Advanced analytical procedures for the characterization of disinfection by-products and methodology for remote sensing of natural organic matter

Deliverable D1.1, WP1

Project Number	**101081980**
Project Title	Climate-resilient management for safe disinfected and non-disinfected water supply systems
Project Acronym	SafeCREW
Project Duration	November 2022 – May 2026
Call identifier	HORIZON-CL6-2022-ZEROPOLLUTION-01
Due date of Deliverable	Month 9, 31.07.2023
Final version date	Month 9, 26.07.2023
Updated version date	Month 21, 12.07.2024
Dissemination Level	PU (Public)
Deliverable No.	D1.1
Work Package	WP1
Task	T1.1 and T1.3
Lead Beneficiary	UFZ
Contributing Beneficiaries	EUT, CAT
Report Author	Maolida Nihemaiti (UFZ), Sonia Jou (EUT), Laura del Val (EUT), Danillo Lange (EUT), Marc Ribalta (EUT)
Reviewed by	Thorsten Reemtsma (UFZ), Irene Jubany (EUT)
Approved by	Anissa Grieb (DVGW-TUHH)
Updated by	Maolida Nihemaiti (UFZ), Sonia Jou (EUT)
Update approved by	Barbara Wendler (DVGW-TUHH)

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

History of Versions		
Version	**Publication Date**	**Change**
0.1	04/07/2023	First draft
0.2	26/07/2023	Final review
1	26/07/2023	Submitted to commission
1.1	12/07/2024	Final review
2	12/07/2024	Submitted to Commission

History of Changes	
Date/section	**Nature of change and reason**
10.07.2024 page 2	To address the reviewers' comment: When updating the deliverable, please include in the history of changes where and what have been added comparing to the initial version of the deliverable. This will facilitate to review the deliverable": Previous "History of Changes" renamed to "History of Versions". "History of Changes" added.
10.07.2024, page 8 (Abstract) and page 19 (section 2.5)	To address the reviewers comments "The report is very descriptive and provides a detailed approach to analytical approaches for DBP analysis (Tasks 1.1 and 1.3). Whilst the results of Task 1.3 are extensively discussed, the report does not sufficiently expand on the results arising from Task 1.1 and particularly the novelty aspects of the analytical approaches are not sufficiently outlined and discussed": A new section "2.5: Application of developed method for analysing samples from CS#3" was added to the report. In this section, the application of newly developed method (freeze-drying enrichment followed by SFC-HRMS analysis) to monitor novel sulfonated DBPs in water samples collected from Tarragona, Spain was described, as well as the results were discussed in the text and illustrated in Figure 2-5. In addition, the following sentence was added to the "Abstract" to justfy the novelty of developed method: "The generic methods to analyse disinfection byproducts, such as mass spectrometry coupled with either gas chromatography (GC-MS) or reversed phase liquid chromatography (RPLC-MS), can not analyse all DBPs, especially the highly polar fractions of DBPs because these DBPs are not volatile enough for GC-MS analysis and can not be retained on RPLC columns. To fill this gap, an advanced analytical method based on supercritical fluid chromatography coupled to high resolution mass spectrometry (SFC-HRMS) was established and further optimized at the UFZ." An additional sentence was also added to the "Abstract" as follows to describe the results from real drinking water samples: "The freeze-drying enrichment method followed by SFC-HRMS analysis developed in this study was applied to analyse the water samples collected from CS#3 in Tarragona, Spain in July 2023. In total eleven novel sulfonated DBPs, which were never reported before, were found present in all disinfected drinking water samples."

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

10.07.2024, page 11, section 2.1	To address the reviewers comments “The title of section 2.1 should be added”: The title of section 2.1 was added as “ 2.1 Material and methods” and section 2.1.1 was renamed as “2.1.1 Chemicals”.
10.07.2024, pages 33 and 36, section 3.6	To address the reviewers comments “The title of Figures 3.7 and 3.8 should be made readable”: The title of each small figure (a, b, c, d) was added above each figure in Figures 3-7 and 3-8 to make them readable.
10.07.2024, pages 24 and 30	To address the reviewers comments: ‘Typos to be corrected page 22, 28”: the pages 24 and 30 (in revised version) were carefully proof read and no typo was found to be corrected.

This project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No 101081980.

Table of contents

Table of contents 4
Abbreviations 6
Abstract 8
1 Introduction 9
1.1 Analytical methods to characterise disinfection byproducts (DBPs) 9
1.2 Methodology for remote sensing of natural organic matter 9
2 Method optimization for DBP analysis 11
2.1 Material and methods 11
2.1.1 Chemicals 11
2.1.2 Lab-scale chlorination of cysteine 11
2.2 Freeze-drying enrichment followed by SFC-HRMS analysis 12
2.3 Optimization of SFC-HRMS parameters 14
2.3.1 Issue of double peaks 14
2.3.2 Issue of peaks overlay 15
2.4 Optimization of sample enrichment process 17
2.5 Application of developed method for analysing samples from CS#3 19
2.6 Transferability of SFC-MS method to the generic RPLC-MS 20
3 Method optimization for NOM analysis using remote sensing 21
3.1 Use of satellite data to estimate surface water quality parameters relevant for DBPs formation 21
3.2 Study site and data availability 23
3.2.1 Study site 23
3.2.2 Satellite data 23
3.2.3 Field data on surface water quality 24
3.3 Methodology developed for the derivation of quality data from satellite images 26
3.3.1 General methodology 26
3.3.2 Images acquisition and initial treatment 27
3.4 Application of water mask 28
3.4.1 Application of cloud mask 28
3.4.2 Deriving water quality parameters 29
3.5 Methodology applied to the case-study 30
3.5.1 Satellite data 30
3.5.2 Combination with field data 31
3.6 Results and discussion 32

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

3.6.1 Water quality parameters obtained from satellite data 32
3.6.2 Comparison of satellite-derived and field parameters 34
3.6.3 Checking potential of satellite data for early warning on organic matter concentration 37
3.7 Limitations 38
3.7.1 Limitations of used satellite data 38
3.7.2 Limitations of field data availability 39
3.7.3 Limitations of the method 39
4 Conclusions 41
5 Bibliography 43

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

Abbreviations

EC	European Commission
WP	Work package
GC-MS	Gas Chromatography-Mass Spectrometry
RPLC	Reversed Phase Liquid Chromatography
SFC	Supercritical fluid chromatography
WAX	Weak Anion Exchanger
HRMS	High resolution mass spectrometry
NMR	nuclear magnetic resonance spectroscopy
RT	Retention time
SPE	Solid phase extraction
ClANSA	Chloroacetonitrilesulfonic acid
Cl_2ANSA	Dichloroacetonitrilesulfonic acid
BrClANSA	Bromochloroacetonitrilesulfonic acid
BrANSA	Bromoacetonitrilesulfonic acid
Br_2ANSA	Dibromoacetonitrilesulfonic acid
ClAcAmSA	Chloroacetamidesulfonic acid
Cl_2AcAmSA	Dichloroacetamidesulfonic acid
BrClAcAmSA	Bromochloroacetamidesulfonic acid
BrAcAmSA	Bromoacetamidesulfonic acid
Br_2AcAmSA	Dibromoacetamidesulfonic acid
ClAcAlSA	Chloroacetaldehydesulfonic acid
Cl_2AcAlSA	Dichloroacetaldehydesulfonic acid
BrClAcAlSA	Bromochloroacetaldehydesulfonic acid
BrAcAlSA	Bromoacetaldehydesulfonic acid
Br_2AcAlSA	Dibromoacetaldehydesulfonic acid
Cl_2MSA	Dichloromethanesulfonic acid
Br_2MSA	Dibromomethanesulfonic acid
DBPs	Disinfection by-products
NOM	Natural organic matter
DOM	Dissolved organic matter
DWDS	Drinking water distribution systems
DWTP	Drinking water treatment plant
CDOM	Chromophoric dissolved organic matter
UV-254	UV-absorbance at 254 nm
T	Temperature
Br	Bromide ion
D	Chlorine dose
t	Reaction time of residual chlorine
CHL-a	Chlorophyll-a
TSS	Total suspended solids
TP	Total phosphorus
DO	Dissolved oxygen
BOD	Biochemical oxygen demand
COD	Chemical oxygen demand
NDTI	Normalized difference turbidity index
USGS	United States Geological Survey
NASA	National Aeronautics and Space Administration

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

ESA	European Space Agency
SAR	Synthetic aperture radar
CHE	Ebro River Waterboard (Confederación Hidrográfica del Ebro)
CAT	Consorci d'Aigües de Tarragona
EB0	Intake of the water treatment plant
TOC	Total organic carbon
CRS	Coordinate Reference System
SLC	Single look complex
MNDWI	Modified normalized difference water index
SCL	Scene Classification
API	Application Programming Interface

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

Abstract

This deliverable describes two methodologies that are relevant for drinking water characterization and will be applied in further tasks of SafeCREW: (1) Advanced analytical methods for disinfection byproduct (DBPs) characterization and (2) a method to measure the natural organic matter using remote sensing.

The generic methods to analyse disinfection byproducts, such as mass spectrometry coupled with either gas chromatography (GC-MS) or reversed phase liquid chromatography (RPLC-MS), can not analyse all DBPs, especially the highly polar fractions of DBPs because these DBPs are not volatile enough for GC-MS analysis and can not be retained on RPLC columns. To fill this gap, an advanced analytical method based on supercritical fluid chromatography coupled to high resolution mass spectrometry (SFC-HRMS) was established and further optimized at the UFZ. This method was applied to identify the extremely polar novel sulfonated DBPs in disinfected drinking water. Various studies were conducted to optimize the SFC parameters (mobile phase, column) in order to get better separation and signal intensity of sulfonated DBPs. Three different sample enrichment methods, such as freeze-drying, solid phase extraction with weak anion exchanger cartridges, and azeotrope evaporation, were tested for the recovery of sulfonated DBPs spiked in tap water. Two sulfonated DBPs with analytical standards, dichloromethane sulfonic acid and dibromomethane sulfonic acid, appeared to have 80-120% apparent recovery rates during these enrichment processes. The freeze-drying enrichment method followed by SFC-HRMS analysis developed in this study was applied to analyse the water samples collected from CS#3 in Tarragona, Spain in July 2023. In total eleven novel sulfonated DBPs, which were not reported before, were found present in all disinfected drinking water samples. This SFC-HRMS method combined with sample enrichment processes will be applied in WP2, WP3, and WP4 for the target analysis of sulfonated DBPs and non-target screening for the identification of unknown DBPs in drinking water treatment plants and distribution networks at lab, pilot, and full-scale. An additional method to characterize sulfonated DBPs with the generic RPLC-MS was also set up and will be accessible to other partners within the SafeCREW project, allowing them to get preliminary data on sulfonated DBPs in their lab without sample shipment to the UFZ.

A methodology is proposed to use satellite data for the derivation of surface water quality parameters related to organic matter content, and thus, to the formation of DBPs. This proposal lays on the necessity for a higher spatial coverage to monitor the evolution of organic matter in surface water bodies used as sources for drinking water, being organic matter the major precursor for DBP formation in drinking water. The proposed methodology used relevant functions found in literature and is applied to the lower part of the Ebro River, between Mequinenza reservoir and Tarragona drinking water intake. Results showed limited application for the specific case study of Tarragona and limited representativeness of the field parameters. However, they also show potential for estimating organic matter-related surface water parameters if some improvements are done. The proposed methodology was therefore extended with a calibration step and will be implemented over a longer time series of data to ensure proper calibration of the functions' parameters.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

1 Introduction

1.1 Analytical methods to characterise disinfection byproducts (DBPs)

When drinking water is disinfected, the natural organic matter and inorganic ions (e.g., bromide) in water can react with disinfectants (e.g., chlorine, chlorine dioxide) to generate various disinfection byproducts (DBPs) (Richardson et al., 2007). The formation of DBPs is undesired as DBP exposure is proven to associate with the risk of bladder cancer and birth defects in humans (Costet et al., 2011). Over 700 DBPs are characterized so far (Richardson and Kimura, 2020). Of these, the (semi)volatile classes are quantified using gas chromatography-mass spectrometry (GC-MS). A small number of DBPs are regulated, such as trihalomethanes (THMs) and haloacetic acids (HAAs). However, many other DBPs, especially nitrogen containing compounds, are more toxic than the regulated DBPs (Wagner and Plewa, 2017). The majority of total organic halogens produced during chlorination are still unknown (Hua and Reckhow, 2007). Therefore, revealing the unknown DBPs is becoming important to identify the toxic potencies in disinfected drinking water.

The generic methods to analyse novel DBPs are either gas chromatography (GC) or reversed phase liquid chromatography (RPLC) coupled with MS. These methods, however, can easily overlook the highly polar fractions of DBPs as they are not volatile enough for GC analysis nor retained by reversed phase columns in RPLC (Reemtsma et al., 2016). The SafeCREW project aims to overcome these limitations by developing advanced analytical tools to identify the unknown polar DBPs in drinking water. An alternative analytical method based on supercritical fluid chromatography (SFC) coupled to MS is optimized and will be applied in SafeCREW.

SFC has a unique mobile phase comprising supercritical CO_2 (nonpolar) and polar modifier (e.g., methanol) and is compatible with both polar and nonpolar columns, allowing SFC to analyze a large number of analytes of a wide polarity range (West, 2018). SFC is widely used in the pharmaceutical industry and food science (Si-Hung and Bamba, 2022). It has recently been applied to characterize the persistent and mobile contaminates and their transformation products in aquatic environments (Schulze et al., 2019; Seiwert et al., 2021).

The UFZ team has recently applied SFC to identify 15 novel sulfonated DBPs in disinfected drinking water and swimming pool waters, including haloacetonitrilesulfonic acids, haloacetamidesulfonic acids, and haloacetaldehydesulfonic acids (Nihemaiti et al., 2023). Based on this preliminary method, the following studies have been conducted during Task 1.1 to:

- Optimize SFC-MS method to improve separation and signal intensity of polar DBPs on SFC chromatogram
- Compare different sample enrichment processes
- Test the transferability of SFC-MS method to the generic RPLC-MS

1.2 Methodology for remote sensing of natural organic matter

Within WP1, the SafeCREW project aims to develop techniques to improve analytics, sampling, and protocols for the comprehensive characterization of DBPs. In this context, Task 1.3 is exploring the potential of satellite imagery for the estimation of natural organic matter (NOM) in the surface water used as a source for drinking water. DBP estimation is mainly based on empirical correlations between organic matter content and chlorine dosages (Sadiq and Rodriguez, 2004; Chowdhury et al., 2009), however, NOM is not always monitored. The possibility of estimating indirectly the trends of NOM in surface water upstream and at the drinking water plant intake, could provide insightful information

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

for the adaptative planning of the drinking water plant operation, possibly mitigating the formation of DBPs.

Satellite images have been successfully correlated with NOM concentrations in surface water, as well as other surface water quality parameters (*3. Method optimization for NOM analysis using remote sensing*). Still, the techniques have not been widely applied because of the many limitations and unknown behaviour in a broad range of circumstances.

To monitor NOM at a catchment scale based on satellite images, SafeCREW aims to develop a method combining satellite and in situ (sensor-based) variables to predict NOM as input to model trends in DBP formation. To achieve this purpose three objectives were set:

- Test the applicability in flowing surface water of approaches found in literature.
- Explore additional variables that could improve the model.
- If feasible, test more advanced statistical methods to improve results.

As a first step towards the global objective, and to comply with Tasks 1.3 of WP1, the best models found in the literature were tested to predict quality parameters measured in the field in an area of interest in one of the case studies, the Tarragona site in Spain (lower catchment of Ebro River, downstream from Mequinenza reservoir).

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

2 Method optimization for DBP analysis

2.1 Material and methods

2.1.1 Chemicals

All chemicals were of analytical grade and used as received without further purification. Carbon dioxide Premium (4.5) was used for SFC. Methanol, acetonitrile, and formic acid were provided by Biosolve (Valkenswaard, Netherlands). Sodium hypochlorite solution (6-14%) was provided by Th.Geyer GmbH (Hamburg, Germany). Cysteine (≥99%) was purchased from Merck (Darmstadt, Germany). Dichloromethanesulfonic acid and dibromomethanesulfonic acid were obtained from Toronto Research Chemicals. Waters Oasis Weak Anion Exchanger (WAX) 6 cc Vac Cartridge (150 mg Sorbent per Cartridge) was used for solid phase extraction. Ultrapure water was obtained from a Merck Milli-Q Integral 5 system (Darmstadt, Germany).

2.1.2 Lab-scale chlorination of cysteine

The mixture of the newly identified sulfonated DBPs (haloacetonitrilesulfonic acids, haloacetamidesulfonic acids, and haloacetaldehydesulfonic acids in Figure 2-1) was prepared by the chlorination of cysteine. Experiments were conducted in ultrapure water in amber glass bottles at room temperature (23±2 ℃) for 23 h. The commercial solution of sodium hypochlorite was standardized by measuring the absorbance of hypochlorite anion at 292 nm (ε = 362 $M^{-1}cm^{-1}$) (Furman and Margerum, 1998). The initial concentrations of cysteine and chlorine were 250 µM and 1.25 mM, respectively. An additional experiment was conducted in the presence of 500 µM of Br^- to prepare the brominated analogues of sulfonated DBPs.

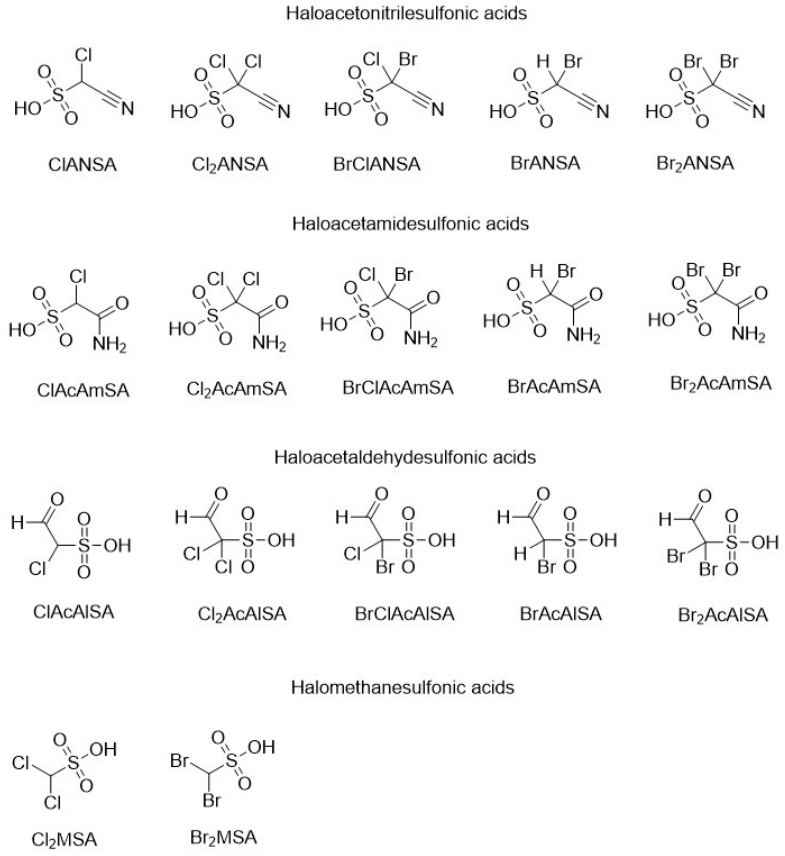


Figure 2-1. The sulfonated DBPs mentioned in this report. ClANSA- chloroacetonitrilesulfonic acid, Cl_2ANSA- dichloroacetonitrilesulfonic acid, BrClANSA- bromochloroacetonitrilesulfonic acid, BrANSA-

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

bromoacetonitrilesulfonic acid, Br_2ANSA- dibromoacetonitrilesulfonic acid, ClAcAmSA-chloroacetamidesulfonic acid, Cl_2AcAmSA- dichloroacetamidesulfonic acid, BrClAcAmSA-bromochloroacetamidesulfonic acid, BrAcAmSA- bromoacetamidesulfonic acid, Br_2AcAmSA-dibromoacetamidesulfonic acid, ClAcAlSA- chloroacetaldehydesulfonic acid, Cl_2AcAlSA-dichloroacetaldehydesulfonic acid, BrClAcAlSA- bromochloroacetaldehydesulfonic acid, BrAcAlSA-bromoacetaldehydesulfonic acid, Br_2AcAlSA- dibromoacetaldehydesulfonic acid, Cl_2MSA-dichloromethanesulfonic acid, Br_2MSA-dibromomethanesulfonic acid

2.2 Freeze-drying enrichment followed by SFC-HRMS analysis

A preliminary method based on freeze-drying enrichment followed by SFC-high resolution MS (HRMS) measurement for analyzing DBPs is available at the UFZ. This method was previously validated by identifying 15 novel sulfonated DBPs (haloacetonitrilesulfonic acids, haloacetamidesulfonic acids, and haloacetaldehydesulfonic acids, Figure 2-1) in disinfected drinking water and swimming pool water (Nihemaiti et al., 2023). In this method, a 40 mL aliquot of a water sample is filled in a 50 mL centrifuge tube and stored at -20 ℃ overnight. The frozen sample is then placed in a freeze-dryer (Alpha1-4, Christ, Germany) at 15 °C and 1.65 mbar for 30 h until dryness. The residue is reconstituted in 400 µL of acetonitrile and water (90:10, by volume), transferred into an Eppendorf tube, and then centrifuged at 13000 min^{-1} for 10 min. The supernatant is then transferred into glass vials and kept at -20 °C until analysis. An ultrapure water blank is prepared following the same procedure.

Samples are analyzed using an ACQUITY UPC2 system coupled with a Synapt GS2 QTOF (Waters, Eschborn, Germany) in negative electrospray ionization mode. Nitrogen and argon are used as cone and collision gases, respectively. Specific parameters are given in Table 2-1 (15 min method). The data is recorded in centroid mode with a 0.08-s scan time over the mass range of *m/z* 50-1200 (resolution approximately 20000). The MS^E acquisition is performed to simultaneously collect two data sets: a low-collision-energy scan to obtain parent ion information and an elevated-collision-energy scan (15-40 eV) to get all fragment ions. The MS/MS spectra of the newly identified DBPs are also recorded using a collision energy ramp of 15 to 40 eV. Data is processed using Waters MassLynx, MarkerLynx, and TargetLynx software. Exact molecular formulas of unknown DBPs are assigned based on the combination of elements $C_{0–20}$, $H_{0–100}$, $N_{0–5}$, $O_{0–10}$, $P_{0–2}$, $S_{0–2}$, $I_{0–4}$, $Br_{0–4}$, $Cl_{0–4}$, and $Na_{0–2}$ within the mass tolerance of 5 ppm. Molecular formulas with Cl and Br are confirmed manually using a simulated isotope model.

Table 2-1. Detailed parameters of SFC-HRMS

	15 min method	**19 min method**
Column	BEH column (3.0 mm × 100 mm, 1.7 µm, Waters, Eschborn, Germany)	BEH column (3.0 mm × 100 mm, 1.7 µm, Waters, Eschborn, Germany)
Column back pressure	1800 psi	2200 psi
Column temperature	55 °C	55 °C
Mobile phase	(A) CO_2 and (B) methanol/water cosolvent (95:5, by volume)	(A) CO_2 and (B) methanol/water cosolvent (95:5, by volume) containing 10 mM ammonium formate

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

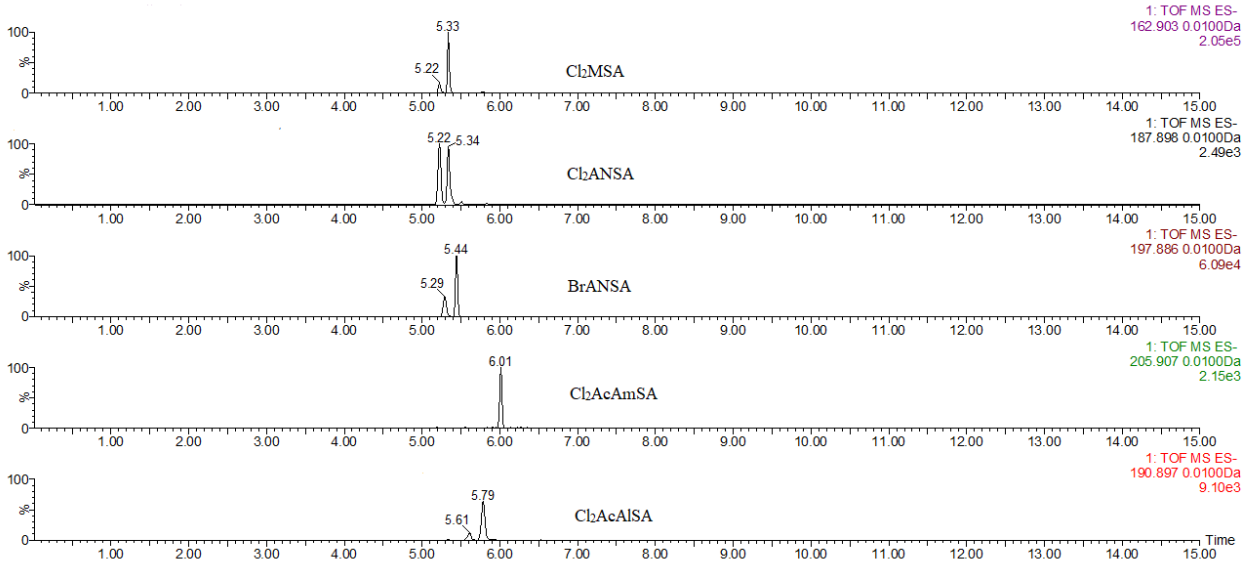
	containing 10 mM ammonium formate	
Mobile phase gradient	0–0.5 min, 1% B; 9–12.5 min, 50% B; 12.6–15 min, 1% B	0–0.2 min, 1% B; 8.5 min, 25% B; 12-17 min, 50% B; 17.1 min, 1%B
Flow rate	1.5 mL/min	1.3 mL/min
Makeup flow	methanol/water (90:10, by volume) with 0.1% formic acid	methanol/water (90:10, by volume) with 0.1% formic acid
Makeup flow rate	0.3 mL/min	0.3 mL/min
Injection volume	5 µL	5 µL
Capillary voltage	-2 kV	-2 kV
Source temperature	140 °C	140 °C
Desolvation temperature	550 °C	550 °C
Cone voltage	20 V	20 V
Source offset	40 V	40 V
Desolvation gas flow	950 L/h	950 L/h

Figure 2-2a illustrates the extracted ion chromatogram of some of the novel sulfonated DBPs detected from a tap water using this method. Structures of DBPs are given in Figure 2-1. Compounds appeared at retention time (RT) between 5 min to 6.5 min with sharp (Gaussian) peak shapes. According to the exact mass and fragmentation pattern obtained from HRMS, the structures of these compounds can be proposed. For instance, the MS/MS spectrum (Figure 2-2b) of *m/z* 197.8863 shows the molecular formula of parent ion $[C_2HNBrSO_3]^-$ and its fragments (Br^- and $[SO_3]^{-\bullet}$), supporting the structure proposal as bromoacetonitrilesulfonic acid (BrANSA).

Based on lab-scale chlorination experiments, amino acid cysteine was found as a precursor of these sulfonated DBPs, suggesting if the drinking water source is affected by wastewater discharge or algal bloom, enriched in biomolecules, then the formation yield of these DBPs may increase upon chlorination (Nihemaiti et al., 2023). As the analytical reference standards of these novel DBPs are currently unavailable, a mixture of them was produced by chlorination of $^{13}C3$-^{15}N-cysteine and analysed using nuclear magnetic resonance (NMR) spectroscopy for further structural confirmation and quantification. The calibration curves were made on SFC-HRMS using this DBPs' mixture with known concentrations and used to quantify the novel sulfonated DBPs in drinking water (Nihemaiti et al., 2023). The approach of synthesizing DBPs by lab-scale chlorination of a known precursor followed by NMR and SFC-HRMS analysis will be applied in the SafeCREW project for the identification and quantification of novel DBPs in the case of no analytical standards availability.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

a)



b)

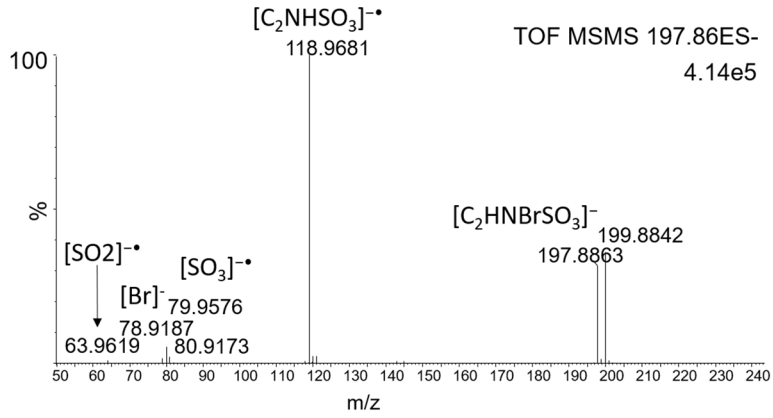


Figure 2-2. a) The extracted ion chromatogram of sulfonated DBPs detected from a tap water. b) The MS/MS spectrum of *m/z* 197.8863 (BrANSA). The sample was analysed using SFC-HRMS in negative ionization mode. (b) was adapted from Nihemaiti et al. (2023). DBPs' structures are given in Figure 2-1.

2.3 Optimization of SFC-HRMS parameters

Within task 1.1, the SFC-HRMS parameters mentioned in section 2.2 (Table 2-1, 15 min method) were further optimized to improve sensitivity and separation of sulfonated DBPs.

2.3.1 Issue of double peaks

Some sulfonated DBPs appeared as double peaks in the chromatogram (Figure 2-2). To avoid this, analytes were tested to be reconstituted in the same solvent (methanol and water) as mobile phase of SFC after freeze-drying enrichment, instead of acetonitrile and water. Only one peak appeared, however, the signal intensity decreased by a factor of >5 (8.43×10^3 *vs.* 4.38×10^4, Figure 2-3). Considering the presence of sulfonated DBPs in real drinking water at low concentrations, the original

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

method using acetonitrile and water will be kept unchanged as it is more sensitive. In the case of the appearance of double peaks, both peaks will be integrated as one.

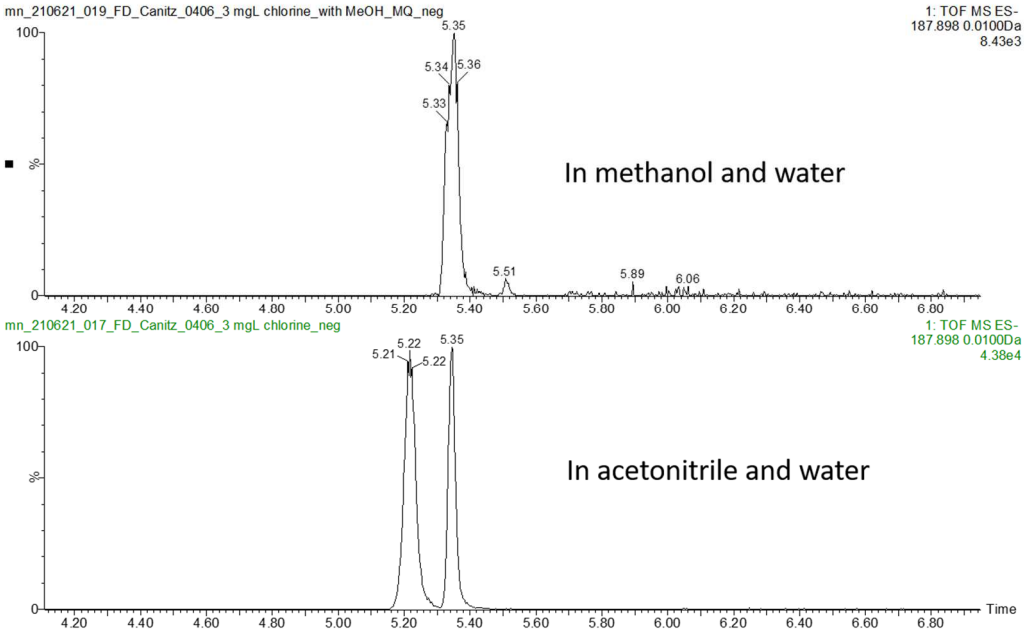


Figure 2-3. The extracted ion chromatogram of *m/z* 187.8980 (Cl_2ANSA) detected from a 40 mL of chlorinated drinking water. The sample was enriched by freezing-drying and reconstituted either in 400 µL of methanol and water (90:10, by volume) or acetonitrile and water (90:10, by volume).

2.3.2 Issue of peaks overlay

In "15 min method", the RT of sulfonated DBPs are very close to each other (Figure 2-2).The overlay of peaks makes it difficult to assign the right mass spectrum to each compound in MS^E mode and likely reduces the signal intensity of analytes. Therefore, different mobile phase and columns were tested within Task 1.1 to improve the separation of compounds on the chromatogram.

Condition 1: The flow rate of mobile phase was reduced from 1.5 mL/min to 1.3 mL/min and the gradinet was updated with an extra curve, allowing the methanol/water cosolvent to slowly reach 50% (19 min method in Table 2-1).

Condition 2: The Torus Diol column (3.0 x 100 mm, 1.7 µm, Waters Acquity UPC^2) replaced the original BEH column and mobile phase conditions were kept the same as condition 1.

Condition 3: The Torus Diol column (3.0 x 100 mm, 1.7 µm, Waters Acquity UPC^2) replaced the original BEH column and mobile phase conditions were the same as condition 1 (19 min run time), except that 10 mM of ammonium hydroxide was added to the methanol/water cosolvent instead of ammonium formate.

The chlorinated solution of cysteine in ultrapure water was injected into SFC-HRMS under these 3 different conditions and the RT of sulfonated DBPs were compared. In Table 2-2, the analytes were sorted from low RT to high RT and the RT difference compared to previous analyte was calculated. The sum of RT difference in conditions 1-3 is bigger than the original method, with condition 2 giving the biggest. This suggests that the sulfonated DBPs separate better on chromatogram when 19 min total run time is applied. Still, there are 2 compounds (BrClANSA, Br_2AcAlSA) that are overlayed with previous compounds (RT difference less than 0.03 min). Torus Diol column was reported to give better

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

sensitivity in negative ionization mode in SFC (Schulze et al., 2020). Although conditions 2-3 with Torus Diol column seem better in separation of sulfonated DBPs, the peak shape of some compounds are bad with large tailing (Figure 2-4). Therefore, condition 1 with BEH colmun and 19 min run time (19 min method in Tabl 2-1) was chosen to be applied in future studies.

Table 2-2. The RT difference of sulfonated DBPs analysed under different SFC conditions. Specific information on these conditions are given in the main text (section 2.3.2).

Original method (15 min method in Table 2-1)		**Condition 1** (19 min method in Table 2-1)		**Condition 2** (Torus Diol column, ammonium formate)		**Condition 3** (Torus Diol column, ammonium hydroxide)	
Analyte	RT difference (min)	Analyte	RT difference (min)	Analyte	RT difference (min)	Analyte	RT difference (min)
Cl_2ANSA		Cl_2ANSA		Cl_2ANSA		Cl_2ANSA	
ClANSA	0.02	ClANSA	0.03	ClANSA	0.08	ClANSA	0.11
BrClANSA	0.08	BrANSA	0.19	BrClANSA	0.15	BrClANSA	0.15
BrANSA	0.02	BrClANSA	0.01	BrANSA	0.05	BrANSA	0.06
Br_2ANSA	0.08	Br_2ANSA	0.21	BrClAcAlSA	0.25	Br_2ANSA	0.25
BrClAcAlSA	0.24	BrClAcAlSA	0.47	Br_2ANSA	0.02	BrClAcAlSA	0.04
Br_2AcAlSA	0.01	Br_2AcAlSA	0.03	Br_2AcAlSA	0.08	Br_2AcAlSA	0.09
Cl_2AcAmSA	0.23	Cl_2AcAmSA	0.43	Cl_2AcAmSA	0.50	Cl_2AcAmSA	0.48
BrClAcAmSA	0.10	BrClAcAmSA	0.13	BrClAcAmSA	0.14	BrClAcAmSA	0.16
Br_2AcAmSA	0.09	Br_2AcAmSA	0.15	Br_2AcAmSA	0.16	Br_2AcAmSA	0.15
sum	0.87	sum	1.65	sum	1.43	sum	1.49

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

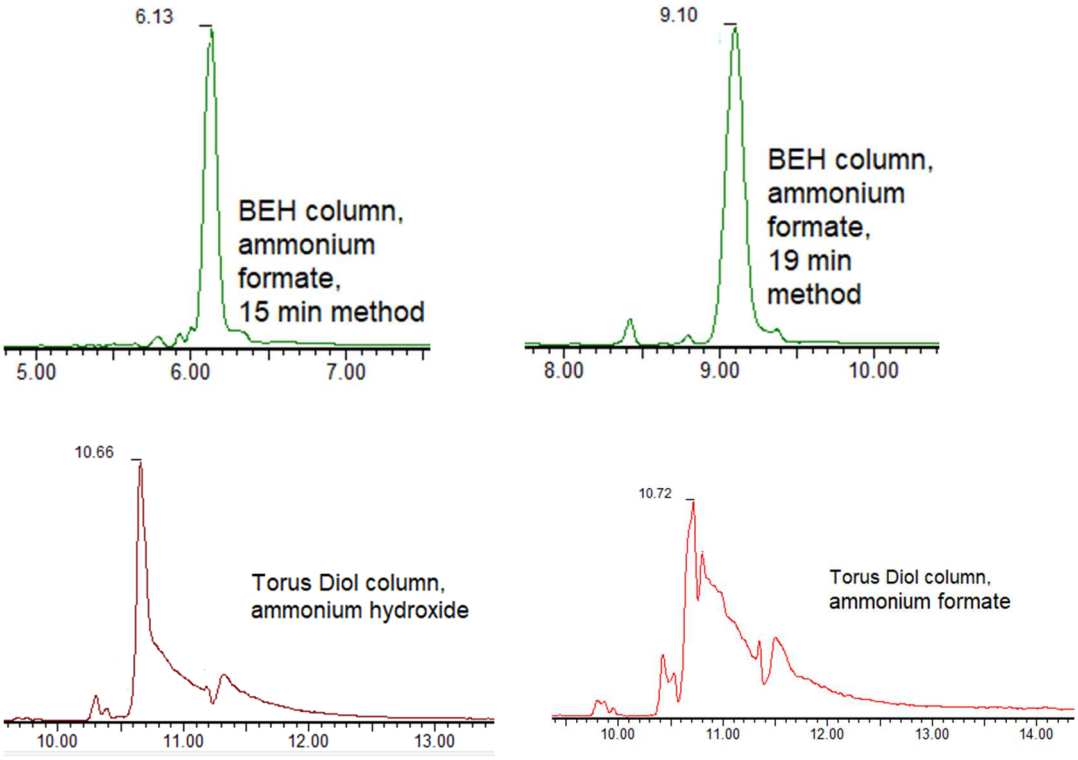


Figure 2-4. The peak shape of BrClAcAmSA under different SFC conditions.

2.4 Optimization of sample enrichment process

DBPs usually occur in ng/L to low µg/L levels in water. Therefore sample enrichment prior to MS analysis is often necessary to enhance the signal intensity of DBPs in order to facilitate their detection and identification. During Task 1.1, solid phase extraction (SPE) using WAX cartridges and azeotrope evaporation were tested in comparison to the preliminary established freeze-drying method.

During each test, an 80 µL aliquot of chlorinated cysteine solution containing the mixture of sulfonated DBPs was spiked into 40 mL of tap water and enriched. Tap water was used to take into account the effect of the water matrix. An 80 µL, 50 µg/L standard solution of each Cl_2MSA and Br_2MSA was also enriched after mixing with 40 mL of tap water. A procedural blank prepared with 80 µL of ultrapure water and 40 mL of tap water was enriched following the same procedure. An 80 µL aliquot of chlorinated cysteine solution or each of Cl_2MSA and Br_2MSA standard solution was mixed with 320 µL acetonitrile and directly injected to SFC-HRMS in order to calculate the apparent recovery as follows:

Apparent recovery = (Area of sulfonated DBPs spiked to tap water before enrichment/Area of sulfonated DBPs in acetonitrile) x 100

The detailed procedure of each enrichment method is given in Table 2-3. The apparent recoveries of Cl_2MSA and Br_2MSA analytical standards were within ~80-120% during all enrichment processes. Thus, all 3 enrichment methods are considered suitable for these two compounds. All other sulfonated DBPs present in chlorinated cysteine solution appeared to have variable and non-reproducible recovery rates. One possible reason is that although the chlorine was fully consumed when taking samples (after 23 hours of reaction), sulfonated DBPs can be continuously produced from intermediate products or transformed to other compounds during enrichment processes. To support this, the apparent

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

recoveries of Cl_2MSA and Br_2MSA in chlorinated cysteine solution were largely different from those obtained from the enrichment of their standard solutions. Furthermore, the evaporation process applied during SPE also is found to affect the composition of sulfonated DBPs. Thus, chlorinated cysteine solution does not seem the best starting material to test the recoveries of novel sulfonated DBPs. While finding solutions to synthesize their reference compounds, an alternative option could be to prepare the mixture of sulfonated DBPs in organic solvent (e.g., WAX cartridge SPE of chlorinated cysteine solution followed by reconstitution in pure acetonitrile). Using this mixture as starting material may prevent the hydrolysis reaction of intermediate products compared to the mixture in water.

During freeze-drying, all non-volatile substances in water can also be enriched with the target DBPs, which may lead to a strong matrix effect. Nevertheless, a wide range of DBPs are expected to be enriched and recovered during freeze-drying. Therefore it is suitable for non-target screening of drinking water to identify unknown DBPs. SPE with WAX cartridge is specifically suitable for anionic DBPs to be enriched but dissolved organic matter and much of inorganic salts to pass through. Therefore, both methods will be applied during SafeCREW depending on the goal of screening studies. Azeotrope evaporation gave promising results on the recovery of two sulfonated DBPs, but it will not be considered as a priority as its enrichment factor is 10 times lower than the other two methods.

Table 2-3. Sample enrichment methods studied in the report

Method name	Procedure	Apparent recovery
Freeze-drying[a]	1) A 40 mL aliquot of water sample is filled in 50 mL of falcon tube and stored at -20 ℃ overnight. 2) The frozen sample is then placed in a freeze-dryer (Alpha1-4, Christ, Germany) at 15 °C and 1.65 mbar for 30 h until dryness. 3) The residue is reconstituted in 400 µL of acetonitrile and water (90:10, by volume), transferred into an Eppendorf tube, and then centrifuged at 13000 min^{-1} for 10 min. The supernatant is then transferred into glass vials and kept at -20 °C until analysis.	Cl_2MSA: 94% Br_2MSA: 79%
SPE with WAX cartridge[b]	1) The cartridge is preconditioned with 4 mL of methanol with 0.1% NH_4OH, followed by 4 mL of methanol, and 4 mL of ultrapure water. 2) A 40 mL sample then passes through the cartridge at 1 drop/sec. 3) The cartridge is then washed with 25 mM of ammonium acetate buffer (pH 4.3) and dried with N_2 for 5 min. 4) The analytes are then eluted with 6 mL of methanol with 0.1% NH_4OH. 5) The elute is then evaporated until dryness under N_2, followed by reconstitution with 400 µL of acetonitrile and water (90:10, by volume).	Cl_2MSA: 91% Br_2MSA: 118%
Azeotrope evaporation[c]	1) A 4 mL aliquot of sample is mixed with 21 mL of acetonitrile in a falcon tube.	Cl_2MSA: 98% Br_2MSA:

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

	2) The mixture is then evaporated (Evaporator system XcelVap, Biotage, Sweden) until dryness at 40 °C under gentle stream of argon. 3) The residue is reconstituted in 400 µL of acetonitrile and water (90:10, by volume), transferred into 0.5 mL of Eppendorf tube and centrifuged for 15 min. An aliquot of 350 µL is filled in glass vials.	92%

[a] adapted from Nihemaiti et al. (2023)
[b] modified from Taniyasu et al. (2008)
[C] modified from Schulze et al. (2020)

2.5 Application of developed method for analysing samples from CS#3

The samples from a drinking water treatment plant (DWTP) and its relevant distribution network at CS#3 in Tarragona, Spain were collected in July 2023. An aliquote (40 mL) of each sample was then freeze-dried and reconstituted in the mixture of acetonitrile and water following the enrichment method in Table 2-3 (freeze-drying). The samples were then analysed using SFC-TOF applying the paramters in Table 2-1 (19 min method).

Results show that ClANSA, Cl_2ANSA, BrANSA, Br_2ANSA, Cl2AcAmSA, BrClAcAmSA, Br2AcAmSA, Cl_2AcAlSA, BrClAcAlSA, Cl_2MSA, and Br_2MSA (structures given in Figure 2-1) were present in almost all disinfected drinking water samples. Figure 2-5 shows the signal intensity of ClANSA as an example. ClANSA was not present in water before chlorination, and rapidly produced after chlorine addition at DWTP. The chlorination booster applied in the distribution network led to the continuous formation of ClANSA. However, subsequent chlorination boosters did not continuously increase its intensity, suggesting that the precursors of ClANSA might be depleted in the water. Another possible explanation is that ClANSA was degraded in the distribution network as haloacetonitriles are known to be unstable due to hydrolysis. More samples will be collected from CS#3 to monitor the seasonal changes in sulfonated DBPs formation potentials applying the method developed in this task.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

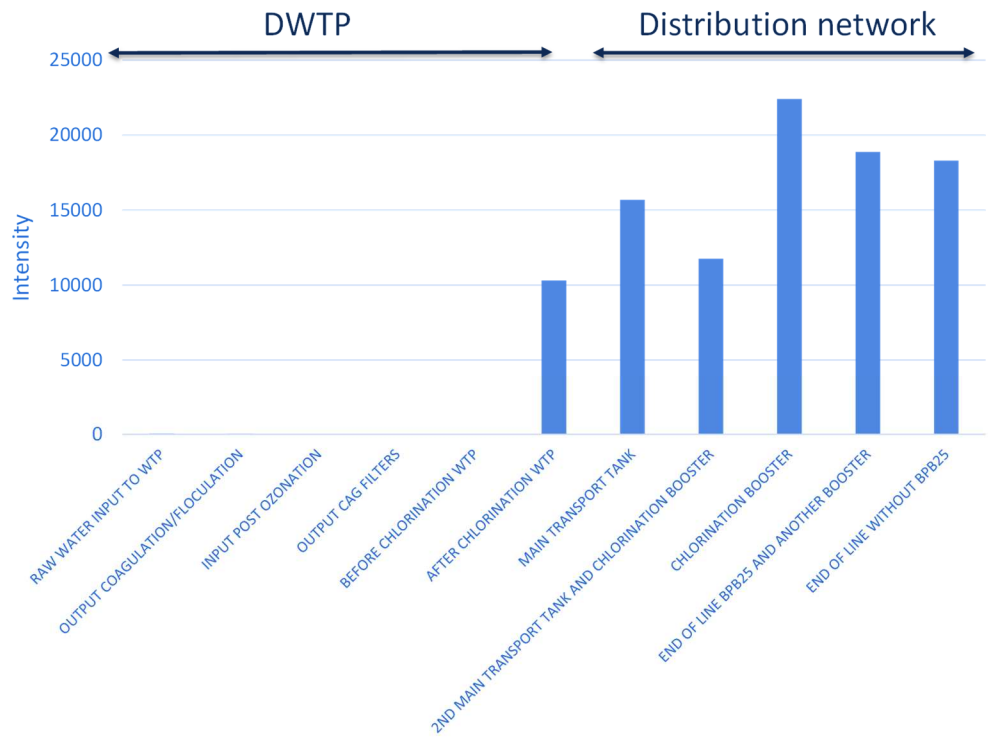


Figure 2-5. The signal intensity of ClANSA (obtained from SFC-TOF) in water samples from CS#3, Tarragona, Spain, in July 2023.

2.6 Transferability of SFC-MS method to the generic RPLC-MS

The transferability of SFC-MS method to the generic RPLC-MS was tested with the purpose of enabling the measurement of sulfonated DBPs by other partners within the SafeCREW project or by the broader scientific community.

The chlorinated cysteine solution containing the mixture of sulfonated DBPs was injected into the ACQUITY UPLC I-Class system coupled with a Synapt GS2 QTOF (Waters, Eschborn, Germany) in negative ionization mode. The MS parameters were similar to the one in Table 2-1, except that the source temperature was set at 120 °C, desolvation temperature was 600 °C, desolvation gas flow rate was 1000 L/h, cone voltage and source offset were 30 V. The sulfonated DBPs are extremely polar and hardly retained by RPLC stationary phases. Thus, the mobile phase parameters were set as follows: flow rate 0.4 mL/min, mobile phase (A) ultrapure water with 0.1% formic acid and (B) methanol. The mobile phase gradient was: 0-1 min: 99.9% of A, 10-11 min: 0.1% of A, 11.1-13 min: 99.9% of A. Two different columns were tested: Waters Atlantis Premier BEH C18 AX Column (1.7 µm, 2.1 x 100 mm) and ACQUITY UPLC HSS T3 Column (1.8 µm, 2.1 mm X 100 mm).

Results showed that the peak shape of sulfonated DBPs was too bad with tailing while applying the Atlantis Premier BEH C18 AX Column. Measurements were conducted by replacing methanol in mobile phase with acetonitrile and by reducing the formic acid concentration in solvent A to 0.001% (pH5.9). However, no improvement in peak shape was observed.

Most sulfonated DBPs appeared with sharp peak shapes while using the ACQUITY UPLC HSS T3 Column, but with low RT, as expected. Some compounds were not separated well from the dead volume (~0.8 min). Nevertheless, Cl_2ANSA, BrClANSA, BrANSA, Br_2ANSA, Cl_2AcAlSA, BrClAcAlSA, and Br_2AcAlSA appeared at RT about 3.3 min (Figure 2-6). Further modification of the mobile phase may help to solve

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

the problem of peak overlay for these compounds. Still, this method with ACQUITY UPLC HSS T3 column can be a preliminary method for other partners to try on their RPLC-MS to characterize the sulfonated DBPs without shipping all samples to the UFZ for SFC-HRMS analysis.

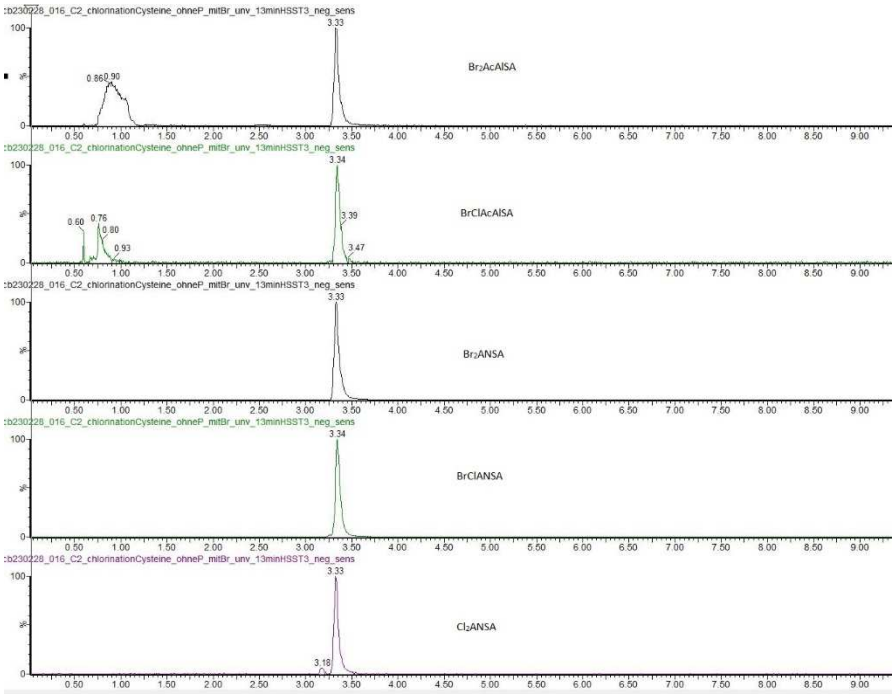


Figure 2-6. The extracted ion chromatogram of some sulfonated DBPs detected by RPLC-MS

3 Method optimization for NOM analysis using remote sensing

3.1 Use of satellite data to estimate surface water quality parameters relevant for DBPs formation

Generation of DBPs in drinking water depends on factors related to the source water entering the drinking water treatment plant (DWTP), like temperature and organic matter, but also on factors related to treatment, such as the type of treatment, the disinfectant type and doses or the retention time of water in the network (Sadiq and Rodriguez, 2004; Chowdhury et al., 2009). There are many models to predict the formation of DBPs based on empirical relations between DBPs concentration and different parameters. One of the most important parameter is the dissolved organic matter (DOM) (Sadiq and Rodriguez, 2004). Therefore, it could be expected that good estimates of DBPs production potential are reached based on organic matter measured in source water. This strong link is important because it opens the door to the application of satellite imagery to the indirect estimation of DPBs formation in drinking water based on the DOM detected with satellites (Chen et al., 2020).

The organic matter content in water in normally expressed as DOM, which is a heterogeneous class of water-soluble compounds that contain reduced (organic) carbon from a variety of biological and geological sources with a wide range of chemical reactivity. In practice, it is the organic matter that passes through a filter (Hartnett, 2018). Another form to express the organic matter content in water is the chromophoric dissolved organic matter (CDOM), which is the optically active fraction of DOM (Li et al., 2016). CDOM in surface waters includes autochthonal (microbial and phytoplankton) and allochthonous sources (terrestrial and anthropogenic inputs) (Li et al., 2016). Because CDOM is optically active, it can easily be related to the reflectance from surface water capture by satellite

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

sensors. Although CDOM and DOM relation is positive, it is not constant between seasons due to the varying contributions from the different sources of DOM (Li et al., 2016).

Because of the complexity and costs for sampling and analysing DOM, field data is scarce, and normally limited to a low temporal and spatial coverage. This is the reason why, despite the limitations to correlate CDOM and DOM, CDOM is increasingly used, as it allows the installation of autonomous optical sensors – with high temporal resolution - or the use of remote sensing - with high spatial resolution - to estimate organic matter in surface waters. The derivation of CDOM from satellite images consists mainly in generating an empirical relation between CDOM measured in samples taken from surface water bodies, and the reflectance captured by satellites like the Landsat 8 OLI or the Sentinel 2 MSI. The relation is then used to invert the satellite signal into CDOM regional maps.

How to better estimate CDOM concentrations from satellite images is the subject of intense research (Olmanson et al., 2020). This includes the use of different temporal windows and averaging algorithms for the satellite image processing, or the use of different machine learning methods to improve the empirical model relation between the reflectance and the measured CDOM (Ruescas et al., 2018; Koll-Egyed et al., 2021). However, all these methods are quite new and case-specific, therefore their broad applicability is still under question. The main limitation is the need to calibrate the algorithm with field data. This could be an obstacle due to the scarce amount of CDOM field data.

Although DOM is known to be the main contributor explaining DPBs production, other parameters can be connected to the production of DBPs. These are UV-absorbance at 254 nm (UV- 254), pH, water temperature (T), concentration of bromide ion (Br), chlorine dose (D), reaction time of residual chlorine (t), chlorophyll-a and fluorescence (Sadiq and Rodriguez, 2004). Some of these variables and others can be estimated based on satellite data, like chlorophyll-a (CHL-a), total suspended solids (TSS), water temperature (T), total phosphorus (TP), dissolved oxygen (DO), biochemical oxygen demand (BOD) and chemical oxygen demand (COD) (Gholizadeh et al., 2016).

As previously mentioned, the overall objective of task 1.3 is to set the basis for the use of satellite data in the estimation of produced DBPs in drinking water distribution systems (DWDS). Based on the cited literature it could be concluded that existing applications and algorithms are case-specific, requiring local field data to calibrate empirical models to the local conditions. A selection of the most relevant algorithms from the mentioned literature was done to derive variables from satellite data that could be related to the potential of production of DBPs and the content of organic matter and in line with the available field measures. The approach was tested, as a first attempt, in case-study number 3, Tarragona site.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

3.2 Study site and data availability

3.2.1 Study site

This task is developed and implemented in case study #3, Tarragona site. This site is a full-scale DWTP with a 400 km distribution network serving 800.000 inhabitants in winter and 1.500.000 in summer. Source water from the Ebro River undergoes pre-ozonation, flocculation, sand filtration, main stage ozonation, activated carbon filtration, ultraviolet disinfection, and final disinfection with NaOCl with several chlorination boosters along the network to keep the required level of residual free chlorine.

Water from the Ebro River is cached from two parallel channels aimed at providing water for agriculture in the Ebro Delta that collects river water in the municipality of Xerta. Upstream there are two dams, one in Flix (58 Km upstream) and one in Mequinenza (73 km upstream).

Surface water from the river Ebro has high seasonal flow and quality variations. Increasing severe droughts and storms affect water quality (organic matter, salinity, algal blooms). Increasing temperature affects the kinetics of DPB formation and bacterial growth in the network and tanks.

3.2.2 Satellite data

Landsat and Sentinel are two prominent Earth observation satellite programs that provide valuable data for monitoring and studying our planet's resources and environment.

Landsat is operated by the United States Geological Survey (USGS) and NASA, and it has been capturing images of Earth since 1972. The Landsat satellites orbit the Earth in a sun-synchronous polar orbit, covering the entire surface every 16 days. Landsat sensors capture images in several spectral bands, including visible, near-infrared, short-wave infrared, and thermal infrared. The spatial resolution of Landsat imagery is 30 meters, meaning that each pixel in the image represents an area of 30 by 30 meters on the ground. Landsat data is freely available, and its long historical archive enables the study of land cover changes over time.

Sentinel, on the other hand, is a satellite program managed by the European Space Agency (ESA) as part of the European Union's Copernicus program. The Sentinel satellites are equipped with various sensors that capture imagery in different spectral bands, including visible, near-infrared, and microwave. The Sentinel constellation comprises multiple satellites with different capabilities, such as Sentinel-2 for high-resolution optical imaging and Sentinel-1 for synthetic aperture radar (SAR) imaging. The frequency of image acquisition varies depending on the specific Sentinel mission. For example, Sentinel-2 typically captures global coverage every five days, while Sentinel-1 SAR data is acquired more frequently, often on a weekly or even daily basis. Sentinel-2 imagery has a spatial resolution of 10 meters for most bands and 20 meters for some, allowing for detailed land cover analysis and monitoring.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

Table 3-1. Sentinel-2 level 2-A bands

Band ID	Description	Spatial resolution (m)	Useful for parameters
B01	Coastal aerosol	60	-
B02	Blue	10	Turbidity, CDOM, TSS, CHL-low
B03	Green	10	Turbidity, CHL-low, Water mask
B04	Red	10	Turbidity, CDOM, CHL-high, Cyanobacteria
B05	Red Edge 1	20	Turbidity, CHL-high, Cyanobacteria
B06	Red Edge 2	20	-
B07	Red Edge 3	20	TSS
B08	Near Infrared (NIR) 1	10	-
B8A	NIR 2	20	-
B09	Water Vapour	60	-
B11	Short-wave IR (SWIR) 1	20	Water mask
B12	SWIR 2	20	-
AOT	Aerosol optical thickness	20	-
WVP	Scene average water vapour	20	-
SCL	Scene classification layer	20	Cloud masking

For this study, both satellites could provide useful data. However, it was decided to first start the analysis by using the Sentinel-2 satellite images (Level-2A images). This choice was motivated by the fact that the Sentinel-2 images have already undergone atmospheric correction, which makes the analysis more efficient. Moreover, the Level 2-A Sentinel-2 images include a Scene Classification Layer (SCL) image helpful to mask and identify clouds. In general, the Sentinel-2 satellite images contain multiple bands, as outlined in **Table 3-1.**, alongside the bands that were useful in this study for estimating parameters.

3.2.3 Field data on surface water quality

In the study area, there are five surface water monitoring points with water quality data available. Four of these (Flix, Ascó, Xerta, and Tortosa) are managed by the CHE (Ebro River Waterboard), and the fifth one is the intake (EB0) of the DWTP operated by the Consorci d'Aigües de Tarragona (CAT), partner of this project. Data was provided by both administrations for the year 2022. Unfortunately, measured parameters and frequency vary enormously between stations (Figure 3-1):

- In Flix, available data include water temperature, pH, conductivity at 25ºC, dissolved oxygen, turbidity, and dissolved mercury every 15 min.
- In Ascó flow rate, dissolved mercury, turbidity, water temperature, pH, conductivity at 20ºC, ammonium, nitrates, and absorbance at 254 nm are measured also continuously with a frequency of 15 min.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

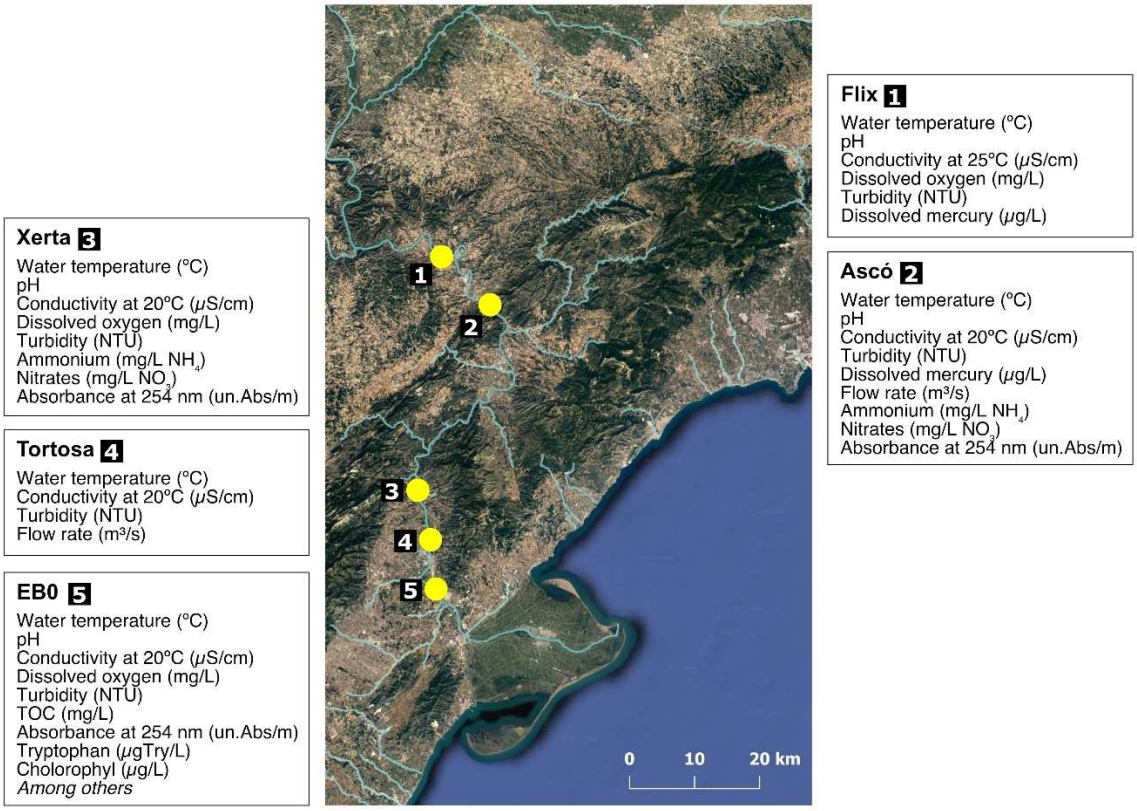


Figure 3-1. Locations of the monitoring points and the variables analyzed in the study site CS#3.

- Xerta station provides information on water temperature, pH, conductivity at 20ºC, DO, absorbance at 254 nm, turbidity, ammonium, and nitrates at a frequency of 15 min.
- Tortosa station records every hour parameters such as water temperature, conductivity at 20ºC, turbidity, and flow rate.
- And finally, at the water intake of the Tarragona DWTP, turbidity, absorbance at 254 nm, tryptophan, total organic carbon (TOC), temperature, conductivity, and DO, continuously. In this station also TOC and CHL-a are analyzed every 15 days and monthly respectively.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

3.3 Methodology developed for the derivation of quality data from satellite images

3.3.1 General methodology

The general methodology used to process the satellite data is summarized in Figure 3-2.

Figure 3- 2. Workflow for the satellite data processing

The methodology consists of collecting satellite data of the area of interest from various scenes using the Sentinel-2 satellite and processing it to end up with a map of the intensity of the parameters of interest. Firstly, the satellite images were selected based on two main characteristics: The date range, which in this study was of the entirety of 2022 (all images available from January to December 2022) and the scene location, composed of two scenes that can be observed in the section **3.5.1**.

Subsequently, the initial treatment of the images was started. The images were downloaded and then cropped and separated to better suit each zone of study, this process is detailed in section **3.3.2**.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

After the initial treatment and considering that the image is suitable for analysis (the pixels are clear and does not show any image quality problems), the water mask was created by selecting the pixels that were identified as water body, according to the method described in section **3.4**.

Another important step was to define the cloud mask of the image and decide if those clouds do not intercept the water pixels previously defined. This step is detailed in section **3.4.1**. Moreover, this step acted as a "filter" to select only the resulting images that were suitable for analysis.

Finally, the processed images served as the basis for extracting water quality parameters, which can be observed in section **3.4.2**. The main processes are described below.

3.3.2 Images acquisition and initial treatment

The downloading method used the Copernicus API and the (Sentinelsat, 2023) library, which made the process smoother than using the provider's servers, which could take longer due to the size of the files (around 1,7GB per image pack) and bandwidth availability. They were stored in hard drive disks for further processing and verification.

For most image matrix operations, Python (2023) programming language was employed, and, more specifically, the libraries Rasterio (2023) and Opencv-Python (2023) were used to open, crop, resize and visualize the images. The cropping process was simple, using a geometric variable called *Polygon*, that contained the bounding of each location of study where field data is available. The images were masked to that specific area, using the appropriated Coordinate Reference System (CRS) for each scene, by using *rasterio's function .mask(),* an example can be observed in **Figure 3-3** using the EB0 zone of study.

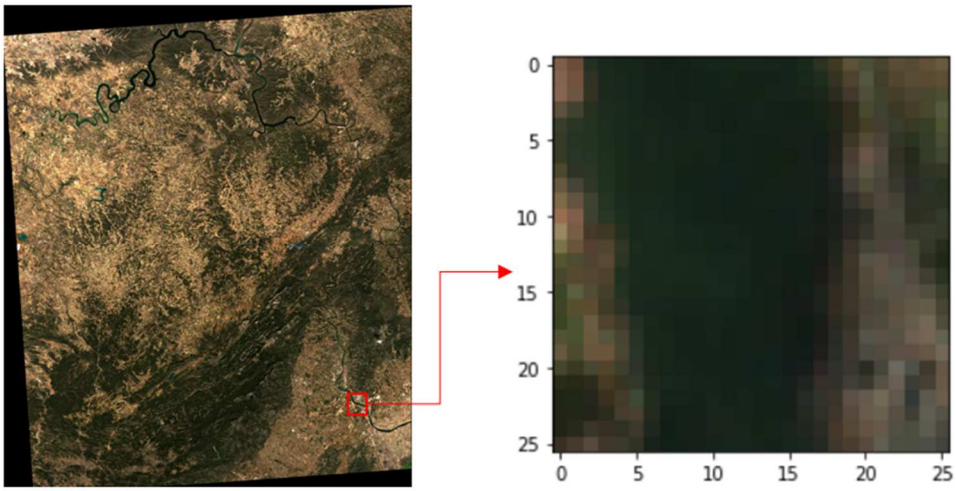


Figure 3-3. Result of the satellite image's cropping process

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

3.4 Application of water mask

The water masking was achieved through the Modified Normalized Difference Water Index (MNDWI). To calculate this equation, two different bands of the available satellite images were used, the green band (B03) and the SWIR band (B11) as can be observed in **Equation 3-1**.

Equation 3-1. MNDWI equation

$$\frac{Green\ (B03) - SWIR\ (B11)}{Green\ (B03) + SWIR\ (B11)}$$

With this equation, a new value per pixel in the image matrix is obtained. Values greater than 0 can be considered water.

Once obtained, the water mask in addition to the cloud mask, helped decide which pixels are considered water and thus suitable for the analysis and the prior water quality parameters estimations. An example of water masking can be observed in Figure 3-*4*.

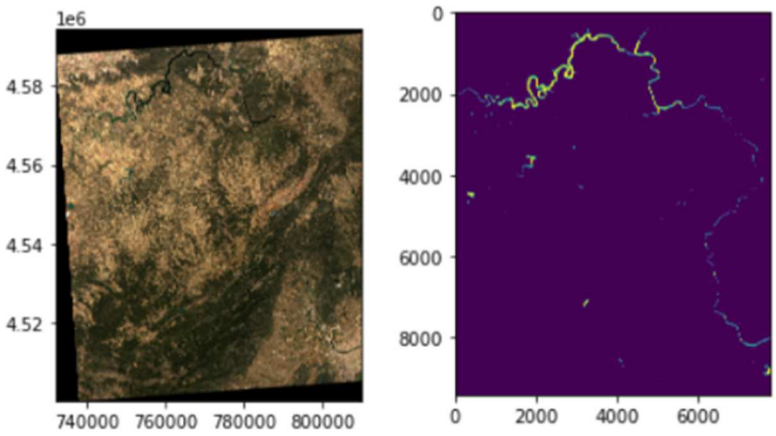


Figure 3-4. Water pixels identification using MNDWI

It is important to note that, as with other equations, the B11 band has originally a lower resolution (20m) when compared to B03 (10m). So, in those cases where bands with different resolutions are needed to be computed together, the lower resolution band (20m) was resized/up sampled to match the highest resolution possible using *opencv-python*'s function *called .resize(),* to its double factor (2x) to match the highest resolution image.

3.4.1 Application of cloud mask

The Scene Classification (SCL) image is the product of the process that Sentinel-2 Level-2A images passes through with the Sen2Cor processor algorithm. More specifically, the SCL image is a pre-processed matrix where each pixel is already labelled as a kind of element. This value ranges between 0 and 11, and can indicate vegetation, water, clouds, snow, and more. The classification values can be observed in Figure 3-*5*.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

Label	Classification
0	NO_DATA
1	SATURATED_OR_DEFECTIVE
2	DARK_AREA_PIXELS
3	CLOUD_SHADOWS
4	VEGETATION
5	NOT_VEGETATED
6	WATER
7	UNCLASSIFIED
8	CLOUD_MEDIUM_PROBABILITY
9	CLOUD_HIGH_PROBABILITY
10	THIN_CIRRUS
11	SNOW

Figure 3-5. SCL classification values

By using this method, the masking was simply a matter of not considering any pixel that would match the classification 9 “CLOUD_HIGH_PROBABILITY”, since those pixels would get in the way of the water making the analysis not possible or possibly wrong. This classification does a correct identification of clouds, as can be observed in Figure 3-*6*.

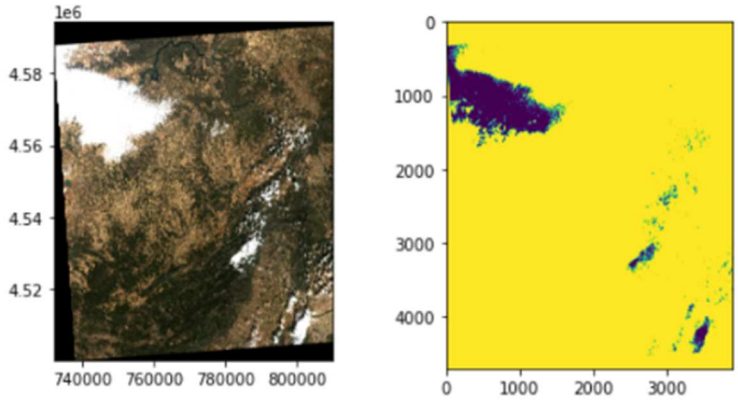


Figure 3-6. Original image (left); SCL Clouds classification image (right)

Additional methods of cloud identification and removal were available, but most of them relied on the band 10 of sentinel-2 image, which was not available in the Level 2-A packages, only on Level 1-C.

3.4.2 Deriving water quality parameters

After the completion of the cropping process, as well as the water and cloud masking, the image array that is specific to the designated study area is ready to be used as a basis for the calculation of the water quality parameters.

Algorithms were selected to estimate turbidity, CDOM, TSS, cyanobacteria and CHL-a. For turbidity, a common estimate derived from satellite data is the Normalized Difference Turbidity Index (NDTI). This index can take negative and positive values, where negative values indicate clear water and positive values indicate increasing turbidity. The potential of this index is that it is physically based, therefore, it doesn’t need calibration of empirical parameters with field data. For the estimation of CDOM, TSS

This project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No 101081980.

and CHL-a a set of functions recently published by (Sòria-Perpinyà et al., 2021) was selected (Table 3-2). The main reason for this choice is its representativeness of the conditions in the study area, as they derived the proposed functions using field data from Spanish Mediterranean reservoirs and lakes.

Table 3-2. Functions used to estimate water quality parameters from satellite data.

Parameter	Units	Function	Range values	Source
NDTI	No units	(R-G)/(R+G)	-1 - 1	Standard well-known indicator
CDOM	ug/L	2.4072 *(B4/B2)+ 0.0709 (Sentinel & Landsat))	0.03-5.30	Sòria-Perpinyà et al. (2021)
TSS	mg/L	14.464 *(B7/B2)+ 16.336 (Sentinel) 14.464 *(B6/B2)+ 16.336 (Landsat)	20-78.82	Sòria-Perpinyà et al. (2021)
CHL-a	mg/m^3	$19.866*(B5/B4)^{2.3051}$ (Sentinel 2) $19.866*(B8/B4)^{2.3051}$ (Landsat)	5.16-674.70	Sòria-Perpinyà et al. (2021)
Cianobacteria	mg/m^3	$21.554*(B5/B4)^{3.4791}$	0.13 - 1040	Sòria-Perpinyà et al. (2021)

These functions are implemented using matrix calculations (using floating point values of 32 bits for better precision), which enable the efficient application to the image array data. The resulting calculations are subsequently saved in CSV format, which facilitates the subsequent analysis and interpretation.

3.5 Methodology applied to the case-study

3.5.1 Satellite data

In total, 139 different Sentinel-2's images were selected as the case-study dataset for 2022. Those images were from two different scenes of the group S2A_ and S2B_ and with product type S2MSI2A. The querying method was through *Copernicu's* API and the respective coordinates of each scene can be observed in **Table 3-3-3**.

Table 3-3. Scene's coordinates

Scene	Type	Coordinates
1	MultiPolygon	((0.65555501440102, 40.50488583658978, 0.710574648674325, 41.49194363396281, -1.2757692614189051e-17, 41.51082844541308, -0.603302, 41.5268622914542, -0.6388855 40.53861750192735, 1.9339596288131015e-17 40.52196886366605, 0.65555501440102 40.50488583658978))
2	MultiPolygon	((-0.5408325, 40.50846282585221, 1.3101912763842559e-17, 40.52201747966694, 0.753813881598681, 40.54090999482793, 0.719948115495856, 41.52923550071836, -5.6985729713309160e-18, 41.51083276065857, -0.5941162, 41.49564643575739, -0.5408325 40.50846282585221))

The developed method was applied. The images were masked to that specific area, using the appropriated CRS for each scene, that in our case was 32630 and 32631, respectively. The approximated area after the zone clipping was around 26 x 26 pixels, with a 10m resolution. All those steps are detailed in section **3.3**.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

3.5.2 Combination with field data

Most of the field parameters measured at the different stations during the year 2022 had a temporal frequency of 15 minutes. To compare this data with the parameters obtained from the satellite data, which has a temporal coverage of 5 days, it was necessary to harmonize the time series. This involved calculating the daily mean of the different variables for each station, enabling comparability between the two data sources.

Once harmonized, the data was normalized and smoothed using a 15-day temporal window. The purpose of this process was to reduce the effects of sensor noise and random fluctuations in the data, while highlighting the underlying trends and patterns. Data normalization has been performed using the calculation of the z-score value, which is a commonly used procedure in data analysis. This involves adjusting the values of a dataset to be centred around zero and has a relative dispersion based on the standard deviation. To achieve this normalization, the mean of the dataset was first calculated, followed by the calculation of the standard deviation. Finally, each value was subtracted from the mean of the dataset and divided by the standard deviation. For the parameters derived from satellite data, the same normalization and smoothing were used, but prior to that, the pixel averages for each study area were calculated.

The resulting dataset consisted of a daily data set of field and satellite-derived parameters at each location. With these processed data, correlation analyses and trend evaluations combining the different parameters at each separate location were performed. The analysis allowed for the identification of possible relationships and patterns between the field-collected data and the data obtained from satellite images. It also allowed for evaluation of the validity of the selected functions.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

3.6 Results and discussion

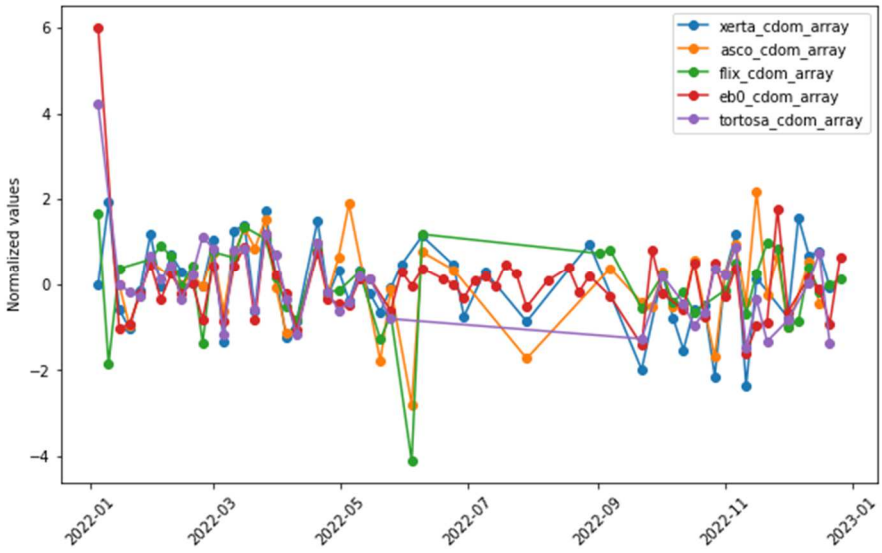
3.6.1 Water quality parameters obtained from satellite data

Fehler! Verweisquelle konnte nicht gefunden werden. displays the temporal variation of the normalized water quality parameters derived from satellite imagery, including CDOM, CHL-a, TSS, and cyanobacteria, for Flix, Ascó, Xerta, Tortosa, and EB0 monitoring stations. The graphs not only demonstrate a notable consistency among the satellite-derived water quality parameters, but also exhibit similar trends across stations.

In the CHL-a graph (**Fehler! Verweisquelle konnte nicht gefunden werden.**b), a peak in the concentrations is observed towards the end of March in the Xerta, Tortosa, and EB0 stations. This same peak is reflected in the cyanobacteria graph (**Fehler! Verweisquelle konnte nicht gefunden werden.**d) for the same dates and stations. This correlation is attributed to the presence of CHL-a, a photosynthetic pigment found in plants, algae, and cyanobacteria. During a cyanobacteria bloom, an increase in CHL-a concentration in the water is expected. This event also demonstrates the consistency among satellite-derived variables.

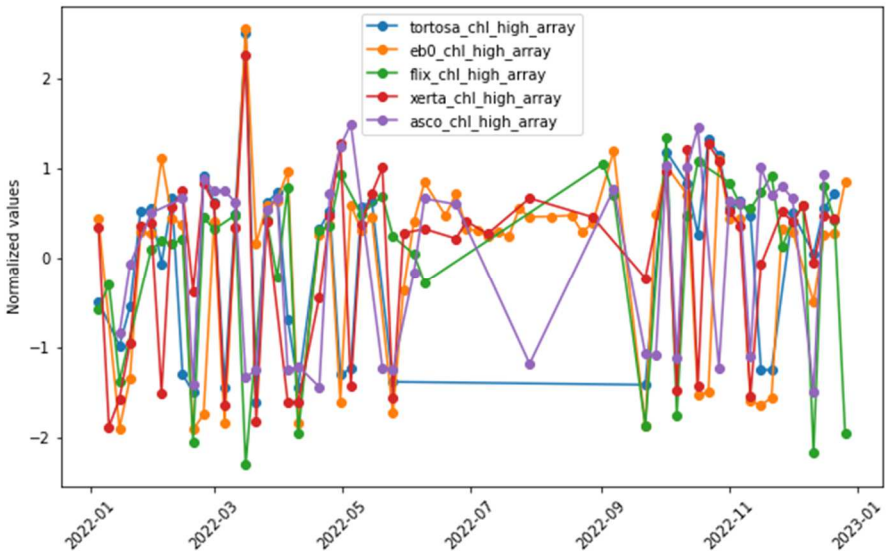
The observed consistency between the parameters derived from satellite imagery at the different locations is a first indication of the method's reliability.

a) CDOM

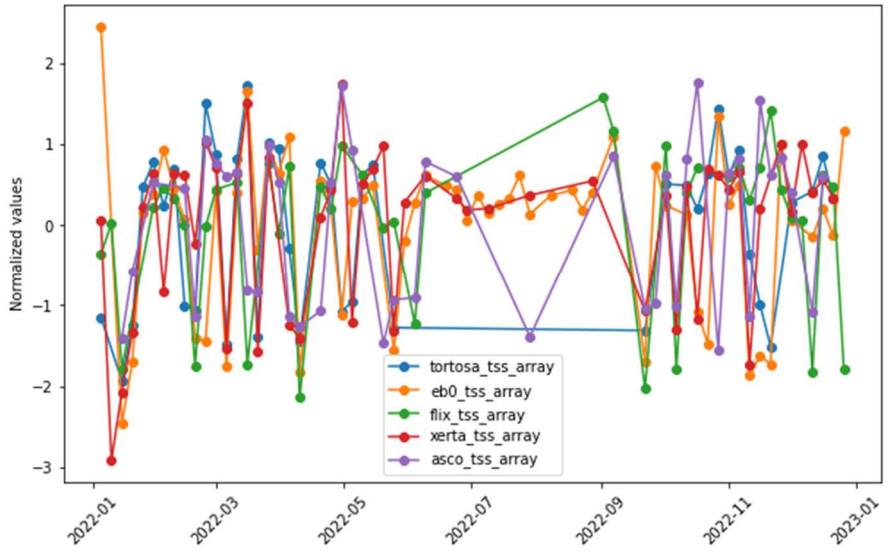


b) Chl-a (High)

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.



c) TSS



d) Cyanobacteria

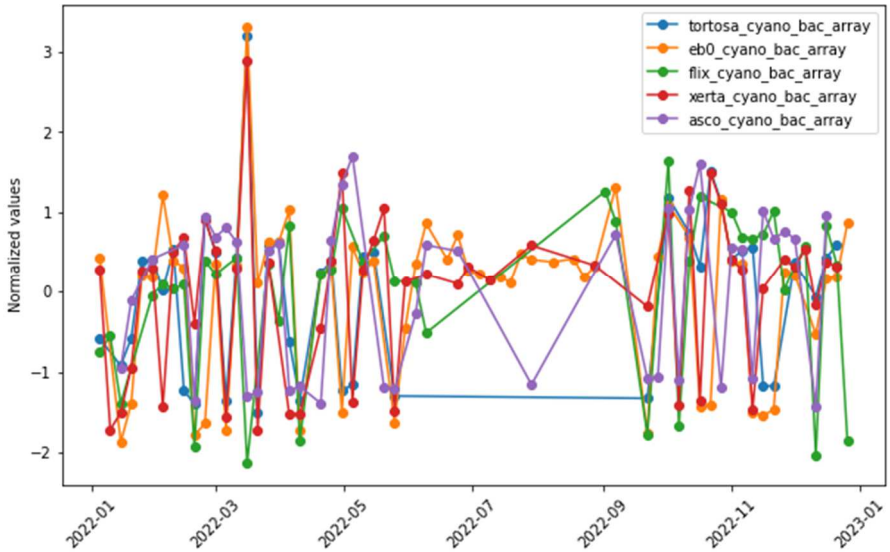


Figure 3-7. Temporal variation of normalized water quality parameters derived from satellite imagery

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

Table 3- 4. Range of values obtained from satellite data for the different parameters during 2022

		CDOM µg/l	CHL-HIGH mg/m^3	CHL-LOW mg/m^3	CYANO mg/m^3	TSS mg/l
ASCÓ	**MIN**	1,93	6,34	0,68	4,61	23,13
	MAX	2,72	23,89	0,91	28,76	33,56
	MEAN	2,37	15,10	0,86	15,64	28,01
XERTA	**MIN**	1,99	6,49	0,64	5,14	19,39
	MAX	2,62	27,47	0,91	36,48	33,61
	Media	2,34	16,00	0,87	16,81	28,29
FLIX	**MIN**	1,64	4,53	0,62	2,95	22,89
	MAX	2,68	22,72	0,91	28,04	32,71
	MEAN	2,38	16,01	0,87	17,13	28,54
TORTOSA	**MIN**	2,13	6,27	0,71	4,69	21,74
	MAX	3,11	28,79	0,91	39,32	32,98
	MEAN	2,39	15,08	0,86	15,73	27,68
EB0	**MIN**	2,07	7,22	0,68	4,98	21,67
	MAX	3,59	27,77	0,91	36,90	35,55
	MEAN	2,39	16,00	0,88	16,51	28,62

When comparing the values of the satellite-derived parameters with those provided by Sòria-Perpinyà et al. (2021) for field measured data in 2 lakes and 50 reservoirs in Spain (0.03 – 17.14 µg/l for CDOM, 0.53 – 704.97 mg/m3 for Chl_a, and 0.67 – 78.82 mg/l for TSS), it can be observed that data obtained for the Ebro river are in the same range. This indicates the functions have been correctly implemented with respect to the original publication, and results remain under realistic ranges.

3.6.2 Comparison of satellite-derived and field parameters

To verify the validity of the selected functions, the estimated satellite-derived water quality parameters were compared with available field measured parameters at each location. Since parameters measured in the field vary among monitoring stations, it was not possible to use exclusively the field data from the four parameters deduced from satellite data. Instead, it was necessary to use a broader approach, comparing the satellite-derived variables with all available parameters at each location. To do so, correlation matrixes were performed for each location, and trend comparisons with normalized data. The results revealed various relationships between field-collected data and data obtained from satellite imagery (**Fehler! Verweisquelle konnte nicht gefunden werden.**).

At Flix monitoring point, an inverse correlation can be observed between CHL-a, cyanobacteria, and measured DO (**Fehler! Verweisquelle konnte nicht gefunden werden.**a and b). This phenomenon can be attributed to the fact that, as the concentration of cyanobacteria increases, the levels of DO may decrease due to oxygen consumption during photosynthesis and subsequent decomposition of cyanobacteria.

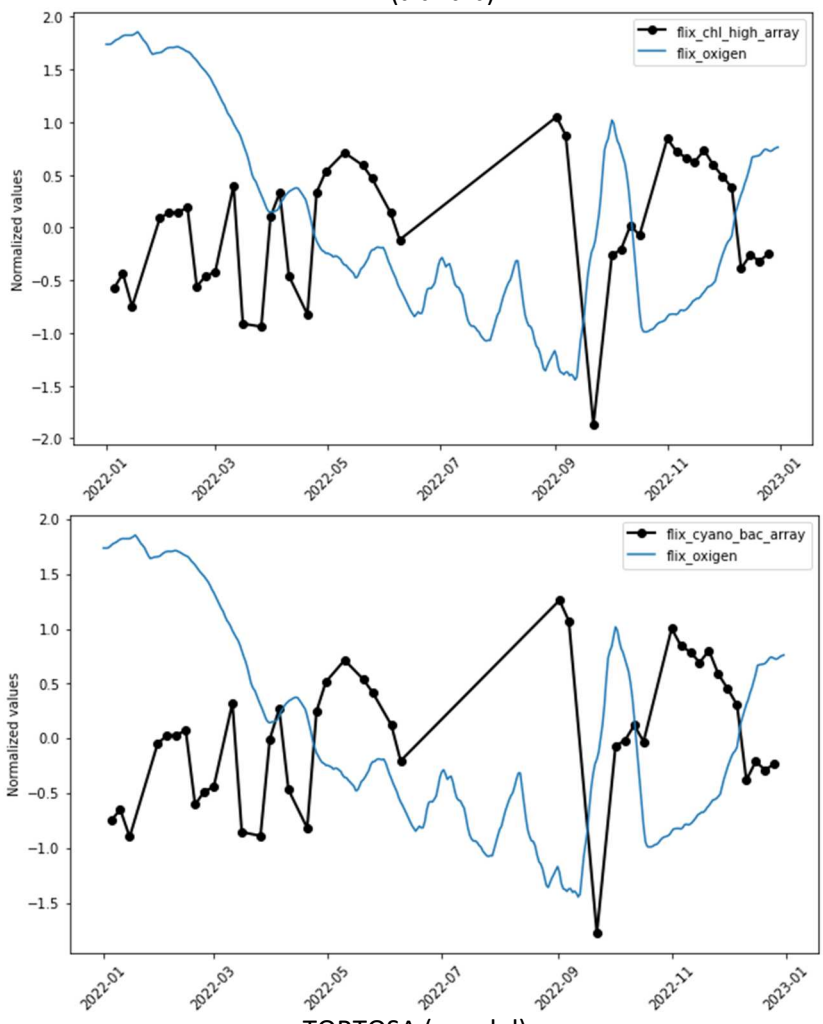
In Tortosa, a positive correlation was identified between turbidity and CDOM (**Fehler! Verweisquelle konnte nicht gefunden werden.**c). This correlation can be explained by the influence of organic matter which contributes to both parameters. Turbidity is a measure of suspended particles in water, such as sediments, organic matter, and microorganisms, while CDOM represents the dissolved fraction of organic matter that is capable of absorbing light in the visible spectrum. Therefore, an increase in CDOM concentration is expected to be associated with an increase in water turbidity. Additionally,

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

data was also compared with alternative variables like water flux. In the case of Tortosa, CDOM correlates with it (**Fehler! Verweisquelle konnte nicht gefunden werden.**d).

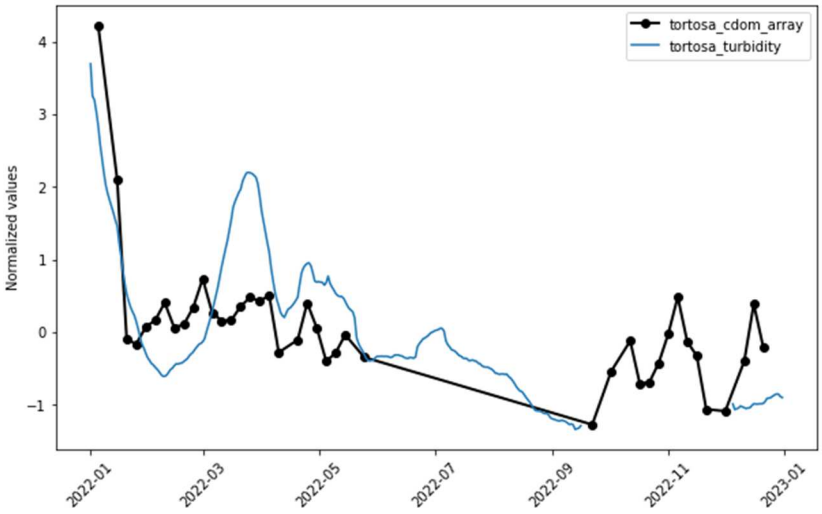
At the drinking water plant intake (EB0), two positive correlations were identified, between UV-254 and CDOM (**Fehler! Verweisquelle konnte nicht gefunden werden.**e), and between tryptophan and CDOM (**Fehler! Verweisquelle konnte nicht gefunden werden.**f). The correlation between UV-254 and CDOM is expected to be strong, as UV-254 is used as a proxy for CDOM concentration in the water (Huang et al., 2019). UV-254 quantifies the amount of ultraviolet light absorbed by the water at that specific wavelength. Given that CDOM is a significant source of light absorption in the UV range, UV-254 nm is commonly used as an indirect indicator of CDOM concentration. Higher CDOM concentration should correspond to higher UV-254. The correlation between fluorescence calibrated with the substance tryptophan and CDOM can be explained as tryptophan is part of the dissolved organic matter in water and contributes to the fluorescence signal of CDOM. As a specific component of CDOM, tryptophan can also be used to estimate CDOM in water using fluorescence techniques.

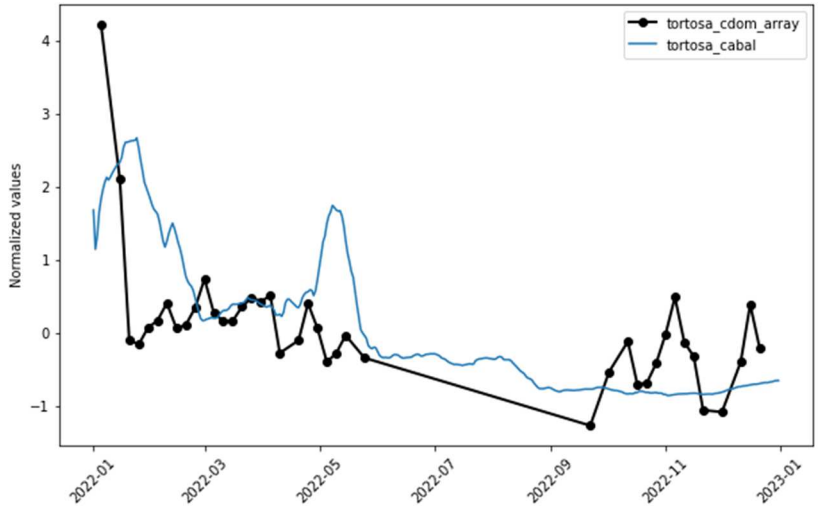
FLIX (a and b)



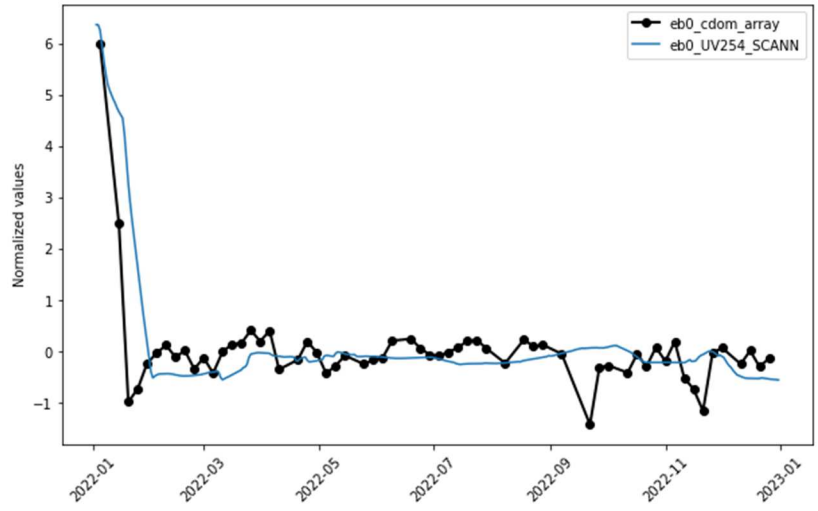
TORTOSA (c and d)

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.





EB0 (e and f)



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

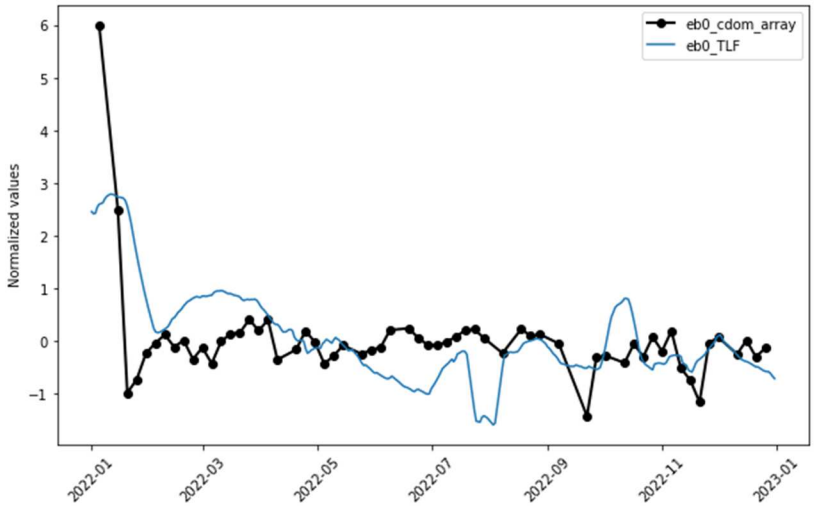


Figure 3-8. Main correlations over time between field-collected data and data obtained from satellite imagery

The correlations described indicate potential use of satellite images to derive surface water parameters relevant to estimate potential of DBPs production. However, those relations were found after normalizing the data, meaning, relations between absolute values were almost inexistent. Moreover, lack of abundant satellite derived data, due to the removal of images because of the interference of clouds during long periods, and high noise of the resulting time series, complicated the analysis enormously. Still, the incipient relations found are plausible, indicating potential for use of satellite data for the intended purpose, if original functions are calibrated with a longer period of field data. In the following section, the limitations of the overall approach are analyzed in further detail.

3.6.3 Checking potential of satellite data for early warning on organic matter concentration

One of the potential objectives to achieve with the satellite derived quality parameters was to provide information for an early warning source of information able to detect high concentrations of organic matter upstream with enough time for the drinking water treatment plant to adjust their treatment scheme to avoid too high DPBs formation. So, a first analysis of the field parameters was done to estimate the residence time between Mequinenza reservoir and the intake of the drinking water plant. To do so, conductivity was used as a conservative tracer to determine the transit time between Ascó and Xerta stations, as well as between Xerta and Tortosa. If residence time along this section of the river is lower than the time resolution of the satellite-derived data, it won't be possible to use satellite-derived information for this purpose.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

Figure 3-9. Conductivity profile over time for the stations of Ascó and Xerta (a) and Xerta and Tortosa (b) .

Figure 3-*9* shows the conductivity profile, expressed in µS/cm, over time. In a, Ascó is represented in blue, and Xerta in red. A shift of two days was applied to the Xerta time series to match peaks with the Asco data. In Figure 3-6b, Xerta data is shown in red, while Tortosa is represented in green, with a one-day shift. Considering an approximate distance of 45 km between Ascó and Xerta, and a lag between peaks of two days, one can have an indication to estimate the transit time between Mequinenza reservoir and the drinking water plant intake at Xerta, which is 3.3 days considering that the distance between these points is 73 km.

The combined temporal resolution of Sentinel's sensors A and B is 5 days. Therefore, satellite images cannot be employed as an early warning system for detecting upstream organic matter sudden changes, that would require adjustments in treatment at the CAT DWTP.

3.7 Limitations

3.7.1 Limitations of used satellite data

During the implementation of the selected functions for the estimation of satellite-derived parameters, and later in the analysis of the results, we found several limitations of the method related to the use of satellite data for the prediction of source water quality in the Ebro River.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

The first is related to the representativeness of the satellite-derived data with respect to the total mass of water where samples are taken. Since sunlight is only reflected from the top water layer (first 50cm), no information is provided of the water quality below this depth. This has some importance in the case of cyanobacteria, which can move and therefore change concentrations in depth.

An additional factor adding errors to the satellite data is the atmospheric composition influence (e.g. water vapour, aerosol...). There are many ways to correct this influence. In this case, the standard correction used for Sentinel data (Sen2Cor) was used. However, this correction is not optimal for water surfaces that reflect only very little sunlight. This may lead to errors that propagate through the methodology. Alternative atmospheric corrections, more adapted to water surfaces are already available, which could be tested to improve results.

Satellite data is acquired at specific times (e.g. 11 am), which are constant for one satellite and region. This could be important for parameters with high daily oscillations, especially if daily averages are being compared like in the approach presented.

Finally, temporal coverage of satellite data is limited due to the influence of clouds. This is a circumstance from which one can escape but can lead to big gaps of information.

3.7.2 Limitations of field data availability

Field data availability has proven to be the main obstacle in this study. The lack of historic time series of the parameters that can be derived by satellite data is a major obstacle, as it didn't allow direct comparison. Additionally, the field parameters were not the same for all monitoring stations, making more difficult the selection of variables to analyze. Moreover, time series of field parameters had long periods without data, either because the data was pending of validation or because the sensor was damaged. In summary, field data availability reduced the number of satellite points we could compare with field data, reducing the significance of the correlations between time series.

3.7.3 Limitations of the method

For the validation of the proposed functions, it is necessary to compare field parameters with satellite-derived parameters. Moreover, in a further step, functions need to be improved by calibration with real data. Given the lack of temporal coverage in the field data and the limit number of usable satellite images, it became clear from this first approach that the period use of the analysis needs to be extended to several years. This can only be done with historic data.

For the presented approach, functions use to derive parameters from satellite data were not calibrated. Although results remain within realistic ranges, absolute values were not comparable with field data for the selected locations. It is well-known that this type of functions are extremely case specific and dataset specific. Therefore, it would be necessary to do so for our cases and dataset. The main difficulty we will face is the lack of the field data on CDOM, cyanobacteria or CHL-a. Therefore, focus will be put on searching for new functions for variables related to organic matter content with field data (eg. UV-245 and on-line fluorescence).

Even though results are very preliminary, and calibration is needed, we realized that the satellite data showed high oscillations. This could be because of the noise caused by the satellite itself or by the implemented corrections. In any case, this resulted in low precision, which in practice means that satellite data will only be able to capture high variations in the derived parameters. The use of longer time series could also help to add to the analysis period with clear peaks in concentrations, easy to identify in satellite-derived data.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

With respect to the usefulness of satellite derived for the specific case of the CS#3, two limitations were found. The first is the short residence time from the most important upstream reservoir to the DWTP intake. This residence time is smaller than the revision period of the satellite. Therefore, the data cannot be used as an early warning tool, as it was initially proposed. The second limitation for the use of the satellite data to predict DBPs formation is the specific treatments included in the DWTP. The treatment is very intensive in removing organic matter (GAC filters and ozonation). The resulting water that is chlorinated before entering the distribution system has little organic matter content and very stable in terms of total organic carbon. Therefore, even if organic matter could be estimated in the input water based on satellite data, it could be extremely complicated to relate it to the formation of DBPs in this case. Even though, the satellite data could still provide useful information to understand the DBPs formation due to seasonal variations of the DOM or the effect of extreme events of algal blooms. This is to be explored in this project in further tasks.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

4 Conclusions

GC-MS and RPLC-MS are commonly applied to characterize DBPs. However, these methods are not suitable for polar DBPs. Within SafeCREW project, UFZ has established an alternative method based on SFC-HRMS to comprehensively characterize DBPs in drinking water, including the highly polar sulfonated DBPs. The SFC-HRMS parameters were further optimized for the separation and sensitivity of sulfonated DBPs. Finally, the BEH column, 1.3 mL/min flow rate, and mobile phase gradient of a total 19 min gave the best performance. Three different sample enrichment processes including freeze-drying, SPE with WAX cartridges, and azeotrope evaporation, were compared for the recovery of sulfonated DBPs. Similar recovery rates were found when enriching the analytical standards of Cl_2MSA and Br_2MSA using these methods. Nevertheless, priority will be given to the freeze-drying and SPE with WAX cartridges as they provide a 10 times higher enrichment factor than the azeotropic evaporation. The established method will be used in the following work packages, when screening the drinking water from treatment plants and distribution networks, for the non-target analysis of unknown DBPs and target analysis of sulfonated DBPs. This method will be also applied in lab-scale disinfection studies. The measurement of sulfonated DBPs was also conducted on RPLC-MS. The RT of sulfonated DBPs were as expected very low and some compounds even overlayed in the dead volume of the chromatogram. Still, seven sulfonated DBPs that appeared to have better separation on the chromatogram. Thus, this method can be used by other partners in the future to characterize sulfonated DBPs in their lab, avoiding unnecessary sample shipment to the UFZ. Current work is focusing on the development of quantification methods for sulfonated DBPs using SFC coupled to triple quadrupole MS.

DBPs formation mainly depends on the concentration of organic matter in the water being chlorinated. Therefore, prediction of organic matter in surface water upstream of the DWTP inlet, may be useful to adapt the plant operation to remove it and to predict the DBPs formation. The proposed methodology relied on satellite data for the derivation of water quality parameters that could be related to organic matter content. To do so, a selection of functions published in the literature and applicable to Sentinel data were applied as a first attempt. However, results showed limited application for the specific case study of Tarragona, and limited representativeness of the field parameters.

The main conclusion is that the methodology needs to be complemented with a calibration step of the selected functions, including larger time series of historic data and new data (obtained through the project) of the relevant variables that can be related to organic matter content. Thus, the proposed methodology consists of several steps:

1. Retrieval of satellite data
2. Cloud and water masking
3. Cropping of selected location from which field data is available.
4. Spatial averaging of each scene
5. Calibration of water quality parameter functions with field data for each location
6. Implementation of the calibrated functions in a subset on used for the calibration.

The proposed methodology will be applied in WP 4 as part of Task 4.2 using a longer array of historical data and data obtained during the monitoring task, Task 3.2. Several methods for the functions calibration will be tested, starting from the most simple multivariate statistics, to potentially more advanced methods like deep learning algorithms, Support Vector machines and other non-parametric algorithms.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

Even if the use of such satellite-derived data is limited for the actual project case-study, there are other sites where the characteristics of the water catchment or the DWTP could make them more suitable for this new application.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

5 Bibliography

Chen, Y., Hozalski, R.M., Olmanson, L.G., Page, B.P., Finlay, J.C., Brezonik, P.L. and Arnold, W.A. 2020. Prediction of Photochemically Produced Reactive Intermediates in Surface Waters via Satellite Remote Sensing. Environmental Science & Technology 54(11), 6671-6681.

Chowdhury, S., Champagne, P. and McLellan, P.J. 2009. Models for predicting disinfection byproduct (DBP) formation in drinking waters: A chronological review. Science of The Total Environment 407(14), 4189-4206.

Costet, N., Villanueva, C.M., Jaakkola, J.J.K., Kogevinas, M., Cantor, K.P., King, W.D., Lynch, C.F., Nieuwenhuijsen, M.J. and Cordier, S. 2011. Water disinfection by-products and bladder cancer: is there a European specificity? A pooled and meta-analysis of European case–control studies. Occupational and Environmental Medicine 68(5), 379-385.

Furman, C.S. and Margerum, D.W. 1998. Mechanism of Chlorine Dioxide and Chlorate Ion Formation from the Reaction of Hypobromous Acid and Chlorite Ion. Inorganic Chemistry 37(17), 4321-4327.

Gholizadeh, M.H., Melesse, A.M. and Reddi, L. 2016 A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques.

Hartnett, H.E. (2018) Dissolved Organic Matter (DOM). In Encyclopedia of Geochemistry: A Comprehensive Reference Source on the Chemistry of the Earth. White, W.M. (ed), pp. 375-378, Springer International Publishing, Cham.

Hua, G. and Reckhow, D.A. 2007. Comparison of disinfection byproduct formation from chlorine and alternative disinfectants. Water Research 41(8), 1667-1678.

Huang, J., Wu, M., Cui, T. and Yang, F. 2019 Quantifying DOC and Its Controlling Factors in Major Arctic Rivers during Ice-Free Conditions using Sentinel-2 Data.

Koll-Egyed, T., Cardille, J.A. and Deutsch, E. 2021 Multiple Images Improve Lake CDOM Estimation: Building Better Landsat 8 Empirical Algorithms across Southern Canada.

Li, S., Zhang, J., Mu, G., Ju, H., Wang, R., Li, D. and Shabbir, A.H. 2016 Spatiotemporal Characterization of Chromophoric Dissolved Organic Matter (CDOM) and CDOM-DOC Relationships for Highly Polluted Rivers.

Nihemaiti, M., Icker, M., Seiwert, B. and Reemtsma, T. 2023. Revisiting Disinfection Byproducts with Supercritical Fluid Chromatography-High Resolution-Mass Spectrometry: Identification of Novel Halogenated Sulfonic Acids in Disinfected Drinking Water. Environmental Science & Technology 57(9), 3527-3537.

Olmanson, L.G., Page, B.P., Finlay, J.C., Brezonik, P.L., Bauer, M.E., Griffin, C.G. and Hozalski, R.M. 2020. Regional measurements and spatial/temporal analysis of CDOM in 10,000+ optically variable Minnesota lakes using Landsat 8 imagery. Science of The Total Environment 724, 138141.

Opencv-Python 2023. Obtained from PyPi : https://pypi.org/project/opencv-python/.

Python 2023. About. Obtained from Python: https://www.python.org/about/.

Rasterio 2023. Docs. Obtained from Rasterio: https://rasterio.readthedocs.io/en/stable/index.html.

Reemtsma, T., Berger, U., Arp, H.P.H., Gallard, H., Knepper, T.P., Neumann, M., Quintana, J.B. and Voogt, P.d. 2016. Mind the Gap: Persistent and Mobile Organic Compounds—Water Contaminants That Slip Through. Environmental Science & Technology 50(19), 10308-10315.

Richardson, S.D. and Kimura, S.Y. 2020. Water Analysis: Emerging Contaminants and Current Issues. Analytical Chemistry 92(1), 473-505.

Richardson, S.D., Plewa, M.J., Wagner, E.D., Schoeny, R. and DeMarini, D.M. 2007. Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: A review and roadmap for research. Mutation Research/Reviews in Mutation Research 636(1), 178-242.

Ruescas, A.B., Hieronymi, M., Mateo-Garcia, G., Koponen, S., Kallio, K. and Camps-Valls, G. 2018 Machine Learning Regression Approaches for Colored Dissolved Organic Matter (CDOM) Retrieval with S2-MSI and S3-OLCI Simulated Data.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.

Sadiq, R. and Rodriguez, M.J. 2004. Disinfection by-products (DBPs) in drinking water and predictive models for their occurrence: a review. Science of The Total Environment 321(1), 21-46.

Schulze, S., Paschke, H., Meier, T., Muschket, M., Reemtsma, T. and Berger, U. 2020. A rapid method for quantification of persistent and mobile organic substances in water using supercritical fluid chromatography coupled to high-resolution mass spectrometry. Analytical and Bioanalytical Chemistry 412(20), 4941-4952.

Schulze, S., Zahn, D., Montes, R., Rodil, R., Quintana, J.B., Knepper, T.P., Reemtsma, T. and Berger, U. 2019. Occurrence of emerging persistent and mobile organic contaminants in European water samples. Water Research 153, 80-90.

Seiwert, B., Nihemaiti, M., Bauer, C., Muschket, M., Sauter, D., Gnirss, R. and Reemtsma, T. 2021. Ozonation products from trace organic chemicals in municipal wastewater and from metformin: peering through the keyhole with supercritical fluid chromatography-mass spectrometry. Water Research 196, 117024.

Sentinelsat 2023. About. Obtained from Sentinelsat: https://sentinelsat.readthedocs.io/en/stable/.

Si-Hung, L. and Bamba, T. 2022. Current state and future perspectives of supercritical fluid chromatography. TrAC Trends in Analytical Chemistry 149, 116550.

Sòria-Perpinyà, X., Vicente, E., Urrego, P., Pereira-Sandoval, M., Tenjo, C., Ruíz-Verdú, A., Delegido, J., Soria, J.M., Peña, R. and Moreno, J. 2021 Validation of Water Quality Monitoring Algorithms for Sentinel-2 and Sentinel-3 in Mediterranean Inland Waters with In Situ Reflectance Data.

Taniyasu, S., Kannan, K., Yeung, L.W.Y., Kwok, K.Y., Lam, P.K.S. and Yamashita, N. 2008. Analysis of trifluoroacetic acid and other short-chain perfluorinated acids (C2–C4) in precipitation by liquid chromatography–tandem mass spectrometry: Comparison to patterns of long-chain perfluorinated acids (C5–C18). Analytica Chimica Acta 619(2), 221-230.

Wagner, E.D. and Plewa, M.J. 2017. CHO cell cytotoxicity and genotoxicity analyses of disinfection by-products: An updated review. Journal of Environmental Sciences 58, 64-76.

West, C. 2018. Current trends in supercritical fluid chromatography. Analytical and Bioanalytical Chemistry 410(25), 6441-6457.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081980.