

Supporting Information

Microbiome triggered transformations of trace organic chemicals in the presence of effluent organic matter in managed aquifer recharge (MAR) systems

Karin Hellauer^{a,†}, Jenny Uhl^{b,†}, Marianna Lucio^b, Philippe Schmitt-Kopplin^{b,c}, Daniel Wibberg^d, Uwe Hübner^{*,a}, Jörg E. Drewes^a

^a Technical University of Munich, Chair of Urban Water Systems Engineering, Am Coulombwall 3, 85748 Garching, Germany

^b Helmholtz Zentrum München, Research Unit Analytical BioGeoChemistry, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany

^c Technical University of Munich, Chair of Analytical Food Chemistry, Maximus-von-Imhof-Forum 2, 85354 Freising, Germany

^d CeBiTec, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany

[†] contributed equally.

*Corresponding author:

E-mail address: u.huebner@tum.de, *phone:* +49 89 289 13706 (U. Hübner)

Number of pages: 15

Number of figures: 11

Number of tables: 2

3D-fluorescence spectroscopy

The analyses of the samples were performed under specific settings as listed in Table S 1. The blank was subtracted from the 3D-Excitation-emission-matrices (3D-EEM) of the sample and furthermore, the sample 3D-EEM were corrected by applying the inner filter effect algorithm as well as the first and second order Rayleigh masking as provided by the Aqualog® software. Finally, the 3D-EEM were normalized to Raman units (R.U.)¹.

Table S 1: Settings chosen for the 3D-fluorescence spectroscopy.

Parameter	Setting
Integration time	1 s
Excitation increment	3 nm
Emission increment	1.64 nm
Excitation wavelength	230-599 nm
Emission wavelength	212-621 nm
Charge-coupled device (CCD)	medium

16S ribosomal RNA amplicon sequencing analysis

Merged and combined reads were processed using USEARCH v9.2.61²: i) quality filtering was done using the *fastq_filter* command³ with default settings; ii) denoising was performed by the *derep_fulllength* command, whereby the *sizeout* option was chosen; iii) sequences were sorted by abundance (*sortbysize* command), a minimum size of 2 was set (*minsize* 2); iv) OTUs were clustered *de novo* by applying the *cluster_otus* command; v) the OTU table was created using *usearch_global* command setting an identity threshold of 0.97 and searching for hits on the forward strand only (*strand plus*).

PARAFAC analysis

The PARAFAC analysis was performed as described by Stahlschmidt et al. (2016)¹. Due to interfering signals, only excitation wavelengths in the range of 239 - 599 nm were taken into

account. Overall, 112 3D-EEM data were available of which 16 were excluded due to missing DOC concentrations and 9 were identified as outliers. The PARAFAC model was generated out of 28 and validated by 59 3D-EEM data. The quality of the model was assessed based on the Core Consistency, Total Variance and Split Half Analysis as proposed by Carvajal et al. (2017)⁴.

FT-ICR-MS analysis

Samples were directly injected in the solarix FT-ICR-MS from Bruker (Bruker Daltonik GmbH, Germany) with a flow rate of 120 μ L/h at a nebulizer gas pressure of 2.2 bar and a dry gas flow rate of 4 L/min at 180 °C. A capillary voltage of 3600 V was applied. 300 scans per sample were accumulated within a mass range of m/z 147.4 – 1500 with a time domain of 4 megawords. Spectra were first externally calibrated on clusters of a standard arginine solution and internal calibration was systematically done in the presence of natural organic matter reaching accuracy values lower than 1 ppm. Elemental formulae for each peak were calculated by an in-house written software tool⁵ with respect to sensible chemical constraints: N rule, O/C ratio ≤ 1 , H/C ratio $\leq 2n+2$ (C_nH_{2n+2}), element counts: $C \leq 100$, $H \leq 200$, $O \leq 80$, $N \leq 3$, $S \leq 2$ and mass accuracy window ± 0.5 ppm. Formulae were categorized into groups containing CHO, CHOS, CHNO and CHNOS molecular compositions.

Multivariate analysis

Table S 2: All parameters from the OPLS model: residual sum of squares (SS), degrees of freedom (DF), mean squares (MS), the F-test (F), the p-values (p) and the standard deviation (SD). All the metadata that were included in the model are significant (p-values < 0.05).

Metadata	SS	DF	MS	F	p	SD
DOC [mg/L]						
Total corr	6	6	1			1
Regression	5.99995	5	1.19999	25338.5	0.0047695	1.09544
Residual	4.74E-05	1	4.74E-05			0.006882
UVA₂₅₄ [1/m]						
Total corr	6	6	1			1
Regression	5.99995	5	1.19999	25001.4	0.0048015	1.09544
Residual	4.80E-05	1	4.80E-05			0.006928
Dissolved Oxygen [mg/L]						
Total corr	6	6	1			1
Regression	5.99983	5	1.19997	6859.82	0.0091663	1.09543
Residual	0.0001749	1	0.00017493			0.013226
Nitrate-N [mg/L]						
Total corr	6	6	1			1
Regression	5.99993	5	1.19999	16286.5	0.005949	1.09544
Residual	7.37E-05	1	7.37E-05			0.008584
Benzotriazole [-]						
Total corr	6	6	1			1
Regression	5.99983	5	1.19997	6992.22	0.0090791	1.09543
Residual	0.0001716	1	0.00017161			0.0131
Carbamazepine [-]						
Total corr	6	6	1			1
Regression	5.99987	5	1.19997	9001.11	0.0080021	1.09543
Residual	0.0001333	1	0.00013331			0.011546
Citalopram [-]						
Total corr	6	6	1			1
Regression	5.99998	5	1.2	52451.7	0.003315	1.09544
Residual	2.29E-05	1	2.29E-05			0.004783
Diclofenac [-]						
Total corr	6	6	1			1
Regression	5.99998	5	1.2	76813.8	0.0027393	1.09544
Residual	1.56E-05	1	1.56E-05			0.003952
Gabapentin [-]						
Total corr	6	6	1			1
Regression	5.99998	5	1.2	65804.1	0.0029596	1.09544
Residual	1.82E-05	1	1.82E-05			0.00427

Table S 2 continued

Metoprolol [-]						
Total corr	6	6	1			1
Regression	5.99999	5	1.2	135454	0.0020629	1.09544
Residual	8.86E-06	1	8.86E-06			0.002976
Primidone [-]						
Total corr	6	6	1			1
Regression	5.99985	5	1.19997	8134.9	0.0084174	1.09543
Residual	0.0001475	1	0.00014751			0.012145
Sotalol [-]						
Total corr	6	6	1			1
Regression	5.99999	5	1.2	101601	0.0023819	1.09544
Residual	1.18E-05	1	1.18E-05			0.003437
Sulfamethoxazole [-]						
Total corr	6	6	1			1
Regression	5.99988	5	1.19998	10142	0.0075387	1.09543
Residual	0.0001183	1	0.00011832			0.010877
Tramadol [-]						
Total corr	6	6	1			1
Regression	5.99697	5	1.19939	395.379	0.0381626	1.09517
Residual	0.0030335	1	0.00303353			0.055078
Venlafaxine [-]						
Total corr	6	6	1			1
Regression	5.99966	5	1.19993	3483.48	0.0128627	1.09541
Residual	0.0003445	1	0.00034446			0.01856

Results

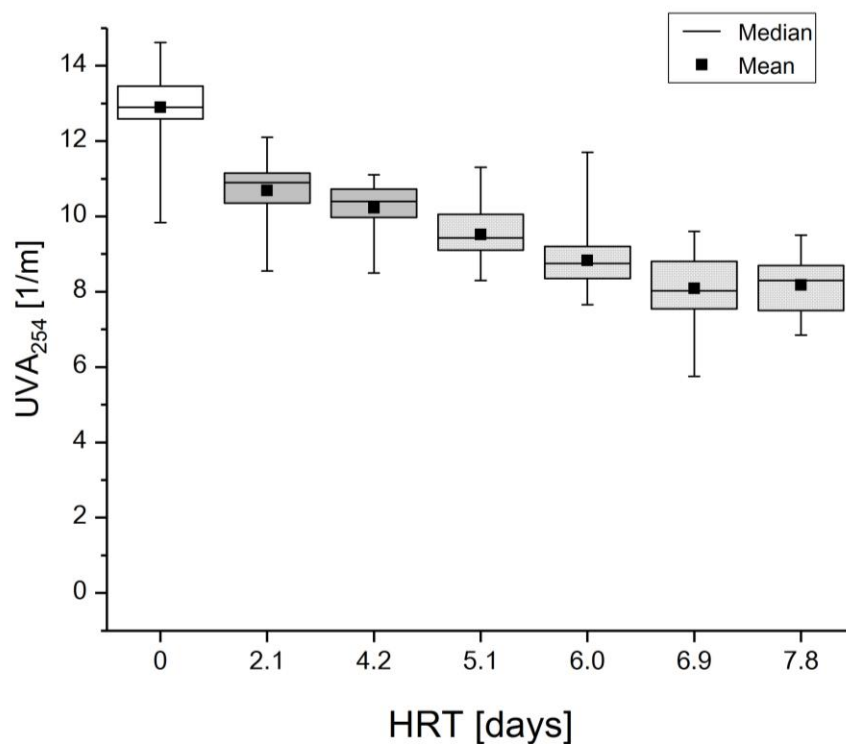


Figure S 1: UV absorbance at 254 nm throughout the system with respect to the HRT ($n \geq 18$). The box represents the 25 – 75 percentiles, the whiskers indicate the maximum and minimum values.

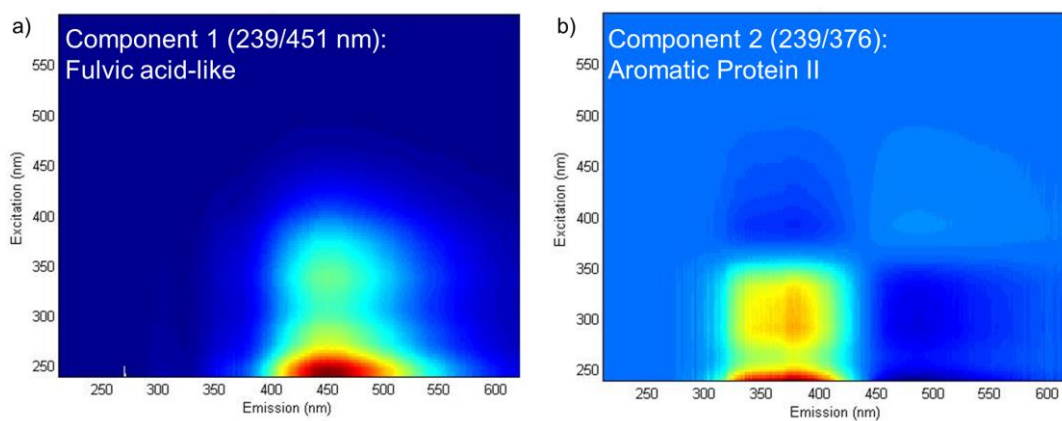


Figure S 2: Two components identified by PARAFAC analysis: a) component 1 fulvic acid-like (λ_{Ex} 239/ λ_{Em} 451 nm), b) component 2 aromatic protein II (239/376 nm) as proposed by Chen et al. (2003)⁶.

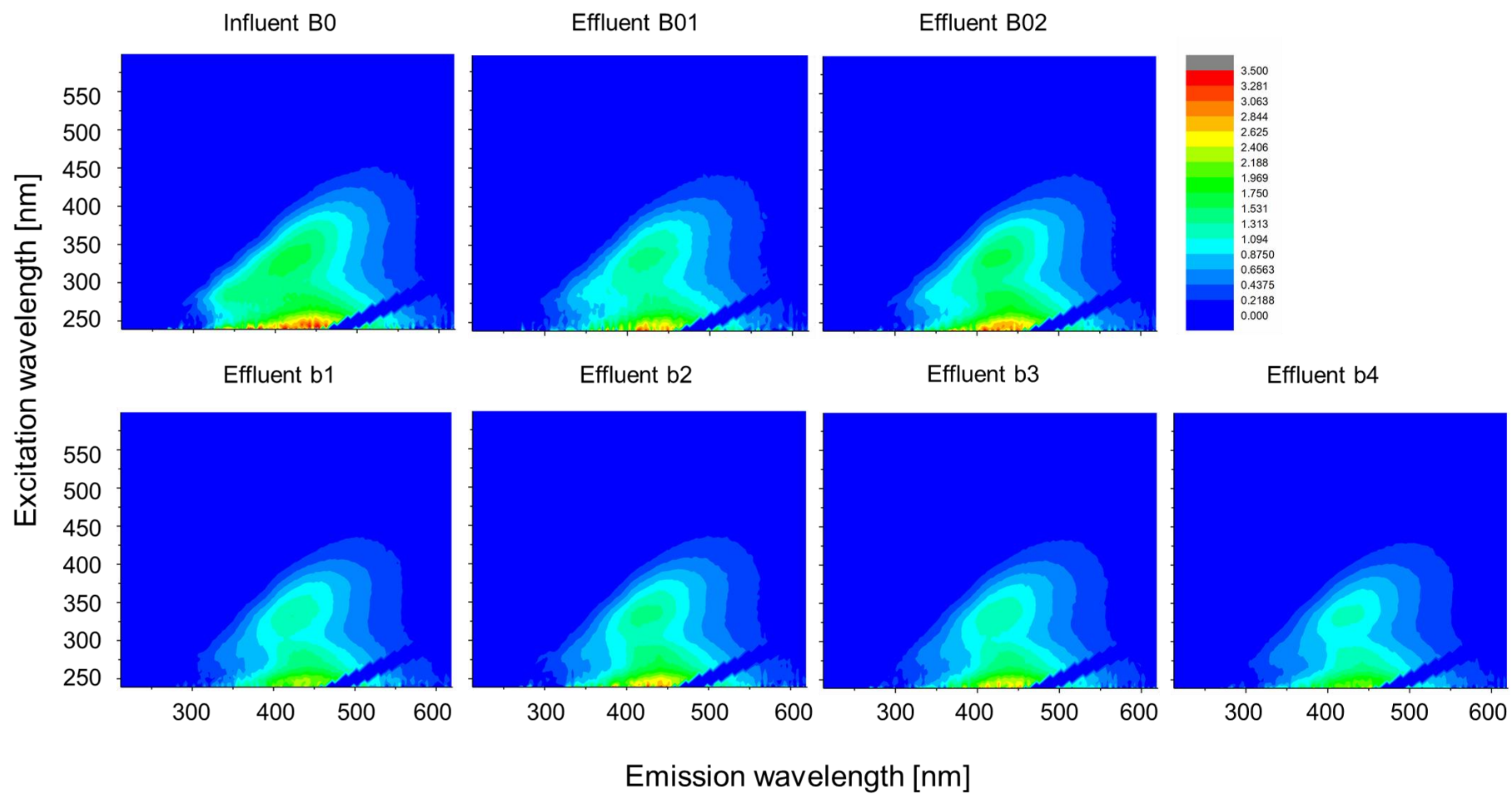


Figure S 3: 3D-Excitation-Emission matrices of all samples throughout the infiltration system used for (-)ESI FT-ICR-MS analyses.

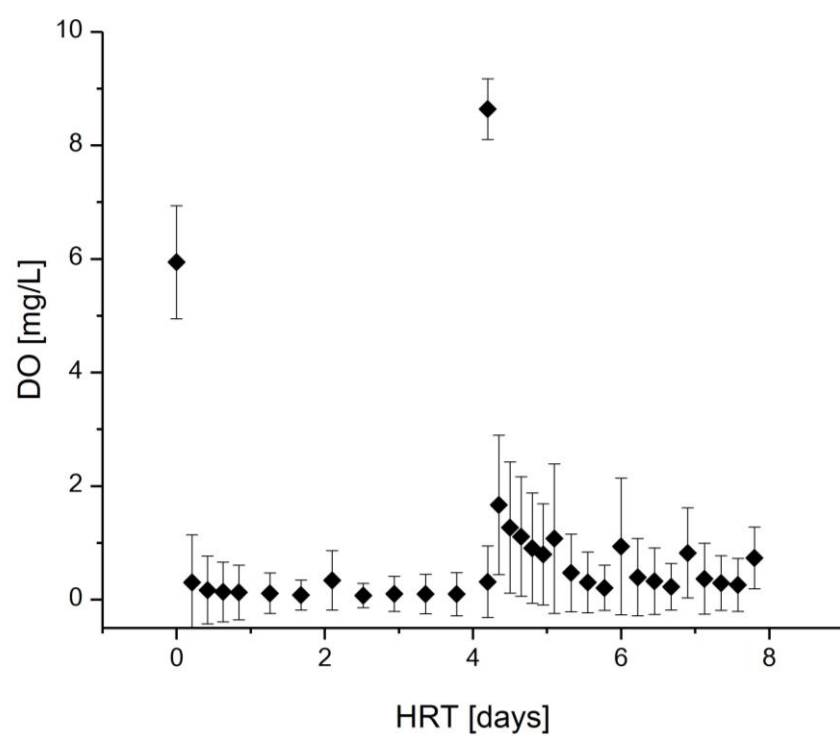


Figure S 4: DO profile ($n \geq 22$) throughout the system with respect to the HRT. Error bars indicate standard deviation.

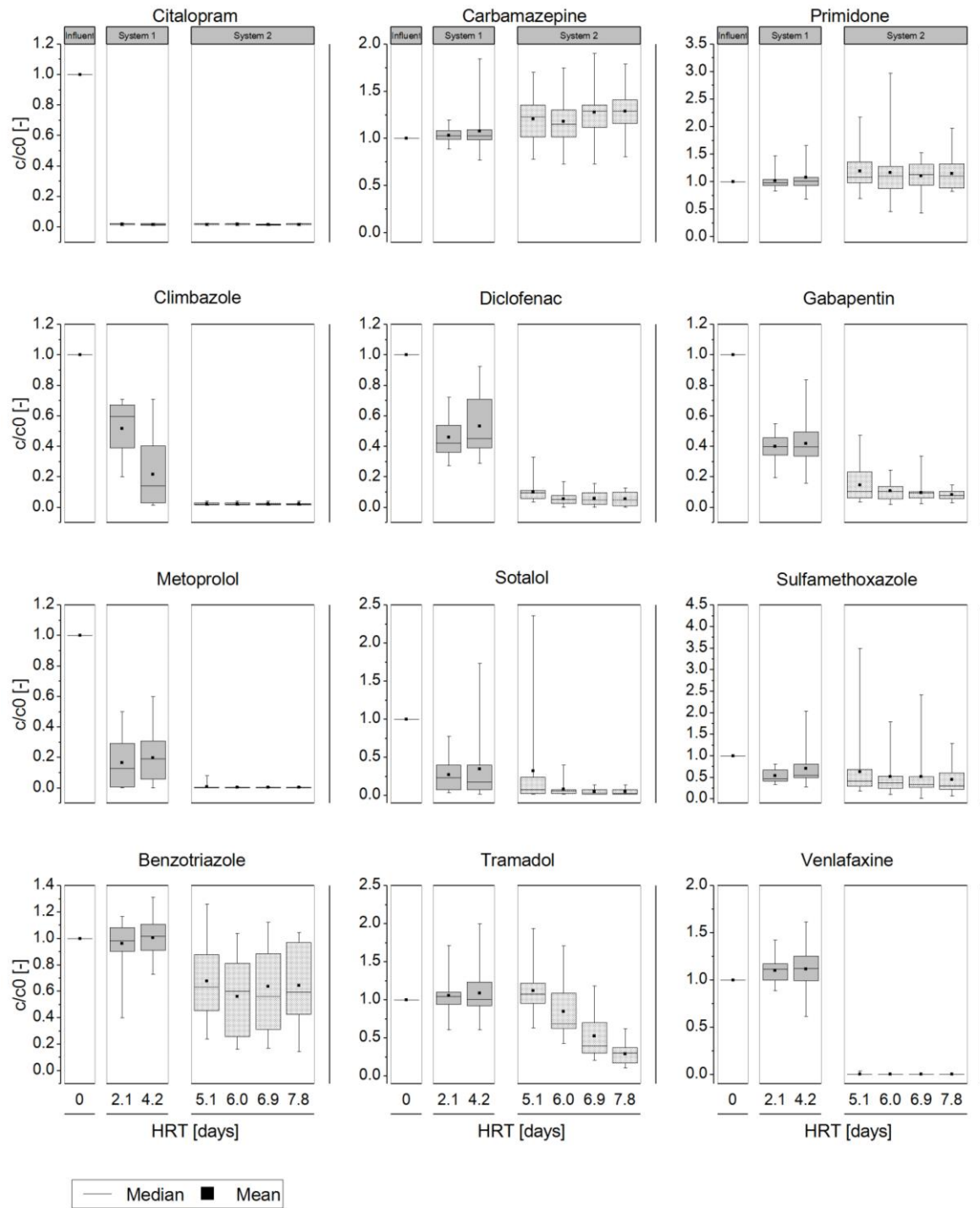


Figure S 5: Relative removal of all targeted TORCs throughout the infiltration system ($n \geq 7$). The box represents the 25 – 75 percentiles. The whiskers indicate maximum and minimum values.

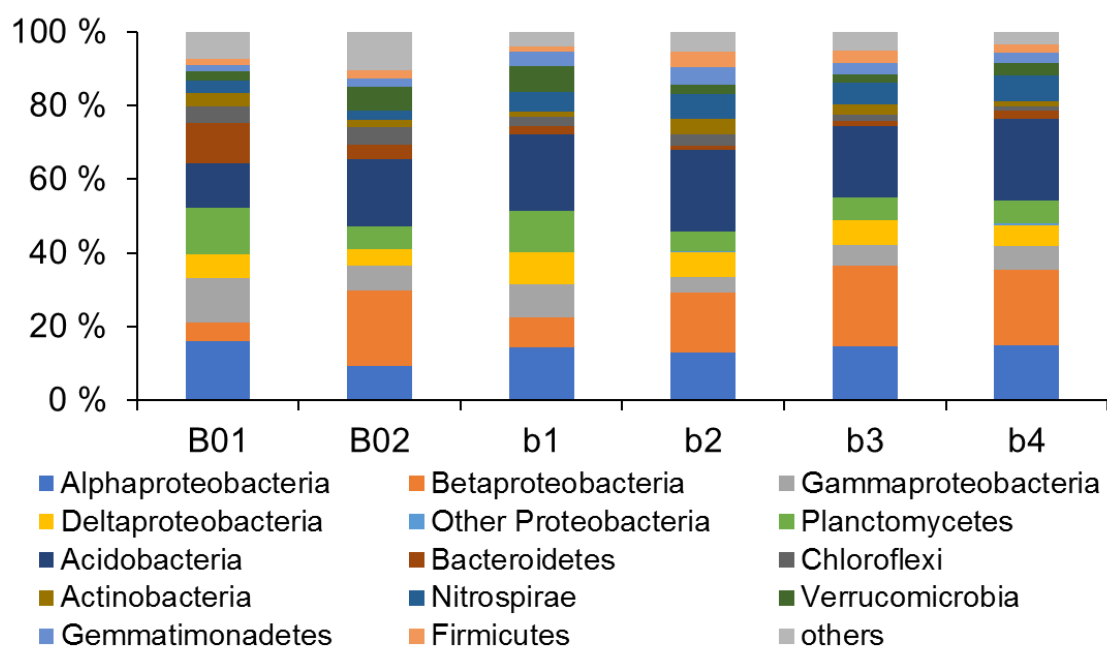


Figure S 6: Percentage of most abundant microbial phyla throughout the system. Proteobacteria are shown in microbial classes.

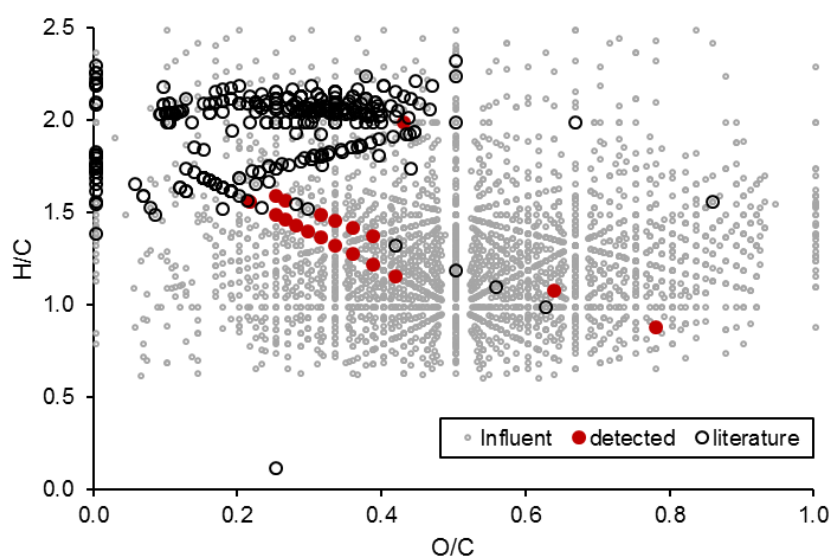


Figure S 7: Surfactants and selected transformation products known from literature⁷⁻¹³ (black circles) and those which were detected in EfOM (red dots) plotted according to their H/C and O/C ratios. Grey dots represent all masses which were detected in the EfOM meaning the influent to the system.

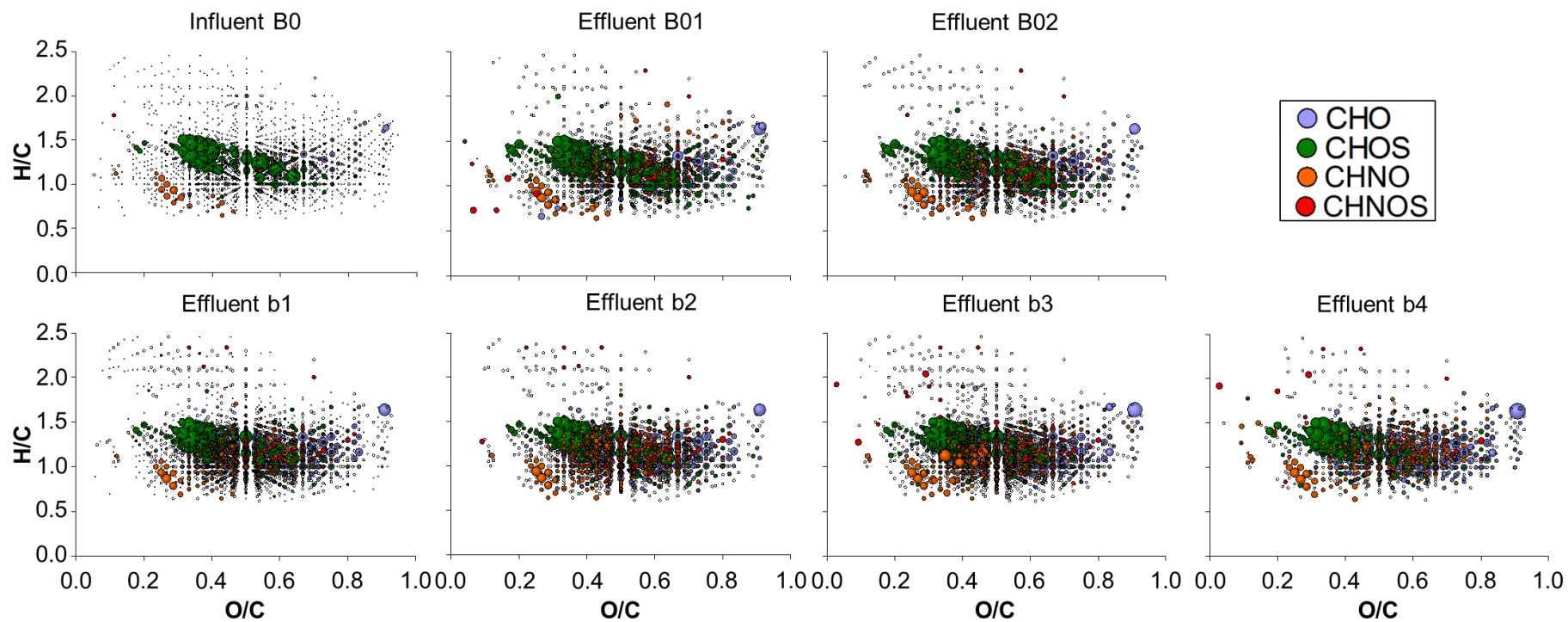


Figure S 8: Results from (-)ESI FT-ICR-MS analyses plotted in van Krevelen diagrams as elemental compositions (CHO, CHOS, CHNO and CHNOS) for masses detected in the influent and effluent samples, respectively. Bubble sizes depict the absolute intensities of each mass.

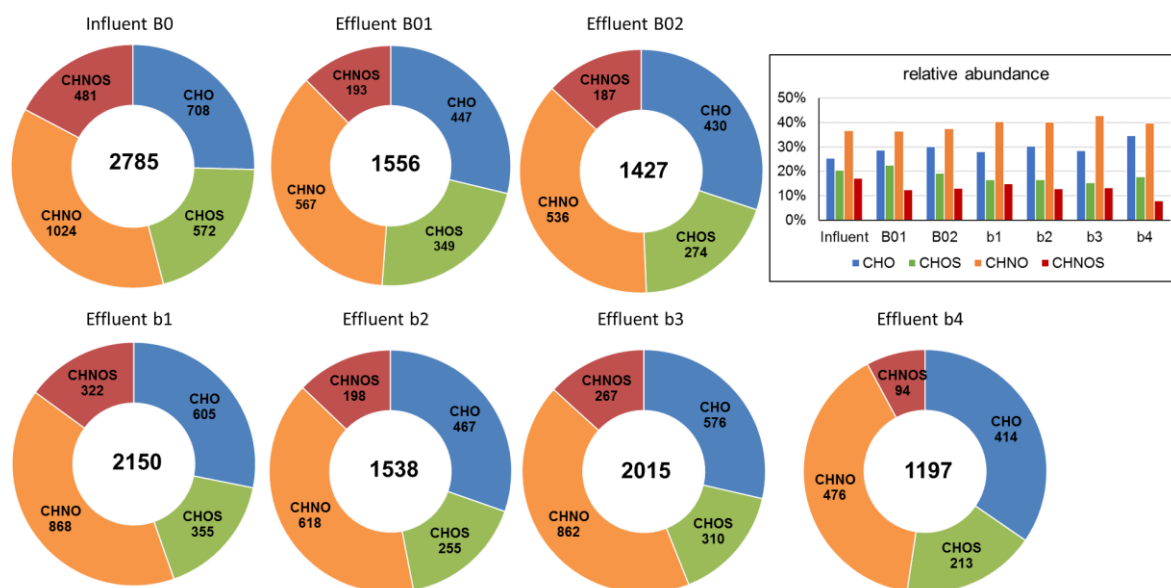


Figure S 9: Molecular composition and the relative abundance of each sample throughout the infiltration based on (-)ESI FT-ICR-MS analyses.

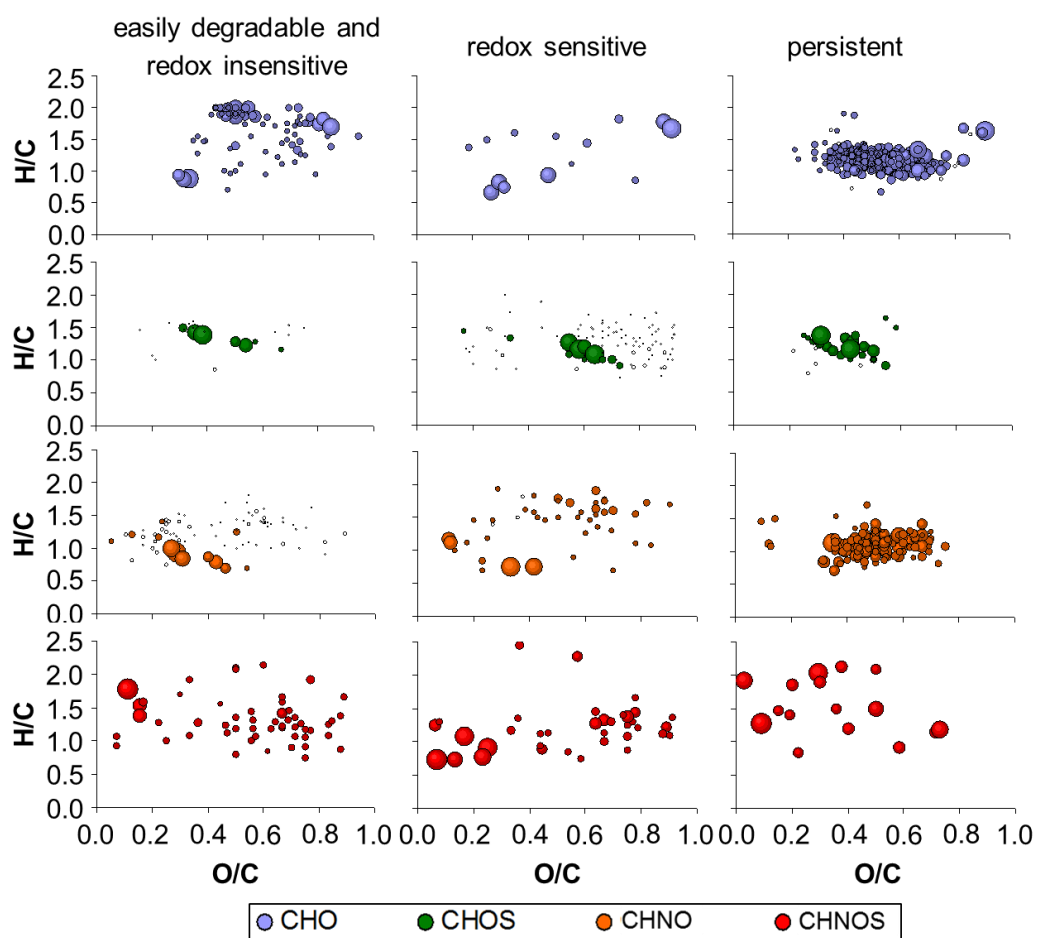


Figure S 10: Van Krevelen plots for each elemental composition (CHO, CHOS, CHNO, CHNOS) of three identified clusters based on MCIA. Bubble sizes depict the absolute intensities of each mass.

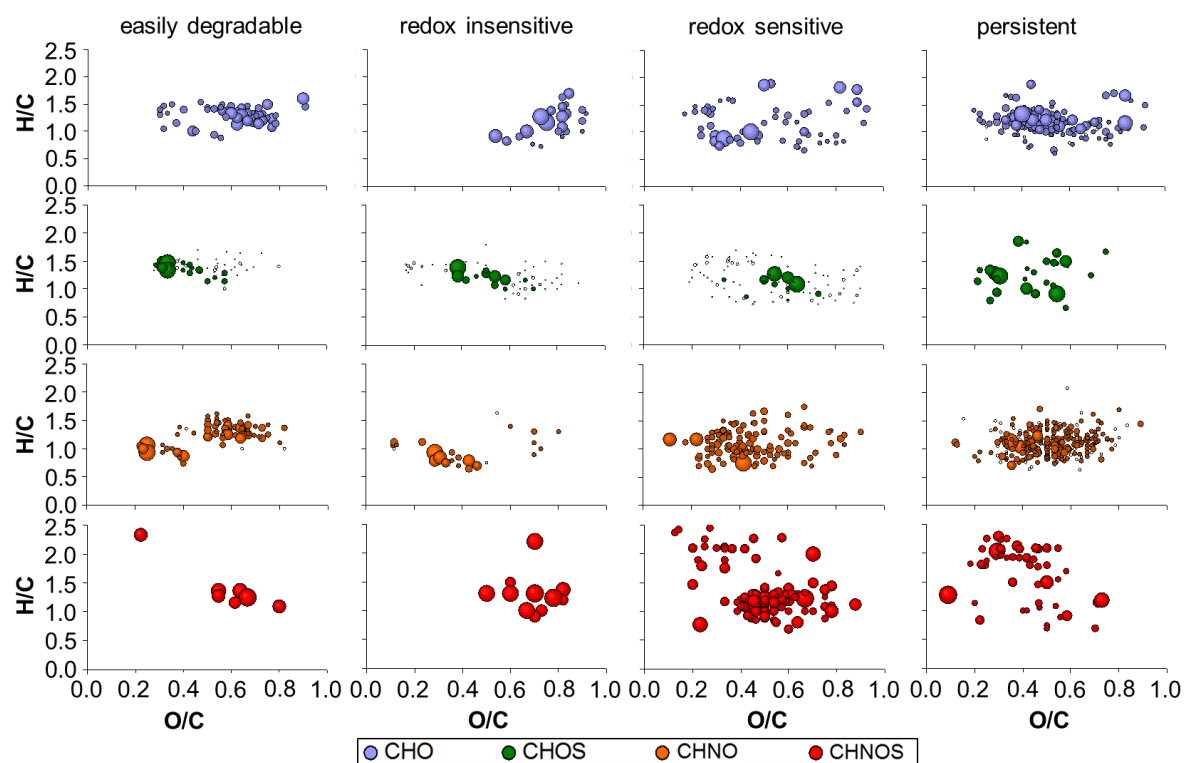


Figure S 11: Van Krevelen plots for each elemental composition (CHO, CHOS, CHNO, CHNOS) of four identified clusters based on OPLS. Bubble sizes depict the absolute intensities of each mass.

References

- (1) Stahlschmidt, M.; Regnery, J.; Campbell, A.; Drewes, J. E. Application of 3D-fluorescence/PARAFAC to monitor the performance of managed aquifer recharge facilities. *J. Water Reuse Desalination* **2016**, *6*, 249–263.
- (2) Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461.
- (3) Edgar, R. C.; Flyvbjerg, H. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* **2015**, *31*, 3476–3482.
- (4) Carvajal, G.; Branch, A.; Michel, P.; Sisson, S. A.; Roser, D. J.; Drewes, J. E.; Khan, S. J. Robust evaluation of performance monitoring options for ozone disinfection in water recycling using Bayesian analysis. *Water Res.* **2017**, *124*, 605–617.
- (5) Tziotis, D.; Hertkorn, N.; Schmitt-Kopplin, P. Kendrick-analogous network visualisation of ion cyclotron resonance Fourier transform mass spectra: Improved options for the assignment of elemental compositions and the classification of organic molecular complexity. *Eur. J. Mass Spectrom. (Chichester)* **2011**, *17*, 415–421.
- (6) Chen, W.; Westerhoff, P.; Leenheer, J. A.; Booksh, K. Fluorescence Excitation–Emission Matrix Regional Integration to Quantify Spectra for Dissolved Organic Matter. *Environ. Sci. Technol.* **2003**, *37*, 5701–5710.
- (7) Terrabase Inc. Surfactants. <http://www.terrabase-inc.com/> (accessed June 29, 2018).
- (8) Riu, J.; Gonzalez-Mazo, E.; Gomez-Parra, A.; Barceló, D. Determination of parts per trillion level of carboxylic degradation products of linear alkylbenzenesulfonates in coastal water by solid-phase extraction followed by liquid chromatography/ion spray/mass spectrometry using negative ion detection. *Chromatographia* **1999**, *50*, 275–281.

- (9) González-Mazo, E.; Honing, M.; Barceló, D.; Gómez-Parra, A. Monitoring Long-Chain Intermediate Products from the Degradation of Linear Alkylbenzene Sulfonates in the Marine Environment by Solid-Phase Extraction Followed by Liquid Chromatography/Ionspray Mass Spectrometry. *Environ. Sci. Technol.* **1997**, *31*, 504–510.
- (10) Crescenzi, C.; Di Corcia, A.; Marchiori, E.; Samperi, R.; Marcomini, A. Simultaneous determination of alkylbenzenesulfonates and dialkyltetralinsulfonates in water by liquid chromatography. *Water Res.* **1996**, *30*, 722–730.
- (11) Trehy, M. L.; Gledhill, W. E.; Orth, R. G. Determination of linear alkylbenzenesulfonates and dialkyltetralinsulfonates in water and sediment by gas chromatography/mass spectrometry. *Anal. Chem.* **1990**, *62*, 2581–2586.
- (12) Trehy, M. L.; Gledhill, W. E.; Mieure, J. P.; Adamove, J. E.; Nielsen, A. M.; Perkins, H. O.; Eckhoff, W. S. Environmental monitoring for linear alkylbenzene sulfonates, dialkyltetralin sulfonates and their biodegradation intermediates. *Environ. Toxicol. Chem.* **1996**, *15*, 233–240.
- (13) Field, J. A.; Leenheer, J. A.; Thorn, K. A.; Barber, L. B.; Rostad, C.; Macalady, D. L.; Daniel, S. R. Identification of persistent anionic surfactant-derived chemicals in sewage effluent and groundwater. *J. Contam. Hydrol.* **1992**, *9*, 55–78.