

Characterization of dissolved organic matter in wastewater using statistical non-target analysis of liquid chromatography-high resolution mass spectrometry (LC-HRMS) data

Yaroslav Verkh^a and Mira Petrovic^{a,b}

^a *Catalan Institute for Water Research (ICRA), Girona, Spain*

^b *Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain*

Introduction

Dissolved organic matter (DOM) in wastewater is a highly complex matrix and composed of a multitude of substances. These can include xenobiotic compound classes like pharmaceuticals, biocides, personal care products. Furthermore, treatment of wastewater can form biologically active transformation products or metabolites. Many of the substances can be dangerous for the biota even at low concentrations and thus should not be released to the environment. Despite the topic's importance monitoring of the pollutants has been elaborate and time-consuming until now. The control of the wastewater DOM is predominantly restricted to a small group of micro-pollutants of interest, which is used to estimate the removal efficiency of a wastewater treatment. However, such analysis does not express the true complexity and performance of the entire system. In particular wastewater treatment technologies, which generate transformation- and by-products are not represented well by the targeted analysis. This approach neglects hundreds of active substances, both coming from the influent and those that emerged during the treatment, thus it is crucial to develop an analytical methodology for the whole DOM in wastewater and bring it in connection with an applied treatment technology. Such fingerprinting will help to address the rising concerns about insufficiently treated wastewater and the quality of wastewater as a reusable resource. Our approach focuses on the development and a real case application of a methodology for a non-targeted analysis of wastewater using liquid chromatography–high resolution mass spectrometry (LC-HRMS) data. We equally examine all the substances detected in the LC-MS spectrum of DOM (unknown unknowns) and using the approach detect 10^3 - 10^4 distinct compounds in wastewater samples. To obtain sensible information from the spectral data, the signals are extracted and analyzed in the free software packages *MzMine 2* and *R* with multiple exploratory statistical tools that reveal patterns for groups and classes rather than individual substances. We successfully test the methodology on a reverse osmosis treatment to illustrate the merits of our approach. We characterize the specific transformation of DOM after the treatment for that system. The non-targeted analysis of the wastewater DOM will help to determine the actual performance of treatment technologies. Also, in future applications it can be connected to the measurements of the environmental impact, which will show how the entire effluent of a wastewater treatment impacts the environment instead of relying on a limited set of proxies.

Experimental procedure

The samples were taken from a pilot WWTP in Quart (Spain). Samples consisted of the feed (influent), concentrate and permeate (effluent) of a reverse osmosis (RO) treatment. Additionally a solvent blank was prepared with HPLC grade water. The organic content in the replicates was pre-concentrated with SPE using Bond Elut-PPL cartridges (100 mg, 3 mL) from Agilent Technologies (USA). Aria TLX-1 chromatographic system (Thermo Fisher Scientific) coupled to a hybrid linear ion trap-Fourier Transform Mass Spectrometry analyzer (LTQ-Orbitrap VelosTM, Thermo Fisher Scientific) with an electrospray ionization source (ESI) was used to record the spectra. In HPLC positive ionization (PI) mode we used methanol and 10 mM aqueous solution of formic acid/ammonium formate (pH = 3.2) and in the negative ionization (NI) mode acetonitrile and 5 mM aqueous solution of ammonium acetate/ ammonia (pH = 8.0). The mass tolerance for the processing of LCMS spectra was set to 0.001 Da or 5 ppm. The duration boundaries of the extracted ion chromatograms were 0.030 – 0.800 min.

Two consecutive formula prediction runs were performed with the following atomic ranges: First run: C₁₋₆₀, H₃₋₈₀, O₀₋₂₀, N₀₋₁₀, P₀₋₁, S₀₋₂, Cl₀₋₄, Br₀₋₂, F₀; Second run: C₁₋₈₀, H₃₋₁₀₀, O₀₋₂₀, N₀₋₁₅, P₀₋₂, S₀₋₂, Cl₀₋₄, Br₀₋₂, F₀₋₂.

Results and Discussion

We established three major findings for the non-target analysis of the RO treatment. The RO rejected the majority of DOM substances and it showed a bias for heavy and unsaturated organic substances.

2331 of 2763 substances (84 % of influent DOM) were rejected by the RO. 309 features in the effluent out of 741 appeared in course of the treatment (Figure 1). Previously transformations of organic matter in RO were found to be caused by a minor bacterial activity on the membrane. Therefore we attribute the appearance of new features to these minor biological processes. Only limited transformations of DOM were involved, since they did not reduce the mass of appearing substances considerably compared to the features, which disappeared during the treatment.

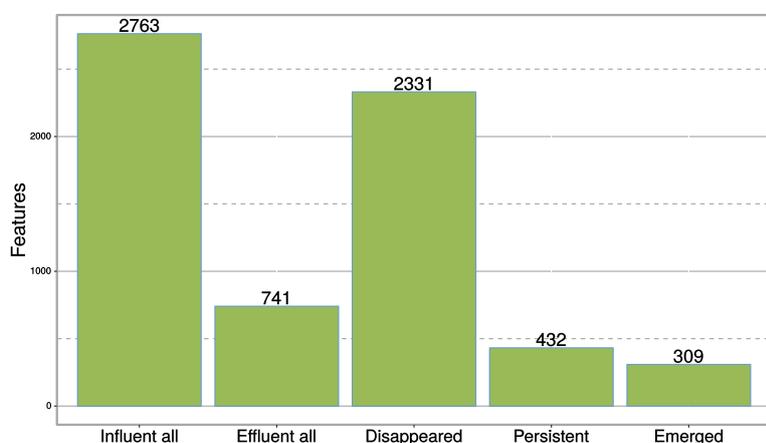


Figure 1: The number of features in RO identified by our methodology. The last two columns describe the persistent and appearing features in the effluent.

The median feature intensity reduction was twenty times (Figure 12). The removal of the total intensity of identified features was 89 %. Total DOM concentration measurements showed on average 94 ± 1 % of rejection for this RO system. Accounting for the ion suppression, which reaches 25 % of intensity, the removal calculated with our method rises to ≤ 91 % of the total intensity. Since the values of the total feature intensity and DOM concentration removal are very similar, we consider that the feature intensity here is a good concentration estimate.

We showed that the RO rejects compounds above 500 Da better than those below the threshold. A two sample t-test identified a mass decrease of 45 Da for the effluent compared to the influent for a 95% confidence interval. The mass and size of organic compounds correlate, therefore we conclude that the RO is good to reject large DOM molecules. Our analysis also found the actual rejection threshold of ≈ 500 Da. This is an efficient way to determine the true molecular size constraints for the RO membrane.

RO shows a rejection dependency on the desaturation of the compounds, since we found that unsaturated substances are preferentially removed compared to the saturated ones (Figure 2~~Error! Reference source not found.~~). The degree of unsaturation (DU) represents the number of double bonds and rings in organic molecules as based on their molecular formula. In turn the value correlates with the polarity of organic substances. We confirmed the decrease in DU of 2.5 for the effluent with a two sample t-test for a 95% confidence interval.

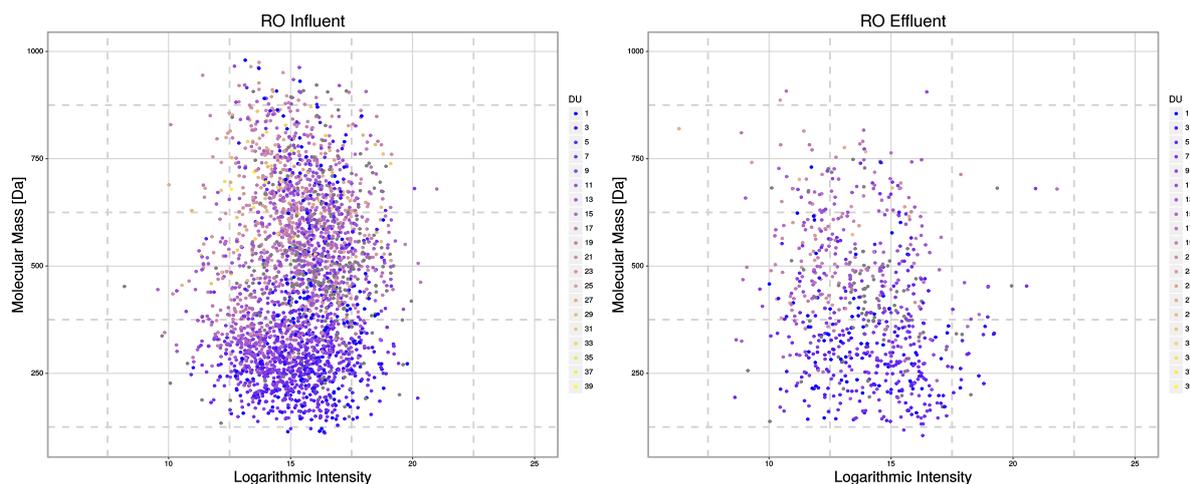


Figure 2: Comparison of mass, absolute logarithmic intensity and the degree of unsaturation for the identified features in the influent and effluent of RO. DU was calculated for features with an assigned elemental composition.

The findings described above were confirmed by another peer reviewed sources and helped to estimate the real performance of a RO treatment for DOM. Our non-targeted methodology was successfully applied to uncover these crucial changes caused by the RO treatment.

Van Krevelen plot is one of the most important graphical exploratory analysis techniques for DOM with the chemical information from LC-MS data. Hereby the atomic ratio X/C , where X is an element of interest, is plotted against H/C . The application of the van Krevelen plot in petroleomics and natural DOM chemistry helped to uncover groups of similarly structured compounds. These correspond to certain regions in the H/C vs. O/C plot, as chemical properties of substances often depend on their atomic content. Also, a comparison of multiple samples can reveal big scale processes and reactions in the system (e.g. oxidation of DOM).

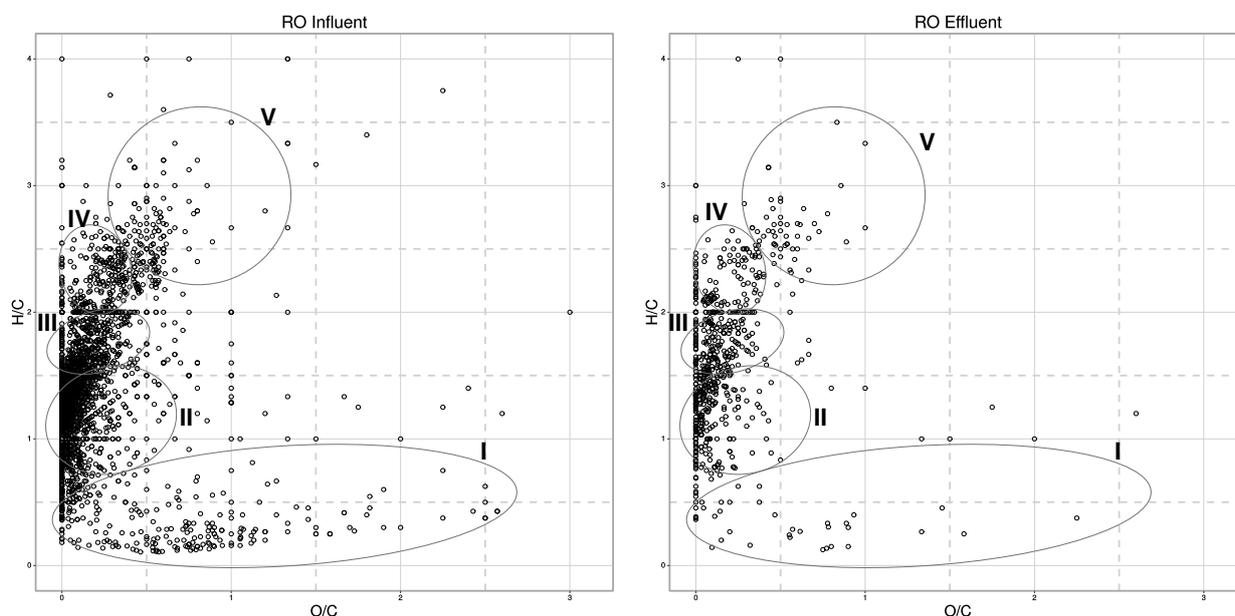


Figure 3: Atomic ratios O/C vs. H/C (van Krevelen plot) of DOM features with assigned elemental formulas. The numbered circles signify structural classes of DOM.

Here in a similar fashion we elucidated chemical classes from the van Krevelen plot as well (Figure 3). Wastewater DOM can have multiple origins and often contains synthetic compounds, which are not as predictable as the highly structured natural DOM.

Despite the heterogeneous matrix we uncovered distinct chemical regions in the van Krevelen plot. The aromatic nature of the region I was corroborated by the high DU and low H/C values and contains many heavy substances. The region II describes mixed oxygen-poor aromatics and aliphatics and contains the majority of identified DOM. The region III made of oxygen-rich aliphatics contains substances with a small mass and a moderate amount of N compared to other groups. The aliphatic region IV with a very high H/C and a moderate oxygen content contains mainly substances with a high mass and a low nitrogen content. Highly aliphatic region V with a variable oxygen content contains rather small masses and a high percentage of nitrogen. The van Krevelen plot allows us to assess the major chemical changes of DOM on a big scale and prioritize farther research.

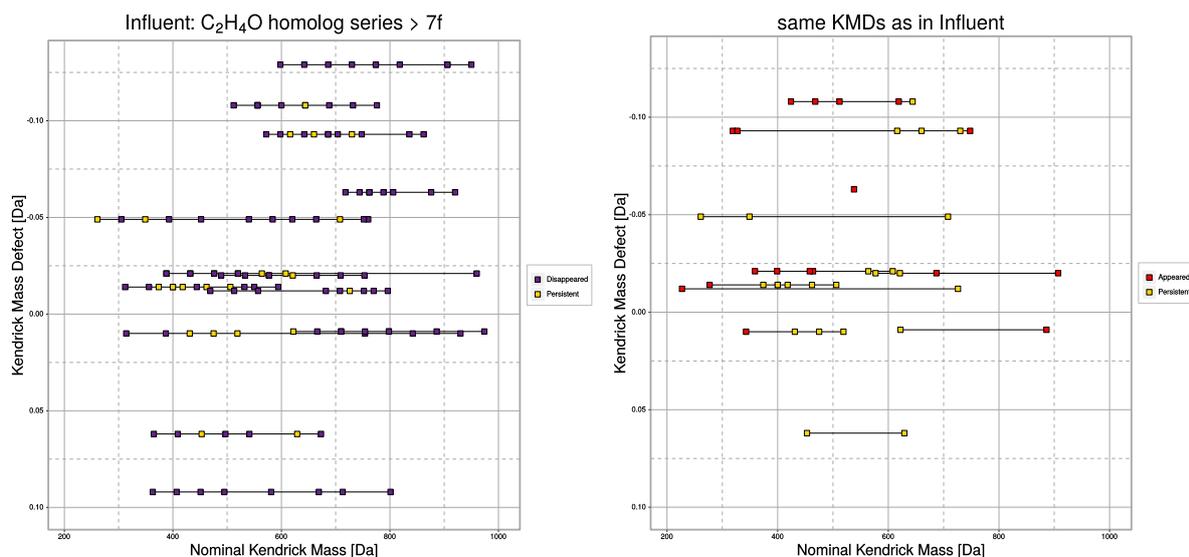


Figure 4: C_2H_4O -homolog series in Influent and their development in the Effluent of RO.

We successfully applied the Kendrick Mass Defect statistics (KMD) to find series of homologs in the complex wastewater DOM matrix. KMD statistics normalize the mass of substances to a certain molecular moiety, which leads to the same KMD for all substances that contain one or more of these building blocks. Thus, it is easy to graphically distinguish propagating homolog series in a complex sample. Homologs occur in natural DOM, but more interestingly in the synthetic organics as well, e.g. priority substances like surfactants, polyfluorinated compounds or chlorine substitute series like PCBs. Figure 4 shows numerous $-C_2H_4O-$ series, which often correspond to surfactants e.g. polyethoxylated series in municipal wastewater treatment. KMD calculation offers a good bridge between the non-target and suspect screenings and can be used to rapidly screen wastewater for priority substances.

The methods and tools described above are routinely used for data analysis in metabolomics and proteomics and were successfully applied here. This confirmed that our methodology is well suited to perform a comprehensive non-target analysis of a wastewater treatment.

Acknowledgment

This work has been supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 642904 - TreatRec ITN-EID project.