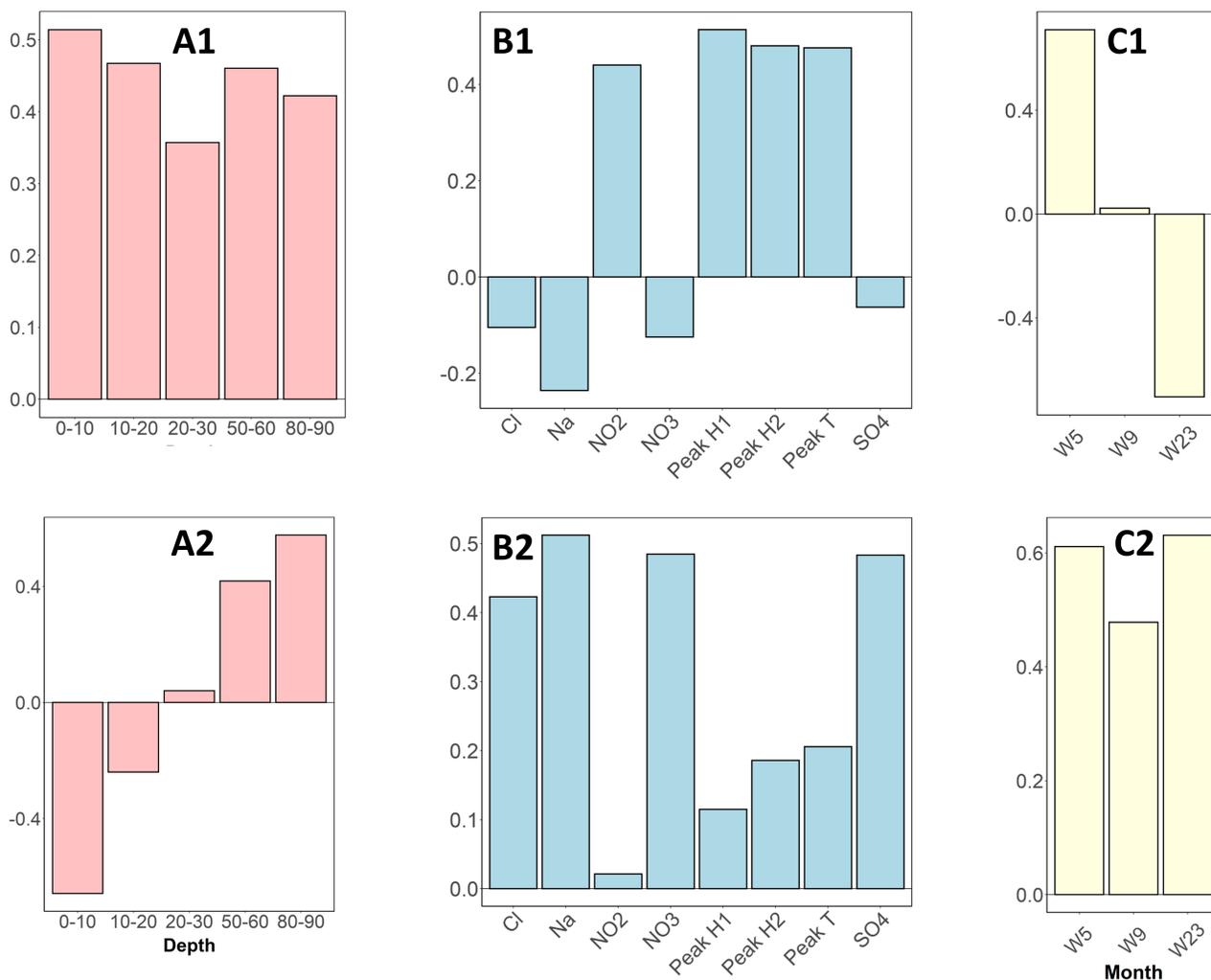


Figure 5.8: Tucker3 [2,2,2]-model loading plots from data set Week with depth loading plot of component 1 (A1), depth loading plot of component 2 (A2), variables loading plot of component 1 (B1), variables loading plot of component 2 (B2), month loading plot of component 1 (C1), and month loading plot of component 2 (C2).

Table 5.2: *Meaningful interactions for elements g111 and g222 of the Tucker3 [2,2,2]-model of data set Week.*

g111 (negative)	A	B	C
Large A1, +B1, -A2	All depths	C1, Peak T, peak H1 peak H2, and NO ₂ ⁻	W23
Large A1, -B1, +A2	All depths	No high magnitude	W5
g222 (negative)	A	B	C
Large -A, B, C	0-10 cm + 10-20 cm	Na ⁺ , Cl ⁻ , NO ₃ ⁻ , SO ₄ ²⁻ , (Peak H1, H2, T)	All weeks

From Figure 5.8 and Table 5.2 it appeared that the variables can be arranged in two groups. Group 1: most anions and the cation variables (Na⁺, Cl⁻, NO₃⁻, and SO₄²⁻) are present in the pore water in all weeks with some form of stratification. Group 2: the other variables (peak T, peak H1, peak H2, and NO₂⁻) are most influenced by the weeks with the highest concentrations at week 23. This grouping of variables is somewhat similar to that of the correlation matrix in Figure 5.5, with the only difference that in the correlation matrix no significant correlation was present for the peak H1 and H2 with nitrite.

Group 1 variables

The first group of variables that showed similar behaviour are sodium, chloride, sulphate, and nitrate. Their concentrations in the pore water of the 90 cm filter over time are given in Figure 5.9. All anions were present at each filter depth and in each week these compounds were in an equilibrium with the mobile phase or an equilibrium was building up. For chloride this concentration was higher than the influent concentration in the top 10 cm of the filter at week 5 (Figure 5.9 A and Figure 5.10 A), while it was lower at all other depths. In the remaining weeks the chloride concentration was for most depths slightly higher than the influent concentration. In Chapter 3, Section 3.3.2 (pg. 70) the significant removal of chloride at the first weeks of operation was discussed, followed by the displacement by nitrate (Chen et al. 2015, Mubita et al. 2019). This displacement resulted in the increased concentrations of chloride in the immobile phase of the micropore that eventually diffuses towards the mobile pore of the larger pores and into the water that exits the filter. This again, resulted in an effluent concentrations higher than influent concentrations.

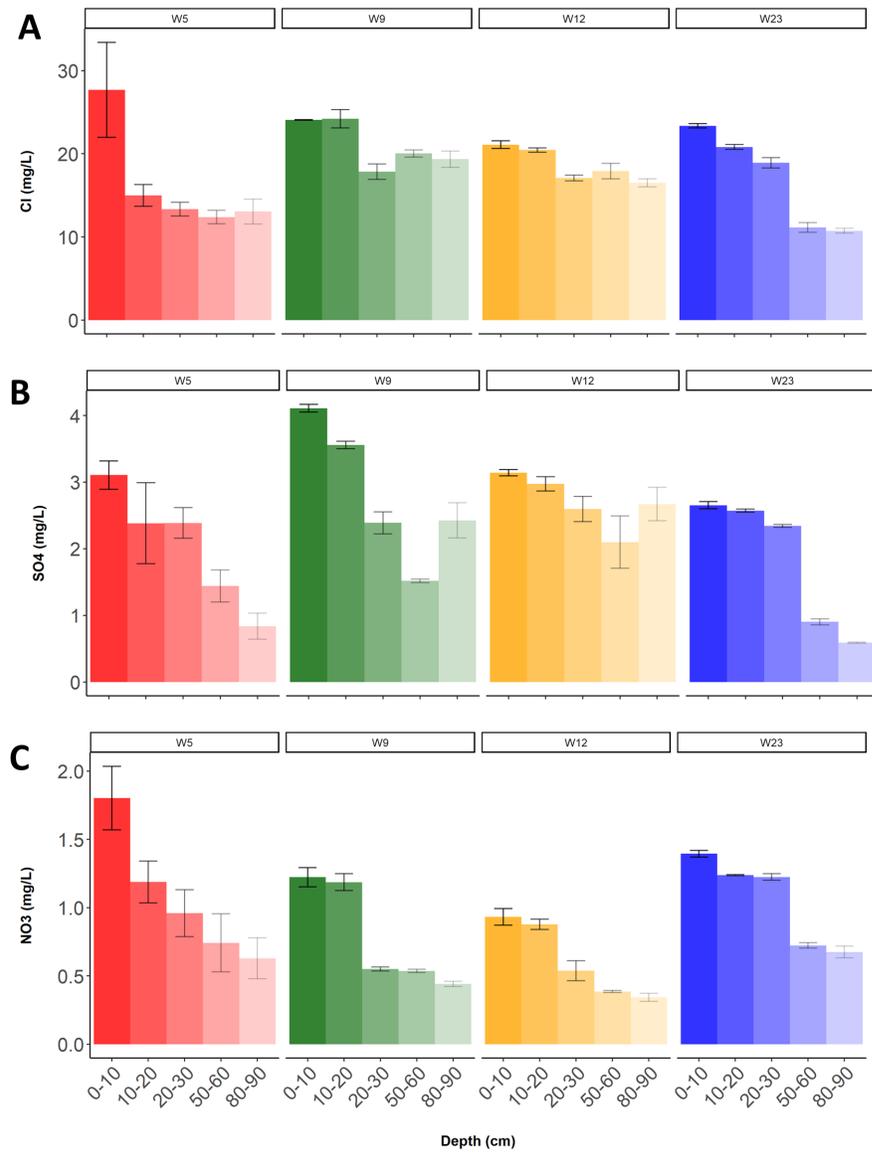


Figure 5.9: Concentration Cl^- (A), SO_4^{2-} (B), NO_3^- (C) in the pore water of various depths (mean \pm s.e., $n = 3$ replicates) of the slow-flow BAC filter over several timepoints.

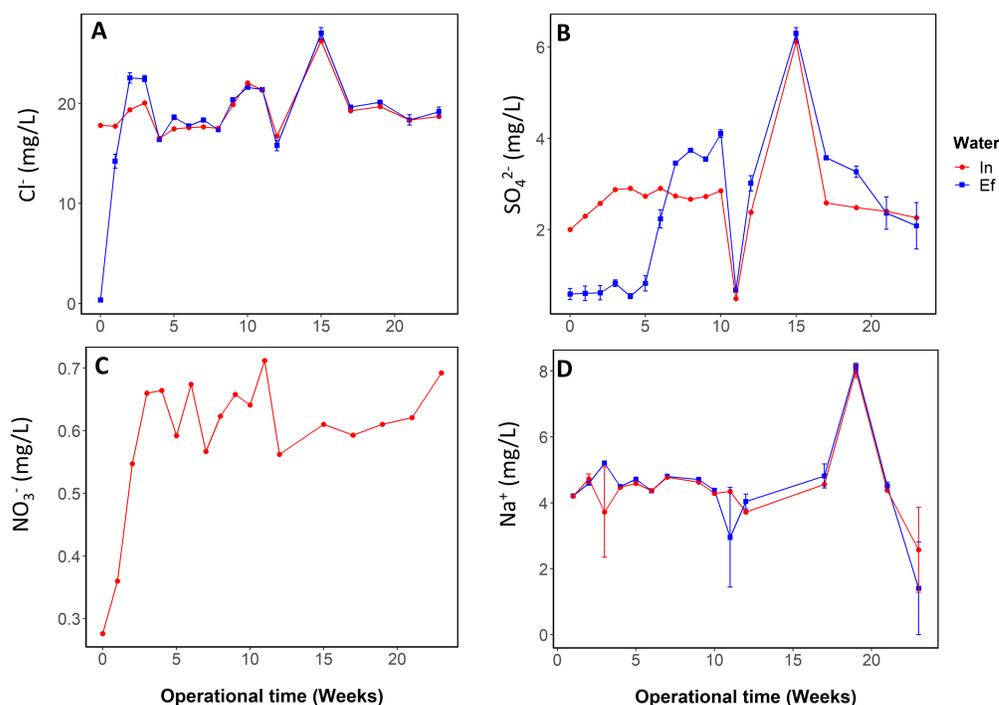


Figure 5.10: Repeated graphs from chapter 3. In- and effluent concentrations of the group 1 variables Cl^- (A), SO_4^{2-} (B), NO_3^- (C) and Na^+ (D) (mean \pm s.e., $n = 3$ replicates) of the 90 cm filter during 23 weeks of operation.

Sulphate showed stratification at week 5 with significant higher concentration in the top 10 cm compared to the bottom 10 cm (Dunn's posthoc on KW, $H(4) = 10.47$, $p_{\text{adj}} = .033$). For sulphate the concentration in the pores at week 5 was lower than that of the influent (Figure 5.9 B and Figure 5.10 B). Chapter 3, Section 3.3.2 (pg. 70) discussed that sulphate was also be removed during the first weeks of the experiment. Indeed, the low concentrations in the pores cause the diffusion to go from mobile phase to immobile phase of the micropores. In the remaining weeks the sulphate concentrations of the various depths varied. The top demonstrated higher concentrations compared to the bottom depths.

Similar to chloride, sulphate was highly likely being substituted by nitrate. Also here, the displacement caused the sulphate concentration in the immobile phase to be higher than in the mobile phase driving the diffusion towards to mobile phase and into the effluent. As a result sulphate was leaching from the filter as observed in Chapter 3.

The concentration of nitrate in the pore water appeared to differ among filter depths (Figure 5.9 C) with highest concentrations in the top, however this was not significant. The concentration in the top pores was higher than the influent, while at lower depths the concentration was lower (Figure 5.10 C). In the top of the filter, the pores were filled with nitrate and excess would slowly diffuse back to the mobile phase again and travel down. There it would meet pores with immobile water with low concentrations and diffusion would bring the molecules into these pores at lower depth. As this process slowly continues, the pore water concentration increased in all depth from a concentration below the influent until a higher concentration, which again could result in desorption of nitrate molecules. At a certain point all pores are saturated, and the nitrate concentration will have a breakthrough. The 30, and 60 cm filters started to show this breakthrough in the last weeks of the experiment (see Appendix A, Figure A.8, pg. 188).

Group 2 variables

Figure 5.11 shows that peak H1 and H2 have a similar pattern through depth over time, which corresponds with the very good correlated (Pearson correlation, $r(13) = .952$, $p < .001$). The concentration of two humic-like fluorophores was below or around zero at week 5 and their concentration increased between week 5 and 9. At week 9, peak H1 and H2 only showed a detectable signal for the 5 and 15 cm sections, which were not significantly different from each other. Between week 9 and 23, peak H1 and H2 kept increasing in all filter depths, but mostly in the top.

The influent water only showed the presence of humic-like fluorescence (peak H1 and H2), while the presence of the biodegradable peak T and peak B were not present (see Appendix C, Section C.1, pg. 197) for the fluorescence EEM PARAFAC model of the in- and effluent). Concentrations of peak H1 and H2 were lower in the pore water than the influent for all filter depth at all timepoints (Figure 5.11 and Figure 5.12). This suggests that the humic substance molecules were driven into the immobile phase of the pores at all depths and all weeks. In other words, GAC seemed not to be fully saturated in any depths at any point of the experiment.

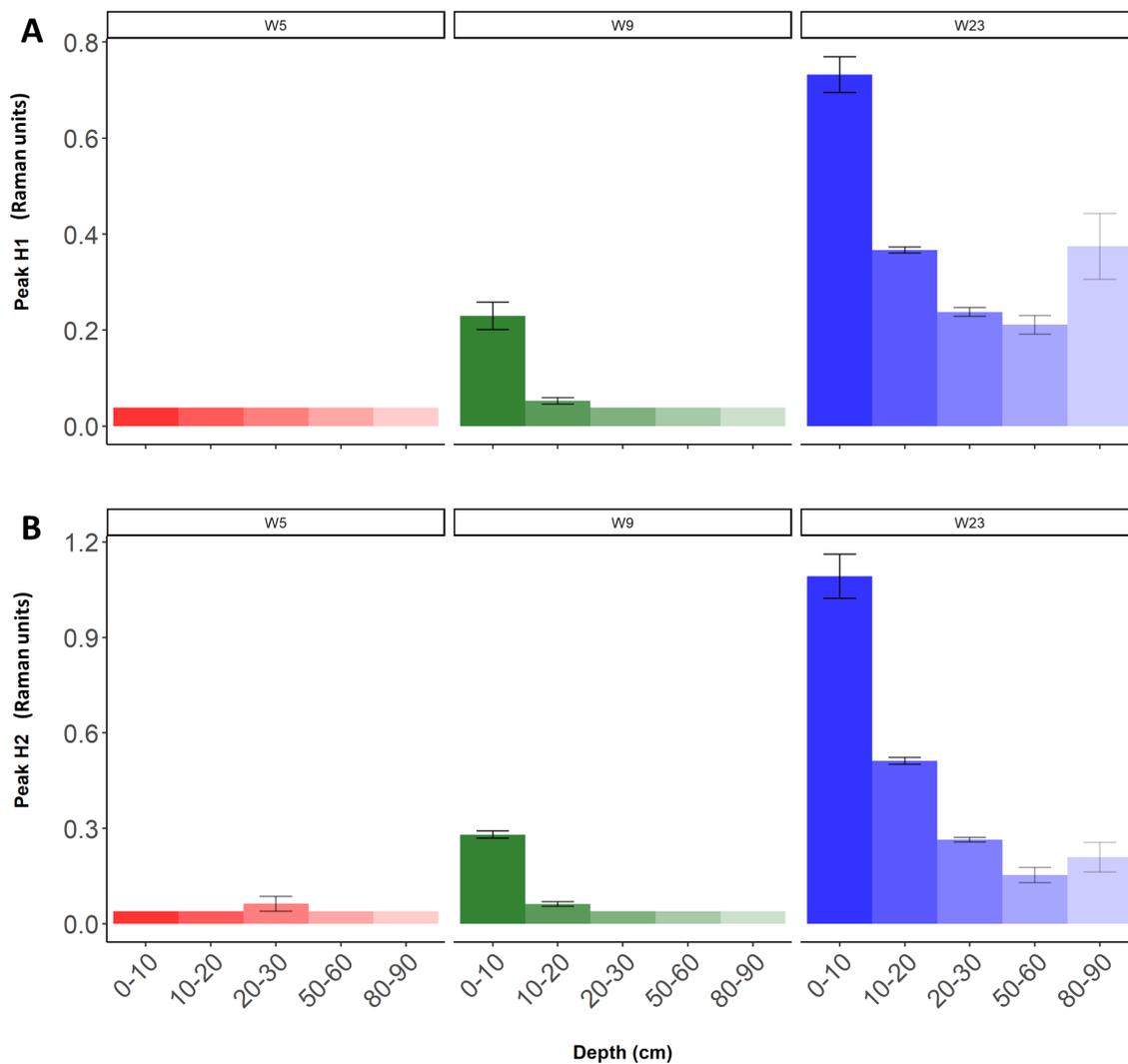


Figure 5.11: The fluorescence signal of fluorophore peak H1 (A) and peak H2 (B) in the pore water of various depths (mean \pm s.e., $n = 3$ replicates) of the slow-flow BAC filter over several timepoints.

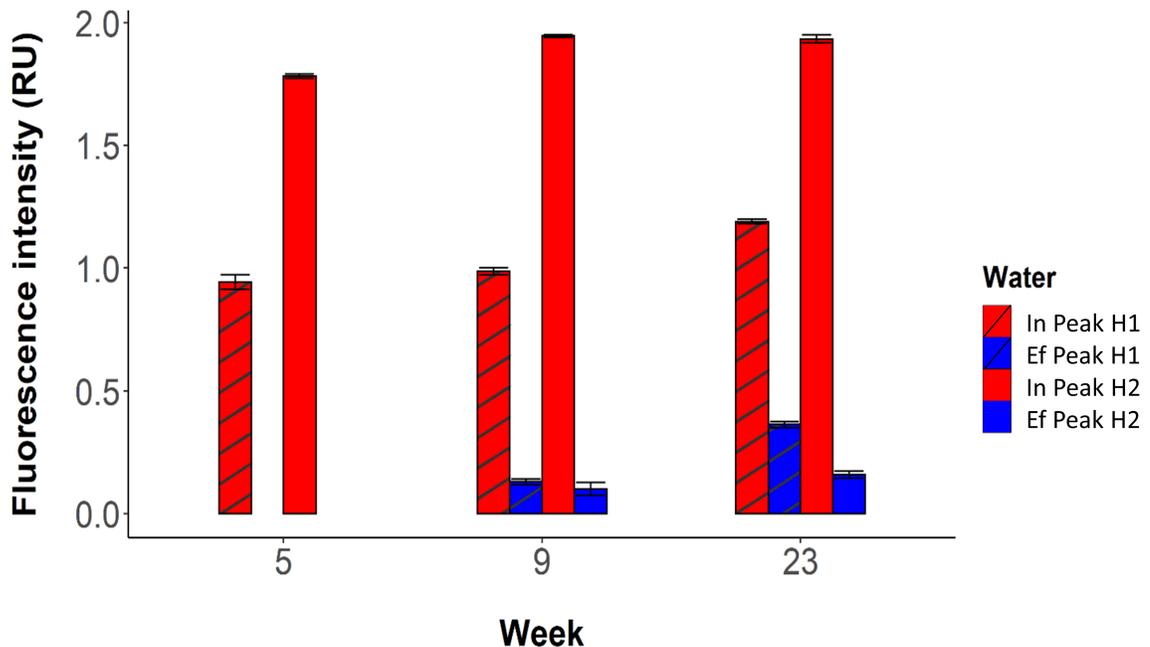


Figure 5.12: Fluorescence signal of Peak H1 (stripes) and Peak H2 (unstriped) in the influent (red) and effluent (blue) water of the biofilter at various weeks (mean \pm s.e., $n = 3$ replicates).

Chapter 3, Section 3.3.2 (pg. 72) extensively discusses the removal of DOC by the slow-flow BAC filter as well as the processes driving this removal. After a rapid fall in adsorption efficiency to a lower level when the nonselective adsorption sites are occupied, adsorption actually continues, apparently unchanged, over an extended period of time, also known as tailing or slow adsorption (Kołodziej et al. 2014, Nguyen & Do 2000, Pohlman 1940). This slow-adsorption can last for months. While chapter 3 section 3.3.2 (pg. 72) talks about these removal processes of DOC, these processes are also applicable to the humic substances. Indeed, the influent water had a high humic substance concentration (Appendix C, Figure C.7, pg. 199) and as a result peak H1 and H2 can be used as a DOC proxy (Carstea et al. 2020). So, in the pore water the initial adsorption is visible by the low signal in all depths, while the filling of the smaller pores mostly takes place in the top of the filter and travels down creating the higher signals.

Fluorescence index (defined as the ratio of fluorescence intensity at 450 nm to fluorescence intensity at 500 nm emission wavelength, both at 370 nm excitation wavelength) has been used to indicate the degree of aromaticity (McKnight et al. 2001). The influent as well as the water in the top pores had a fluorescence index above 1 meaning that the presence of H2 > H1. Deeper in the filter as well

as the water in the bottom pores had a fluorescence index below 1 meaning that the presence of H2 < H1. H1 fluorescence material has a high aromaticity, while H2 fluorescence are less aromatic fractions (McKnight et al. 2001). In other words, the filter preferentially removed the more aromatic fractions, leaving a higher concentration of less aromatic fractions to travel down the filter and enter the effluent. These results confirm the results found in chapter 3, that the adsorption still played a significant role even up to week 23.

Despite the fact that microbial activity must play a role in the removal of the DOM components, because biomass was found at different depths of the biofilter (Chapter 3, Section 3.3.2, Figure 3.10, pg. 76) and bacteria require DOM as an energy source, the extent of this form of removal was unfortunately not further revealed with help of the pore water fluorescence signal.

The protein-like fluorophore, peak T, and nitrite had a signal below the LOQ and therefore their plots (Figure 5.13 A and B, respectively) are absolute values instead of normalized plot. At week 5 nitrite showed the first concentrations above the LOQ in the top three depths, while peak T was not present yet. From week 5 to week 9 the Peak T increased in all depths, with the highest concentrations in the top. From week 9 to week 23 the signal increased for depths 10-20 to 80-90, while 0-10 remained the same. The stratification at week 23 was significant higher in the top three depths compared to the bottom two depths (Dunn's posthoc on KW, $H(4) = 34.40$, $p_{adj} < .001$).

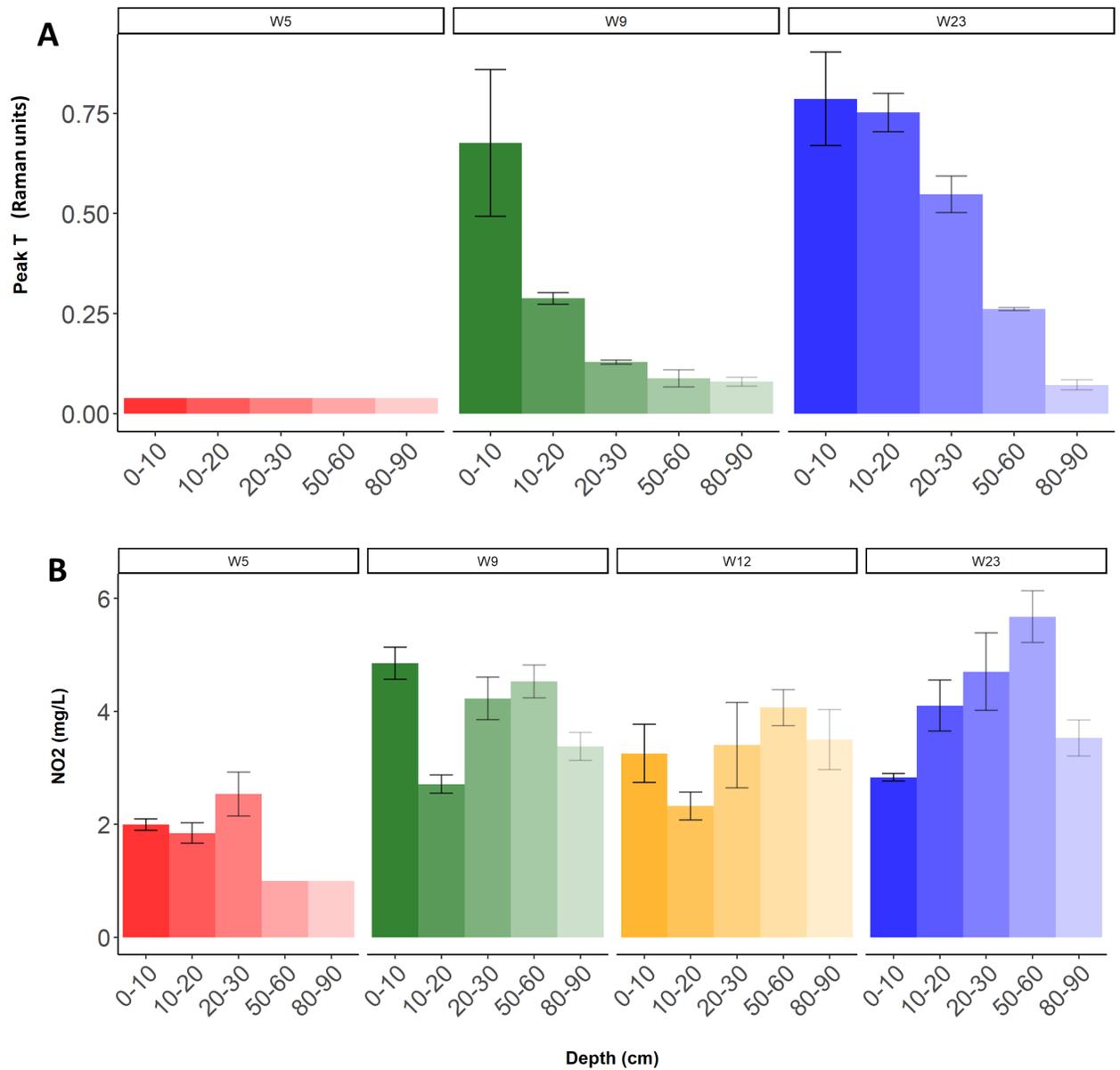


Figure 5.13: Fluorescence signal of peak T (A) and the nitrite concentration (B) at various depths (mean \pm s.e., $n = 3$ replicates) of the biofilter measured at various timepoints.

The nitrite concentration increased in various depths over time. At week 5 the top three depths showed nitrite concentration, while the deeper sections had no detectable concentration. At week 9 these deeper sections almost overtook the higher sections with respect to nitrite concentration. Also at week 23 the lower sections visually had a higher nitrite concentration than the top. The stratification appeared to have reversed in W9.

In Figure 5.13 Peak T and nitrite appear to have a different behaviour which was confirmed by their low correlation for filter length 90 cm over time and depth (Pearson correlation, $r(13) = .454$, $p = .089$). However they are linked to each other by the Tucker3 model (Figure 55). This is probably caused by the fact that both analytes had low concentrations at week 5 which increased over time. Overall, the group 1 variables are present in high concentrations through filter depth and over time with stratification noticeable through depth. Their change in concentration has most likely to do with the diffusion constantly happening through the length and impacted by influent concentration. However, this effect will decrease with decreasing concentration. Group 2 variables on the other hand are highly dependent on time. Additionally, for these variables, stratification plays a role in their concentration gradient. After 23 weeks the filter does not seem to be in equilibrium yet. Especially peak H1 and peak H2 clearly show that adsorption still is taking place and impacting the biofilter processes greatly.

One observed irregularity is unexpected lower concentrations of the anions and the higher concentration of H1 and H2 at week 23 for the 50-60 and 80-90 cm depths. The cause of this remains unclear. The standard error or the measurements are low, meaning that it is not caused by an outlier of a biological replicate. Possibly the system was not drained properly diluting the pore water with the short circuiting water that changed the chemical water composition of the bottom filter depth.

5.3.4 Pore water chemistry for different filter lengths

At the end of the experiment, the top 30 cm appear to behave more similarly for the group 1 variables, while the group 2 variables differences are still apparent. The question is if filter length (30, 60, or 90 cm) also has an influence on the behaviour of the chemical parameters in the top 30 cm of the biofilters. For example, pathogen removal monitored by DQ did not show a difference in filter length with removal processes happening in the top 30 cm. For this, a Tucker3 analysis was applied to the data set Length.

The length dataset was best described by a [2,2,2]-model, accounting for 82% variance. However, it did not give an appropriate core body diagonalization. Therefore the [1,2,2]-model was used with one component in mode A (depths), two components in mode B (the chemical variables), and two components in mode C (lengths) and explaining 75.72% of the variations.

With most of the variance covered by g111 (explaining 58% of information) and g122 (explaining 18% of information), these elements together describe almost the full G core, which simplified leads to a trilinear model with three terms, see Equation 5.4. Term 1 of Equation 5.4 (explaining 55% of information) has a negative core element, g111, while term 2 of Equation 5.4 (explaining 16% of information) also has a positive core element, g122. To which interactions the g111 and g122 are important is shown in Table 5.3.

$$x_{ijk} = a_{i1}b_{j1}c_{k1}g_{111} + a_{i1}b_{j2}c_{k2}g_{122} + e_{ijk} \quad (5.4)$$

Table 5.3: *Meaningful interactions for the elements g111 and g122 of the Tucker3 [1,2,2]-model of data set Length.*

g111 (negative)	A	B	C
+A1, +B1, -C1	0-10 cm	NO ₂ ⁻	L30
-A1, +B1, +A2	20-30 cm	NO ₂ ⁻	L90
g122 (negative)	A	B	C
+A, +B, +C	0-10 cm + 10-20 cm	Peak H1, H2, T, Na ⁺ , Cl ⁻ , NO ₃ ⁻ , SO ₄ ²⁻ , (NO ₂ ⁻)	All weeks

Figure 5.14 A1 presents the loading plots of one component of the depth's mode, while Figure 5.14 B1 and B2 present the loading plots of the two components of the chemical variable's mode, and Figure 5.14 C1 and C2 present the loading plots of the two components of the lengths mode.

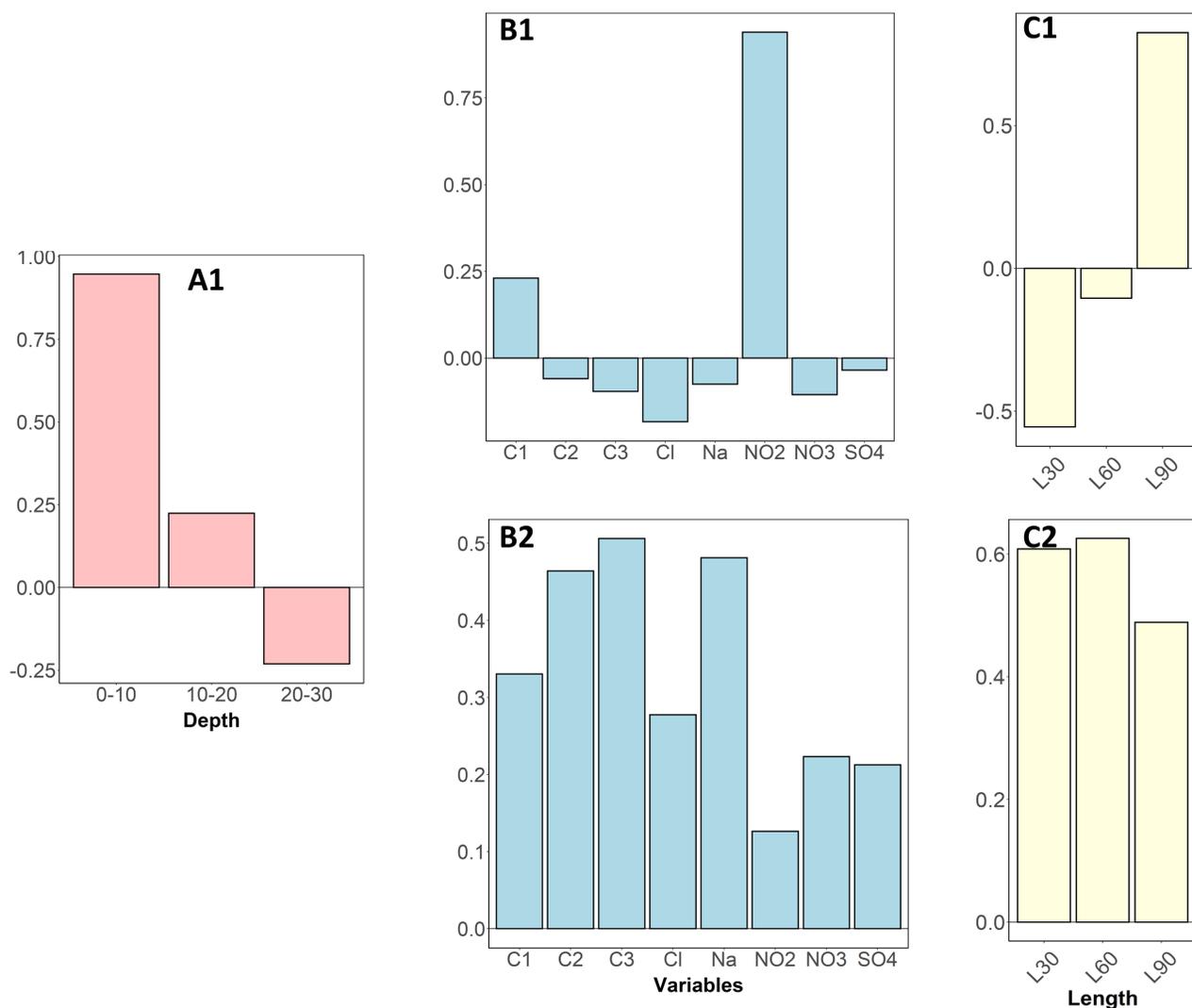


Figure 5.14: Tucker3 [1,2,2]-model loading plots from data set Length with depth loading plot of component 1 (**A1**), variables loading plot of component 1 (**B1**), variables loading plot of component 2 (**B2**), month loading plot of component 1 (**C1**), and month loading plot of component 2 (**C2**).

The two most meaningful interpretations are as follows: (1) negative g111: nitrite has the highest concentration in the top 0-10 cm in the 30 cm filter, while nitrite concentration in the 90 cm filter is higher in the 20-30 cm. (2) positive g122: all variables but mostly the Peak T, H1, H2 and sodium have the highest concentrations in the top of the filter in all filter lengths. From Figure 5.14 it appeared that most of the variables behaved similarly in all filters, with the exception of nitrite. For nitrite, the top 10 cm of the 30 cm filter has the highest concentration, while for the 90 cm filter the first 10 cm has the lower concentration.

It therefore seems that the three filters behaved similarly in the top 30 cm, apart from nitrite. The extra centimetres of the 60 and 90 cm therefore probably only extend the chemical processes happening in the top, but do not alter them.

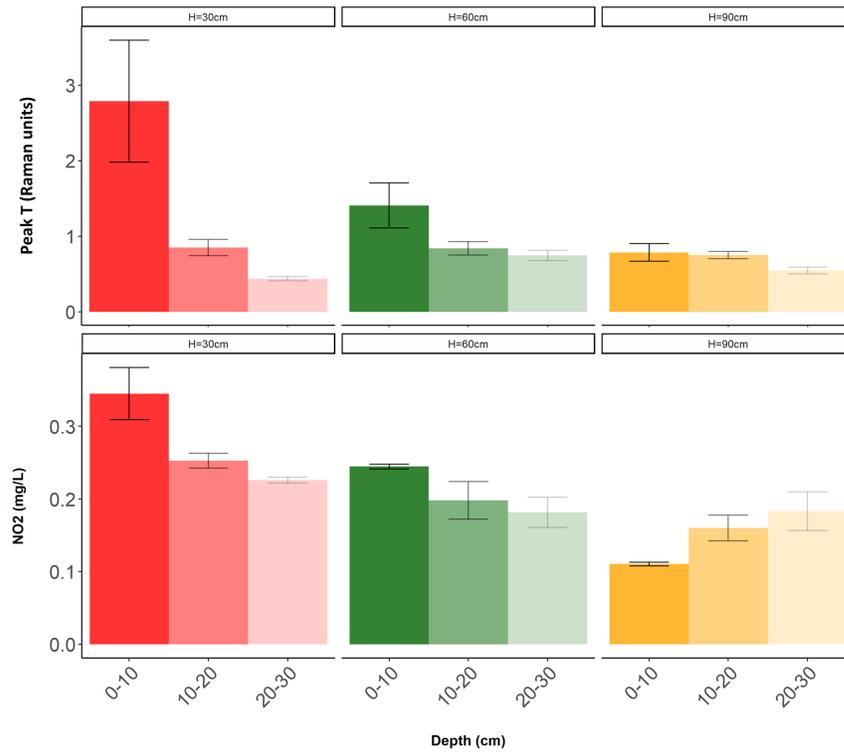


Figure 5.15: Concentrations of fluorophore peak T (A) and nitrite (B) at different depths (mean \pm s.e., $n = 3$ replicates) of the biofilter for various filter lengths.

5.3.5 Microbial activity or adsorption

Protein-like fluorescence signal has recently been ascribed to microbial activity (Fox et al. 2017). Other generally accepted methods for monitoring microbial activity are ATP and Quantitative Polymerase Chain Reaction (PCR) measurements, which were also done throughout the experiment by MV and DQ (Appendix C, Figure C.8 and C.9, pg. 200). The correlation among the three proxies is found in Figure 5.16 and demonstrates a good correlation. Of these techniques, fluorescence spectroscopy is the easiest to perform. Only pore water extraction is a time-consuming step, while the fluorescence scan is simple and easy. As the excitation/emission wavelength of peak T is known (280-350 nm) a quick first screening of microbial activity by analysing the fluorescence signal at 280-350 nm can be done without the use of PARAFAC scaling down the analysis process.

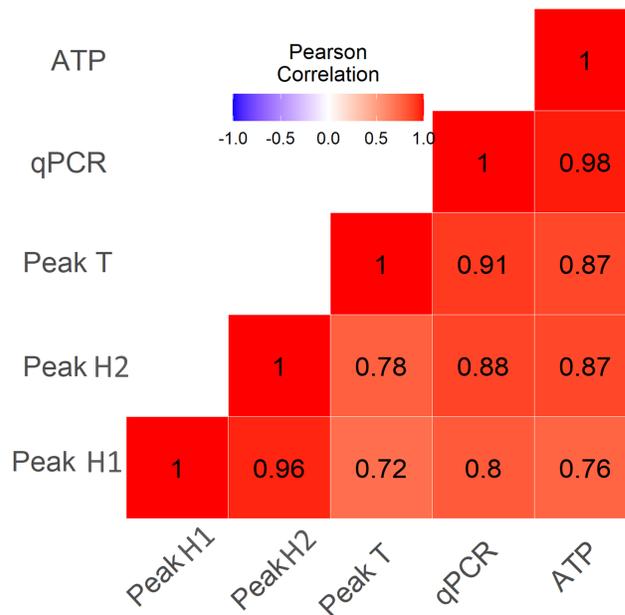


Figure 5.16: Significant correlations ($-0.6 > R > 0.6$) of fluorophores and other biomass proxies measured in the biofilter over time and depth for various lengths.

However, not all peak-T signal came from microbial activity. The total peak-T concentration included the peak-T present within the cells, Extracellular polymeric substances, or colloidal/particulate retained by the GAC filter, or a combination. Though, Chen et al. (2016) and Fox et al. (2017) showed that the majority of the protein-like fluorescence (> 75%) is intracellular in origin, as a result of the building of proteins for growth and metabolism. This suggests that most of the signal in the biofilter samples was caused by peak-T present within the cells. In other words, a clear presence of microbial activity was measured in the biofilter .

The highest concentrations were measured in the top of the filter and over time the concentration through filter depth homogenised. This stratification in microbial activity (or biomass) has been demonstrated by previous studies (Boon et al. 2011, Chen et al. 2016, Liao et al. 2013, Moll et al. 1999, Persson & Wedborg 2001, Urfer et al. 1997, Velten et al. 2011a, Wang et al. 1995, Zhang et al. 2016). The high microbial activity in the top was a direct result of presence of the highest DOM concentration in the top niches Velten et al. (2011a). Overall, the present study demonstrated that microbial activity was taking place and consequently removed DOM. However, the contribution of microbial DOM breakdown to the overall DOM removal remains unclear.

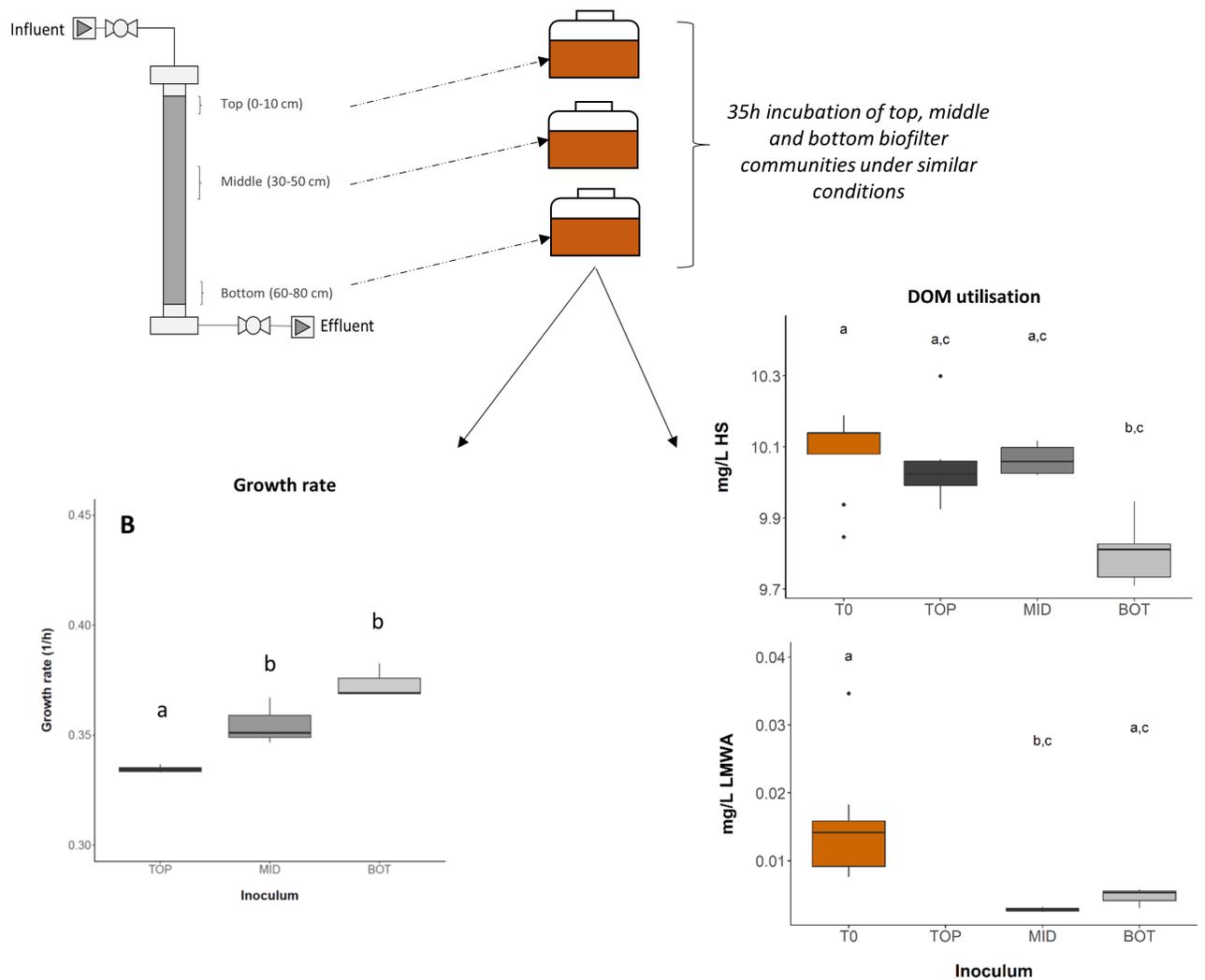
5.4 Conclusion

The objective of this chapter was to further understand removal processes of DOM and other chemical parameters by slow-flow BAC filtration using the analysis of pore water. Results from this study suggest that:

- The two-stage adsorption (1. attachment to pore walls and 2. migration into the micropores) as observed in chapter 3 was supported by the pore water analysis. Lower sections already started to adsorb molecules, while the top was not fully saturated yet.
- Filter length did not impact the processes happening in the top of the filter. It only extends these processes to deeper in the filter.
- Adsorption and diffusion processes influenced the pore water chemistry. As these processes play a large role up to the end of the experimental run, it remains unclear to what extent microbial activity contributed to DOM removal.
- This role of microbial communities at various depths of the filter on DOM removal is further investigated in chapter 6.
- As demonstrated by DQ and MV, a stratification of microbial activity is present throughout the filter with most activity taking place in the top of the filter and signal increasing over time.
- This stratification is similar to that of the peak-T fluorescence. Peak T can be used as a proxy for microbial activity or cell count.
- Microbial stratification is similar to that of the humic-like fluorescence. As humic-like fluorescence is a proxy for DOC, a nutrient stratification is present within the biofilter as described by Boon et al. (2011), which might play a role in the colonization of microbes throughout the filter.

Chapter 6

Differential utilisation of dissolved organic matter compound groups by different biofilter microbial communities



Contribution

This chapter is based on a submitted paper: ‘Differential utilisation of dissolved organic matter compound groups by different biofilter microbial communities’ written by Jeanine Lenselink^{1*}, Marta Vignola^{1*}, Dominic Quinn¹, Umer Zeeshan Ijaz¹, Ryan Pereira², Steve Joyce¹, William Sloan¹, Stephanie Connelly¹, Graeme Moore³, Caroline Gauchotte-Lindsay¹, Cindy Smith¹. (¹ James Watt School of Engineering, University of Glasgow, Glasgow, G12 8LT, United Kingdom; ² The Lyell Centre Heriot-Watt University, Edinburgh, EH14 4AP, United Kingdom; ³ Scottish Water, 6 Castle Drive, Dunfermline, KY11 8GG, United Kingdom; * Joint first author). The GAC used for this work was part of the biofiltration work done by MV, DQ, SJ, and myself. The batch experiment was done together with MV. Cell counts were done by MV, DNA analysis was done by DQ, and DOM sample preparation was done by me and sent to the Lyell Centre at Herriot Watt University where Ryan Pereira analysed the samples with LC-OCD. Data analysis with respect to cell count and DNA analysis were done by MV, while LC-OCD analysis was done by me. Introduction, materials & methods, and results & discussion were written by MV and myself, in which I focussed on the DOM and MV on the cell count microbial data.

6.1 Introduction

Diverse microbial communities have been found to populate biofilters and significant steps have been made to increase the understanding of their composition (Hu et al. 2020, Lautenschlager et al. 2014, Oh et al. 2018, Palomo et al. 2016) and their role in the treatment process (Bai et al. 2013, Pinto et al. 2012, Ye et al. 2001). However, the true challenge is still identifying the ecological mechanisms responsible for their assembly and how this might affect their functions and the quality of the final drinking water (Chen et al. 2021, Li et al. 2021).

Boon et al. (2011) suggested that the nutrient gradients in a GAC biofilter fed with ozonated-water contributed to the establishment of different environmental niches across the different depths of the filter bed, which selected for communities differing in species richness and evenness. Furthermore, they observed an increase in the normalised DOM removal efficiency with filter depths; with the communities in the lower filter sections being the most efficient. The authors suggested that the different functionalities shown by the communities at different sections could be the result of their different microbial richness, evenness and dynamics and that microbial community composition drives biological DOM removal efficiencies. The results of chapter 5 also demonstrated that the biofilter had a

nutrient gradient through filter depth which changed over time. For example, the highest concentration of DOM was present in the pore water of the top section of the biofilter. Microbial communities have been shown to degrade various DOM compound groups (Boon et al. 2011, Chen et al. 2016, Logue et al. 2016, Spray et al. 2021, Wu et al. 2018). Recently, several studies have tried to understand the impact of microbial community composition on the ability to degrade the DOM in water: particularly how this could impact the processing of the different fractions composing the DOM pool (Logue et al. 2016, Wu et al. 2018). Logue et al. (2016) were able to show a direct link between microbial communities' composition and their ability to degrade DOM in aquatic environments. They showed how the kinetics of DOM degradation and the DOM fractions preferentially utilized by the microorganisms in freshwater communities depended strictly on their compositions. Furthermore, DOM composition has been shown to be responsible for driving microbial succession and selection in aquatic communities. Wu et al. (2018) observed that microbial communities tend to use the more labile fractions (fast turnover by bacteria) in a DOM pool first; once these are depleted and the DOM pool enriched in recalcitrant carbon (slow turnover by bacteria), microbial communities undergo further selection and bacteria capable of processing a wide variety of carbon groups tend to become dominant.

As the water flows through the filter, the microbial assemblies and their functions at a given depth of the filter are intimately linked with processes happening above, analogously to the river continuum concept theory developed by Vannote et al. (1980), but in an engineered system. The communities at the top of the filters receive water rich in heterogeneous assemblages of labile and refractory DOM. These communities have the luxury of choosing which compound to use; the labile compounds are quickly used, as also observed by Wu et al. (2018), while the more refractory compounds are left in the stream for use by the communities at the bottom. Hence, the communities at the bottom of the filter are forced to adapt and survive on what is left, resulting in the enrichment of microorganisms capable of growing on more complex DOM compounds. This should result in a stratification of the composition of filter communities as well as in a stratification of their ability to process the different DOM fractions.

The objective of this chapter was therefore to examine the relationship between degradation of DOM and filter microbial communities at the different depths of the biofilter. For this, microbial communities from the top (TOP), middle (MID) and bottom (BOT) of the 12-week operational laboratory GAC biofilter were harvested and their growth rate and use of DOM compound groups on the same freshwater sample during 35h incubation compared. Moreover, their community composition was compared at the end of the incubation period.

6.2 Materials Methods

6.2.1 Glassware preparation

All glassware was prepared following the protocol of Hammes & Egli (2005) to reduce the AOC contamination. In short, all glassware and screw caps were washed with a common detergent and rinsed three times in Milli-Q water then submerged overnight in 0.2 M HCl and again rinsed three times with Milli-Q water. Borosilicate vials were subsequently heated in a Muffle oven to 550 °C for at least 6 h.

6.2.2 AOC-free *Inocula* preparation

For preparation of the three *inocula* (TOP, MID, BOT) the GAC from the three biofilters that were deconstructed in week 12 was used (see materials and methods of chapter 5 for the deconstruction of the biofilter). 500 mg of GAC was collected from the top 10 centimeters at 0-2, 2-4, 4-6, 6-8, 8-10 cm depths separately, and transferred to sterile 5 mL containers (TOP) for a different study; 2500 mg of GAC was collected in sterile 5 mL containers from 30-50 cm (MID) and 60-80 cm depths (BOT). 3 mL of filtered (0.22 µm filter) Milli-Q water was added to all the GAC samples and vortexed for 5 seconds to detach biofilm. A higher GAC to Milli-Q water ratio was chosen for the MID and BOT sections to take into account lower biomass concentration. 500 µL was withdrawn from each of the 5 extractions of 0 to 10 cm and combined, while 1 mL were withdrawn from the 30-50 and 60-80 samples. This procedure might enrich for certain microbial communities as they can proliferate more easily in the inoculum. However, using the GAC material with microbes still attach was not possible, the the remaining adsorption capacity would interfere with the DOM carbon pool.

The procedure was repeated for each of the three biofilters; extracts of the same depth from the three reactors were pooled together (3x 0-10 cm, 3x 30-50 cm, 3x 60-80 cm) in 15 mL sterile Polypropylene tubes stored in the dark at 4 °C and used within 4h to prepare three (TOP, MID, BOT) AOC free *inocula* as described in Hammes & Egli (2005). For this, 200 µL from each pooled extract was added to 50 mL filtered (0.1 µm, Millex-GP, Millipore) freshwater in 100 mL AOC-free borosilicate vials. Although pooling can affect the microbial composition of the three individual replicates, this step was necessary to harvest enough DNA material. Moreover, as the three depths were treated similarly, the effect of pooling was present in all three depths.

Vials were incubated without further amendments at 30 °C for 14 days. At the end of the 14 days

of incubation, the content of the vials was transferred to 50 mL sterile Polypropylene centrifuge tubes. The cells were harvested by centrifugation (10 min, 3000 g), resuspended in Milli-Q water amended with a mineral buffer as described by Lechevallier et al. (1993) and then poured into fresh 100 mL AOC-free borosilicate vials. Finally, the vials were incubated for a further 7 days to ensure that all residual organic was depleted. *Inocula* were kept in the dark and at 4 °C until their use.

6.2.3 Experimental set-up

The experiment was performed applying a batch culture approach in which the same filtered fresh water was inoculated independently with each of the three AOC-free *inocula* (TOP, MID and BOT); cell growth and associated degradation of DOM was measured as described below. For the tracking of DOM degradation fluorescence EEM spectroscopy was selected among other techniques. However, due to miscommunication this analysis was forgotten during the experiment.

6.2.4 Cell Abundance Specific growth rate

For cell abundance, 100 ml AOC-free glass vials, sealed with Teflon coated screw caps (Fisherbrand™ with PTFE liner material) were prepared in triplicate. Vials were filled with 60 mL of 0.1 µm filtered freshwater (0.1 µm, Merck™ Stericup™) to remove autochthonous microorganisms as well as particulate organic matter. Water was inoculated with either TOP, MID or BOT *inoculum* to a similar concentration of cells ($4.23 \times 10^3 \pm 1 \times 10^2$ cells/mL). Alongside the three experimental treatments, a negative control, filtered fresh water only with no added *inocula*, was incubated. All the vials were incubated at 30 °C, in the dark. Samples were collected from triplicates at 0, 10, 12, 14, 16, 18, 20, 22, 23, 36 h (negative control was measured at 0 h and 23 h only). At each timepoint, vials were opened, and 1 mL of sample was withdrawn using a sterile pipette; the samples were immediately fixed with 1 mL of 1% v/v Glutaraldehyde, stored at 4 °C in the dark and measured within 2 h by flow cytometry using a BD Accury™ C6 Plus flow cytometer equipped with a laser emitting at 488 nm (66 µl/min flowrate; 50 µl sample analysed). Samples (1 mL) were stained with 10 µl/mL SYBR Green I (1:100 dilution in Tris-EDTA buffer solution, pH 8.0) and incubated in the dark at 37 °C for at least 13 min before measurement (Hammes 2008, Vignola et al. 2018). The specific growth in each sample were determined as follows:

$$\mu = \frac{\ln(x_{T18}) - \ln(x_{T10})}{\delta t} \quad (6.1)$$

where x_{T18} and x_{T10} are the concentrations measured after 18 and 10 h of incubation and t the time interval between the two points.

6.2.5 DNA extraction and 16S rRNA gene sequencing

At the end of the 36 h incubation, replicates from each depth treatment (TOP, MID or BOT) were pooled, and cells filtered onto 0.2 µm membrane filters (Whatman) and frozen at -80 °C until DNA extraction. DNA was extracted with the FastDNA Spin Kit for soil kit (MP Biomedicals) following the manufacturer's protocol. Extracted DNA was quantified using the Qubit dsDNA HS Assay Kit (Life Technologies, Eugene, OR, United States) with a Qubit fluorometer (Invitrogen, Eugene, OR, United States). V3-V4 regions of 16S rRNA gene were amplified for sequencing using the Illumina MiSeq (at GENEWIZ, Inc, South Plainfield, NJ, USA). Sequences were submitted to NCBI.

6.2.6 16S rRNA gene Sequencing Analyses

Raw sequence data were analysed using QIIME2 pipeline and Deblur algorithm (Caporaso et al. 2010). The Deblur algorithm was used to facilitate better discrimination as it is an overlap free algorithm allowing only one set of reads, either forward or reverse reads. The reverse reads were of low quality and therefore discarded from the analyses. Additionally, while selecting trim length for Deblur, we have used https://www.qiita.ucsd.edu/static/doc/html/deblur_quality.html for their suggestions and as such reads were trimmed using the recommendations. The sequences were then trimmed above a Phred Quality score of 20 using qiime2. Amplicon Sequence Variants were produced using Deblur against SILVA v138 gene reference database. Qiime's align-to-tree-mafft-fasttree was then used to generate the rooted phylogenetic tree. The representative sequences were taxonomically classified using TaxAss workflow (<https://github.com/McMahonLab/TaxAss>) which uses an additional database formatted in the same format as SILVA v138 database to resolve ASVs that are not resolved by standard Naïve Bayesian Classifier using the standard database. The biom file for the ASVs was generated by combining the abundance table with taxonomy information using biom utility available in Qiime2 workflow. Results were visualised by a cluster heatmap with rows and columns ordered using hierarchical (average linkage) clustering.

Further analyses on the ASVs table were performed using R (version 4.0.2). Taxonomic alpha diversity indexes – Richness, Shannon and Fisher Alpha - were estimated after rarefying at the lowest sequencing depth of 53,948. These indexes are employed to describe the microbial biodiversity of the communities in terms of richness, as the number of different species present in the samples, and evenness as a distribution of species abundances. While Richness and Fisher Alpha indexes relate to

richness only, Shannon relates to both richness and evenness. The abundance table was then normalized using TSS + CLR (Total Sum Scaling followed by Centered Log Ratio) normalization at the Genus level as per recommendations by Rohart et al. (2017).

6.2.7 LC-OCD-UVD-OND sample preparation

DOM compound groups were analysed using liquid chromatography-organic carbon detection-ultraviolet detection-organic nitrogen detection (LC-OCD-UVD-OND; Huber et al. (2011)). For LC-OCD-UVD-OND analyses, 250 mL AOC-free serum bottles were prepared as described above, with the ratio medium/headspace kept the same as the 100 mL glass vials. The amount of *inoculum* added was changed to have the same initial concentration of cells. Bottles were capped with metal crimp and silicone septa. All the vials were incubated at 30 °C, in the dark. At T₀, 50 mL medium were poured into a 60 mL carbon-free glass vial; at T₂₃, incubated bottles were decapped, and 50 mL sample were poured into a 60 mL carbon-free glass vial. For the two time points, the samples were stored at room temperature in the dark, transported to the Lyell Centre, Edinburgh and measured (three technical replicates per sample) within 24 h. The LC-OCD-UVD-OND separates DOM without prior modification by injecting 1 ml of water onto a size exclusion column (2 ml/min; HW50S, Tosoh, Japan) with a phosphate buffer (potassium dihydrogen phosphate 1.2 g/L plus 2 g/L di-sodium hydrogen phosphate x 2 H₂O, pH 6.58) to identify five different DOM classes (see Table 6.1). All DOM classes were quantified as well as the molecularity (nominal molecular weight) of the humic substance class given (Huber et al. 2011).

The limit of detection of the OND and OCD were calculated as three times the standard deviation of the mean area of the noise for six blank injections (Milli-Q water), and the values were converted to concentration units using calibration curves. The LOQ was calculated as 10 times the standard deviation of the mean area of the noise for six blank injections. The limit of detection for OCD and OND were 82 ppb/C and 53 ppb/N, respectively. The corresponding LOQ was 273 ppb/C and 176 ppb/N, respectively. The reproducibility of the OCD and OND was tested by injecting 0.4 µg of International Humic Substances Society Humic Acid Standard (HS₃S₁₀H) in triplicate. For the OCD, the RSDs were lower than 2.2%. For the OND, the RSDs of Humic Substances was lower than 1.6%.

Table 6.1: LC-OCD-OND DOM fractions and corresponding molecular weights and description.

DOM class	Mw (g/mol)	Description	Reference
Biopolymers (BP)	>10,000	Hydrophobic, largely non-UV-absorbing extracellular polymers with saturated structures, polysaccharides and some contribution of proteins or amino sugars	Huber (2011)
Humic substances (HS)	1000	UV-absorbing aromatic molecular aggregates of relatively small molecules, stabilized by the hydrophobic effect and hydrogen bonds	Gerke (2018)
Building blocks (BB)	300–500	UV-absorbing humics of lower-molecular-weight that have been shown to include microbial breakdown products of humic substances	Huber et al. (2011) Velten et al. (2011)
Low molecular weight neutrals (LMW neutrals)	350	non-UV-absorbing, weakly or uncharged hydrophilic or amphophilic compounds that can include alcohols, aldehydes, ketones, and amino acids hydrophilic	Huber (2011)
Low molecular weight acids (LMW acids)	350		Huber et al. (2011)

6.2.8 Statistical analysis

All statistical analyses were carried out using R (version 4.0.2). For cells abundances and growth rates significant differences were ascertained by ANOVA with data being normal and homogeneously distributed, followed by the Tukey post-hoc test (equal variances, and equal group sizes) to discriminate between treatments with a 95% confidence level. For the LC-OCD-UVD-OND compound groups significant differences were ascertained by a KW is H-test as the data were not normally and homogeneously distributed using an alpha significance level of 0.05, followed by the Dunn's post-hoc test (equal variances, unequal group sizes) to examine the pairwise comparison with p-values adjusted by the Bonferroni method. For both methods, the results were expressed by the mean for each group together with its variance (expressed by the standard error).

6.3 Results and discussion

6.3.1 Increase in cell abundance

The results showed that the microbial communities enriched from three different depths of a biofilter (extracted and transformed into three *inocula*: TOP, MID, BOT) varied in their cell abundance and growth rate. Over the course of the experiment, cell abundances increased in all three treatments (TOP, MID and BOT, Figure 6.1 A). Cell concentrations increased from an average of $4.35 \cdot 10^3 \pm 2 \cdot 10^2$ cells/mL at T_0 for all three *inocula* ($n = 9$) to $1.83 \cdot 10^6 \pm 9 \cdot 10^3$ ($n = 2$ due to a problem with FCM analyses on one replicate); $2.06 \cdot 10^6 \pm 1 \cdot 10^4$ ($n = 3$); $2.15 \cdot 10^6 \pm 7 \cdot 10^3$ ($n = 3$) cells/mL for TOP, MID and BOT inoculum respectively after 36h of incubation. The total amount of cells harvested differed significantly between BOT and the other two treatments after only 10 h of incubation (BOT>TOP-MID, Tukey's post-hoc test on ANOVA; $p_{\text{adj}} < 0.05$); while after 20 h all the treatments differed between each other (BOT>MID>TOP, Tukey's post-hoc test on ANOVA; $p_{\text{adj}} < 0.05$ per each pairwise comparison).

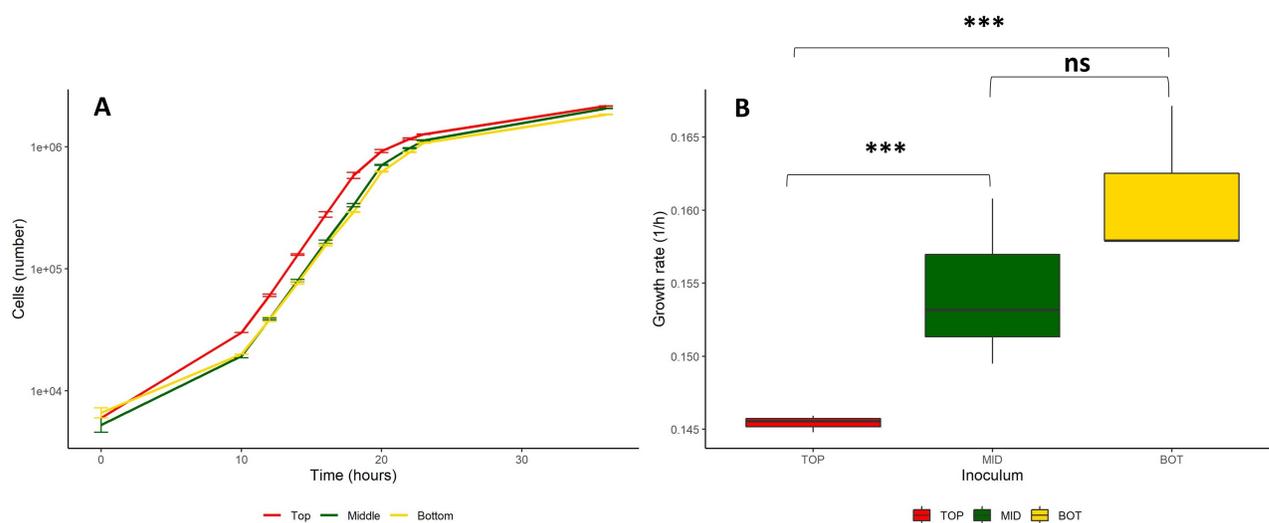


Figure 6.1: Trends of cell abundances (A) and growth-rates (B). Abundances were measured over the course of the experiment for the three treatments (mean \pm s.e., $n = 3$ replicates). Stars and ns in B denote significant differences with regard to growth rates between treatments assessed by Tukey's post-hoc test on ANOVA. ns denote no significant difference, Stars denote significant differences between treatments ($p < 0.05$).

While they were exposed to the same DOM composition and incubated under the same controlled conditions, the MID and BOT communities showed faster growth rates than the TOP and consequently, with a higher abundance of cells at the end of the experiment. (Tukey's post-hoc test on ANOVA; $p_{\text{adj}} < 0.05$) (Figure 6.1 B). Although a difference between MID and BOT was detectable (Figure 6.1 B), it was not statistically significant ($p > 0.05$). This outcome is highly likely a result of high within-group variation of MID treatment presented by the wider error of the boxplot.

Assuming an average content of organic carbon per cells equal to 1.5×10^{-13} g of C/cell (as suggested by Pick et al. (2019), Vrede et al. (2002)), the total net growth of cells at the end of the 23 h incubation accounted for the 1.31% and 1.36% of the total initial DOC content in the TOP and MID treatment, respectively, and for the 1.53% of the BOT treatment. Therefore, we expected to see a similar percentage decrease in the DOC concentration values; however, we did not observe a decrease in DOC concentration for any of the treatments. The expected change in DOC concentration might have fallen within the instrument's error.

6.3.2 Microbial community structure

A total of 180,300 high-quality reads were obtained through sequencing (TOP = 53,947, MID = 71,954, BOT = 54,398), and 1,127 ASVs were detected in this study. Of these, 166 were shared between the three communities; 305 were only found in the TOP, 199 only in the MID and 190 only present in the BOT. The remaining 267 ASVs were either shared between TOP-MID (81), or TOP-BOT (58) or MID-BOT (128). In terms of alpha biodiversity, the TOP community had the highest observed Richness, Shannon and Fisher Alpha indexes compared to BOT and MID communities (Table 6.2). In other words, the TOP community had the greatest diversity of species which were the most evenly distributed. The genera *Ralstonia* and *Pandoraea* were the most abundant in all the three communities accounting for 62% and 15% of the TOP community; 66% and 16% of the MID community and 62% and 16% of the BOT community, respectively. The lineage *betI-B* of *Burkholderiales* and the genus *Curvibacter* were one to two orders of magnitude more abundant in the BOT and MID communities (11% and 1% for BOT; 7% and 0.6% for MID) than in the TOP one (0.08% and 0.01%). The Genus *Sphingobium* and the lineage *alfIV alfIV-B* of *Sphingomonadales* were highly abundant in the TOP (5%, 8% respectively), but almost absent in the other two treatments (0.0% and 0.002% for BOT; 0.003% and 0.0% for MID) (Figure 6.2).

Table 6.2: Alpha diversity parameters for the pooled triplicates Sample Number of High quality reads Richness Fisher Alpha Shannon.

Sample	Number of High quality reads	Richness	Fisher Alpha	Shannon
TOP	53,948	610	96	4.13
MID	71,954	547	85	3.99
BOTTOM	54,398	541	84	4.10

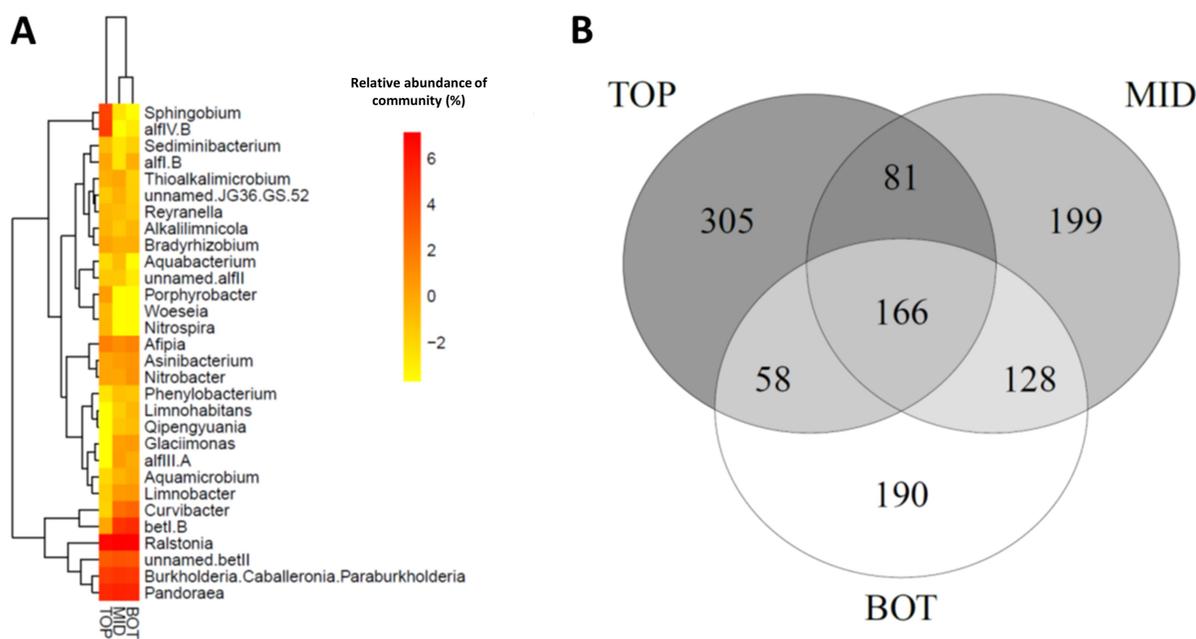


Figure 6.2: Heatmap of the communities in the three treatments at genus level with both rows and columns ordered using hierarchical (average linkage) clustering (A). Venn diagram of the ASVs found in the treatments (B).

6.3.3 DOC concentration

Measured LC-OCD fraction concentrations can be found in Appendix D (pg. 203). DOC concentration before inoculation was 12.19 ± 0.03 mg-C/L; after 23 h of incubation, 12.35 ± 0.04 , 12.34 ± 0.10 , 12.18 ± 0.06 mgC/L were measured in the TOP, MID and BOT treatments respectively. Only the TOP DOC was statistically different from T_0 (Dunn's post-hoc test on KW H-test; $H(3) = 14.50$, $p_{adj} = 0.014$), but no significant differences between final DOC treatments were observed (Figure 6.3). Again, within group variation was noticeable, especially for the TOP and BOT treatment.

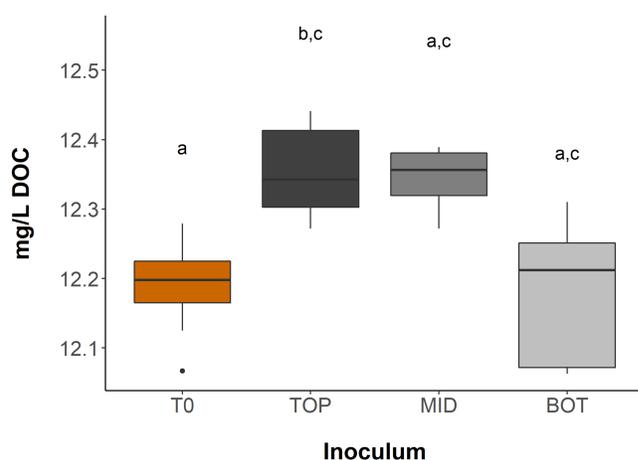


Figure 6.3: Boxplot of the DOC concentration at $T = 0$ and the three treatments (TOP, MID, BOT) after 23h incubation. Letters a, b and c denote significant differences between treatments regarding DOC concentration. Letters (a,b,c) denote significant differences with regard to DOC concentration in the different treatments assessed by Dunn's post-hoc test on KW H-test. Same letters denote no significant difference, different letters denote significant differences between treatments.

6.3.4 DOM degradation

The initial DOC was mainly composed of humic substances ($83 \pm 0.4\%$), but the compound groups building blocks ($8.0 \pm 0.2\%$), LMW neutrals ($5.7 \pm 0.2\%$), LMW acids ($0.1 \pm 0.02\%$) and biopolymers ($3.3 \pm 0.7\%$) were also found. This DOM composition changed slightly in all the samples during the incubation period. While no significant differences were observed among the three treatments at the end of the experiment regarding the different DOM fractions concentrations, significant differences were observed between T_0 and the three treatments after 23h incubation (Dunn's post-hoc test on KW H-test; $p_{\text{adj}} < 0.05$). Although no reduction in humic substance concentration was observed in TOP and MID from T_0 to T_{23} , BOT exhibited a significant decrease in humic substance content compared to T_0 and significant increase in both building block content (Figure 6.4 A-B). Along with a decrease of humic substances, in BOT treatment, we also observed an increase in its molecularity (Figure 6.4 C). The TOP and MID treatment demonstrated no change in the molecularity. Molecularity refers to the nominal molecular weight. In other words, the BOT treatment increased the molecular weight of the humic substance pool. The detailed composition of the humic substance structure is unknown. However, it has been suggested that they are composed of LMW compounds weakly attached to a larger refractory core (Finkbeiner et al. 2020, Gerke 2018). An increase in molecular weight of this carbon group can be the result of the removal of small or the addition of high molecular weight humic substance structures. In this experiment, the BOT community showed microbial degradation of the

humic substance fraction as suggested by the significant decrease in its concentration. It therefore seems most plausible that only the small humic substance molecules were degraded by microorganisms leaving the bigger structures in the DOM pool and increasing the molecularity of the DOM pool.

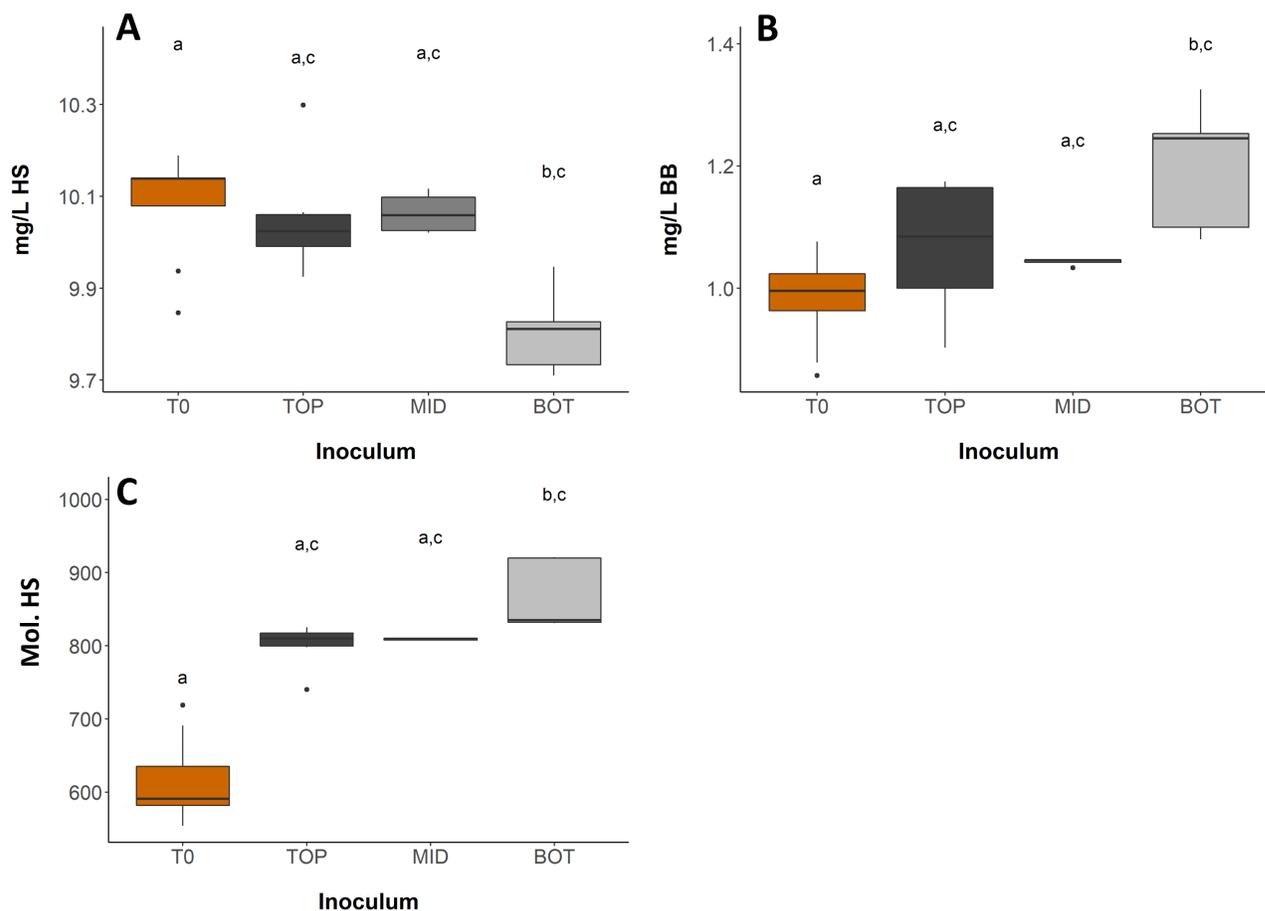


Figure 6.4: Boxplots of humic substance concentration (A), building block concentration (B), and molecularity of the humic substance pool (C) at $T = 0$ and in the three treatments (TOP, MID, BOT) at $T = 23$. Letters (a,b,c) denote significant differences in concentration humic substances (A) and building blocks (B), and molecularity of humic substance pool (C) between the different treatments assessed by Dunn's post-hoc test on KW H-test. Same letters denote no significant difference, different letters denote significant differences between treatments.

Throughout incubation, the LMW neutral carbon fraction did not change (Figure 6.5 A). The LMW acid concentration on the other hand decreased in the TOP and MID treatment. However, the difference between T_0 and T_{23} was only statistically significant for MID (Dunn's post-hoc test on KW H-test; $H(3) = 8.79$, $p_{\text{adj}} = 0.045$); the concentration of LMW acids in TOP fell below the limit of

detection (Figure 6.5 B). Finally, the biopolymers concentration increased significantly only in MID at the end of the experiment (Dunn's post-hoc test on KW H-test; $H(3) = 11.20$, $p_{\text{adj}} = 0.020$) (Figure 6.5 C). T_0 showed a larger variance, while TOP had a significant outlier.

While the present study reports no removal of the LMW neutrals (Figure 6.5 A) and building blocks (Figure 6.5 B), these fractions are generally considered microbially labile and their biological removal has been reported in several studies (Boon et al. 2011, Krzeminski et al. 2019). However, these studies used ozonated waters, in which DOM is oxidised and broken down into smaller (more bioavailable) components. In the study, we used non-ozonated surface water where these compounds are presumed to be more refractory to biological degradation (Amon & Benner 1996, Chen et al. 2016, Vasyukova et al. 2014). The refractory nature of building blocks as well as the fact that building blocks are the product of humic substance breakdown (Huber et al. 2011), caused the building block concentration to increase in the BOT treatment during incubation.

The study also reports no removal of biopolymers (Figure 6.5 C), while these fractions are generally considered easily available for microbial degradation (Chen et al. 2016, Vasyukova et al. 2014). On the contrary, here we observed an increase in the biopolymers concentrations during incubation in all treatments, with only the MID treatment being statistically significant (due to high variances). Biopolymers could have been formed by the microbial cells during their growth. Lautenschlager et al. (2014) also observed a slight increase of the biopolymers concentrations in their reservoir with treated drinking water, where the presence of cells was also reported.

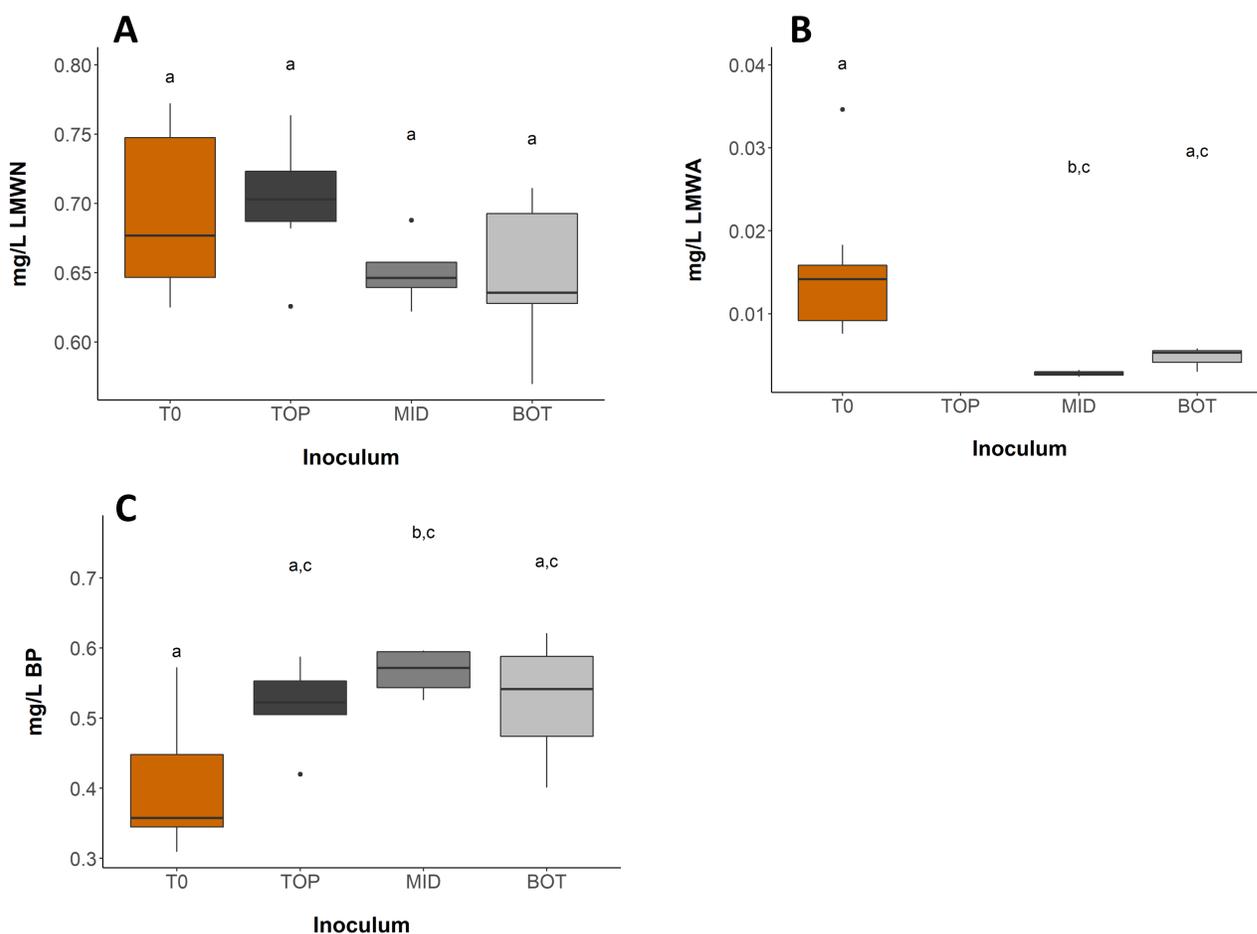


Figure 6.5: Boxplots of the LMW neutrals (A), LMW acids (B), and biopolymers (C) at $T = 0$ and (TOP, MID, BOT) at $T = 23h$. Letters a, b and c denote significant differences between treatments regarding concentrations. Letters (a,b,c) denote significant differences in concentration LMW neutrals (A) and acids (B), and biopolymers (C) between the different treatments assessed by Dunn's post-hoc test on KW H-test. Same letters denote no significant difference, different letters denote significant differences between treatments.

6.3.5 Differences

The results showed that the three microbial communities differed in the type of DOM compound group utilized during the incubation process. This could indicate that the change in DOM distribution at the end of the incubation reflects the utilisation of DOM fractions (as identified by the LC-OCD-UVD-OND) by the three different communities. The TOP and MID treatments removed compounds belonging to the fraction of DOM with a LMW acids (Figure 6.5 B). The BOT community, on the other hand, used the complex and the most abundant fraction of humic substances (Figure 6.4 A). It suggests that only BOT was able to degrade the widely abundant humic substance fractions, which

could explain the higher growth rate observed in the BOT community, as compared to TOP and MID.

The metabolism of the different DOM fractions by heterotrophic bacteria is a complex and controversial subject (Amon & Benner 1996). The traditional view is that small molecules present in the DOM pool (up to 600 Da) are also the most labile (Weiss et al. 1991), since they can be readily taken up by microorganisms across their cell membrane, and they are the C-source preferentially chosen by microorganisms (Münster & Chróst 1990). Bigger and more complex molecules (greater than 600 Da) would require extracellular enzymes for the conversion into smaller and more easily degradable compounds; therefore, they are considered more recalcitrant to microbial activity. Although biologically more recalcitrant than LMW acids and neutrals, several studies have shown that a portion of humic substances is biologically degradable; this has been observed both in natural (Moran & Hodson 1990, Volk et al. 1997) and engineered environments (Boon et al. 2011, Lautenschlager et al. 2014). The humic substance fraction was the most abundant DOM fraction in the experiment accounting for 82.7% of the total dissolved organic carbon present in water. The high concentration of such compounds in freshwater compensates for their slower rates of utilisation (Volk et al. 1997).

Apart from the difference in DOM utilization, the three treatment communities showed different compositions, with TOP sharing the lowest similarity with the other two. The structure of a microbial community impacts its ability to degrade the different DOM fractions (Logue et al. 2016). While the degradation of LMW DOM is believed to be a functional trait common to many microbial communities, the ability to utilize more recalcitrant DOM (such as humic substances and biopolymers with greater molecular weight) is considered a trait less widely distributed (Logue et al. 2016). Therefore it seems that the different structure of the three communities has influenced their ability to degrade the DOM pool. The BOT and MID communities possess the rarer ability to degrade complex organic carbon molecules. The genus *Curvibacter*, which was significantly more abundant in the BOT and MID communities, has been thought to play an important role in the DOM degradation of freshwater since it seems to possess the ability to utilize specific types of carbon source (Wu et al. 2018). Hence, some of the taxa in the *Curvibacter* genus seem to play an important role in the degradation of humic substances and fractions. However, only the BOT community actually demonstrated this ability by significantly lowering the humic substance concentration.

6.4 Conclusion

The objective of this study was to explore the ability of microbial communities extracted from different depths of the 12 week operating laboratory scale biofilter to degrade DOM. Understanding this is essential to improve biofilter performances and DOM removal in drinking water treatments. Results from this study suggest that:

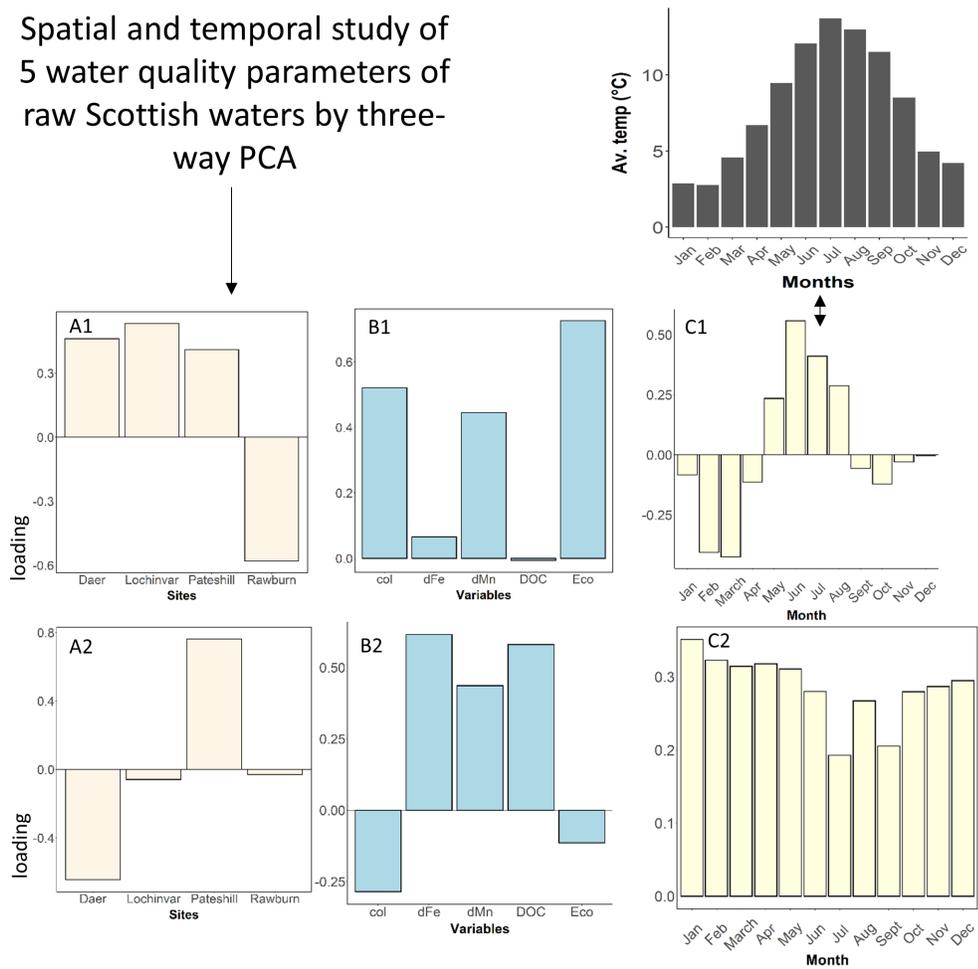
- The differences observed in the three communities' ability to degrade DOM, as well as their composition studied after 35h of incubation, are the consequence of the initial compositional difference of the biofilter communities from which they were derived.
- Such differences are a consequence of the adaptation of the original biofilter communities to the different environmental niches developing at different depths of the biofilter. Lower depths highly likely select for communities capable of degrading more complex DOM compounds.
- Whether this results from their continued exposure to such DOM compound groups during the biofilter operation needs further investigation.

- Looking at the result of Chapter 3, 5 and 6 together the processes within the filters can be better understood. DOM removal is mostly a result of multistage adsorption (initial adsorption in macropores and continued tailing in the micropores). The top of the filter becomes saturated quicker compared to the bottom. This stratification of carbon source results in different ecological niches. The stratification of DOM results not only in a stratification of biomass, but also in a difference in community structure through filter depth. The highest species diversity is present in the top niche, probably as a result of a wider variety of DOM. The bottom section has a lower diversity of microbes, highly likely as a result of DOM restriction. However, as a result of this DOM restriction, only specific microbes are able to inhabit this niche and able to survive on the limited choice of DOM species. Using fresh Scottish water with high concentrations of not-easily degradable humic substances, the communities able to survive in the bottom niche are able to degrade these recalcitrant DOM molecules. This insight highlights the benefits of using long biofilters. Not only does the long biofilter remove DOM to a greater extent for a longer period of time, it also provides the opportunity for humic substances degrading microbes to inhabit the filter.

Chapter 7

Three-way PCA to study spatial and temporal variation in water quality parameters and its impact on biofilter performance

Spatial and temporal study of 5 water quality parameters of raw Scottish waters by three-way PCA



High biology in summer as result of temperature
 Biofilter influent: high DOC, dFe & dMn → better performance for other fresh Scottish water.

7.1 Introduction

Natural water quality varies from place to place, with the seasons, with climate, and with the types of soils and rocks through which water moves (Aitkenhead-Peterson et al. 2007, Liu et al. 2014, Muller & Tankéré-Muller 2012, USGS 2001).

In biofiltration, water quality not only impacts the performance of the biofilter (Basu et al. 2016, Lauderdale et al. 2012, Pharand et al. 2015), but it also influences the composition of bacterial communities within the biofilter (Ma et al. 2020). The results of chapters 3, 5, and 6 showed that the studied biofilters removed DOM and consequently complexed iron via a dual-stage adsorption in combination with some degree of biodegradation. At the end of the experiment the GAC material was not exhausted yet and a significant stratification of DOM was visible within the pores of the filter, highly likely resulting in a stratification of microbial communities. In this apparent steady state the iron concentration in the effluent was above the regulatory limits. Moreover, the DOM concentration was also high, resulting in coloured water, probably in combination with smell, taste and making it a possible carbon source for microbial growth. It must be noted that the concentration of DOC and iron in influent water was relatively high. It was sampled at one particular water treatment work location having its own geochemical and microbial signature. How this signature compares to other water sources and seasonal fluctuation is unknown. Water sources of other regions could have a higher geochemical loading, which can negatively impact the performance of the biofilters even more. Also seasonal variation can alter the geochemical signature of the water and thereby influence the performance of the biofilter performance, the chemical signature within the pore water, and the stratification of microbial communities. Therefore, understanding this spatial and temporal variation of biochemical water signature can help to understand how results of Chapter 3, 5, and 6 will be impacted when using the waters from other regions in Scotland throughout the year.

The objective of this chapter was to apply chemometric methods to study spatial and temporal variation in water quality parameters and evaluate how this might influence the performance of slow-flow BAC filter. First, a multivariate biochemical data set of five water sources in Scotland monitored from January 2015 till December 2019 was analysed by three-way PCA to understand spatial and temporal variation of the biochemical water quality. Second, the biofilter influent water was compared with these various Scottish water sources by normal PCA. In this way, the present study systematically assesses the slow-flow BAC performance in relation to water source and seasonality.

7.2 Materials and methods

7.2.1 Sampling sites

Data from various WTWs from across Scotland ran by Scottish Water were used in this study. Table 7.1 shows coordinates, sampling location, water type and vegetation surrounding the reservoirs/catchment area. The Biofilter Influent (IN) water used for the experiments of chapter 3, 5, and 6 was taken at Pateshill (Table 7.1) and prefiltered with a 10 μm polypropylene cartridge filter.

7.2.2 Sampling period

The biochemical analysis of the various water treatment work water samples was done at the Scottish Water laboratory and spanned the period January 2015 until December 2019. IN water samples were analysed in-house once a week in the period October 2018 and December 2018, and once every two weeks from January 2019 until March 2019.

7.2.3 Variables

Different variables were measured in-house for the IN samples and in the Scottish Water lab for the water treatment work samples (Table 7.1). The methods used for the in-house analysis of the variables are given in the materials and methods of Chapter 3, Section 3.2 (pg. 53).

7.2.4 Data pre-treatment

WTW data required pre-treatment to get the required data format. A full overview of all pre-treatment steps including the number of samples remaining per step are given in Appendix E, Table E.1 (pg. 205). The most important steps were as follows: (1) Data entries from January 2015 until December 2019 were extracted. (2) "Greater than" symbols were removed from its value, as further data processing did not allow for "greater than symbols". (3) Values with a smaller than symbols were replaced by a random number between zero and value itself. (4) Measurements from the same day were considered technical replicates and the mean was taken. (5) Variables DOC, dissolved iron (dFe), dissolved manganese (dMn), *E. coli* and total *coliform* were selected, because WTW and IN had these variables in common. (6) Measurements of the same month over several years were averaged by taking the median to compensate for unmeasured entries. The median was chosen to minimise the impact of outliers. (7) Daer, Lochinvar, Rawburn and Pateshill had only <10% unmeasured entries, while this was 40% for Forehill, Rosebery and Penwhapple. Only water sources with <10% unmeasured

Table 7.1: Description of the Scottish sampling sites with the variables analysed over a time period.

Water source	Coordinates ^a	Type of water	Surrounded vegetation ^b	Sampling period	Analysed variables	Analysed by
Daer	55.35, -3.62	Loch	Grassland > peatland > coniferous plants	Jan '15-Dec '19	TOC, DOC, manganese, iron, dissolved manganese, dissolved iron, total <i>coliform</i> , <i>E. coli</i> , Alkalinity, UV absorbance, UV transmittance,	Scottish water
Lochinvar	55.14, -4.11	Loch	coniferous plants >peatland		UV transmittance, colony count, total cell count, Enterococci, clostridium, SUVA, pH	191
Pateshill	55.82, -3.62					
Rawburn	55.80, -2.51	Loch	Heather moor			
Biofilter influent (IN)	55.82, -3.62	Prefiltered Pateshill		Oct '18-March '19	TOC, DOC, nitrate, nitrite, ammonium, phosphate, various dissolved metals, <i>E. coli, coliform</i>	In-house

^a Latitude and Longitude coordinates.

^b Vegetation is based on The Landscape of Scotland 1988 (land use research institute 1988).

variables were kept, being: Dear, Lochinvar, Pateshill, Rawburn. The data set included four sample sites (Dear, Lochinvar, Pateshill, Rawburn), 5 variables (DOC, dFe, dMn, *E. coli* and total *coliform*) and 12 monthly measurements (January – December). This data set (data set Waterquality) helped to understand the spatial and temporal variation of WTWs.

7.2.5 Substituting missing data

Missing data as a result of measurement absence needed to be replaced before three-way PCA could be applied. In normal PCA, samples with missing parameters can be excluded. In three-way PCA, however, a full data array is required. The data set missed the DOC, dFe, and dMn concentration of Pateshill in February, despite the averaging between 2015 and 2019. The missing values as a result of measurement absence were substituted by applying nonparametric multivariate imputation on the Pateshill subset using the *zCompositions* package for R (version 4.0.2) (Palarea-Albaladejo & Martín-fernández 2015).

7.2.6 Normality distribution and outliers

The normal distribution of each variable was tested. Shapiro–Wilk test was applied to check the distribution pattern of the variables. Almost none of the variables demonstrated a normal distribution, which is a common phenomenon in environmental data (Leardi et al. 2000). This was confirmed by frequency histograms demonstrating skewness. A logarithmic transformation was therefore applied to the variables to transform the data into a (pseudo-) normal distribution. Zero values (as a result of counting *E. coli* and total *coliform*) were substituted with 0.1 before the log transform was applied. Although the use of log transformation is debated for counting data (Hara & Kotze 2010, Ives 2015), normality distribution after log transformation did not show severe outliers after log transforming these variables. Outliers were evaluated by the Hotelling T^2 – Q residual plot generated from an initial PCA model (Bro & Amigo 2020, Goueguel 2020) as described in Chapter 5, Section 5.2 (pg. 45). Outliers were inspected, but none were found based on their cut-off values (Appendix E, Figure E.1, pg. 206).

7.2.7 Data pre-processing

The data sets required pre-processing which homogenises scales and units without altering the differences among the sample sites and among the sampling times. For PCA, this problem is solved by autoscaling in the variable (column) direction. For three-way PCA, this problem is solved by performing a *j*-scaling. The three-way array *X* (with *I*, *J*, and *K* modes) was matricized to a two-way matrix

Xb having I x K modes in the row direction and the J mode in the column direction. On this matrix autoscaling was performed in the variable (column) direction. As a result, the global variance of each variable was set to one, and the differences among the objects and the conditions are preserved. J-scaling calculates averages over two modes.

7.2.8 PCA

Traditional PCA was applied on the data set using the MixOmics toolbox ® (Rohart et al. 2017). For this, the data array was transformed to a matrix. This operation was carried out by unfolding the data array. The measured variables were kept in the columns and the sampling sites measured at various months were placed in the rows, resulting in a matrix of 60 samples (4 sites × 12 months) × 5 variables.

7.2.9 Tucker3

Secondly, three-way PCA, more precise Tucker3, was applied to data set. A detailed explanation of the Tucker3 concept is given in Chapter 2, Section 2.4.1 (pg. 45). For the Tucker3 analysis the ThreeWay toolbox was used in R (version 4.0.2) (Giordani et al. 2015). The T3 code provided an interactive Tucker3 analysis with the following main steps: (1) provide the number of entities in the P, Q, and R mode. (2) determine the type of centring and normalization (already done, so not executed). (3) PCA's of super matrices with slices of the 3way array next to each other, thus 3 super matrices are analysed by PCA, and for each component matrix and eigenvalues is found. The results are used to create a generalized scree plot, which can be used to find the number of components for each mode. (4) specify convergence criterion is set to $1 \cdot 10^{-6}$ by default. The statistical validity of the obtained component matrices was assessed by a bootstrap procedure for computing the confidence intervals for the solution (Giordani et al. 2015). The analysis is based on 500 bootstrap samples and matching via optimal transformation towards full solutions.

7.2.10 Other statistical analyses

Further statistical analyses were carried out using R (version 4.0.2). The correlations were determined by Pearson correlation using an alpha significance level of 0.05.

7.3 Results and Discussion

7.3.1 Spatial and temporal variation within Scotland

Spatial and temporal variation of biochemical water quality parameters in Scottish raw waters were studied. The data set consisted of 5 sampling sites, 5 measured biochemical variables and over 12 (averaged between 2015 - 2019) months (Table 7.1). Therefore, the data were arranged as a three-dimensional array with modes: 5 (sites) \times 5 (biochemical variables) \times 12 (months).

PCA

Traditional PCA applied on the data set of samples (4 sites \times 12 months) \times 5 variables required 2 principal components to explain 91% variance (Figure 7.1). The biplots obtained after PCA is reported in Appendix E, Figure E.2 (pg. 206). The separation of sampling sites is well visible in the bi-plot. The four sites (indicated by the four colours) are well grouped and can be described by the variables. However, information about the sampling time (indicated by the symbols) are mixed and clear patterns cannot be found. Moreover, the simultaneous interpretation of sampling sites and time is not possible in the PCA biplot.

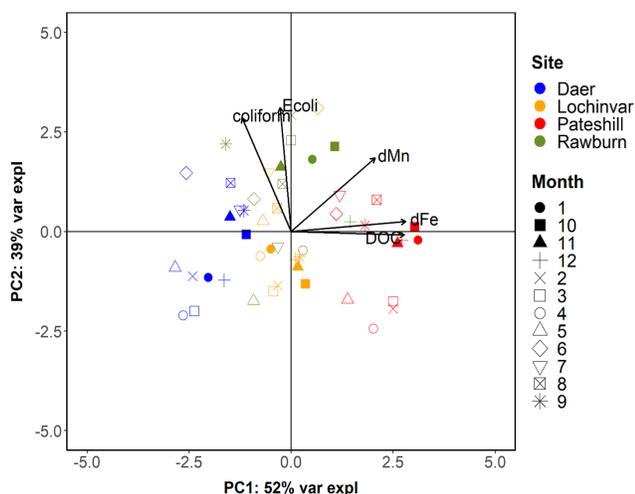


Figure 7.1: PCA biplot of the data set Waterquality with principal component 1 explains 52% of the data and principle component 2 39%. The scores grouped by colour, months by symbols and the loadings are presented by the vectors.

Tucker3

To interpret the biochemical variables over time for various sampling sites simultaneously, Tucker3 was applied to the data set. For Tucker3, the data set was arranged in a three-way data array with modes: 4 (sampling sites) x 5 (biochemical variables) x 12 (months). To find the appropriate number of components for each mode, initial Tucker3 models were created with up to 4, 5, and 12 components in respectively the p, q, and r mode. The scree plot was used to visualise how many additional percent of variance each model explained when adding extra components Appendix E, Figure E.3 (pg. 207). Generally, the optimal complexity of the Tucker3 model is the one that requires the smallest number of components, but still describes relatively high fraction of data variance.

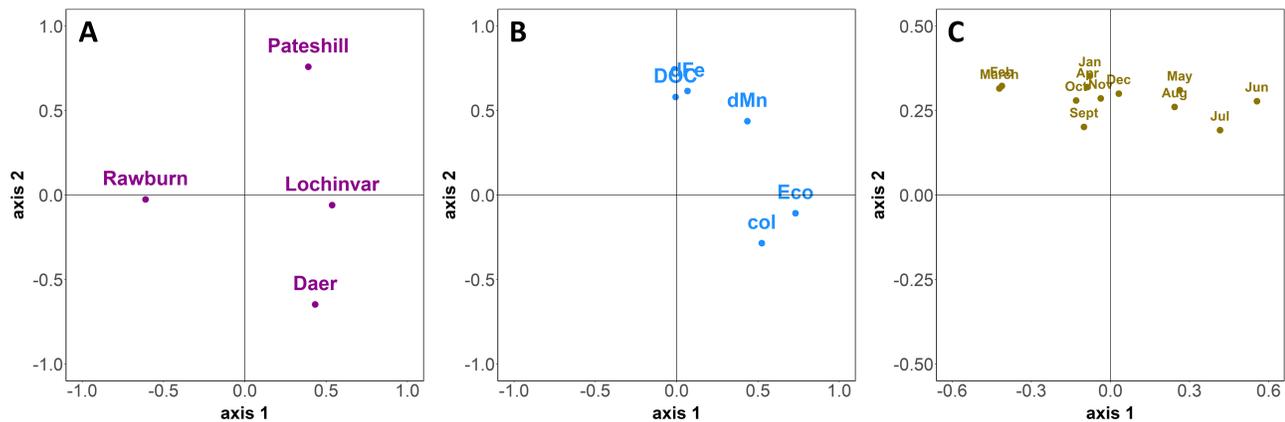


Figure 7.2: Loading plots of the Tucker3 [2,2,2]-model from the data set Waterquality with sampling sites (A), biochemical variables (B), months (C).

The complexity that describes as much variance but is still interpretable is the model with 2 components in each mode, also known as a [2,2,2]-model, accounting for 67% variance. The statistical validity of the bootstrap computing confidence intervals for the current fit was between 64 and 88%. Such a medium variance with environmental data is not unusual as a result of the very high noise related to the great variability of environmental and sampling conditions (Leardi et al. 2000). It has two components in mode P (sampling site), two components in mode Q (the biochemical variables) and two components in mode R (months). Results of the Tucker3 model are reported in Figure 7.2 with loading plots of Samples (P), Variables (Q) and Months (R). Each of these diagrams may be inspected separately, e.g. in Figure 7.2A shows a clear discrimination between the Rawburn and the other three sampling sites, while in Figure 7.2 B DOC and dFe form a clear variable cluster as well as *E. coli* and *coliform*.

More information, however, can be expected by the joint interpretation of all loading plots. For this, the information in G core is required. The G core of the [2,2,2]-model is shown in Figure 7.3 A together with its unfolded two frontal planes (time slices) in Figure 7.3 B. In total the core matrix of the [2,2,2]-model has eight elements (g111 up to g222). The two parts of this unfolded matrix correspond to the first and second components of ‘Time’, while the rows relate to the components of ‘Sampling sites’ and the columns within each block refer to the components of ‘Biochemical variables’. From the core matrix it becomes clear that a diagonalization has been obtained, because the two major parts of the variance are covered by the body diagonal elements, g111 and g222 (Henrion 1993). The remaining elements have minor importance.

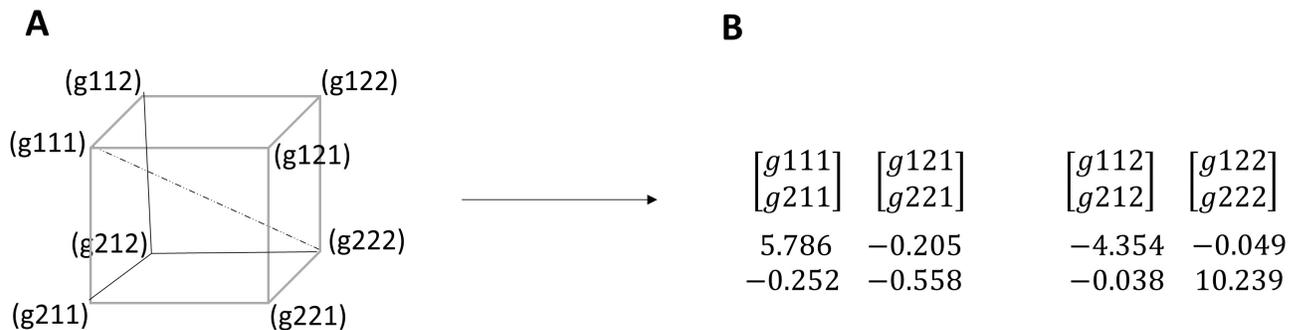


Figure 7.3: G core of the Tucker3 [2,2,2]-model (A) and its unfolded core matrix equivalent represented by the frontal planes (B). G111 has a high positive value, G222 a high negative value, while the other values are of minor importance.

With most of the variance covered by g111 (explaining 42% of information) and g222 (explaining 18% of information), these elements together describe almost the full G core of Figure 7.3 A, which simplified leads to a trilinear model with three terms:

$$x_{ijk} = a_{i1}b_{j1}c_{k1}g_{111} + a_{i2}b_{j2}c_{k2}g_{222} + e_{ijk} \quad (7.1)$$

Loading plots can be interpreted simultaneously with respect of their respective signs and magnitudes. This means that only large magnitude loadings with the right combination of signs can be interpreted together. For a better visual interpretation, the loading of the two components for each mode in Figure 7.2 are plotted in individual plots in Figure 7.4. Figure 7.4 A1 and A2 present the loading plots of the two components of the sampling site mode, while Figure 7.4 B1 and B2 present the loading plots of the two components of the biochemical variable’s mode, and Figure 7.4 C1 and C2 present the loading

plots of the two components of the month's mode.

An example of the interpretation of the Tucker3 model is as follows: the first term of the [2,2,2]-model, g111, is the combination of the first components in all modes. In other words, the loadings plot A1, B1, and C1 are interpreted together. G111 has a positive sign (Figure 7.2) and therefore the combination of loadings of the first components from Figure 7.4 should together be positive in order to create this positive g111. The positive element g111 is formed by the combination of positive A1 loadings together, all B1 loadings and positive C1 loadings. This means that positive large A1 (Daer, Lochinvar and Pateshill), positive large B1 (*E. coli*, *coliform*, and dMn), and large positive C1 (June, July, and August) are meaningful together. The positive g111 element can also be constructed by large negative A1 (Rawburn), large positive B1 (*E. coli*, *coliform* and dMn) together with a large negative C1 (February and March). In a similar way the g222 term can be analysed. An overview of all meaningful combinations is given in Table 7.2.

Table 7.2: *Meaningful interactions for the elements G111, G112, and G222 of the Tucker3 [2,2,2]-model of the SW data set.*

G111 (positive)	A	B	C
Large +A1, +B1, +C1	Daer, PH, Loch	<i>E. coli</i> , <i>coliform</i> , dMn	May, Jun, Jul, Aug
Large -A1, +B1, -C1	Raw	<i>E. coli</i> , <i>coliform</i> , dMn	Feb, Mar
G222 (positive)	A	B	C
Large +A2, +B2, +C2	PH	dFe, dMn, DOC	All months, highest in winter
Large -A2, -B2, -C2	Daer	<i>E. coli</i> , <i>coliform</i> , dMn	All months, highest in winter
G122 (negative)	A	B	C
Large -A2, +B2, +C2	Rawburn	<i>E. coli</i> , <i>coliform</i> , dMn	All months, highest in winter

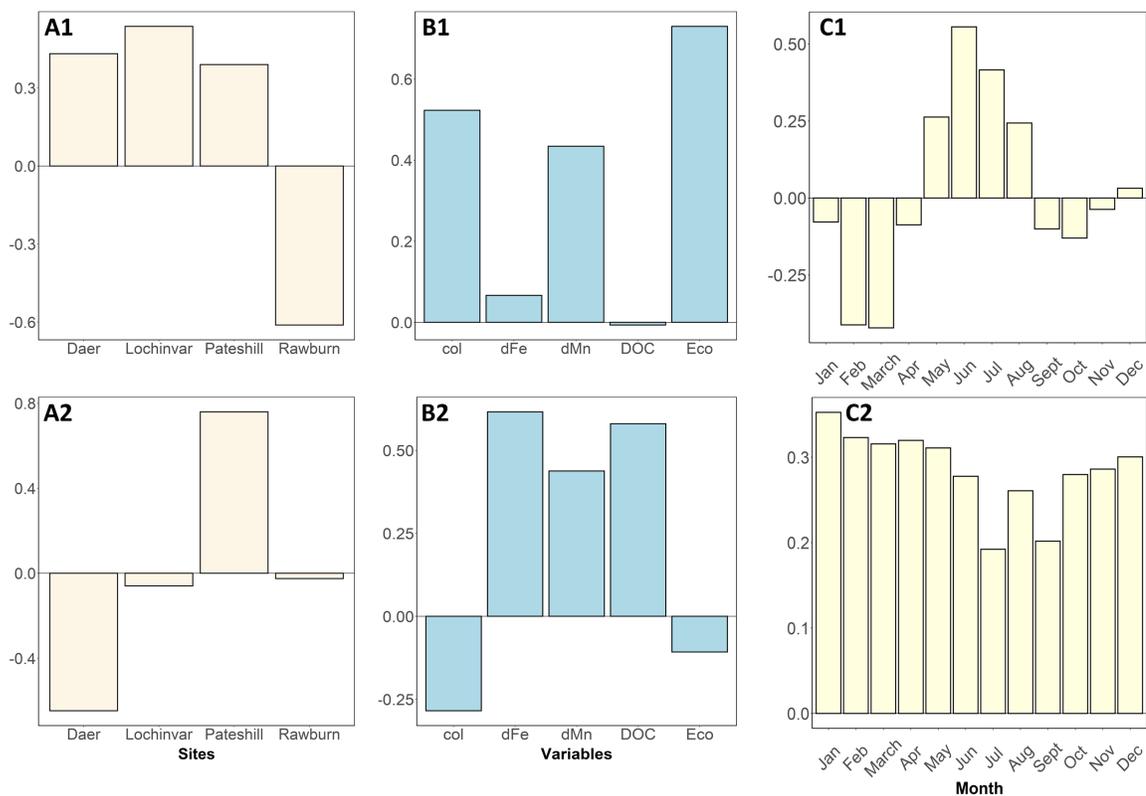


Figure 7.4: Tucker3 [2,2,2]-model loading plots with sample loading plot of component 1 (A1), sample loading plot of component 2 (A2), variables loading plot of component 1 (B1), variables loading plot of component 2 (B2), time loading plot of component 1 (C1), and time loading plot of component 2 (C2).

From Figure 7.4 and Table 7.2 it becomes clear that the components B1 and B2 of the Tucker3 model relate to the different way the biology and chemistry of the water behave over time. Moreover, DOC and dFe have similar loadings which was also demonstrated by a high correlation (Pearson correlation, $r(45) = .925$, $p < .001$). This high correlation between DOC and dFe was also found in Chapter 3, Section 3.3.2 (pg. 72) and was explained by the complex formation between DOC and dFe in water (Soulsby et al. 2001).

There seems to be a seasonality related to the biochemical signature of the water sources, which also varied among sampling sites. Sampling sites Daer, Lochinvar, and Pateshill have high concentrations of *E. coli*, *coliform* and dMn in the summer months, while Rawburn has the highest microbial content in the winter. Looking at the correspondent average monthly temperature and precipitation in Figure 7.5 it becomes clear that the months with the highest temperatures demonstrated the highest microbial content for most sampling sites. The relationship between temperature and microbial growth is

generally understood, with warm water temperatures of 15-20° C during summer months in Scotland providing better conditions for bacteriological growth compared to winter (Grose et al. 1998, van der Kooij & van der Wielen 2011). Rawburn, however, showed the higher microbial content in the winter months.

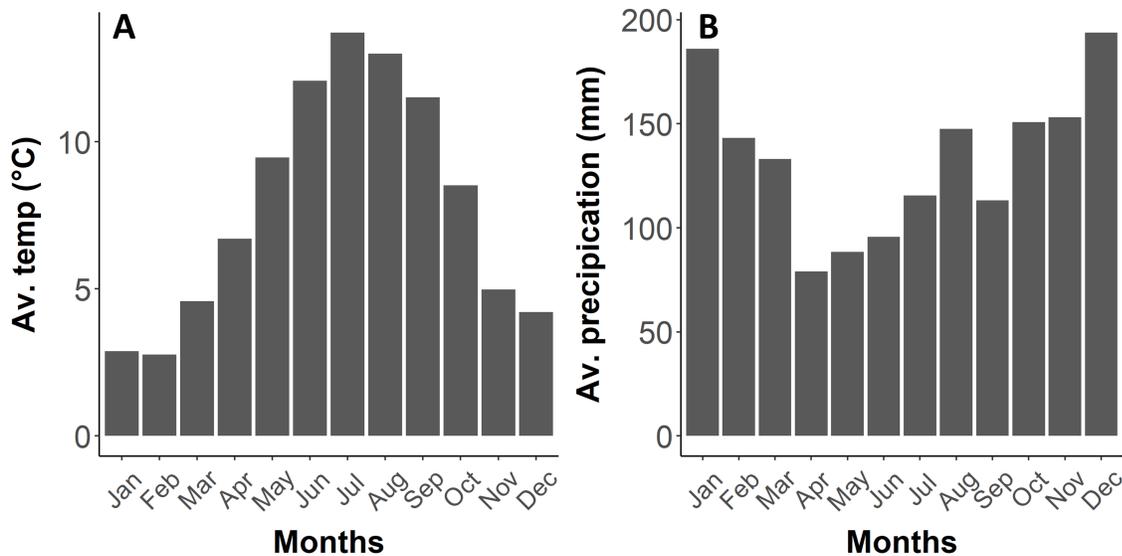


Figure 7.5: Average temperature (A) and precipitation (B) in Scotland throughout the year. Recorded by Metoffice and available at <https://www.metoffice.gov.uk>.

Pateshill samples demonstrated similar *E. coli* and *coliform* content as Lochinvar and Daer, but much lower than Rawburn. DOC, dFe on the other hand were much higher than the other sampling sites with a slight increased concentration during the winter months. In the study on Irish streams, Liu et al. (2014) found the highest DOC concentrations in autumn and early winter with spring having the lowest concentrations. Soulsby et al. (2001), who studied seasonal water quality trends in four streams in the Cairngorm Mountains in Scotland, reported TOC maxima/peaks in January, February, May, August, September, and November. A possible explanation for this pattern is the annual cycle of plant growth and decomposition (Muller & Tankéré-Muller 2012), together with precipitation (Figure 7.5). DOC concentration in water increases with increased hydrological flow through enriched organic and mineral soils countering the effect of dilution by rainwater (Aitkenhead-Peterson et al. 2007, Muller & Tankéré-Muller 2012, Soulsby et al. 2001). In a study by Soulsby et al. (2001) the winter period – and in particular snowmelt – strongly influences both the hydrological regime and hydrochemical seasonality of streams in the western Cairngorms.

Also, DOC concentration depends on the topography of the land (Aitkenhead-Peterson et al. 2007, Liu et al. 2014). According to Liu et al. (2014), it increases in the following order: arable land, grassland, peatland and forestland. In autumn and winter, DOC export is linked to peatland, while spring is best described by grassland and arable and summer by heather, montane and conifer (Aitkenhead-Peterson et al. 2007). Interestingly, the surroundings of Pateshill were predominantly covered by grassland and conifer trees (Table 7.1), which does not explain the higher DOC concentrations in autumn and winter.

Overall, a seasonal variability can be expected for the microbial and chemical water quality parameters as well as variation among water sources, with the highest concentrations of DOC and dFe found at Pateshill. The variables DOC and dissolved iron had similar loadings and therefore monitoring their behaviour could be done by one of the variables. In this case, DOC is the simplest analytical method and therefore the best proxy. Also the variables *E. coli* and *coliform* described similar water quality aspect. For these variables also one of the pair can be chosen as proxy.

While independent PCA plots were also able to show similarity between the variable loadings as well as the grouping of the sample sites, Tucker3 made it possible to simultaneously interpret time and sampling sites. This made it possible to see that Rawburn behaved differently to the other sites with respect to the seasonality of the microbial water content. Tucker3 therefore demonstrates to be beneficial for data sets where multiple monitoring conditions are used and can be seen as the future of multivariate and multidimensional data analysis. For this type of data analysis, systematic monitoring of various sampling sites is necessary. If, for example, pH was monitored on a regular basis for the various water sources it could have included in the data array. pH plays an important role in the adsorption of NOM on GAC (Bond & Digiano 2004, Cardenas 2008), and observing possible fluctuation could explain possible changes in biofilter performance.

7.3.2 IN and Pateshill comparison

The previous paragraph demonstrated the influence of spatial and seasonal variation for various water sources in Scotland. To be able to discuss how spatial and seasonal variation affects biofilter performance, we must first know how the INfluent water of the biofiltration (IN) compares to the other water sources. Due to the absence of data from April till September, the influent was not included in the Tucker3 model and direct comparison of water quality between IN and the other water sources is not possible. Though, IN was sampled at Pateshill, and in case similar results of Pateshill can be seen as

a proxy for IN, therefore the Tucker3 results can be used to discuss the impact of spatial and temporal variation in water quality and the impact on the biofilter.

For the microbial and chemical variables DOC, dFe, dMn, and *E. coli*, IN and Pateshill were significant different, while *coliform* was not statistically significantly different (Mann-Whitney-Wilcoxon, $p_{\text{adjust}} < 0.05$). IN was sampled at Pateshill and therefore similar microbial and chemical water quality was expected. A possible explanation for the noticed difference could be that the measurements from IN (Oct 2018 and March 2019) were lower than averaged Pateshill datapoints (Oct - March averaged over 2015 - 2019), while Pateshill Oct 2018 - March 2019 could be similar to that of IN. However, Figure 7.6 demonstrates that this is not the case. In all boxplot the IN data (described in red) is indeed positioned below the boxplot of Pateshill. However, Pateshill measurements done in the similar months and years as IN are not positioned at the bottom of the boxplot but varied. This indicates that the averaging of the years for Pateshill did not cause the difference between IN and Pateshill.

An explanation for the lower values of IN could be the rough prefiltering done for the IN waters. This filter step could have retained part of the microbial content, possibly attached to particulate matter. Also, the IN water was stored for a period of time before used in the biofiltration experiment which could have altered the water and thereby decreasing the microbial content. For dMn and dFe concentration, the difference between Pateshill and IN could be a result of the difference of sample pre-treatment done in-house and the Scottish water laboratory. In the Scottish water lab the samples were filtered through a 0.45 μm filter before the DOC, dFe and dMn analyses, while in-house a 0,2 μm filters were used before analysis of the variables resulting in a significantly lower dFe and dMn concentration. Finally, systematic variations between the analytical instruments used in-house and in the Scottish water laboratory can explain the observed differences between IN and Pateshill.

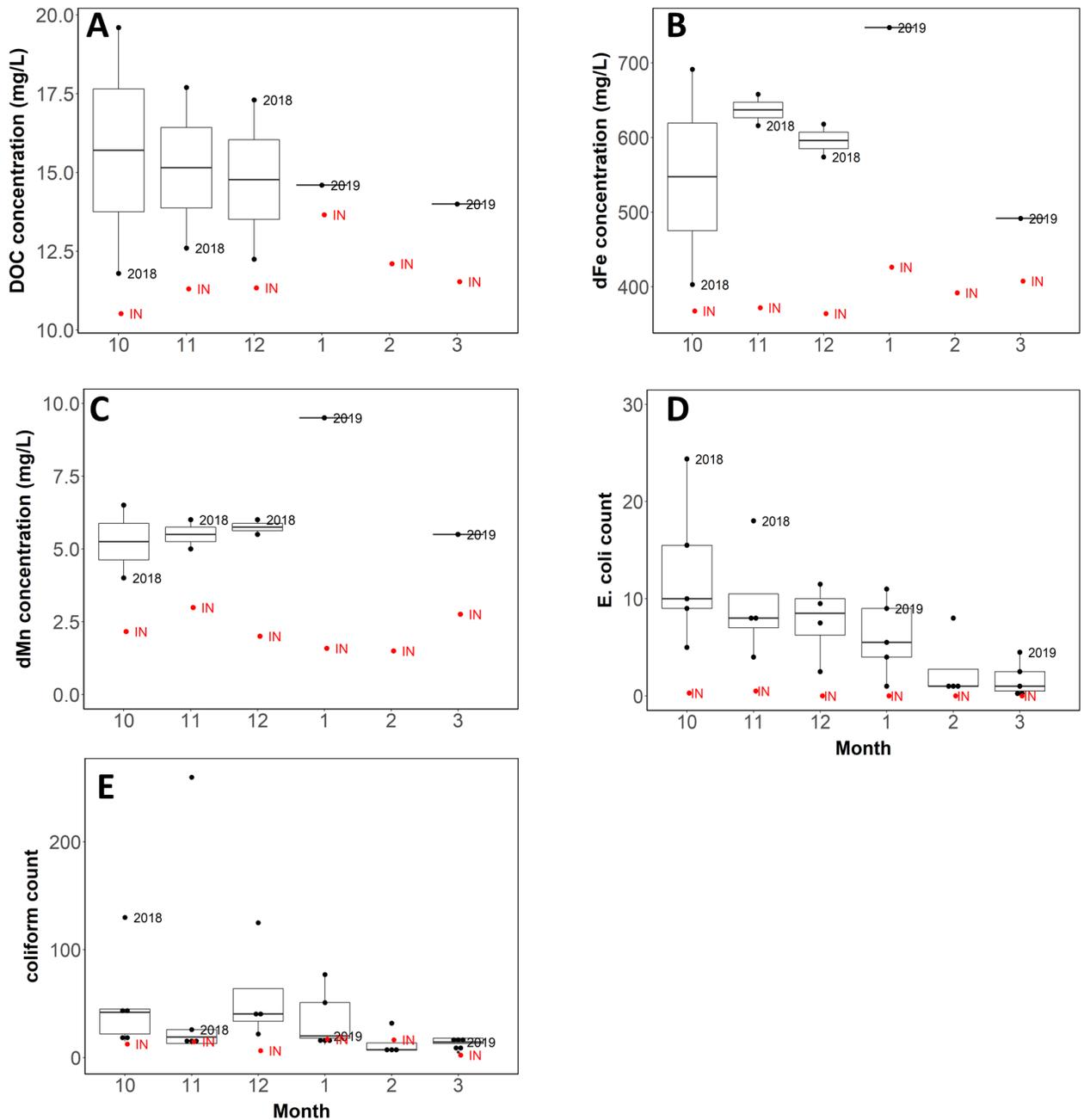


Figure 7.6: Boxplot of all measurements done by Scottish Water between 2015 and 2019 between or variables DOC (A), dFe (B), dMn (C), E. coli (D), and coliform (E). Datapoints of Pateshill in Oct-Dec2018 (10-12) and Jan-March 2019 (1-3) are pointed out with labels and data from IN labelled in red.

7.3.3 Impact of spatial and temporal variation

To what extent can we discuss the impact of temporal and seasonal variation on the biofilter performance with only the variable *coliform* being similar between Pateshill and IN? Strictly speaking, only *coliform* was in similar quantities present in IN and Pateshill. However, for the scope of this thesis the DOC is of greater interest. Despite the lower DOC concentration in IN compared to Pateshill, both water sources had much higher DOC concentration than Daer, Lochinvar, and Rawburn. We will therefore discuss the spatial and temporal variation of DOC, while keeping in mind that IN does differ from Pateshill.

Pateshill and IN both have a complex water chemistry. Organic matter in the soil and surrounding vegetation determines the amount and composition of DOC produced in watersheds (Aitkenhead-Peterson et al. 2007). Large soil carbon pools and areas of peat result in high DOC concentrations and exports (Aitkenhead-Peterson et al. 2007). Pateshill is water storage, which receives water from a populated area characterized by anthropogenic pressure due to industrial and agricultural activities, while Daer, Lochinvar and Rawburn are natural lochs, which receive rainwater coming down from hillsides.

In Chapter 3, the relatively high DOC concentration in the effluent of the pilot-scale biofilter demonstrated to be an issue. Therefore, longer filters were advised to achieve the best removal. Using other water sources in Scotland with lower DOC loading will impact the biofilter performance. Lowering the influent DOC concentration decreases the rate at which adsorption sites are occupied and thereby lower the advancement of the adsorption zone and elongate the breakthrough time (Bond & Digiano 2004, Zachman & Summers 2010). Adsorption kinetics is modelled by a dual Langmuir kinetic model meaning a series of two consecutive processes: nonselective adsorption of molecules in meso- and supermicropores followed by the movement of such adsorbed molecules into a small micropores through the pore mouth barriers (Nguyen & Do 2000). A lower DOC concentration not only elongates the time required for the nonselective adsorption of molecules in mesopores but also the slow adsorption by the movement of such adsorbed molecules into a small micropores. It moves forward and elongates the start of the initial breakthrough (when mesopores become filled), but also moves forward and stretches the movement of the adsorbed molecules into the small micropores. This second stage, also referred to as slow adsorption (Peel & Benedek 1980), is positioned at a higher removal efficiency compared to the IN water used during biofiltration (Cardenas 2008).

If the biofiltration would have continued for a longer period of time using the IN water, the filter would run in spring and summer. A season effect could therefore be expected to influence the biofilter. Pateshill demonstrated a seasonality for the DOC concentration (Figure 7.4). If the biofiltration experiment would have continued for more months, the influent DOC concentration would have decreased. Consequently, the biofilter would have been loaded with a lower concentration of DOC. With slow adsorption still taking place, a lower concentration would cause the removal efficiency to increase.

A lower DOM influent concentration as a result of a different water source or seasonal fluctuation not only impacts the performance of the biofilter. With similar DOM composition expected in different Scottish water throughout the season, a difference in the speed the adsorption takes place and thereby the DOM profile over time. In chapter 5 the first humic-like fluorescence signal was found in the pore water at week 9. With a lower DOM concentration the presence of this signal can be delayed by weeks. Consequently, these different environmental niches throughout filter depth at week 9 will also impact the composition of microbial communities at that time found in chapter 6. Again, further DOM characterization at various sampling sites throughout the year as well as similar measurements done within the biofilter could lead to a better understanding of the spatial and seasonal impact of DOM on biofilter performance, stratification and consequent microbial structuring can be found.

7.4 Conclusion

The objective of this study was to understand how water source and seasonal changes can impact the slow-flow BAC filter. Results show that:

- Compared to other sampling locations, the influent water used in Chapter 3, 5, and 6 had a high chemical loading and low microbial loading.
- Despite seasonal variation in concentrations, the filter would highly likely perform better in terms of chemical water quality parameters when applied across Scotland or during summer months. Moreover, it would slow down pore filling and possibly impact microbial composition at the various niches through filter depth.
- Three way-PCA via Tucker3 is not commonly used in this field of expertise but proved invaluable in giving insights when dealing with multiway data. For example, the difference in seasonal variability of Pateshill and Rawburn would not have been so clearly visualized by PCA.
- For Scottish Water this chemometric tool can help to monitor a large number of sampling sites over time by looking at multiple variables at the same time. It does not only give a clear overview on how these sites behave, but it also informs on processes that are not easily found by interpretation of individual plots.
- If the WTW data collection would have been complete resulting in a data array without missing values, more variables and sampling sites could have been included in the analysis making it more comprehensive. This highlights the importance of structured data collection.

Chapter 8

Conclusion

8.1 Restatement of research aim and objectives

The aim of this thesis was to evaluate slow-flow BAC filtration as possible application for small scale water treatment systems by looking at the chemical water quality aspects with help of chemometric methods. To achieve the aims of this project, the following objectives were identified:

1. To characterize DOM and other chemical water quality parameters removal by pilot slow flow BAC filters during the first months of operation and see how filter length impacts this process.
2. To develop a simple cost-effective fluorescence by excitation-emission spectroscopy method combined to PARAFAC analysis using a plate reader for the analysis of the DOM in objective 3.
3. To further understand removal processes of DOM and other chemical parameters by slow flow BAC filtration using the analysis of pore water.
4. To examine the degradation of DOM by filter microbial communities at the different depths of a biofilter.
5. To understand how water source and seasonal changes can impact the slow-flow BAC filter.

8.2 Summary of key findings of chapter 3 to 7

8.2.1 Characterization of DOM and other chemical water quality parameters removal by laboratory-scale slow-flow BAC biofilters

An initial understanding of the removal of chemical water quality parameters in a pilot-scale slow-flow BAC filter from virgin to steady state for various filter lengths was studied. Slow-flow BAC filtration of 90 cm was found to remove 20% higher DOC concentration in apparent steady state compared to the 30 cm filter. Apparent steady state was found to depend on processes such as slow-adsorption by GAC or biosorption for the removal of TOC. Longer filters have more GAC material for this slow adsorption. Overall, it was shown that the BAC filter did not meet the criteria for providing clean drinking water, as it was not able to filter iron until a concentration below the EU standards. Moreover, the DOC concentration was shown to increase over time to such high concentration that colour was visible. With all microbial processes taking place within the first 30 cm of the filter, no added value of filter length was demonstrated for pathogen removal. The choice for filter length was therefore demonstrated to be dependent on the treatment goal for the Scottish off-grid system.

8.2.2 Fluorescence excitation–emission measurements of DOM using a plate reader coupled to PARAFAC analysis

A simple cost-effective method for the analysis of fluorescence by EEM spectroscopy combined to PARAFAC using plate reader was developed. It was demonstrated that the method is fit for purpose. Using two reference standard and various concentrations the PR method produces a high signal linearity ($R > 0.998$), good repeatability ($RSD < 6.5\%$) and intermediate precision ($RSD < 18\%$). Also the recovery was within acceptable limits (97 - 118%). Applied to environmental samples the results showed high similarity with the results of the golden standard : correlation among the retrieved fluorophores was good ($R > 0.89$) and signal recovery was within acceptable limits (85-98%).

8.2.3 Understanding removal processes of DOM and other chemical parameters by slow-flow BAC filtration using the analysis of pore water

The removal process of DOM and other chemical parameters by slow-flow BAC filtration using the analysis of pore water was studied. It was demonstrated that the two-stage adsorption (1. attachment to pore walls and 2. migration into the micropores) as observed in chapter 3 was supported by the analysis of the pore water. Results showed that lower sections already started to adsorb molecules,

while the top was not fully saturated yet. It was demonstrated that filter length did not impact the processes in the top of the filter. It only extends these processes to deeper in the filter. The extent to which microbial activity was taking place in the biofilter could not be given. It was demonstrated that the biomass stratification found by DQ and MV was similar to that of the peak-T fluorescence and that peak T can be used as a proxy for microbial activity or cell count. It was also demonstrated that microbial stratification was similar to that of the humic-like fluorescence and thereby to DOC. This showed that a nutrient stratification is present within the biofilter, which could impact bacterial community structures within the biofilter.

8.2.4 Differential utilisation of dissolved organic matter compound groups by different biofilter microbial communities

The ability of microbial communities extracted from different depths of a biofilter at week 12 to degrade DOM was studied. It was demonstrated that the differences observed in the three communities' ability to degrade DOM, as well as their composition studied after 35h of incubation, are the consequence of the initial compositional difference of the biofilter communities from which they were derived. Such differences were expected to be a consequence of the adaptation of the original biofilter communities to the different environmental niches developing at different depths of a biofilter. It was suggested that lower depths might select for communities capable of degrading more complex DOM compounds.

8.2.5 Three-way PCA to study spatial and temporal variation in water quality parameters and its impact on slow-flow BAC performance

How spatial and temporal variations in the water quality can influence the performance results of the pilot-scale slow-flow BAC filter was studied. It was demonstrated that the influent water used in the experiments had a high chemical loading and low microbial loading. Despite seasonal variation in concentrations, the filter was expected to perform better in terms of chemical water quality parameters when applied across Scotland or during summer months. It slows down pore filling and possibly impacts microbial composition at the various niches through filter depth.

8.3 Conclusion and recommendations for future research

This thesis introduces the use of a slow-flow BAC filter system for the application of a small-scale filter system in remote areas in Scotland. It demonstrates to remove the problematic high TOC and dissolved iron concentrations, especially during the first weeks of treatment. The water used during the experiment was of high chemical loading which consequently caused the concentration of the treated water not to remain under the regulatory limit during apparent steady state. When water in summer months or from other water sources is used, an improved removal during apparent steady state is expected, most likely under the regulatory limits. It is recommended to prefilter the water to remove the particulate matter that can include high concentrations of iron, manganese, and TOC. Moreover, it is recommended that the filter is thoroughly flushed before usage to remove dissolved ash present in the GAC and to prevent pH swings as a result of GAC protonation.

This thesis highlights the importance of slow adsorption, also known as tailing, in apparent steady state. This process is often confused with biodegradation resulting in an overestimated microbial DOC removal in apparent steady state. It is therefore advised not to look at a plateau in the TOC removal curve, but look at the number of bed volumes processed in combination with the theoretically expected TOC removal (between 10-20%) when describing the filter process (Figure 8.1). The tailing of DOM removal can continue for months and using a long filter with more absorptive material can remove DOC for an extensive period of time.

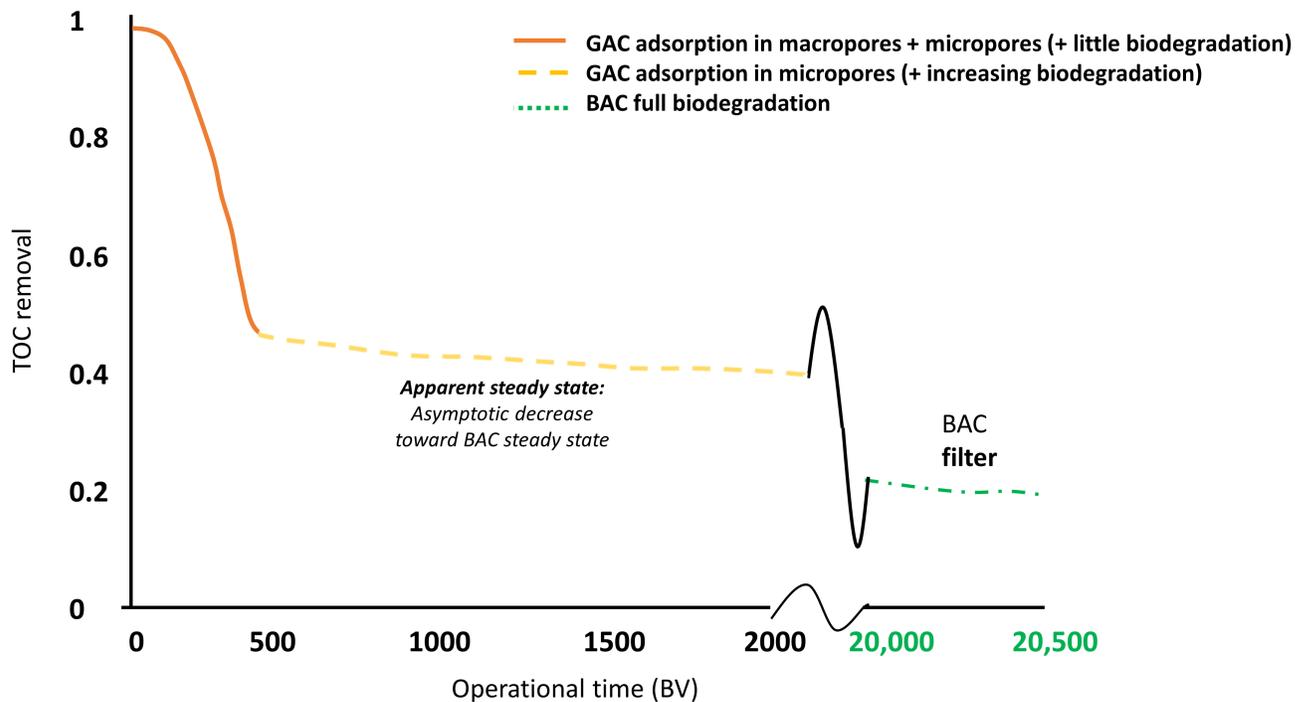


Figure 8.1: *TOC removal as function of bed volumes for a GAC filter that becomes a BAC filter once the adsorption capacity is fully exhausted. Initial removal is dominated by multistage adsorption (macropore filling and micropore filling). Removal will asymptotically reach the steady state (at around 20,000 BV) where removal process is mostly dependent on biological removal. Based on Figure 2.13.*

This thesis also introduces the benefit of using longer filter lengths. While pathogen removal has been demonstrated to happen in the first 30 cm of the biofilter, the highest removal of TOC and dissolved iron happens in apparent steady state of the 90 cm filter. This 90 cm filter removes significantly more TOC and dissolved iron compared to the 30 cm filter. Moreover, also nitrate was removed best by the longest filters. When applied to waters with increased nitrogen content, for example as a result of nearby livestock, the longest filter makes sure that nitrate is removed sufficiently. Using the increased filter length does not influence the chemical processes taking place in the top 30 cm of the filter, but it extends the removal process and thereby the lifetime of the filter.

Another benefit of the long filter introduced by this thesis is the selection of microbial communities capable of removing recalcitrant DOC, such as humic substances. These microbial communities were positioned deep in the filter, highly likely as a result of nutrient limitation. These long filters therefore cultivate microbial communities important for the humic rich Scottish waters.

To conclude, when installed with a rough pre-treatment filter and sufficiently flushed GAC, a 90 cm BAC filter will be a good alternative for the treatment of Scottish fresh water to provide good quality drinking water for remote areas in Scotland, low in microbial and chemical water quality parameter concentrations. For further application the pilot scale system needs to be scaled up to provide sufficient drinking water per household. Up-scaling can be achieved by increasing the size of the BAC filter or placing multiple filters parallel to each other. While using one large filter with increased filter width sounds as the easiest solution it might behave differently compared to the pilot scale filters, for example, it might affect the short circuiting paths.

In hindsight, several aspects of the study should have been chosen differently. First, the GAC should have been flushed thoroughly to diminish the impact of GAC settling and leaching. Second, DOM should be characterized in more detail. For example, the combination of fluorescence EEM spectroscopy, LC-OCD, and AOC/BDOC measurement will give a more comprehensive impression of the DOM removal and/or transformation within the slow-flow BAC filter. Using more sophisticated methods requires a higher volume of porewater than extracted during the present study. This problem could be overcome by pooling the water samples of the biological replicates and focussing on the DOM characterization and leave out the anion and cation analysis. Also larger sections can be cut to increase the harvested pore water volume. Third, the filter experiment should have been extended for enough EBV for the filter to reach a true steady state where biodegradation is fully taking over the removal process. If this takes too much time, a smaller scale of the filter should be used to shorten this process. Alternatively, exhausted GAC can be used that was treated with the influent water used for the experiment. With this exhausted GAC the steady state DOM removal can be monitored directly.

Another recommendation to the biofilter study is the measurement of oxygen throughout the filter. This could contribute to the understanding of the limiting factor of microbial growth deeper in the filter. Is this a lack of substrates and nutrients or also a result of a shortage of oxygen?

Finally, it is recommended to repeat batch experiment using more biological replicates and monitoring the DOM transformation at multiple timepoints instead only at the start of the experiment and 23 hours. The increased number of replicates can strengthen the statistics and the findings. Using multiple timepoints can help to reveal in which order the DOM fractions are being utilized. Adding to the experiment fluorescence EEM spectroscopy will help to further characterize the DOM pool and its changes during incubation.

Appendix A

Appendix chapter 3

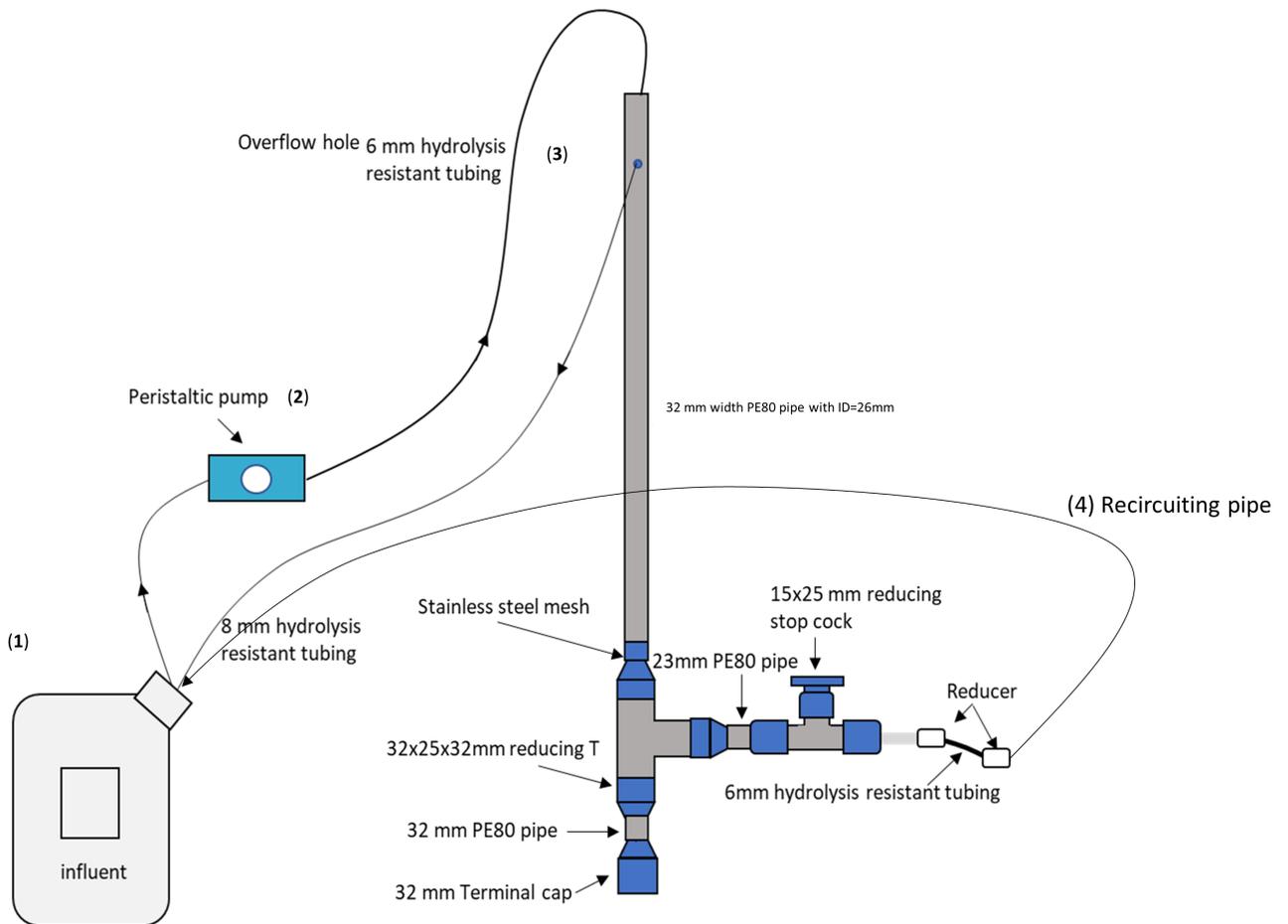


Figure A.1: Schematic representation of the biofilter when working in recirculating mode. The water is pumped from the influent jerrycan (1), peristaltic pump brings the influent water into the filter (2), overflow hole runs back the excess of water that is pumped in the filter (3), and after the water is run through the filter it is brought back to the jerrycan.(4)

TOC-VCSN Flow Diagram

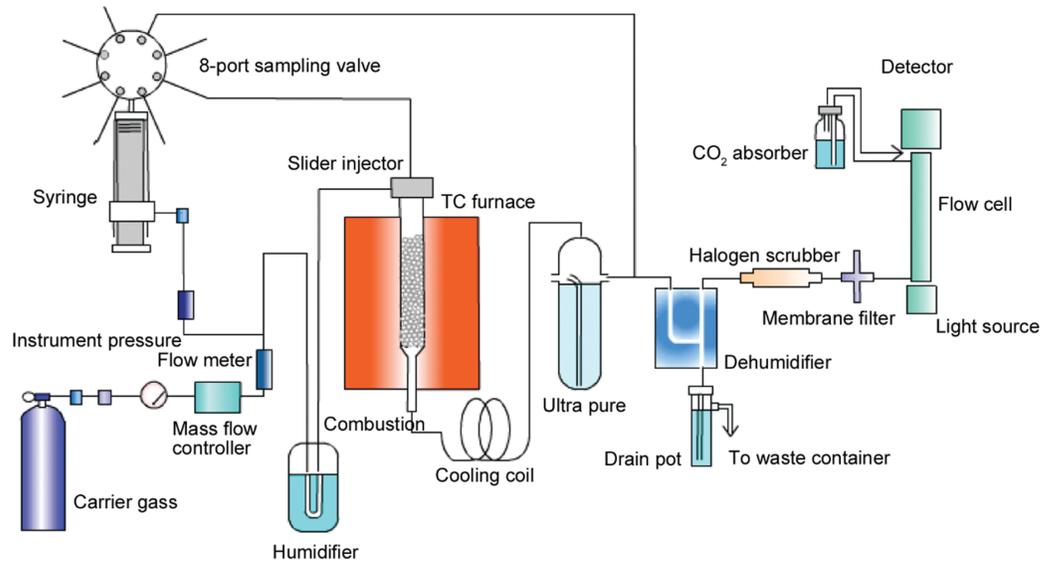


Figure A.2: Schematic representation of the carbon combustion method. From Qualls (2015).

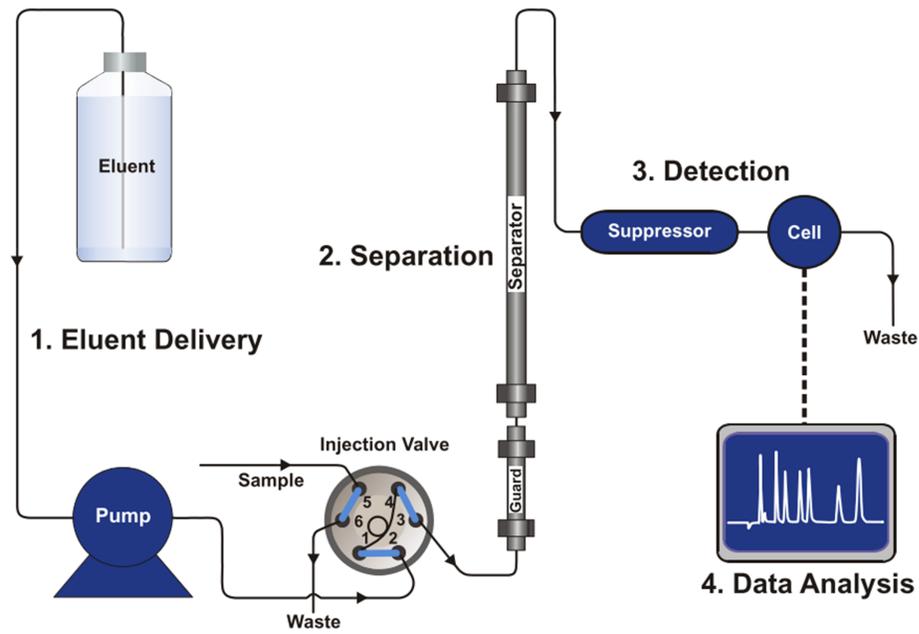


Figure A.3: Schematic representation of a typical ion chromatography (IC) analysis process. From Thermo Fisher Scientific (2012a).

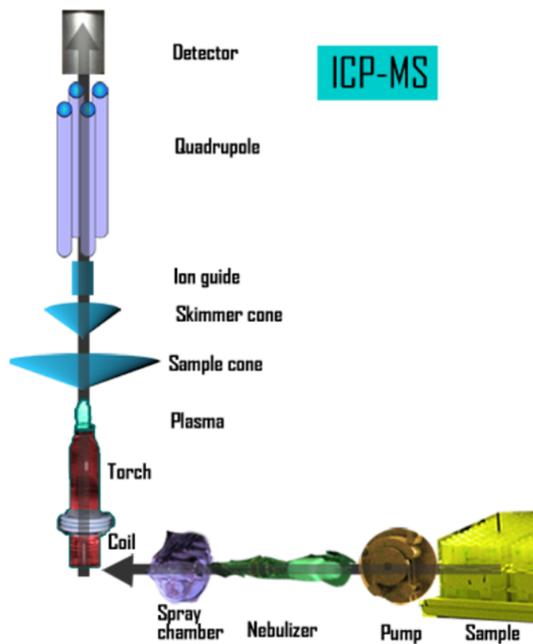


Figure A.4: Schematic representation of ICP-MS. From Radboud University (n.d.).

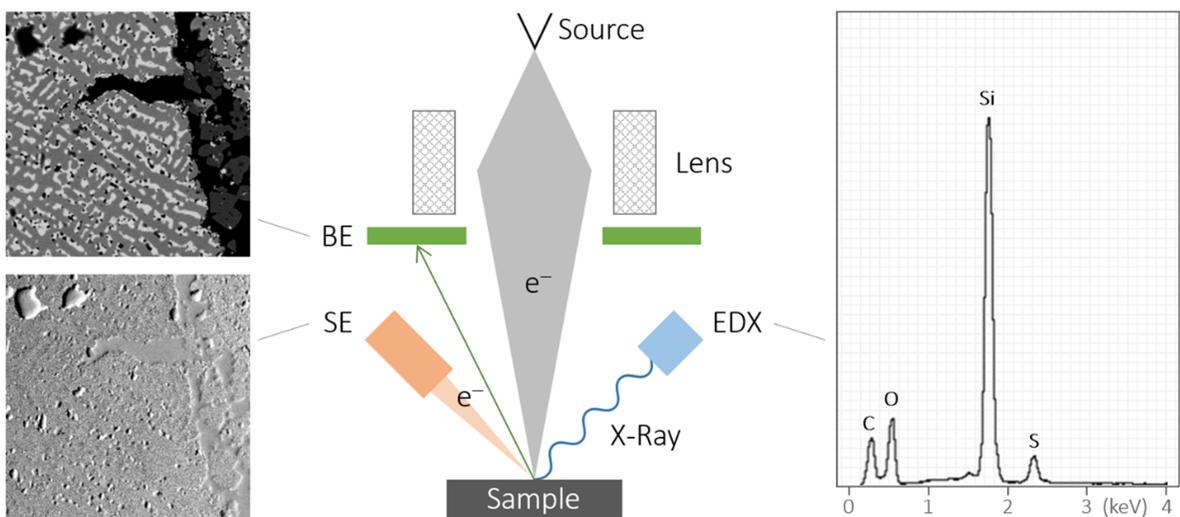


Figure A.5: Schematic representation of SEM-EDX. From RJL Micro & Analytics (n.d.).

Parameter	Prescribed Concentration or Value (PCV)	Total No. of Samples	No. Failed Samples	No. Zones with Failures	% Compliance in 2020	% Compliance in 2019	% Compliance in 2018
Key Parameters							
Bacteria							
Coliform Bacteria	0 number/100ml	14,832	20	18	99.87	99.75	99.72
<i>E. coli</i>	0 number/100ml	14,832	2	2	99.99	99.99	99.98
<i>Clostridium perfringens</i>	0 number/100ml	5,185	0	0	100.00	99.94	99.96
Total bacteria		34,849	22	18	99.94	99.88	
Metals							
Aluminium	200µg Al/l	5,213	2	2	99.96	100.00	99.89
Iron	200µg Fe/l	5,213	21	16	99.60	99.29	99.34
Lead (10)	10µg Pb/l	772	2	2	99.74	99.53	99.06
Manganese	50µg Mn/l	5,213	15	13	99.71	99.81	99.69
Nickel	20µg Ni/l	772	7	7	99.09	99.93	99.87
Total metals		17,183	47	35	99.73	99.70	
Other key parameters							
Colour	20mg/l Pt/Co	5,238	0	0	100.00	100.00	100.00
Hydrogen ion (pH)	6.5-9.5 pH	5,248	1	1	99.98	99.98	99.91
Nitrite	0.5mg NO ₂ /l	2,752	0	0	100.00	99.98	99.89
Odour	No abnormal change	5,247	0	0	100.00	99.90	99.91
Radon	100Bq/l	3,106	0	0	100.00	100.00	95.12
Taste	3 dilutions	5,243	0	0	100.00	99.96	99.98
Total Trihalomethanes	100µg/l	1,478	1	1	99.93	99.80	99.50
Turbidity	4 NTU	5,249	1	1	99.98	99.98	99.95
1,2 Dichloroethane	3µg/l	1,477	0	0	100.00	100.00	100.00
Aldrin	0.1µg/l	0	0	0	N/A	-	-
All Other Individual Pesticides	0.1µg/l	4,606	0	0	100.00	99.98	100.00
Ammonium	0.5mg NH ₄ /l	5,249	0	0	100.00	100.00	100.00
Antimony	5µg Sb/l	1,454	0	0	100.00	100.00	100.00
Arsenic	10µg As/l	1,454	0	0	100.00	100.00	100.00
Benzene	1µg/l	1,478	0	0	100.00	100.00	100.00
Benzo 3,4 Pyrene	0.01µg/l	1,498	0	0	100.00	100.00	99.94
Boron	1mg B/l	1,487	0	0	100.00	100.00	100.00
Bromate	10µg BrO ₃ /l	1,486	0	0	100.00	100.00	100.00
Cadmium	5µg Cd/l	1,454	0	0	100.00	100.00	100.00
Chloride	250mg Cl/l	1,483	0	0	100.00	100.00	100.00
Chromium	50µg Cr/l	1,454	0	0	100.00	100.00	100.00
Conductivity	2500µS/cm at 20°C	5,249	0	0	100.00	100.00	100.00
Copper	2mg Cu/l	772	0	0	100.00	100.00	100.00
Cyanide	50µg CN/l	1,485	0	0	100.00	100.00	100.00
Dieldrin	0.1µg/l	0	0	0	N/A	100.00	100.00
Enterococci	0 number/100ml	1,487	1	1	99.93	99.93	100.00
Fluoride	1.5mg F/l	1,486	0	0	100.00	100.00	100.00
Gross alpha activity*	0.1Bq/l	1,572	0	0	100.00	100.00	99.94
Gross beta activity*	1Bq/l	1,534	0	0	100.00	100.00	100.00
Heptachlor	0.03µg/l	0	0	0	N/A	-	100.00
Heptachlor epoxide	0.1µg/l	0	0	0	N/A	100.00	100.00

Parameter	Prescribed Concentration or Value (PCV)	Total No. of Samples	No. Failed Samples	No. Zones with Failures	% Compliance in 2020	% Compliance in 2019	% Compliance in 2018
Mercury	1µg Hg/l	1,487	0	0	100.00	100.00	100.00
Nitrate	50mg NO ₃ /l	2,752	0	0	100.00	100.00	100.00
Nitrite/Nitrate formula	<1mg/l	2,752	0	0	100.00	100.00	100.00
PAH - Sum of 4 Substances	0.1µg/l	1,516	0	0	100.00	100.00	100.00
Pesticides - Total Substances	0.5µg/l	1,138	0	0	100.00	100.00	100.00
Selenium	10µg Se/l	1,454	0	0	100.00	100.00	100.00
Sodium	200mg Na/l	1,487	0	0	100.00	100.00	100.00
Sulphate	250mg SO ₄ /l	1,487	0	0	100.00	100.00	100.00
Tetrachloroethene/Trichloroethene	10µg/l	1,478	0	0	100.00	100.00	100.00
Tetrachloromethane	3µg/l	1,478	0	0	100.00	100.00	100.00
Total other parameters		52,088	1	1	100.00	99.996	
Scotland total		137,681	73	52**	99.95	99.92	99.91

* Gross alpha and gross beta activity are not included in the Total other parameters or Scotland total figures and are not used to calculate the total compliance figure. This is because there is no PCV value for these variables. They form part of a monitoring program to detect any radionuclides in accordance with The Public Water Supplies (Scotland) Amendment Regulations 2017. The total number of samples including monitoring for radon, which is taken at WTW final water but applied across Water Supply Zones.

** A supply zone can fail for more than one parameter. This means that the total number of zones that failed for *at least* one parameter is less than the sum of the No. Zones with Failures column.

Figure A.6: *Parameters of water quality at consumers' taps provided by the drinking water quality regulator (2020).*

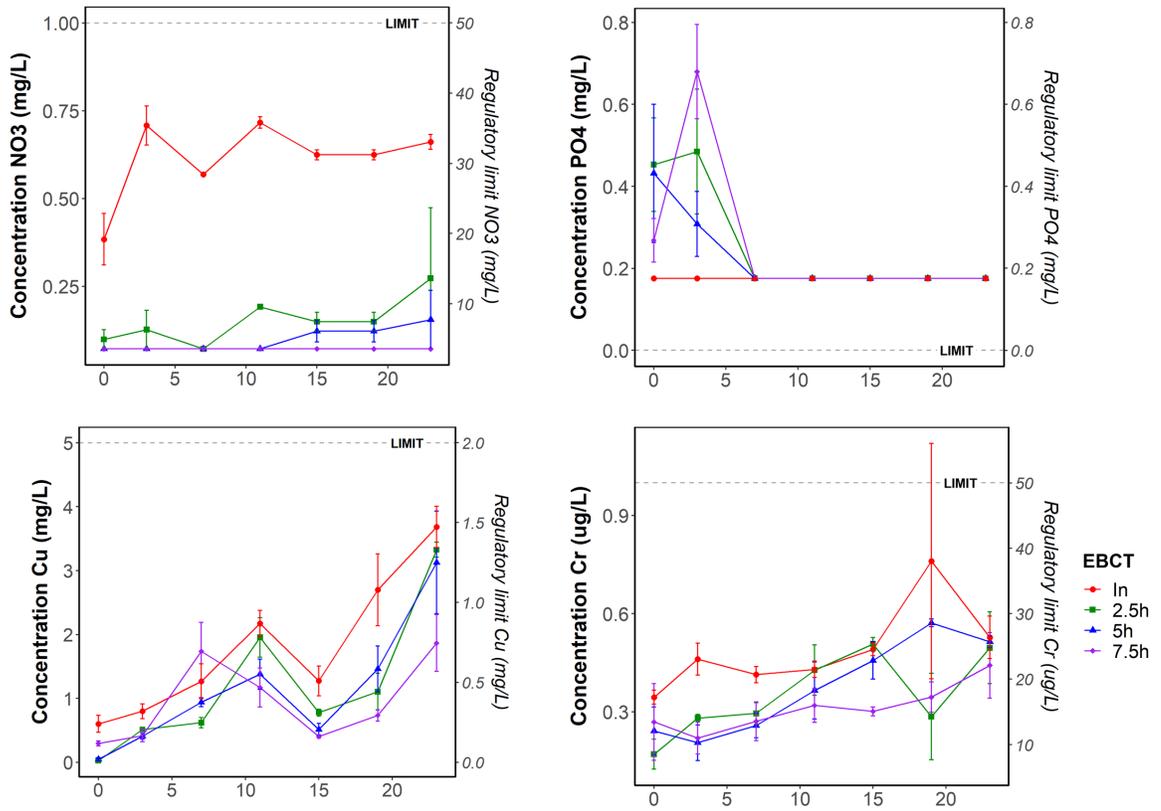


Figure A.7: Concentration of nitrate (Top left), phosphate (Top right), copper (Bottom left), chromium (Bottom right). Left y-axis shows the measured concentration of the influent and the $h=30, 60, 90$ cm effluent (mean \pm s.e., $n=3$ replicates) over 23 weeks of operation. The right axis together with the dashed line (when present) highlights the regulatory limits.

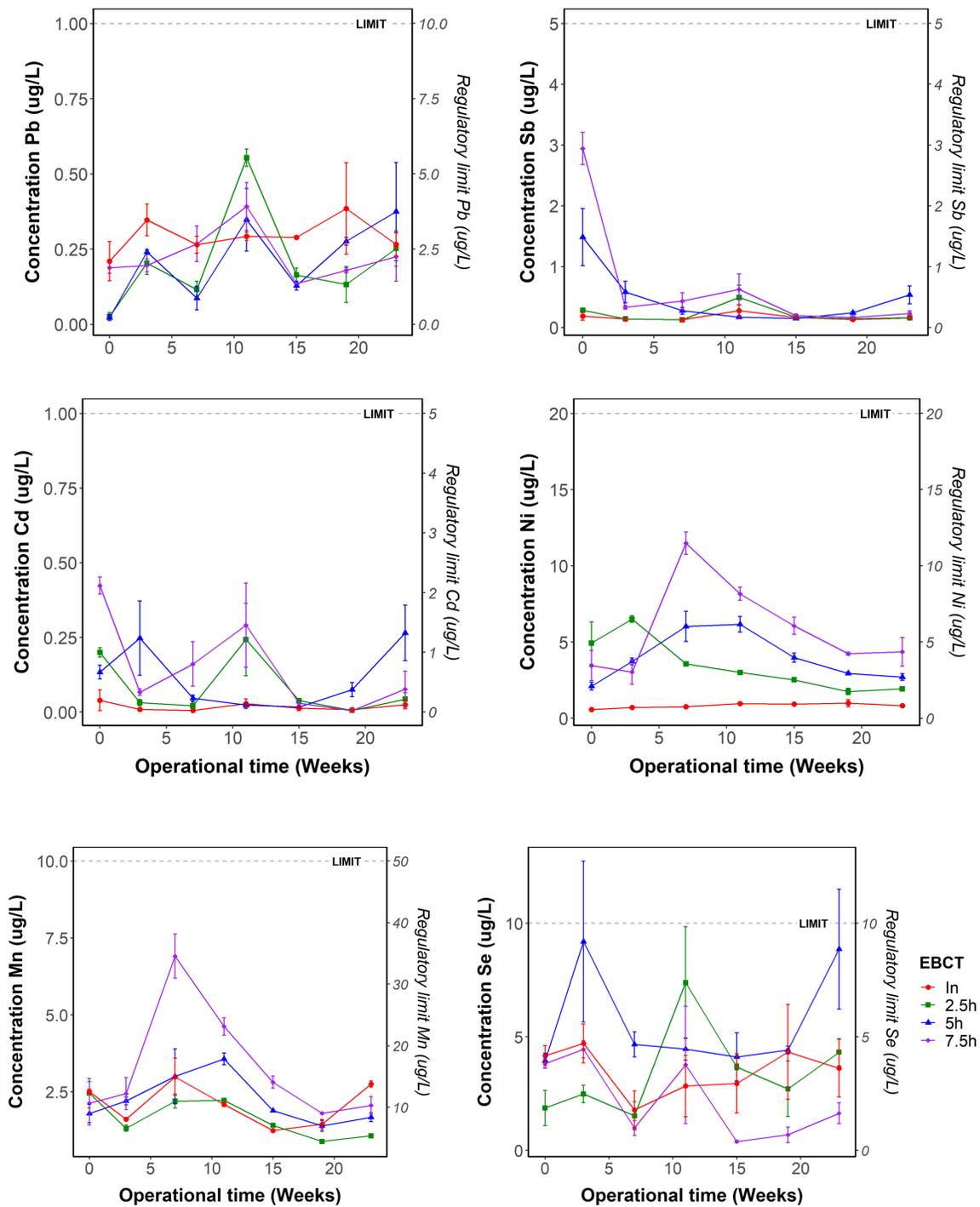


Figure A.8: Concentration of lead (Top left), antimony (Top right), cadmium (Middle left), nickel (Middle right), manganese (Bottom left), and selenium (Bottom right). Left y-axis shows the measured concentration of the influent and the $h=30, 60, 90$ cm effluent (mean \pm s.e., $n=3$ replicates) over 23 weeks of operation. The right axis together with the dashed line (when present) highlights the regulatory limits.

Table A.1: *Heatmap sample names*

Week	Length	EBCT	BV	Name
0	30	2.5	0	W0-30cm-BV0
0	60	5.0	0	W0-60cm-BV0
0	90	7.5	0	W0-90cm-BV0
3	30	2.5	180	W3-30cm-BV180
3	60	5.0	90	W3-60cm-BV90
3	90	7.5	60	W3-90cm-BV60
7	30	2.5	470	W7-30cm-BV470
7	60	5.0	240	W7-60cm-BV240
7	90	7.5	160	W7-90cm-BV160
11	30	2.5	720	W11-30cm-BV720
11	60	5.0	360	W11-60cm-BV360
11	90	7.5	240	W11-90cm-BV240
15	30	2.5	1000	W15-30cm-BV1000
15	60	5.0	500	W15-60cm-BV500
15	90	7.5	330	W15-90cm-BV330
19	30	2.5	1260	W19-30cm-BV1260
19	60	5.0	630	W19-60cm-BV630
19	90	7.5	420	W19-90cm-BV420
23	30	2.5	1520	W23-30cm-BV1520
23	60	5.0	630	W23-60cm-BV630
23	90	7.5	510	W23-90cm-BV510

Appendix B

Appendix chapter 4

B.1 Spectral correction

The emission spectral correction vector was measured according to the method of Lakowicz (2006), Parambath (2016), Ryder et al. (2017). European pharmaceutical reference standards (Phenyl-alanine (PA), Tyrosine (TY), Tryptophan (TR), and Quinine Sulphate (QS)) (Sigma Aldrich, UK) were used; reagents and materials were weighted with a precise electronic balance (Ohaus Explorer, Cole-Parmer); all standards except quinine sulphate had a concentration of 10^{-4} to 10^{-5} M with no added buffer and pH 7; quinine sulphate in 0.02 M $\text{H}_2\text{SO}_4^{2-}$ had a concentration of 15 mg/L and pH 1. Samples were analysed within 8 h.

The fluorescence emission signals of Phenyl-alanine, Tyrosine, Tryptophan and Quinine Sulphate (PA, Ty, Tr and QS, respectively) were measured by the plate reader (see chapter 4 section 4.2 pg. 82 for settings) at excitation wavelength 260, 270, 280, 345.6 nm, respectively, over the emission wavelength range of 290 - 320, 300 - 350, 310 - 450 and 380 - 500 nm, respectively. For every analyte, the signals was normalized with respect to its emission peak. Afterwards, the correction factors were calculated for each wavelength by dividing the measured intensity by the theoretical intensity. CF at emission wavelength 290 - 300 nm (PA), 300 - 340 nm (TY) and 340 - 410 nm (TR) and QS (410 - 500 nm) were selected and aligned to produce one curve according to the method of a Certified Calibration Kit - Spectral Fluorescence Standards Hoffmann & Monte (2006). CF of 280 and 510 - 550 nm were determined by extrapolation of the produced curve.

Most of the emission correction was needed in the ultraviolet region (280 - 400 nm). The measured (circles scatter) and theoretical (triangle scatter) spectra of phenyl-alanine (red), tyrosine (yellow), tryptophan (green) and quinine sulphate (purple) are given in Figure B.1. The calculated combined emission correction curve is in Figure B.2. These emission correction values are used in further PARAFAC analysis by the PR method.

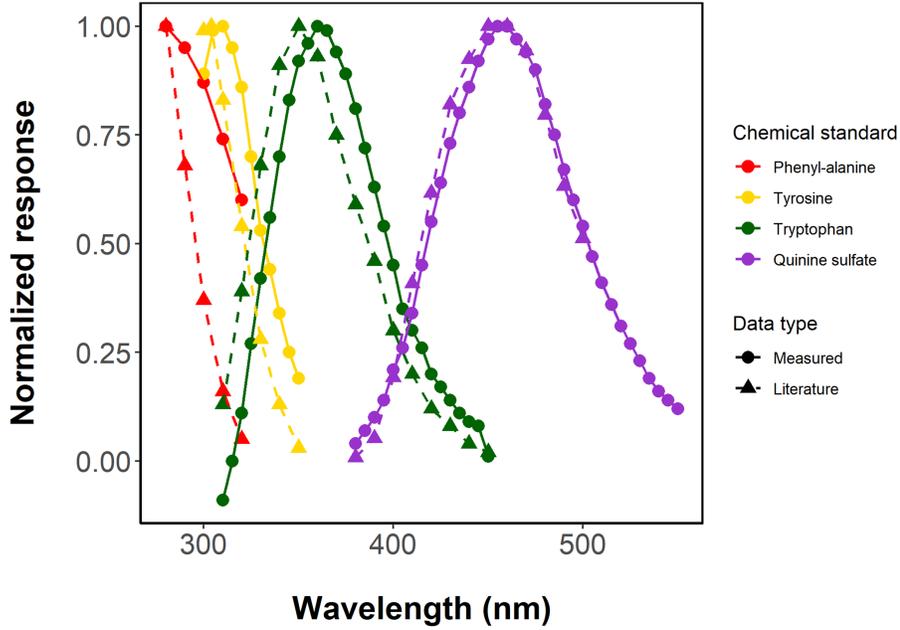


Figure B.1: Normalized fluorescence emission spectra of 4 standards analysed by the Tecan M200 Pro plate reader (measured) and Chen (1967) (theoretical).

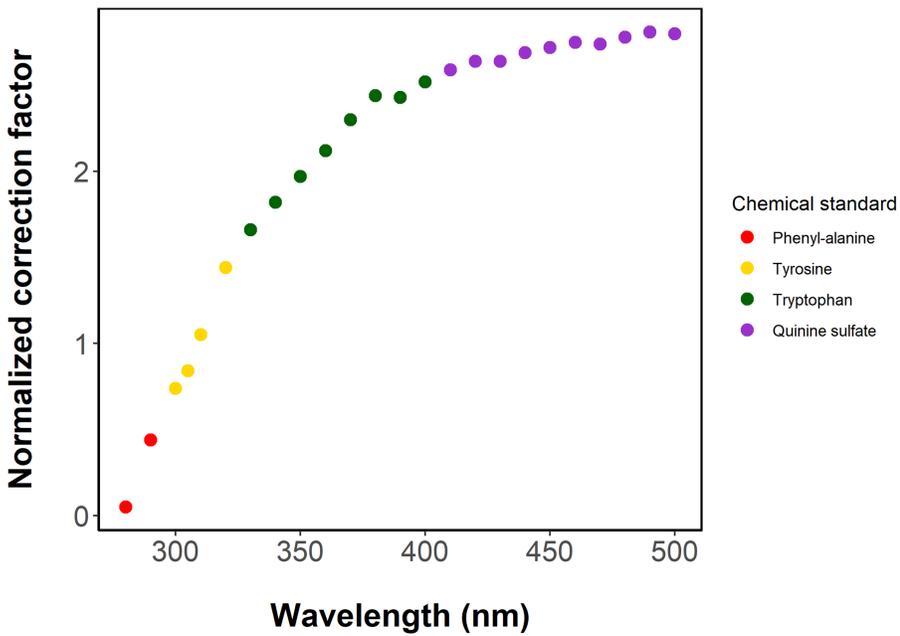


Figure B.2: Combined emission correction values.

B.2 EEMQual

Figure B.3 A - E, respectively. Model 3 shows a clear drop at component 3, meaning that a 2-component model is the correct choice. The EEMQual of 4 components is too small to be correct. However the three component model needed to be inspected. For all models, the three component models included the modeling of background noise, while the two component models clearly included analyte signal. Therefore the two component model was chosen.

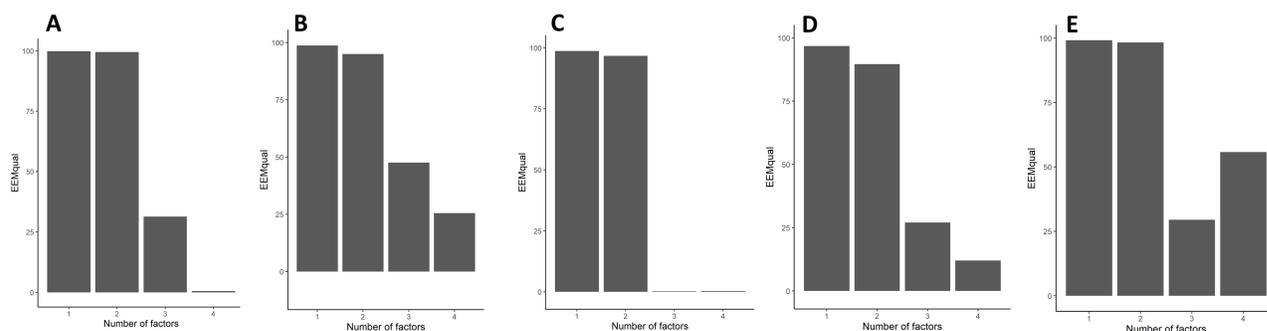


Figure B.3: *EEMQual* for models 1-5 given by A-E, respectively.

The EEMQual of validation models 6 and 7 are given in Figure B.4 A - B, respectively. Model 6 shows a clear drop at 4 components, meaning that a 3-component model is the correct choice. For model 7, a clear drop was not present, but a step-wise decline of EEMQual over number of components. The three and four component model needed to be inspected. Both the 3 and 4 component model included the modeling of background noise, while the two component models clearly included analyte signal. Therefore the two component model was chosen.

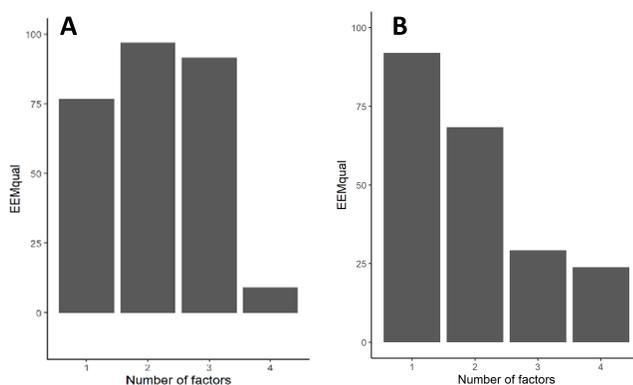


Figure B.4: *EEMQual* of models 6 and 7 given by A-B, respectively.

Appendix C

Appendix chapter 5

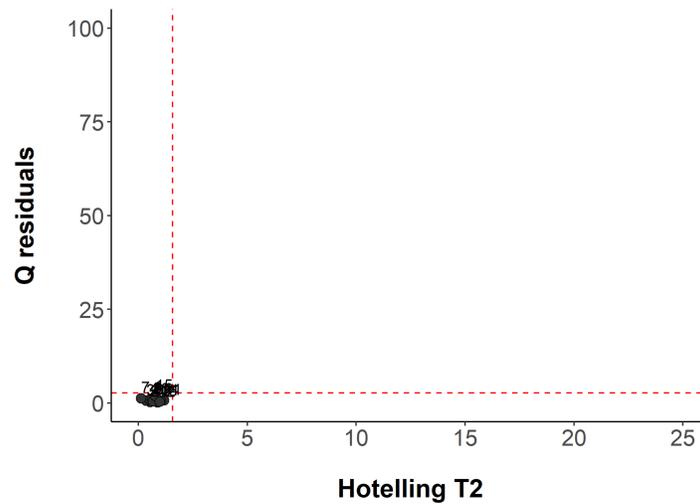


Figure C.1: $T^2 - Q$ residual plot for the detection of outliers in data set *Time*. No outliers found

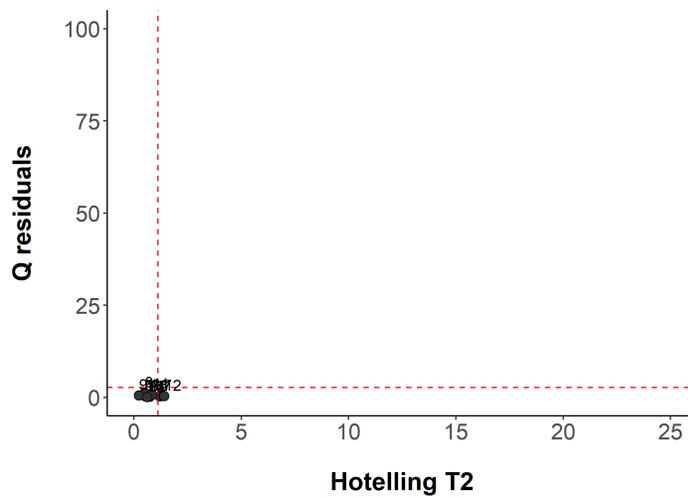


Figure C.2: $T^2 - Q$ residual plot for the detection of outliers in data set *Length*. No outliers found

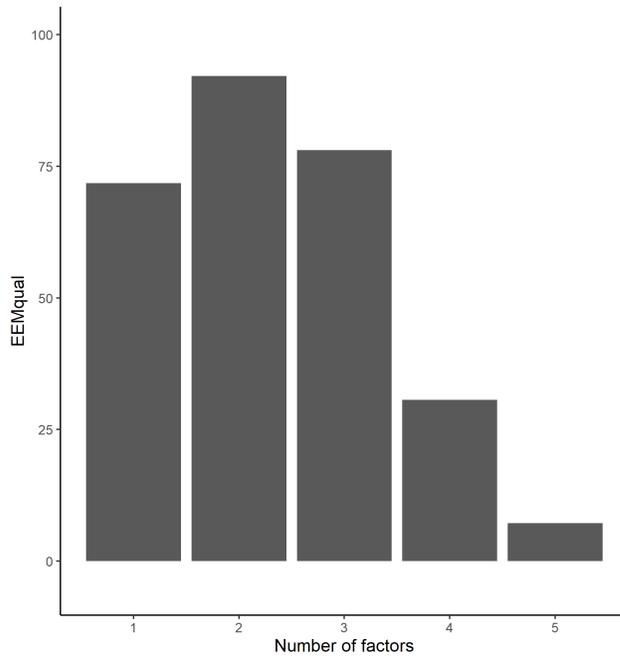


Figure C.3: *EEMqual* for the pore water data set. High *EEMqual* is shown for up to 3 component, while at the 4th component the *EEMqual* drops. As a result *EEMqual* indicates 3 components as the correct number.

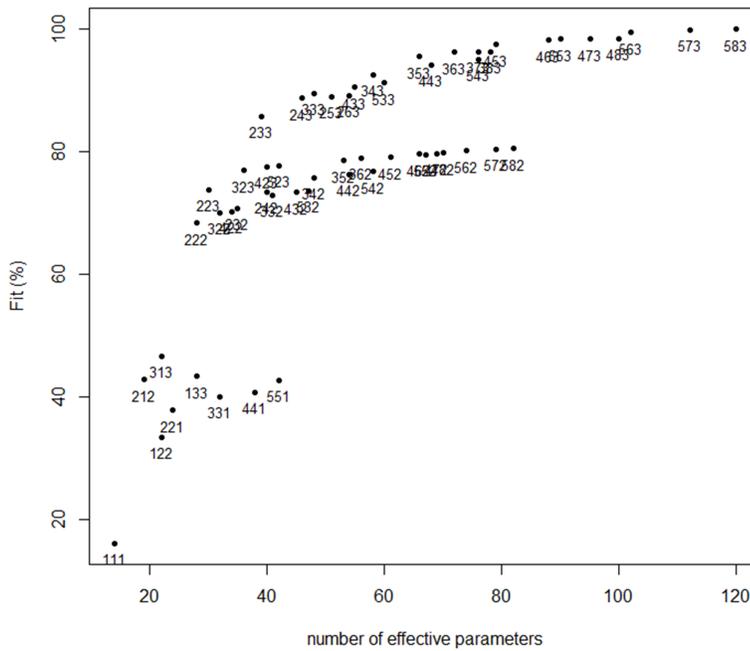


Figure C.4: Scree plot of the Tucker3 model of data set Week with the percentage variance explained by the model when using up to 5, 8, and 3 components in the *P*, *Q*, and *R* modes.

C.1 PARAFAC in- and effluent water

PARAFAC model of the in- effluent biofilter water samples of week 5,9 23 ($n = 54$). EEMQual data showed that the number of components is 3 or 4 (Figure C.5). Further inspection of the 3 and 4 component models showed that the 4 component model was similar to that of the three component model. The EEMs of the 3 component model are given in Figure C.6.

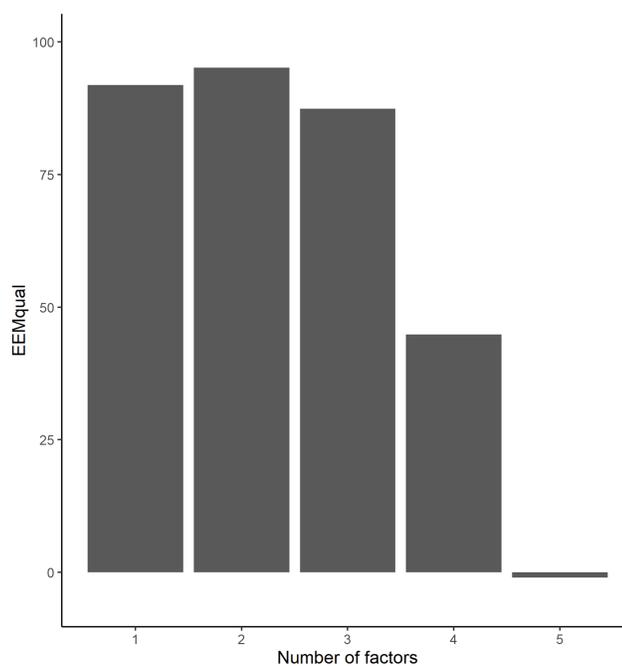


Figure C.5: *EEMQual* of the biofilter in- effluent data set.

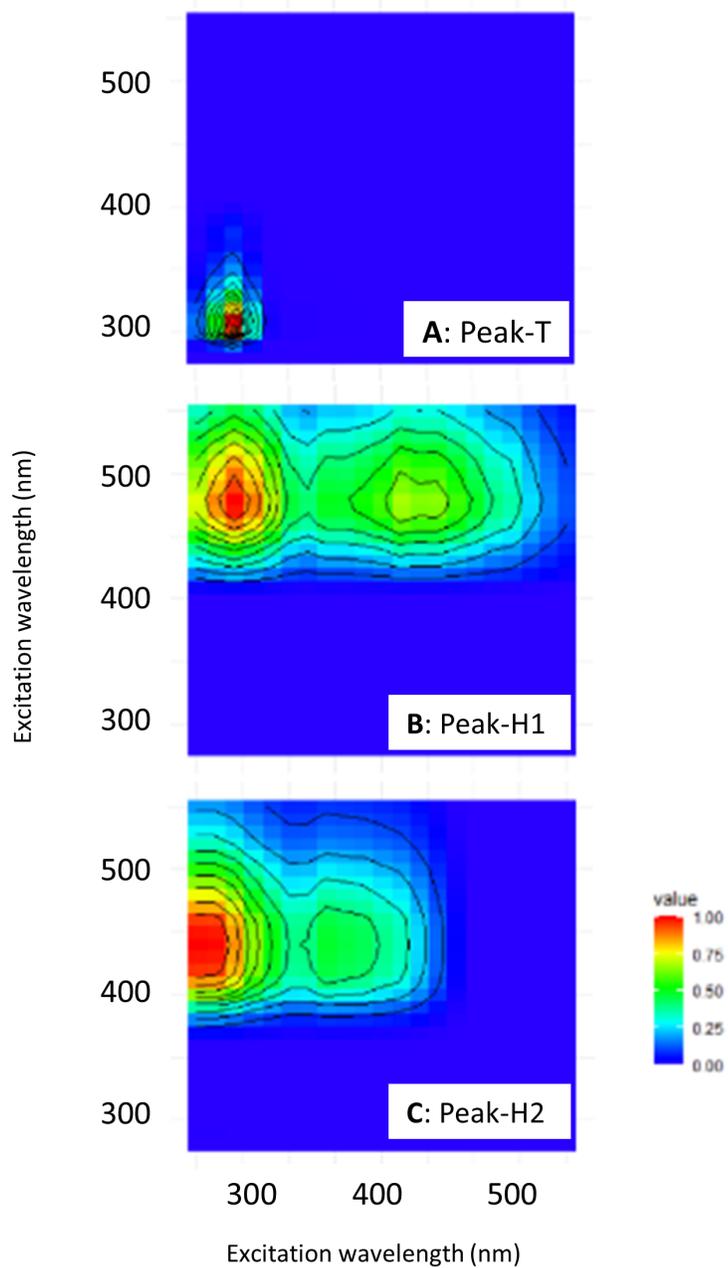


Figure C.6: EEM of the PARAFAC component 1 Peak-T (A), component 2 Peak-H1 (B), and component 3 Peak-H2 (C) present in the in effluent water.

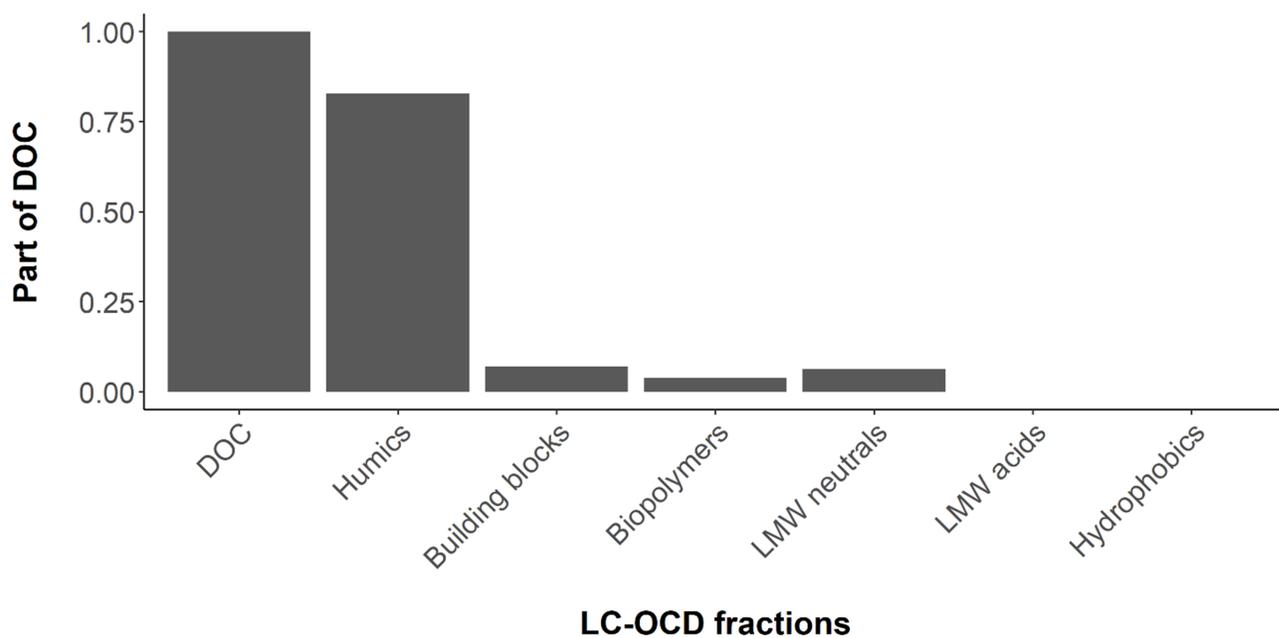


Figure C.7: LC-OCD fractions and their contribution to the whole DOC content of the biofilter influent water.

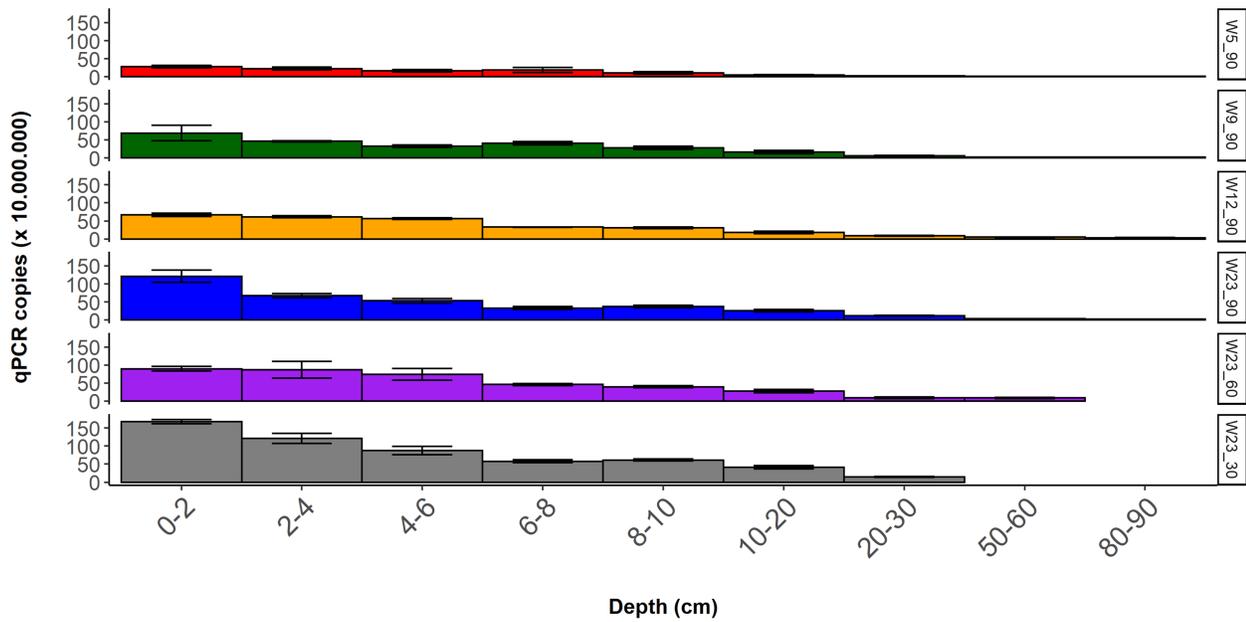


Figure C.8: *qPCR copies measured at the various depths of the biofilter (mean \pm s.e., $n = 3$ replicates) at different weeks (5, 9, 12, and 23) and lengths (week 23: 30, 60, and 90 cm).*

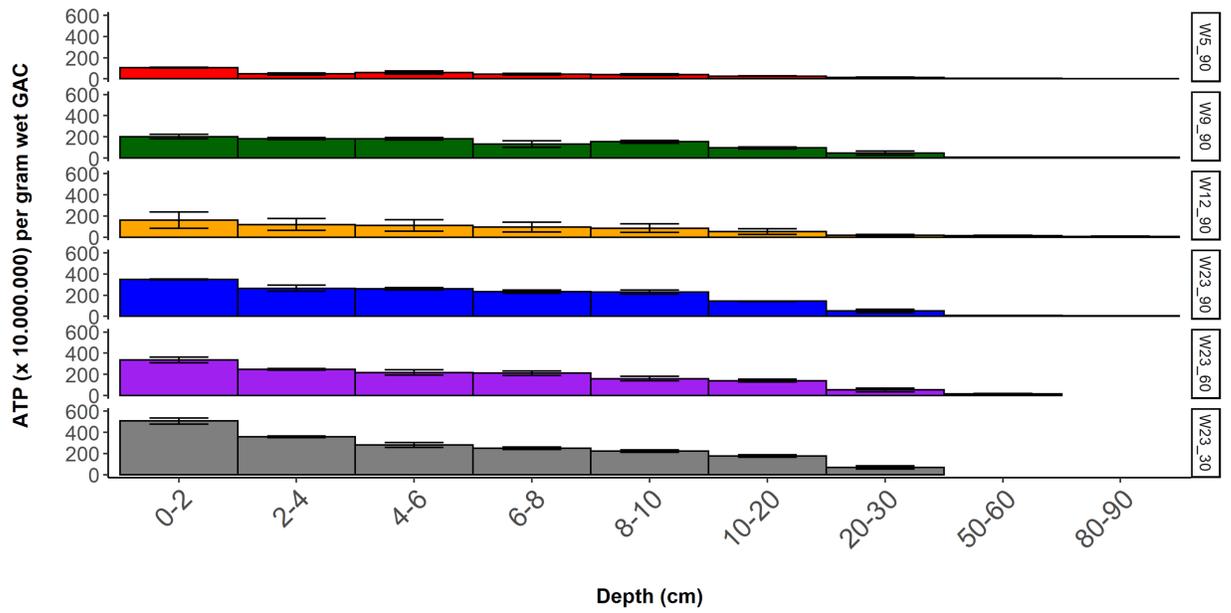


Figure C.9: ATP per gram wet GAC measured at the various depths of the biofilter (mean \pm s.e., $n = 3$ replicates) at different weeks (5, 9, 12, and 23) and lengths (week 23: 30, 60, and 90 cm).

Appendix D

Appendix chapter 6

Table D.1: *Measured LC-OCD fraction concentration (mean \pm se, n=3) at $T = 0$ and in the three treatments (TOP, MID, BOT) at $T = 23$.*

	DOC	HS	MW_HS	BB	LMWN	LMWA	BP
T0	12.19 \pm 0.03	10.08 \pm 0.06	0.62 \pm 0.02	0.98 \pm 0.04	0.69 \pm 0.02	0.01 \pm 0.004	0.40 \pm 0.03
T	12.35 \pm 0.04	10.05 \pm 0.07	0.80 \pm 0.02	1.07 \pm 0.07	0.70 \pm 0.03	0.000	0.52 \pm 0.03
M	12.41 \pm 0.10	8.69 \pm 1.97	0.81 \pm 0.06	2.46 \pm 2.03	0.72 \pm 0.07	0.003 \pm 0.001	0.44 \pm 0.12
B	12.18 \pm 0.06	9.81 \pm 0.05	0.87 \pm 0.03	1.20 \pm 0.06	0.65 \pm 0.03	0.005 \pm 0.001	0.53 \pm 0.05

Appendix E

Appendix chapter 7

Table E.1: *data set Waterquality pre-treatment*

Nr. samples	Pre-treatment steps
24486	
↓	Identical entries removed
24371	
↓	Entries with NA removed
24353	
↓	Keep raw water data only
9832	
↓	Keep 2015-2019
6772	
↓	Remove < and > sign
6772	
↓	Take mean of data from same day
6550	
↓	Unfold data to different structure with variables in column
1005	
↓	Keep variables of interest
1005	
↓	Take median of measurement done in the same month over the years
84	
↓	Remove Forehill
42	

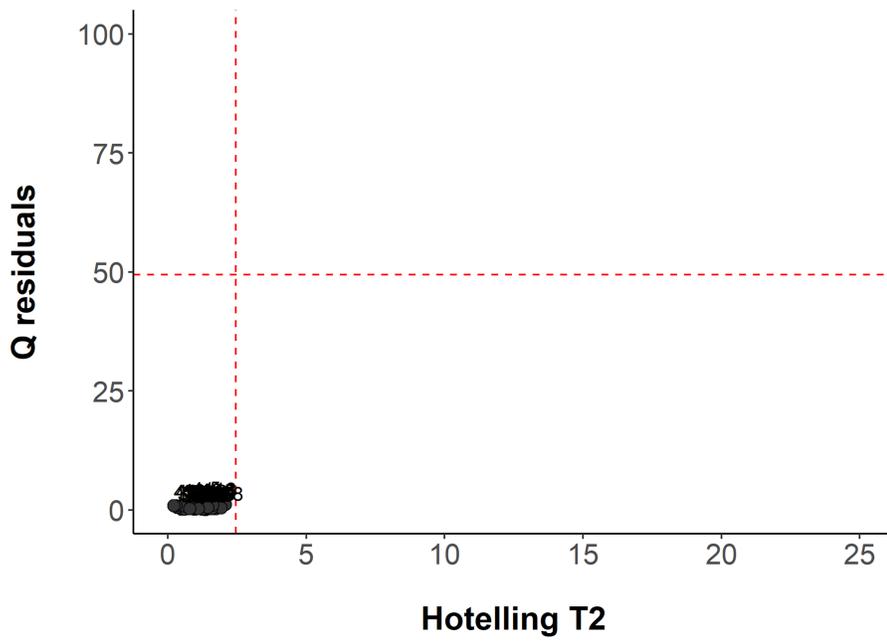


Figure E.1: *Hotelling T^2 - Q residuals plot for the SW data set; no outliers detected.*

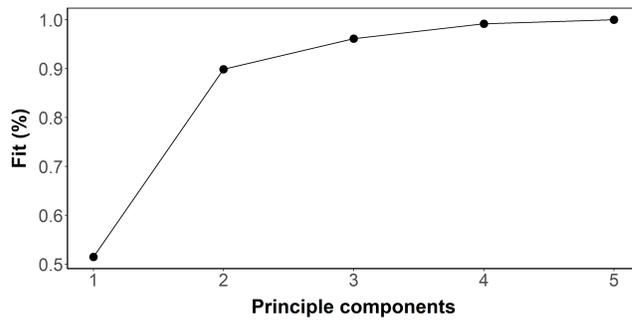


Figure E.2: *PCA scree plot of the data set Waterquality in which the increase of principle components results in a higher fit of the data.*

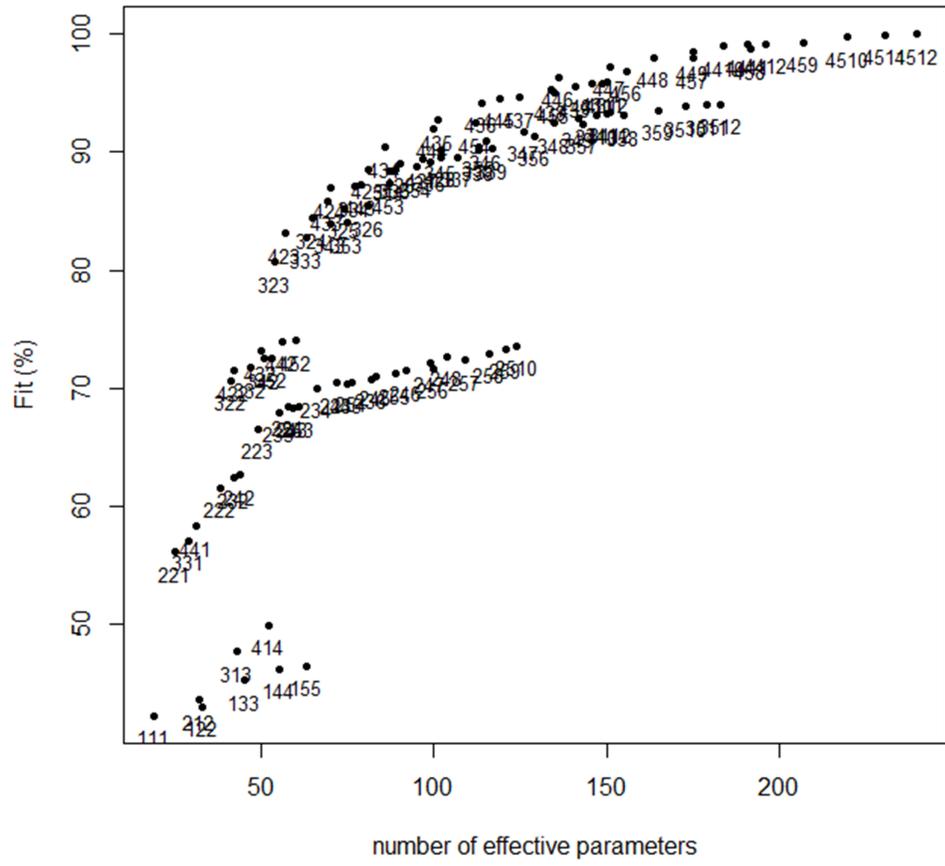


Figure E.3: Tucker3 model scree plot of the SW data set with the percentage variance explained by the model when using up to 4, 5, and 12 components in the P, Q, and R modes.

Bibliography

- Adusei-Gyamfi, J., Ouddane, B., Rietveld, L., Cornard, J.-p. & Criquet, J. (2019), 'Natural organic matter-cations complexation and its impact on water treatment. A critical review.', *Water Research* **160**, 130–147.
- Aitkenhead-Peterson, J. A., Smart, R. P., Aitkenhead, M. J., Cresser, M. S. & McDowell, W. H. (2007), 'Spatial and temporal variation of dissolved organic carbon export from gauged and ungauged watersheds of Dee Valley, Scotland: Effect of land cover and C:N', *Water Resources Research* **43**(5), 1–11.
- Aktaş, & Çeçen, F. (2006), 'Effect of activation type on bioregeneration of various activated carbons loaded with phenol', *Journal of Chemical Technology and Biotechnology* **81**(7), 1081–1092.
- Aktaş, & Çeçen, F. (2007a), 'Adsorption, desorption and bioregeneration in the treatment of 2-chlorophenol with activated carbon', *Journal of Hazardous Materials* **141**(3), 769–777.
- Aktaş, & Çeçen, F. (2007b), 'Bioregeneration of activated carbon: A review', *International Biodeterioration and Biodegradation* **59**(4), 257–272.
- Al-Attas, O. G., Narbaitz, R. M. & Douglas, I. (2018), 'Iron and NOM interactions in GAC groundwater treatment', *Water Quality Research Journal* **53**(3), 105–117.
- Álvarez-Torrellas, S., Rodríguez, A., Ovejero, G. & García, J. (2016), 'Comparative adsorption performance of ibuprofen and tetracycline from aqueous solution by carbonaceous materials', *Chemical Engineering Journal* **283**, 936–947.
- American Water Works Association (1999), *Water quality & treatment*, 5 edn, McGraw-Hill, New York.
- Amigo, J. M. (2021), 'Data mining, machine learning, deep learning, chemometrics: Definitions, common points and trends (Spoiler Alert: VALIDATE your models!)', *Brazilian Journal of Analytical Chemistry* **8**(32), 22–38.
- Amon, R. M. & Benner, R. (1996), 'Bacterial utilization of different size classes of dissolved organic matter', *Limnology and Oceanography* **41**(1), 41–51.
- Andersson, C. A. (1998), 'Interactive introduction to multi-way analysis in MATLAB'.
URL: <http://www.models.life.ku.dk/courses/parafac/chap8tuckerintro.htm>

Ando, N., Matsui, Y., Kurotobi, R., Nakano, Y., Matsushita, T. & Ohno, K. (2010), 'Comparison of natural organic matter adsorption capacities of super-powdered activated carbon and powdered activated Carbon', *Water Research* **44**(14), 4127–4136.

URL: <http://dx.doi.org/10.1016/j.watres.2010.05.029>

Arnold, M., Batista, J., Dickenson, E. & Gerrity, D. (2018), 'Use of ozone-biofiltration for bulk organic removal and disinfection byproduct mitigation in potable reuse applications', *Chemosphere* **202**, 228–237.

URL: <https://doi.org/10.1016/j.chemosphere.2018.03.085>

Baghoth, S. A., Sharma, S. K. & Amy, G. L. (2011), 'Tracking natural organic matter (NOM) in a drinking water treatment plant using fluorescence excitation-emission matrices and PARAFAC', *Water Research* **45**(2), 797–809.

URL: <http://dx.doi.org/10.1016/j.watres.2010.09.005>

Bai, Y. H., Liu, R. P., Liang, J. S. & Qu, J. H. (2013), 'Integrated Metagenomic and Physiochemical Analyses to Evaluate the Potential Role of Microbes in the Sand Filter of a Drinking Water Treatment System', *Plos One* **8**(4).

Baker, M. (1948), *The Quest for Pure Water: The History of Water Purification from the Earliest Records to the Twentieth Century*, 2 edn, American Water Works Association.

Barsotti, F., Ghigo, G. & Vione, D. (2016), 'Computational assessment of the fluorescence emission of phenol oligomers: A possible insight into the fluorescence properties of humic-like substances (HULIS)', *Journal of Photochemistry and Photobiology A: Chemistry* **315**, 87–93.

URL: <http://dx.doi.org/10.1016/j.jphotochem.2015.09.012>

Basu, O. D., Dhawan, S. & Black, K. (2016), 'Applications of biofiltration in drinking water treatment - a review', *Journal of Chemical Technology and Biotechnology* **91**(3), 585–595.

Bhatnagar, A., Hogland, W., Marques, M. & Sillanpää, M. (2013), 'An overview of the modification methods of activated carbon for its water treatment applications', *Chemical Engineering Journal* **219**, 499–511.

URL: <http://dx.doi.org/10.1016/j.cej.2012.12.038>

Bhatnagar, A. & Sillanpää, M. (2017), 'Removal of natural organic matter (NOM) and its constituents from water by adsorption – A review', *Chemosphere* **166**, 497–510.

- Bielefeldt, A. R. (2011), *Appropriate and sustainable water disinfection methods for developing communities*, number January 2011.
- Bieroza, M., Baker, A. & Bridgeman, J. (2011), 'Exploratory analysis of excitation – emission matrix fluorescence spectra with self-organizing maps — A tutorial', *Education for Chemical Engineers* **7**(1), 22–31.
URL: <http://dx.doi.org/10.1016/j.ece.2011.10.002>
- Biswas, S. & Mishra, U. (2015), 'Continuous Fixed-Bed Column Study and Adsorption Modeling : Removal of Lead Ion from Aqueous Solution by Charcoal Originated from Chemical Carbonization of Rubber Wood Sawdust', *Journal of Chemistry* **2015**.
- Bjelopavlic, M., Newcombe, G. & Hayes, R. (1999), 'Adsorption of NOM onto activated carbon: Effect of surface charge, ionic strength, and pore volume distribution', *Journal of Colloid and Interface Science* **210**(2), 271–280.
- Boehm, H. (1994), 'Some aspects of the surface chemistry of carbon blacks and other carbons', *Carbon* **32**(5), 759–769.
- Bond, B. Y. R. G. & Digiano, F. A. (2004), 'Evaluating GAC performance using the ICR dataset', *AWWA* **96**(6).
- Boon, N., Pycke, B. F. G., Marzorati, M. & Hammes, F. (2011), 'Nutrient gradients in a granular activated carbon biofilter drives bacterial community organization and dynamics', *Water Research* **45**(19), 6355–6361.
URL: <http://dx.doi.org/10.1016/j.watres.2011.09.016>
- Bourgin, M., Beck, B., Boehler, M., Borowska, E., Fleiner, J., Salhi, E., Teichler, R., von Gunten, U., Siegrist, H. & McArdell, C. S. (2018), 'Evaluation of a full-scale wastewater treatment plant upgraded with ozonation and biological post-treatments: Abatement of micropollutants, formation of transformation products and oxidation by-products', *Water Research* **129**, 486–498.
- Bridgeman, J., Bieroza, M. & Baker, A. (2011), 'The application of fluorescence spectroscopy to organic matter characterisation in drinking water treatment', *Reviews in Environmental Science and Biotechnology* **10**(3), 277–290.
- Bro & Amigo (2020), 'Lecture notes of the online chemometrics course from the University of Copenhagen'.

- Bro, R. (1997), 'PARAFAC . Tutorial and applications', **38**, 149–171.
- Bro, R. & Kiers, H. A. (2003), 'A new efficient method for determining the number of components in PARAFAC models', *Journal of Chemometrics* **17**(5), 274–286.
- Bro, R. & Vidal, M. (2011), 'EEMizer: Automated modeling of fluorescence EEM data', *Chemometrics and Intelligent Laboratory Systems* **106**(1), 86–92.
URL: <http://dx.doi.org/10.1016/j.chemolab.2010.06.005>
- Brouwer, P. & Kroonenberg, P. M. (1991), 'Some notes on the diagonalization of the extended three-mode core matrix', *Journal of Classification* **8**(1), 93–98.
- Brown, J., Upadhyaya, G., Nyfennegger, J., Pope, G., Evans, A. N., Carter, J. T., Black, S., Alito, C., Lauderdale, C., Hooper, J., Stephens, L. W., Dickenson, E., Wert, E. & Weinrich, L. (2020), Biofiltration Guidance Manual for Drinking Water Facilities Biofiltration Guidance Manual for Drinking Water Facilities, Technical Report Project nr. 4719.
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. & Knight, R. (2010), 'QIIME allows analysis of high-throughput community sequencing data', *Nature Methods* **7**(5), 335–336.
- Cardenas, J. (2008), Granular Activated Carbon Adsorption of Dissolved Organic, PhD thesis, University of Florida.
- Carlson, K. H. & Amy, G. L. (1998), 'BOM removal during biofiltration', *Journal / American Water Works Association* **90**(12), 42–52.
- Carlson, K. H. & Amy, G. L. (2001), 'Ozone and biofiltration optimization for multiple objectives', *Journal / American Water Works Association* **93**(1), 88–98.
- Carroll, J. & Chang, J. D. (1970), 'Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition', *Psychometrika* **35**, 283–319.
- Carstea, E. M., Bridgeman, J., Baker, A. & Reynolds, D. M. (2016), 'Fluorescence spectroscopy for wastewater monitoring: A review', *Water Research* **95**, 205–219.
URL: <http://dx.doi.org/10.1016/j.watres.2016.03.021>

- Carstea, E. M., Popa, C. L., Baker, A. & Bridgeman, J. (2020), 'In situ fluorescence measurements of dissolved organic matter: A review', *Science of the Total Environment* **699**.
URL: <https://doi.org/10.1016/j.scitotenv.2019.134361>
- Cattell, R. B. (1944), 'Parallel proportional profiles and other principles for determining the choice of factors by rotation', *Psychometrika* **9**, 267–283.
- Chen, F., Peldszus, S., Elhadidy, A. M., Legge, R. L., Van Dyke, M. I. & Huck, P. M. (2016), 'Kinetics of natural organic matter (NOM) removal during drinking water biofiltration using different NOM characterization approaches', *Water Research* **104**, 361–370.
URL: <http://dx.doi.org/10.1016/j.watres.2016.08.028>
- Chen, L., Zhai, Y., van der Mark, E., Liu, G., van der Meer, W. & Medema, G. (2021), 'Microbial community assembly and metabolic function in top layers of slow sand filters for drinking water production', *Journal of Cleaner Production* **294**.
- Chen, R. F. (1967), 'Fluorescence quantum yields of tryptophan and tyrosine', *Analytical Letters* **1**(1), 35–42.
- Chen, W., Westerhoff, P., Leenheer, J. A. & Booksh, K. (2003), 'Fluorescence Excitation-Emission Matrix Regional Integration to Quantify Spectra for Dissolved Organic Matter', *Environmental Science and Technology* **37**(24), 5701–5710.
- Chen, Z., Zhang, H., Wu, C., Wang, Y. & Li, W. (2015), 'A study of electrosorption selectivity of anions by activated carbon electrodes in capacitive deionization', *Desalination* **369**, 46–50.
- Chowdhury, Z., Traviglia, A., Carter, J., Brown, T., Summers, R. S., Corwin, C., Zearley, T., Thurman, M., Ferrara, I., Olson, J. & Barron, P. (2010), 'Cost-Effective Regulatory Compliance With GAC Biofilters'.
- Coble, P. G. (1996), 'Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy', *Marine Chemistry* **51**(4), 325–346.
- Coble, P., Spencer, R., Baker, A. & Reynolds, D. (2014), Aquatic Organic Matter Fluorescence, in P. Coble, J. Lead, A. Baker, D. Reynolds & R. Spencer, eds, 'Aquatic Organic Matter Fluorescence', Cambridge University Press, pp. 75–122.
- Crittenden, C., Rhodes Trussell, R., Hand, D., Howe, K. & Tchobanoglous, G. (2005), *Water treatment: principles and design*, 2 edn, Wiley.

- Croue, J. P., Korshin, G. V. & Benjamin, M. M. (1999), *Isolation, fractionation and characterization of natural organic matter in drinking water*, AWWA Research Foundation and.
- Cuss, C. W. & Guéguen, C. (2015), 'Characterizing the labile fraction of dissolved organic matter in leaf leachates: Methods, indicators, structure, and complexity', *Labile Organic Matter—Chemical Compositions, Function, and Significance in Soil and the Environment* (62), 237–274.
- De Jonge, R. J., Breure, A. M. & Van Aniel, J. G. (1996), 'Bioregeneration of powdered activated carbon (PAC) loaded with aromatic compound', *Water Research* **30**(4), 875–882.
- De Smedt, F. & Wierenga, P. J. (1979), 'Mass transfer in porous media with immobile water', *Journal of Hydrology* **41**(1-2), 59–67.
- Demiral, H. & Gündüzoğlu, G. (2010), 'Removal of nitrate from aqueous solutions by activated carbon prepared from sugar beet bagasse', *Bioresource Technology* **101**(6), 1675–1680.
- Deng, L., Lu, B., Li, J., Lv, G., Du, S., Shi, J. & Yang, Y. (2017), 'Effect of pore structure and oxygen-containing groups on adsorption of dibenzothiophene over activated carbon', *Fuel* **200**, 54–61.
URL: <http://dx.doi.org/10.1016/j.fuel.2017.03.018>
- Dong, J. D., Zhang, Y. Y., Zhang, S., Wang, Y. S., Yang, Z. H. & Wu, M. L. (2010), 'Identification of temporal and spatial variations of water quality in Sanya Bay, China by three-way principal component analysis', *Environmental Earth Sciences* **60**(8), 1673–1682.
- Drinking Water Quality Regulator (2020a), *Drinking Water Quality in Scotland 2020 - Public Water Supply*, Technical report.
- Drinking Water Quality Regulator (2020b), *Supporting Information 2020*, Technical report.
- Eden, R. e. a. (2014), *Encyclopedia of Food Microbiology*, 2 edn.
- El Gamal, M., Mousa, H. A., El-Naas, M. H., Zacharia, R. & Judd, S. (2018), 'Bio-regeneration of activated carbon: A comprehensive review', *Separation and Purification Technology* **197**, 345–359.
- Elliott, S., Lead, J. R. & Baker, A. (2006), 'Characterisation of the fluorescence from freshwater, planktonic bacteria', *Water Research* **40**(10), 2075–2083.
- Emelko, M. B., Huck, P. M., Coffey, B. M. & Smith, E. F. (2006), 'Effects of media, backwash, and temperature on full-scale biological filtration', *Journal / American Water Works Association* **98**(12), 61–73.

- Engle, M., Gallo, M., Schroeder, K., Geboy, N. & Zupancic, J. (2014), 'Three-Way Compositional Analysis of Water Quality Monitoring Data', *Environmental and Ecological Statistics* .
- Escobar, I. C. & Randall, A. A. (2001), 'Assimilable organic carbon (AOC) and biodegradable dissolved organic carbon (BDOC): Complementary measurements', *Water Research* **35**(18), 4444–4454.
- Faria, P. C. C., Orf, J. J. M. & Pereira, M. (2004), 'Adsorption of anionic and cationic dyes on activated carbons with different surface chemistries', *Water Research* **38**, 2043–2052.
- Faris, W. G. & Copeland, R. A. (1997), 'Wavelength dependence of the Raman cross section for liquid water', *Appl. Opt.* **36**, 2686–2688.
- Farmer, R., Dussert, B. & Kovacic, S. (2015), 'Improved granular activated carbon for the stabilization of wastewater pH'.
- Farre, M. J., Reungoat, J., Argaud, F. X., Rattier, M. & Gernjak, W. (2011), 'Fate of N-nitrosodimethylamine, trihalomethane and haloacetic acid precursors in tertiary treatment including biofiltration', *Water Research* **5**, 5695–5704.
- FDA (2020), 'Methods, Method Verification and Validation, ORA Laboratory Manual Volume II', **II**.
- Feng, Y., Yu, Y., Duan, Q., Tan, J. & Zhao, C. (2010), 'The characteristic research of ammonium removal in grain-slag biological aerated filter (BAF)', *Desalination* **263**(1-3), 146–150.
URL: <http://dx.doi.org/10.1016/j.desal.2010.06.051>
- Finkbeiner, P., Moore, G., Pereira, R., Jefferson, B. & Jarvis, P. (2020), 'The combined influence of hydrophobicity, charge and molecular weight on natural organic matter removal by ion exchange and coagulation', *Chemosphere* **238**, 124633.
URL: <https://doi.org/10.1016/j.chemosphere.2019.124633>
- Foo, K. Y. & Hameed, B. H. (2010), 'Insights into the modeling of adsorption isotherm systems', *Chemical Engineering Journal* **156**(1), 2–10.
- Fox, B. G., Thorn, R. M., Anesio, A. M. & Reynolds, D. M. (2017), 'The in situ bacterial production of fluorescent organic matter; an investigation at a species level', *Water Research* **125**, 350–359.
URL: <https://doi.org/10.1016/j.watres.2017.08.040>

- Frisbie, S. H., Mitchell, E. J., Dustin, H., Maynard, D. M. & Sarkar, B. (2012), 'World health organization discontinues its drinking-water guideline for manganese', *Environmental Health Perspectives* **120**(6), 775–778.
- Galvis, G., Latorre, J. & Galvis, A. (2002), 'Multi-stage filtration technology'.
URL: https://www.sswm.info/sites/default/files/reference_attachments/SMET2002Multi-stageFiltrationTechnologyChapter16.pdf
- Gauchotte-Lindsay, C., Aspray, T. J., Knapp, M. & Ijaz, U. Z. (2019), 'Systems biology approach to elucidation of contaminant biodegradation in complex samples-integration of high-resolution analytical and molecular tools', *Faraday Discussions* **218**, 481–504.
- Gauden, P. A., Terzyk, A. P., Furmaniak, S., Włoch, J., Kowalczyk, P. & Zieliński, W. (2014), 'MD simulation of organics adsorption from aqueous solution in carbon slit-like pores. Foundations of the pore blocking effect', *Journal of Physics Condensed Matter* **26**(5).
- Gauden, P. A., Terzyk, A. P., Jaroniec, M. & Kowalczyk, P. (2007), 'Bimodal pore size distributions for carbons: Experimental results and computational studies', *Journal of Colloid and Interface Science* **310**(1), 205–216.
- Gerke, J. (2018), 'Concepts and misconceptions of humic substances as the stable part of soil organic matter: A review', *Agronomy* **8**(5).
- Gerrity, D., Gamage, S., Holady, J. C., Mawhinney, D. B., Quiñones, O., Trenholm, R. A. & Snyder, S. A. (2011), 'Pilot-scale evaluation of ozone and biological activated carbon for trace organic contaminant mitigation and disinfection', *Water Research* **45**(5), 2155–2165.
- Ghosh, U., Weber, A. S., Jensen, J. N. & Smith, J. R. (1999), 'Granular activated carbon and biological activated carbon treatment of dissolved and sorbed polychlorinated biphenyls', *Water Environment Research* **71**(2), 232–240.
- Gibert, O., Lefèvre, B., Fernández, M., Bernat, X., Paraira, M., Calderer, M. & Martínez-Lladó, X. (2013), 'Characterising biofilm development on granular activated carbon used for drinking water production', *Water Research* **47**(3), 1101–1110.
- Giordani, P., Kiers, H. & Del Ferraro, M. (2015), 'Package 'ThreeWay'', **57**(7).
- Giussani, B., Monticelli, D., Gambillara, R., Pozzi, A. & Dossi, C. (2008), 'Three-way principal component analysis of chemical data from Lake Como watershed', **88**, 160–166.

- Goletz, C., Wagner, M., Grübel, A., Schmidt, W., Korf, N. & Werner, P. (2011), ‘Standardization of fluorescence excitation-emission-matrices in aquatic milieu’, *Talanta* **85**(1), 650–656.
URL: <http://dx.doi.org/10.1016/j.talanta.2011.04.045>
- Goueguel, C. (2020), ‘Multivariate Outlier Detection in High-Dimensional Spectral Data’.
URL: <https://towardsdatascience.com/multivariate-outlier-detection-in-high-dimensional-spectral-data-45878fd0ccb8>
- Greiner bio-one (2020), ‘Microplate Selection Guide’.
URL: https://www.gbo.com/fileadmin/user_upload/Downloads/Brochures/BrochuresBioScience/F073048
- Grieco, S. (2021), ‘Implementing Granular Activated Carbon Systems : Important Design And Start-Up Considerations’.
- Grose, A. B., Smith, A. J., Donn, A., O’Donnell, J. & Welch, D. (1998), ‘Supplying high quality drinking water to remote communities in Scotland’, *Desalination* **117**, 107–117.
- Guchi, E. (2015), ‘Review on Slow Sand Filtration in Removing Microbial Contamination and Particles from Drinking Water’, *American Journal of Food and Nutrition* **3**(2), 47–55.
- Hammes, F. (2008), ‘A comparison of AOC methods used by the different TECHNEAU partners’.
- Hammes, F. A. & Egli, T. (2005), ‘New method for assimilable organic carbon determination using flow-cytometric enumeration and a natural microbial consortium as inoculum’, *Environmental Science and Technology* **39**(9), 3289–3294.
- Hanaki, K., Saito, T. & Matsuo, T. (1997), ‘Anaerobic treatment utilizing the function of activated carbon’, *Water Science and Technology* **35**(8), 193–201.
URL: [http://dx.doi.org/10.1016/S0273-1223\(97\)00167-4](http://dx.doi.org/10.1016/S0273-1223(97)00167-4)
- Hara, R. B. O. & Kotze, D. J. (2010), ‘Do not log-transform count data’, *Methods in Ecology and Evolution* **1**, 118–122.
- Harshman, R. A. (1970), ‘Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multi-modal factor analysis’, *UCLA Working Papers in Phonetics* **16**, 1–84.
- Harshman, R. A. & Sarbo, W. S. D. (1984), An Application of PARAFAC to a Small Sample, Demonstrating Preprocessing, Orthogonality Constraints, and Split-Half Diagnostic Techniques, in H. G. Law, C. W. Snyder, J. A. Hattie & R. P. McDonald, eds, ‘Research methods for ultimode data analysis’, Praeger, New York, pp. 602–642.

- Hart, B. T. & Davies, H. R. (1981), 'Trace Metal Speciation in the Freshwater and Estuarine Regions of the Yarra River , Victoria', *Estuarine, Coastal and Shelf Science* **12**, 353–374.
- Heal, K. V. (2001), 'Manganese and land-use in upland catchments in Scotland', *The Science of the Total Environment* **265**, 169–179.
- Henderson, R. K., Baker, A., Murphy, K. R., Hambly, A., Stuetz, R. M. & Khan, S. J. (2009), 'Fluorescence as a potential monitoring tool for recycled water systems: A review', *Water Research* **43**(4), 863–881.
URL: <http://dx.doi.org/10.1016/j.watres.2008.11.027>
- Henrion, R. (1993), 'Body diagonalization of core matrices in three-way principal components analysis: Theoretical bounds and simulation', *Journal of Chemometrics* **7**(6), 477–494.
- Herzberg, M., Dosoretz, C. G., Tarre, S. & Green, M. (2003), 'Patchy biofilm coverage can explain the potential advantage of BGAC reactors', *Environmental Science and Technology* **37**(18), 4274–4280.
- Hijnen, W., Schurer, R., Martijn, B., Bahlman, J., Hoogenboezem, W. & van der Wielen, P. (2014), 'Removal of easily and more complex biodegradable NOM by full-scale BAC filters to produce biological stable drinking water'.
- Hitchcock, F. L. (1927), 'The expression of a tensor or a polyadic as a sum of products', *J. Math. Phys.* **6**, 164–189.
- Hoffmann, D. P. K. & Monte, A. H. C. (2006), 'The Calibration Kit Spectral Fluorescence Standards — A Simple and Certified Tool for the Standardization of the Spectral Characteristics of Fluorescence Instruments', *J Fluoresc (2006)* **16**, 581–587.
- Hozalski, R. M., Goel, S. & Bouwer, E. J. (1995), 'TOC removal in biological filters', *Journal / American Water Works Association* **87**(12), 40–54.
- Hu, J. Y., Song, L. F., Ong, S. L., Phua, E. T. & Ng, W. J. (2005), 'Biofiltration pretreatment for reverse osmosis (RO) membrane in a water reclamation system', **59**, 127–133.
- Hu, W., Liang, J., Ju, F., Wang, Q., Liu, R., Bai, Y., Liu, H. & Qu, J. (2020), 'Metagenomics Unravels Differential Microbiome Composition and Metabolic Potential in Rapid Sand Filters Purifying Surface Water Versus Groundwater', *Environmental science & technology* **54**(8), 5197–5206.
- Huber, S. A., Balz, A., Abert, M. & Pronk, W. (2011), 'Characterisation of aquatic humic and non-humic matter with size-exclusion chromatography - organic carbon detection - organic nitrogen

- detection (LC-OCD-OND)', *Water Research* **45**(2), 879–885.
URL: <http://dx.doi.org/10.1016/j.watres.2010.09.023>
- Hudson, N., Baker, A., Ward, D., Reynolds, D. M., Brunson, C., Carliell-Marquet, C. & Browning, S. (2008), 'Can fluorescence spectrometry be used as a surrogate for the Biochemical Oxygen Demand (BOD) test in water quality assessment? An example from South West England', *Science of the Total Environment* **391**(1), 149–158.
- Ishii, S. K. & Boyer, T. H. (2012), 'Behavior of reoccurring parafac components in fluorescent dissolved organic matter in natural and engineered systems: A critical review', *Environmental Science and Technology* **46**(4), 2006–2017.
- Ives, A. R. (2015), 'For testing the significance of regression coefficients, go ahead and log-transform count data', *Methods in Ecology and Evolution* **6**, 828–835.
- Jeyakumar, R., Parimalarenganayaki, S. & Elango, L. (2017), 'River bank filtration for natural treatment of water in India: A review', *International Journal of Civil Engineering and Technology* **8**(8), 1203–1212.
- Juhna, T. & Melin, E. (1985), 'Ozonation and Biofiltration in Water', *Science and Technology* **6**(3), 123–127.
- Ka, K., Lü, M., Edith, N., Stefan, K. & Marc, W. (2020), 'Slow sand filtration of raw wastewater using biochar as an alternative filtration media', *Nature Research scientific reports* pp. 1–11.
- Kalkan, , Yapsakli, K., Mertoglu, B., Tufan, D. & Saatci, A. (2011), 'Evaluation of Biological Activated Carbon (BAC) process in wastewater treatment secondary effluent for reclamation purposes', *Desalination* **265**(1-3), 266–273.
URL: <http://dx.doi.org/10.1016/j.desal.2010.07.060>
- Karanfil, T. & Dastgheib, S. A. (2004), 'Trichloroethylene adsorption by fibrous and granular activated carbons: Aqueous phase, gas phase, and water vapor adsorption studies', *Environmental Science and Technology* **38**(22), 5834–5841.
- Kerndorff, H. & Schnitzer, M. (1980), 'Sorption of metals on humic acid *', *Geochimica et Cosmochimica Acta* **44**.
- Kitis, M., Karanfil, T., Wigton, A. & Kilduff, J. E. (2002), 'Probing reactivity of dissolved organic matter for disinfection by-product formation using XAD-8 resin adsorption and ultrafiltration fractionation', *Water Research* **36**(15), 3834–3848.

- Klimenko, N., Winther-Nielsen, M., Smolin, S., Nevynna, L. & Sydorenko, J. (2002), 'Role of the physico-chemical factors in the purification process of water from surface-active matter by biosorption', *Water Research* **36**(20), 5132–5140.
- Kolda, T. G. & Bader, B. W. (2009), 'Tensor Review', *SIAM Review* **51**(3), 455–500.
- Kołodziej, A., Fuentes, M., Baigorri, R., Lorenc-Grabowska, E., García-Mina, J. M., Burg, P. & Gryglewicz, G. (2014), 'Mechanism of adsorption of different humic acid fractions on mesoporous activated carbons with basic surface characteristics', *Adsorption* **20**(5-6), 667–675.
- Korotta-Gamage, S. M. & Sathasivan, A. (2017), 'Potential of a biologically activated carbon treatment to remove organic carbon from surface waters', *International Biodeterioration and Biodegradation* **124**, 82–90.
URL: <http://dx.doi.org/10.1016/j.ibiod.2017.05.025>
- Kothawala, D. N., Murphy, K. R., Stedmon, C. A., Weyhenmeyer, G. A. & Tranvik, L. J. (2013), 'OCEANOGRAPHY : METHODS Inner filter correction of dissolved organic matter fluorescence', pp. 616–630.
- Kowalczyk, P., Gauden, P. A., Terzyk, A. P. & Furmaniak, S. (2010), 'Microscopic model of carbonaceous nanoporous molecular sieves—anomalous transport in molecularly confined spaces', *Physical Chemistry Chemical Physics* **12**(37), 11351.
URL: <http://xlink.rsc.org/?DOI=b926206g>
- Krasner, S., Croue, J.-P., Buffle, J. & Perdue, E. (1996), 'Three approaches for characterizing NOM (Gary Amy).pdf', **88**(6), 66–79.
- Kroonenberg, P. M. (1983), *Three-mode principal component analysis*, DSWO Press.
- Kroonenberg, P. M. (2009), 'Analysing three-way profile data using the PARAFAC and Tucker3 models illustrated with views on parenting 1', *Applied Multivariate Research* **13**(1), 5–41.
- Krzeminski, P., Vogelsang, C., Meyn, T., Köhler, S. J., Poutanen, H., de Wit, H. A. & Uhl, W. (2019), 'Natural organic matter fractions and their removal in full-scale drinking water treatment under cold climate conditions in Nordic capitals', *Journal of Environmental Management* **241**(January), 427–438.
URL: <https://doi.org/10.1016/j.jenvman.2019.02.024>

- Kulin, M., Kazaz, T., De Poorter, E. & Moerman, I. (2021), 'A survey on machine learning-based performance improvement of wireless networks: PHY, MAC and network layer', *Electronics (Switzerland)* **10**(3), 1–64.
- Lakowicz, J. R. (2006), *Principles of Fluorescence Spectroscopy*, 3rd edn, Springer Science, New York.
- land use research institute, M. (1988), *The Land Cover of Scotland 1988*, Technical report.
- Larsson, T., Wedborg, M. & Turner, D. (2007), 'Correction of inner-filter effect in fluorescence excitation-emission matrix spectrometry using Raman scatter', *Analytica Chimica Acta* **583**(2), 357–363.
- Lauderdale, C., Chadik, P., Kirisits, M. J., Brown & Jess (2012), 'Engineered biofiltration: Enhanced biofilter performance through nutrient and peroxide addition', *Journal - American Water Works Association* **104**(5).
- Lautenschlager, K., Hwang, C., Ling, F., Liu, W. T., Boon, N., Köster, O., Egli, T. & Hammes, F. (2014), 'Abundance and composition of indigenous bacterial communities in a multi-step biofiltration-based drinking water treatment plant', *Water Research* **62**, 40–52.
- Lawaetz, A. J. & Stedmon, C. A. (2009), 'Fluorescence intensity calibration using the Raman scatter peak of water', *Applied Spectroscopy* **63**(8), 936–940.
- Leardi, R., Armanino, C., Lanteri, S. & Alberotanza, L. (2000), 'Three-mode principal component analysis of monitoring data from Venice lagoon', *Journal of chemometrics* **14**, 187–195.
- Lechevallier, M. W., Becker, W. C., Schorr, P., Lee, R. G., Lechevallier, M. W., Becker, W. C., Schorr, P. & Lee, R. G. (1992), 'Eval of Rapid Filters', **84**(4), 136–146.
- Lechevallier, M. W., Shaw, N. E., Kaplan, L. A. & Bott, T. L. (1993), 'Development of a Rapid Assimilable Organic Carbon Method for Water Development of a Rapid Assimilable Organic Carbon Method for Water', *Applied and environmental microbiology* **59**(5), 1526–1531.
- Lee, M., Crittenden, J., Snoeyink, V., Members, ASCE & Ari, M. (1983), 'Design of carbon beds to remove humic substances', *Journal of Environmental Engineering* **109**(3), 631–645.
- Leenheer, J. A. & Huffman, E. J. (1979), 'Analytical method for dissolved-organic carbon fractionation', *Water-resources investigations* **79**(4).

Leenheer, J. & Croue, J.-P. (2003), 'Characterizing Aquatic Dissolved Organic Matter', *Environmental Science & Technology* **37**(1), 18–26.

URL: <http://scholar.google.com/scholar?hl=enbtnG=Searchq=intitle:Characterizing+Dissolved+Aquatic+Organic>

Li, A. & DiGiano, F. (1983), 'Availability of sorbed substrate for microbial degradation on granular', *Water pollution control federation* **55**(4), 1–4.

URL: <http://iaspub.epa.gov/tdb/pages/treatment/treatmentOverview.do?treatmentProcessId=2074826383>

Li, F., Yuasa, A., Ebie, K. & Azuma, Y. (2003), 'Microcolumn test and model analysis of activated carbon adsorption of dissolved organic matter after pre-coagulation: Effects of pH and pore size distribution', *Journal of Colloid and Interface Science* **262**(2), 331–341.

Li, L., Ning, D., Jeon, Y., Ryu, H., Santo Domingo, J. W., Kang, D. W., Kadudula, A. & Seo, Y. (2021), 'Ecological insights into assembly processes and network structures of bacterial biofilms in full-scale biologically active carbon filters under ozone implementation', *Science of the Total Environment* **751**, 141409.

Li, L., Zhu, W., Zhang, P., Zhang, Q. & Zhang, Z. (2006), 'AC/O₃-BAC processes for removing refractory and hazardous pollutants in raw water', *Journal of Hazardous Materials* **135**(1-3), 129–133.

Liang, X., Chi, J. & Yang, Z. (2018), 'The influence of the functional group on activated carbon for acetone adsorption property by molecular simulation study', *Microporous and Mesoporous Materials* **262**, 77–88.

URL: <https://doi.org/10.1016/j.micromeso.2017.06.009>

Liao, X., Chen, C., Wang, Z., Wan, R., Chang, C. H., Zhang, X. & Xie, S. (2013), 'Pyrosequencing analysis of bacterial communities in drinking water biofilters receiving influents of different types', *Process Biochemistry* **48**(4), 703–707.

URL: <http://dx.doi.org/10.1016/j.procbio.2013.02.033>

Liao, X., Chen, C., Zhang, J., Dai, Y., Zhang, X. & Xie, S. (2014), 'Operational performance, biomass and microbial community structure: Impacts of backwashing on drinking water biofilter', *Environmental Science and Pollution Research* **22**(1), 546–554.

Liu, B., Finkel, M. & Grathwohl, P. (2021), 'Mass transfer principles in column percolation tests: Initial conditions and tailing in heterogeneous materials', *Materials* **14**(16).

- Liu, W., Xu, X., McGoff, N. M., Eaton, J. M., Leahy, P., Foley, N. & Kiely, G. (2014), 'Spatial and seasonal variation of dissolved organic carbon (DOC) concentrations in Irish streams: Importance of soil and topography characteristics', *Environmental Management* **53**(5), 959–967.
- Logue, J. B., Stedmon, C. A., Kellerman, A. M., Nielsen, N. J., Andersson, A. F., Laudon, H., Lindström, E. S. & Kritzberg, E. S. (2016), 'Experimental insights into the importance of aquatic bacterial community composition to the degradation of dissolved organic matter', *ISME Journal* **10**(3), 533–545.
- Ma, B., Lapara, T. M., Evans, A. N. & Hozalski, R. M. (2020), 'Effects of geographic location and water quality on bacterial communities in full-scale biofilters across North America', *FEMS Microbiology Ecology* **96**, 1–9.
- MacDonald, J. A. F. & Evans, M. J. B. (2002), 'Adsorption and enthalpy of phenol on BPL carbon', *Carbon* **40**(5), 703–707.
- Mahajan, O., Moreno-Castilla, C. & Walker JR., P. (1980), 'Surface-Treated Activated Carbon for Removal of Phenol from Water', *Separation science and technology* **15**(10), 1773–1752.
URL: <https://doi.org/10.1080/01496398008055619>
- Martin-Fernandez, J., Barcelo-Vidal, C. & Pawlowsky-Glahn, V. (2003), 'Dealing With Zeros and Missing Values in Compositional Data Sets Using Nonparametric', *Mathematical Geology* **35**(3).
- Matilainen, A., Gjessing, E. T., Lahtinen, T., Hed, L., Bhatnagar, A. & Sillanpää, M. (2011), 'An overview of the methods used in the characterisation of natural organic matter (NOM) in relation to drinking water treatment', *Chemosphere* **83**(11), 1431–1442.
URL: <http://dx.doi.org/10.1016/j.chemosphere.2011.01.018>
- Matilainen, A., Vepsäläinen, M. & Sillanpää, M. (2010), 'Natural organic matter removal by coagulation during drinking water treatment: A review', *Advances in Colloid and Interface Science* **159**(2), 189–197.
URL: <http://dx.doi.org/10.1016/j.cis.2010.06.007>
- Matsui, Y., Nakao, S., Sakamoto, A., Taniguchi, T., Pan, L., Matsushita, T. & Shirasaki, N. (2015), 'Adsorption capacities of activated carbons for geosmin and 2-methylisoborneol vary with activated carbon particle size: Effects of adsorbent and adsorbate characteristics', *Water Research* **85**, 95–102.
URL: <http://dx.doi.org/10.1016/j.watres.2015.08.017>

- McCue, M. E. & McCoy, A. M. (2017), 'The scope of Big Data in one medicine: Unprecedented opportunities and challenges', *Frontiers in Veterinary Science* **4**, 1–23.
- McKnight, D. M., Boyer, E. W., Westerhoff, P. K., Doran, P. T., Kulbe, T. & Andersen, D. T. (2001), 'Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity', *Limnology and Oceanography* **46**, 38–48.
- Moll, D. M., Summers, R. S., Fonseca, A. C. & Matheis, W. (1999), 'Impact of Temperature on Drinking Water Biofilter Performance and Microbial Community Structure', *Environmental Science & Technology* **33**(14), 2377–2382.
URL: <http://pubs.acs.org/doi/abs/10.1021/es9900757>
- Moona, N., Holmes, A., Wu, U. J., Pettersson, T. J. R. & Murphy, K. R. (2021), 'Full-Scale Manipulation of the Empty Bed Contact Time to Optimize Dissolved Organic Matter Removal by Drinking Water Biofilters', **1**, 1117–1126.
- Moona, N., Murphy, K. R., Bondelind, M., Bergstedt, O. & Pettersson, T. J. (2018), 'Partial renewal of granular activated carbon biofilters for improved drinking water treatment', *Environmental Science: Water Research and Technology* **4**(4), 529–538.
- Moran, M. A. & Hodson, R. E. (1990), 'Bacterial production on humic and nonhumic components of dissolved organic carbon', *Limnology and Oceanography* **35**(8), 1744–1756.
- Moreno-Castilla, C. (2004), 'Adsorption of organic molecules from aqueous solutions on carbon materials', *Carbon* **42**(1), 83–94.
- Mounier, S., Nicolodelli, G., Redon, R. & Milori, D. M. B. P. (2017), 'Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy Direct solid surface fluorescence spectroscopy of standard chemicals and humic acid in ternary system', *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **177**, 79–85.
URL: <http://dx.doi.org/10.1016/j.saa.2017.01.017>
- Mubita, T. M., Dykstra, J. E., Biesheuvel, P. M., van der Wal, A. & Porada, S. (2019), 'Selective adsorption of nitrate over chloride in microporous carbons', *Water Research* **164**, 114885.
URL: <https://doi.org/10.1016/j.watres.2019.114885>
- Muller, F. L. & Tankéré-Muller, S. P. (2012), 'Seasonal variations in surface water chemistry at disturbed and pristine peatland sites in the Flow Country of northern Scotland', *Science of the Total*

Environment **435-436**, 351–362.

URL: <http://dx.doi.org/10.1016/j.scitotenv.2012.06.048>

Münster, U. & Chróst, R. J. (1990), Origin, Composition, and Microbial Utilization of Dissolved Organic Matter, in J. Overbeck & R. J. Chróst, eds, 'Aquatic Microbial Ecology', Springer-Verlag, New York, chapter 2, pp. 8–46.

Murphy, K. R., Bro, R. & Stedmon, C. A. (2014), *Chemometric Analysis of Organic Matter Fluorescence*.

Murphy, K. R., Butler, K. D., Spencer, R. G., Stedmon, C. A., Boehme, J. R. & Aiken, G. R. (2010), 'Measurement of dissolved organic matter fluorescence in aquatic environments: An interlaboratory comparison', *Environmental Science and Technology* **44**(24), 9405–9412.

Murphy, K. R., Stedmon, C. A., Graeber, D. & Bro, R. (2013a), 'Decomposition routines for Excitation Emission Matrices', *Analytical Methods* pp. 1–29.

Murphy, K. R., Stedmon, C. A., Graeber, D. & Bro, R. (2013b), 'Fluorescence spectroscopy and multi-way techniques. PARAFAC', *Analytical Methods* **5**(23), 6557–6566.

Murphy, K. R., Stedmon, C. A., Wenig, P. & Bro, R. (2014a), 'OpenFluor- An online spectral library of auto-fluorescence by organic compounds in the environment', *Analytical Methods* **6**(3), 658–661.

Nam, S. W., Choi, D. J., Kim, S. K., Her, N. & Zoh, K. D. (2014), 'Adsorption characteristics of selected hydrophilic and hydrophobic micropollutants in water using activated carbon', *Journal of Hazardous Materials* **270**, 144–152.

URL: <http://dx.doi.org/10.1016/j.jhazmat.2014.01.037>

Nath, K. & Bhakhar, M. S. (2011), 'Microbial regeneration of spent activated carbon dispersed with organic contaminants: Mechanism, efficiency, and kinetic models', *Environmental Science and Pollution Research* **18**(4), 534–546.

Nemani, V. A., McKie, M. J., Taylor-Edmonds, L. & Andrews, R. C. (2018), 'Impact of biofilter operation on microbial community structure and performance', *Journal of Water Process Engineering* **24**(February), 35–41.

Newcombe, G., Drikas, M. & Hayes, R. (1997), 'Influence of characterised natural organic material on activated carbon adsorption: II. Effect on pore volume distribution and adsorption of 2-methylisoborneol', *Water Research* **31**(5), 1065–1073.

- Nguyen, C. & Do, D. D. (1999), 'A New Method for the Characterization of Ion-Exchange Membranes', *Langmuir* **15**, 3608–3615.
- Nguyen, C. & Do, D. D. (2000), 'Dual Langmuir kinetic model for adsorption in carbon molecular sieve materials', *Langmuir* **16**(4), 1868–1873.
- Oh, S., Hammes, F. & Liu, W.-T. (2018), 'Metagenomic characterization of biofilter microbial communities in a full-scale drinking water treatment plant', *Water Research* **128**, 278–285.
- Ouakouak, A. & Youcef, L. (2017), 'Nitrate Adsorption from Synthetic Solutions Using Granular Activated Carbon', *Materials Focus* **6**(5), 525–530.
- Owen, D., Amy, G., Chowdhury, Z., Paode, R., McCoy, G. & Viscosil, K. (1995), 'American Water Works Association', **87**(1), 46–63.
- Palarea-Albaladejo, J. & Martín-fernández, J. A. (2015), 'zCompositions — R package for multivariate imputation of left-censored data under a compositional approach', *Chemometrics and Intelligent Laboratory Systems* **143**, 85–96.
URL: <http://dx.doi.org/10.1016/j.chemolab.2015.02.019>
- Palomo, A., Jane Fowler, S., Gülay, A., Rasmussen, S., Sicheritz-Ponten, T. & Smets, B. F. (2016), 'Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of *Nitrospira* spp.', *ISME Journal* **10**(11), 2569–2581.
- Parambath, M. (2016), *Development of new fluorescence spectroscopy approaches for the study of silica*.
- Pardo, R., Vega, M., Barrado, E., Castrillejo, Y. & Sánchez, I. (2013), 'Three-way principal component analysis as a tool to evaluate the chemical stability of metal bearing residues from wastewater treatment by the ferrite process', *Journal of Hazardous Materials* **262**, 71–82.
URL: <http://dx.doi.org/10.1016/j.jhazmat.2013.08.031>
- Paredes, L., Fernandez-Fontaina, E., Lema, J. M., Omil, F. & Carballa, M. (2016), 'Understanding the fate of organic micropollutants in sand and granular activated carbon biofiltration systems', *Science of the Total Environment* **551-552**, 640–648.
URL: <http://dx.doi.org/10.1016/j.scitotenv.2016.02.008>
- Peel, R. & Benedek, A. (1983), 'Biodegradation and adsorption within activated carbon adsorbers', *Water Pollution Control Federation* **55**(9), 1168–1173.

- Peel, R. G. & Benedek, A. (1980), 'Dual Rate Kinetic Model of Fixed Bed Adsorber', *Journal of the Environmental Engineering Division* **106**(4).
- Peleato, N. M., McKie, M., Taylor-Edmonds, L., Andrews, S. A., Legge, R. L. & Andrews, R. C. (2016), 'Fluorescence spectroscopy for monitoring reduction of natural organic matter and halogenated furanone precursors by biofiltration', *Chemosphere* **153**, 155–161.
- Pendleton, P., Wu, S. H. & Badalyan, A. (2002), 'Activated carbon oxygen content influence on water and surfactant adsorption', *Journal of Colloid and Interface Science* **246**(2), 235–240.
- Persson, F., Heinicke, G., Uhl, W., Hedberg, T. & Hermansson, M. (2006), 'Performance of Direct Biofiltration of Surface Water for Reduction of Biodegradable Organic Matter and Biofilm Formation Potential', *Environmental Technology* **27**(9), 1037:1045.
- Persson, T. & Wedborg, M. (2001), 'Multivariate evaluation of the fluorescence of aquatic organic matter', *Analytica Chimica Acta* **434**(2), 179–192.
- Peterson, E. S. & Summers, R. S. (2021), 'Removal of effluent organic matter with biofiltration for potable reuse : A review and meta-analysis', *Water Research* **199**.
- Pharand, L., Van Dyke, M. I., Anderson, W. B., Yohannes, Y. & Huck, P. M. (2015), 'Full-Scale Ozone – Biofiltration : Seasonally Related Effects on NOM Removal', *AWWA* pp. 425–435.
- Pick, F. C., Fish, K. E., Biggs, C. A., Moses, J. P., Moore, G. & Boxall, J. B. (2019), 'Application of enhanced assimilable organic carbon method across operational drinking water systems', *PLoS ONE* **14**(12), 1–24.
- Pietsch, W. (2015), 'Aspects of theory-ladenness in data-intensive science', *Philosophy of Science* **82**(5), 905–916.
- Pinto, A. J., Xi, C. & Raskin, L. (2012), 'Bacterial Community Structure in the Drinking Water Microbiome Is Governed by Filtration Processes', *Environmental Science and Technology* **46**(16), 8851–8859.
- Pipe-Martin, C. (2010), Dissolved Organic Carbon Removal by Biological treatment, in C. A. Brebbia, D. P. Rico & Y. Villacampa, eds, 'Water pollution IX', 9 edn, WIT press Southam, Southampton, Boston.
- Pohlman, G. G. (1940), 'Soil Science Society of America', *Soil Science Society of America Journal* **4**(C), 446–447.

Pucher, M., Wunsch, U., Weigelhofer, G., Murphy, K., Hein, T. & Graeber, D. (2019), 'staRdom : Versatile Software for Analyzing Spectroscopic Data of Dissolved Organic Matter in R', *Water* **11**(2366), 1–19.

Qualls, R. G. (2015), *Dissolved organic matter*.

Quintelas, C., Silva, B., Figueiredo, H. & Tavares, T. (2010), 'Removal of organic compounds by a biofilm supported on GAC: Modelling of batch and column data', *Biodegradation* **21**(3), 379–392.

Radboud University (n.d.), 'ICP-MS'.

URL: <https://www.ru.nl/science/gi/facilities-activities/elemental-analysis/icp-ms/>

Rattier, M., Reungoat, J. & Gernjak, W. (2012), 'Organic Micropollutant Removal by Biological Activated Carbon Filtration : A Review. Urban Water Security Research Alliance Technical Report No . 53'.

Rattier, M., Reungoat, J., Gernjak, W., Joss, A. & Keller, J. (2012a), 'Investigating the role of adsorption and biodegradation in the removal of organic micropollutants during biological activated carbon filtration of treated wastewater', *Journal of Water Reuse and Desalination* **2**, 127–139.

Reaume, M. J., Seth, R., McPhedran, K. N., da Silva, E. F. & Porter, L. A. (2015), 'Effect of Media on Biofilter Performance Following Ozonation of Secondary Treated Municipal Wastewater Effluent: Sand vs. GAC', *Ozone: Science and Engineering* **37**(2), 143–153.

Reungoat, J., Escher, B. I., Macova, M. & Keller, J. (2011), 'Biofiltration of wastewater treatment plant effluent : Effective removal of pharmaceuticals and personal care products and reduction of toxicity', *Water Research* **45**(9), 2751–2762.

URL: <http://dx.doi.org/10.1016/j.watres.2011.02.013>

Rivera-Utrilla, J., Bautista-Toledo, I., Ferro-Garca, M. A. & Moreno-Castilla, C. (2001), 'Activated carbon surface modifications by adsorption of bacteria and their effect on aqueous lead adsorption', *Journal of Chemical Technology and Biotechnology* **76**(12), 1209–1215.

RJL Micro & Analytics (n.d.), 'SEM-EDX image'.

URL: <https://www.rjl-microanalytic.de/en/analysenservice-mit-iso-17025-akkreditierung/sem-edx-analytik-von-strukturen-und-werkstoffen/>

Rohart, F., Gautier, B., Singh, A. & Lê Cao, K. A. (2017), 'mixOmics: An R package for 'omics feature selection and multiple data integration', *PLoS Computational Biology* **13**(11), 1–19.

- Rohold, L. & H., P. (1993), 'Degradation of non-diffusible organic matter in biofilter reactors', *Water Research* **27**(11), 1689–1691.
- Ryder, A. G., Stedmon, C. A., Harrit, N. & Bro, R. (2017), 'Calibration, standardization, and quantitative analysis of multidimensional fluorescence (MDF) measurements on complex mixtures (IUPAC Technical Report)', *Pure and Applied Chemistry* **89**(12), 1849–1870.
- Sanchez, N. P., Skeriotis, A. T. & Miller, C. M. (2013), 'Assessment of dissolved organic matter fluorescence PARAFAC components before and after coagulation-filtration in a full scale water treatment plant', *Water Research* **47**(4), 1679–1690.
URL: <http://dx.doi.org/10.1016/j.watres.2012.12.032>
- Scholz, M. & Martin, R. J. (1997), 'Ecological equilibrium on biological activated carbon', *Water Research* **31**(12), 2959–2968.
- Scottish Statutory Instruments (2015), WATER SUPPLY - The Public Water Supplies (Scotland) Regulations 2014, Technical Report 364.
- Scottish Water (2012), Shaping the future of your water and waste water services Draft Strategic Projections, Technical report.
URL: www.scottishwater.co.uk/
- Scottish Water (2015), Water quality standards explained, Technical report.
- Scottish Water (2018), 'Innovative solutions sought for £450,000 water treatment challenge'.
URL: <https://www.scottishwater.co.uk/business/about-us/media-centre/latest-news/can-do-challenge>
- Scottish Water (2019), Sustainability report, Technical report.
- Servais, B. P. (1994), 'Biological colonization of granular activated carbon filters in drinking-water treatment', *Journal of Environmental Engineering* **120**(4), 888–899.
- Shimadzu (2012), TOC-L CPH/CPN user's manual, Technical report.
- Sillanpää, M., Matilainen, A. & Lahtinen, T. (2014), *Characterization of NOM*, Elsevier Inc.
URL: <http://dx.doi.org/10.1016/B978-0-12-801503-2.00002-1>
- Simon, F. X., Penru, Y., Guastalli, A. R., Esplugas, S., Llorens, J. & Baig, S. (2013), 'NOM characterization by LC-OCD in a SWRO desalination line', *Desalination and Water Treatment* **51**(7-9), 1776–1780.

- Simpson, D. R. (2008), 'Biofilm processes in biologically active carbon water purification', *Water Research* **42**(12), 2839–2848.
- Sirotkin, A. S., Koshkina, L. Y. & Ippolitov, K. G. (2001), 'The BAC-process for treatment of waste water containing non-ionogenic synthetic surfactants', *Water Research* **35**(13), 3265–3271.
- Soulsby, C., Malcolm, R., Gibbins, C. & Dilks, C. (2001), 'Seasonality, water quality trends and biological responses in four streams in the Cairngorm Mountains, Scotland', *Hydrology and Earth System Sciences* **5**(3), 433–450.
- Spray, J., Wagner, T., Bischoff, J., Trojahn, S., Norouzi, S., Hill, W., Brasche, J., James, L. & Pereira, R. (2021), 'Unravelling Light and Microbial Activity as Drivers of Organic Matter Transformations in Tropical Headwater Rivers', *Biogeosciences Discussions* (April), 1–26.
- Stedmon, C. A. & Bro, R. (2008), 'Characterizing dissolved organic matter fluorescence with parallel factor analysis: A tutorial', *Limnology and Oceanography: Methods* **6**(11), 572–579.
- Stewart, M. H., Wolfe, R. L. & Means, E. G. (1990), 'Assessment of the bacteriological activity associated with granular activated carbon treatment of drinking water', *Applied and Environmental Microbiology* **56**(12), 3822–3829.
- Świetlik, J., Dabrowska, A., Raczyk-Stanisławiak, U. & Nawrocki, J. (2004), 'Reactivity of natural organic matter fractions with chlorine dioxide and ozone', *Water Research* **38**(3), 547–558.
- Terry, L. G. & Summers, R. S. (2018), 'Biodegradable organic matter and rapid-rate biofilter performance: A review', *Water Research* **128**, 234–245.
URL: <https://doi.org/10.1016/j.watres.2017.09.048>
- Terzyk, A. P. (2004), 'Molecular properties and intermolecular forces - Factors balancing the effect of carbon surface chemistry in adsorption of organics from dilute aqueous solutions', *Journal of Colloid and Interface Science* **275**(1), 9–29.
- Terzyk, A. P., Furmaniak, S., Gauden, P. A., Harris, P. J. & Kowalczyk, P. (2012), *Virtual Porous Carbons*, Elsevier Ltd.
URL: <http://dx.doi.org/10.1016/B978-0-08-097744-7.00003-X>
- Thermo Fisher Scientific (2012a), 'Dionex ICS-1100 Ion Chromatography System Operator ' s Manual'.

- Thermo Fisher Scientific (2012*b*), ‘Dionex ICS-900 Ion Chromatography System Operator ’ s Manual’, (065215), 144.
- Thermo Fisher Scientific (n.d.), ‘SEM: Types of Electrons and the Information They Provide’.
URL: <https://www.thermofisher.com/blog/materials/sem-signal-types-electrons-and-the-information-they-provide/>
- Thiel, P., Nolan, C. P., Scott, D., Hiller, B., Masters, D., Zappia, L., Warton, B., Nolan, P., Alessandrino, M., Franzmann, P. & Heitz, A. (2006), ‘Activated Carbon Vs Anthracite As Primary Dual Media Filters – a Pilot Plant Study. Activated Carbon Vs Anthracite As Primary Dual Media Filters – a Pilot Plant Study’, *Exhibition Centre – Bendigo* **5**(5), 8–14.
- Tränckner, J., Wricke, B. & Krebs, P. (2008), ‘Estimating nitrifying biomass in drinking water filters for surface water treatment’, *Water Research* **42**(10-11), 2574–2584.
- Tucker, L. R. (1963), *Implications of factor analysis of three-way matrices for measurement of change, in Problems in Measuring Change*, University of Wisconsin Press.
- Tucker, L. R. (1966), ‘Some mathematical notes on three-mode factor analysis’, *Psychometrika* **31**, 279–311.
- Urfer, D. & Huck, P. M. (2001), ‘Measurement of biomass activity in drinking water biofilters ysubf a respirometric method’, *Water Research* **35**(6), 1469–1477.
- Urfer, D., Huck, P. M., Booth, S. D. & Coffey, B. M. (1997), ‘Biological filtration for BOM and particle removal: A critical review: The authors review key parameters and engineering variables influencing biological filtration and identify areas requiring further research’, *Journal / American Water Works Association* **89**(12), 83–98.
- USGS (2001), ‘A Primer on Water Quality’.
- Uyguner-Demirel, C. S. & Bekbolet, M. (2011), ‘Significance of analytical parameters for the understanding of natural organic matter in relation to photocatalytic oxidation’, *Chemosphere* **84**(8), 1009–1031.
URL: <http://dx.doi.org/10.1016/j.chemosphere.2011.05.003>
- van der Kooij, D. & van der Wielen, P. W. (2011), *Microbial Growth in Drinking Water Supplies Problems, Causes, Control and Research needs*, IWA Publishing, London.
- VanLoon, G. & Duffy, S. (2005), *Environmental chemistry*, Oxford university press, Oxford.

- Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R. & Cushing, C. E. (1980), 'The River Continuum Concept', *Canadian Journal of Fisheries and Aquatic Sciences* **37**(1), 130–137.
- Vasyukova, E., Proft, R. & Uhl, W. (2014), Evaluation of dissolved organic matter fractions removal due to biodegradation, in N. Nakamoto, N. G. Gimbe, M. R. Collin & R. Gimbel, eds, 'Progress in Slow Sand and Alternative Biofiltration Processes: Further Developments and Applications', 1 edn, Iwa Publishing, London, chapter 7, pp. 59 – 66.
URL: <https://books.google.co.uk/books?id=mbgDBAAAQBAJprintsec=frontcoverdq=Vasyukova+f=false>
- Velten, S., Boller, M., Köster, O., Helbing, J., Weilenmann, H. U. & Hammes, F. (2011), 'Development of biomass in a drinking water granular active carbon (GAC) filter', *Water Research* **45**(19), 6347–6354.
- Velten, S., Hammes, F., Boller, M. & Egli, T. (2007), 'Rapid and direct estimation of active biomass on granular activated carbon through adenosine tri-phosphate (ATP) determination', *Water Research* **41**(9), 1973–1983.
- Velten, S., Knappe, D., Traber, J., Kaiser, H., von Gunten, U., Boller, M. & Meylan, S. (2011a), 'Characterization of natural organic matter adsorption in granular activated carbon adsorbers', *Water Research* **45**(13), 3951–3959.
- Vignola, M., Werner, D., Wade, M. J., Meynet, P. & Davenport, R. J. (2018), 'Medium shapes the microbial community of water filters with implications for effluent quality', *Water Research* **129**, 499–508.
- Volk, C. J., Volk, C. B. & Kaplan, L. A. (1997), 'Chemical Composition of biodegradable dissolved organic matter in stream water', *Limnol. Oceanogr.* **42**(1), 39–44.
- Vrede, K., Heldal, M., Norland, S. & Bratbak, G. (2002), 'Elemental composition (C, N, P) and cell volume of exponentially growing and nutrient-limited bacterioplankton', *Applied and Environmental Microbiology* **68**(6), 2965–2971.
- Wang, J. Z., Summers, R. S. & Miltner, R. J. (1995), 'Biofiltration performance: part 1, relationship to biomass', *Journal / American Water Works Association* **87**(12), 2018.
- Weiss, M., Abele, U., Weckesser, J., Welte, W., Schiltz, E. & Schulz, G. (1991), 'Molecular architecture and electrostatic properties of a bacterial porin', *Science* **254**(5038), 1627–1630.

- Wu, F. C., Evans, R. D. & Dillon, P. J. (2003), '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, H. & Xie, Y. F. (2005), 'Effects of EBCT and water temperature on HAA removal using BAC', *Journal / American Water Works Association* **97**(11).
- Wu, X., Wu, L., Liu, Y., Zhang, P., Li, Q., Zhou, J., Hess, N. J., Hazen, T. C., Yang, W. & Chakraborty, R. (2018), 'Microbial interactions with dissolved organic matter drive carbon dynamics and community succession', *Frontiers in Microbiology* **9**, 1–12.
- Xiaojian, Z., Zhansheng, W. & Xiasheng, G. (1991), 'Simple combination of biodegradation and carbon adsorption-the mechanism of the biological activated carbon process', *Water Research* **25**(2), 165–172.
- Yamamura, S. (2014), Overview on the current condition of slow sand filtration and its challenges in Japan, in N. Nakamoto, N. Graham, M. Collins & R. Gimbel, eds, 'Progress in Slow Sand and Alternative Biofiltration Processes: Further Developments and Applications', iwa edn, London.
- Yang, L., Hur, J. & Zhuang, W. (2015), 'Occurrence and behaviors of fluorescence EEM-PARAFAC components in drinking water and wastewater treatment systems and their applications: A review', *Environmental Science and Pollution Research* **22**(9), 6500–6510.
- Yapsakli, K. & Çeçen, F. (2010), 'Effect of type of granular activated carbon on DOC biodegradation in biological activated carbon filters', *Process Biochemistry* **45**(3), 355–362.
- Yapsaklı, K., Çeçen, F., Aktaş, & Can, Z. S. (2009), 'Impact of Surface Properties of Granular Activated Carbon and Preozonation on Adsorption and Desorption of Natural Organic Matter', *Environmental Engineering Science* **26**(3), 489–500.
URL: <http://www.liebertonline.com/doi/abs/10.1089/ees.2008.0005>
- Ye, C., Lei, Z., Wang, X., Gong, A. & Zheng, H. (2001), 'Multimedia environmental behavior of herbicide atrazine', *Huan jing ke xue = Huanjing kexue* **22**(2), 69–73.
- Zachman, B. A. & Summers, R. S. (2010), 'Modeling TOC Breakthrough in Granular Activated Carbon Adsorbers', *Journal of Environmental Engineering* **136**(2), 204–210.
- Zearley, T. L. & Summers, R. S. (2012), 'Removal of trace organic micropollutants by drinking water biological filters', *Environmental Science and Technology* **46**(17), 9412–9419.

- Zepp, R. G., Sheldon, W. M. & Moran, M. A. (2004), 'Dissolved organic fluorophores in southeastern US coastal waters: Correction method for eliminating Rayleigh and Raman scattering peaks in excitation-emission matrices', *Marine Chemistry* **89**(1-4), 15–36.
- Zhang, H., Gu, L., Liu, B., Gan, H., Zhang, K., Jin, H. & Yu, X. (2016), 'An improved biofilter to control the dissolved organic nitrogen concentration during drinking water treatment', *Environmental Science and Pollution Research* **23**(18), 18137–18144.
- Zhang, S., Gitungo, S. W., Axe, L., Raczko, R. F. & Dyksen, J. E. (2017), 'Biologically active filters – An advanced water treatment process for contaminants of emerging concern', *Water Research* **114**, 31–41.
URL: <http://dx.doi.org/10.1016/j.watres.2017.02.014>
- Zhang, S. & Huck, P. M. (1996), 'Parameter estimation for biofilm processes in biological water treatment', *Water Research* **30**(2), 456–464.
URL: <https://www.sciencedirect.com/science/article/pii/004313549500162X>
- Zhu, I. X., Getting, T. & Bruce, D. (2010), 'Review of biologically active filters in drinking water applications', *Journal / American Water Works Association* **102**(12), 67–77.