Disinfection By-Product Formation Potential Algorithm

Deliverable 2.2

WP2 SENSORS – Innovative Monitoring Systems



Author: Christoph Wagner

Date: 28.05.2024

Funded by the European Union

GRANT AGREEMENT NUMBER	101081728		
FULL TITLE / ACRONYM	intoDBP		
START DATE	1.12.2022	**DURATION**	48 months
END DATE	30.12.2026		
PROJECT URL	https://intodbp.eu/		
DELIVERABLE	D2.2		
WORK PACKAGE	WP2		
CONTRACTUAL DATE OF DELIVERY	31.05.2024		
ACTUAL DATE OF DELIVERY	29.05.2024		
NATURE	REPORT	**DISSEMINATION LEVEL**	SEN
LEAD BENEFICIARY	SCAN		
RESPONSIBLE AUTHOR	Christoph Wagner		
CONTRIBUTIONS FROM	Dieter Steiner, Maria José Farre		
ABSTRACT	This report outlines the creation of algorithms predicting the potential for disinfection by-product (DBP) formation. Water samples from various case study locations underwent shock chlorination in a laboratory setting. We then measured DBP formation through both gas chromatography-mass spectrometry (GC-MS) and ultraviolet-visible (UV/VIS) spectrophotometry. The collected data informed the development of these predictive algorithms.		

Disclaimer

Any dissemination of results reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains.

Copyright message

© intoDBP Consortium, 2024
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation, or both. Reproduction is authorised provided the source is acknowledged.

TABLE OF CONTENTS

EXECUTIVE SUMMARY 4
LIST OF TABLES 5
LIST OF FIGURES 5
ABBREVIATIONS 6
1.0 EXPERIMENTAL DESIGN 6
2.0 EXPERIMENTAL SETUP 7
2.1 Mechanical design 7
2.2 Experimental procedure 8
2.3 Calculating the Disinfection By-Product formation potential 10
2.4 Technical challenges 11
2.5 Sample selection for algorithm development 12
2.6 Collected data 13
3.0 ALGORITHM DEVELOPMENT 13
3.1 Per water source 14
3.1.1 Ter Plant (Barcelona) 15
3.1.2 Irish Surface Water 17
3.1.3 Abrera Water 19
3.2 One combined model from all data points 20
3.3 Model based only on unchlorinated water samples 23
3.4 THM prediction models 23
3.5 HAN prediction models 25
3.6 Delta spectrum of unchlorinated minus final spectrum 27
3.7 Difference between high and low bromide concentrations 28
4.0 SUMMARY 29

EXECUTIVE SUMMARY

This research report investigates the capability of UV-Visible (UV/VIS) sensors to predict the formation potential of disinfection by-products (DBPs), specifically trihalomethanes (THMs) and haloacetonitriles (HANs), in various water samples collected from Ireland and two locations around Barcelona. The study's primary objective was to evaluate the efficacy of chemometric algorithms in predicting the formation potential of these DBPs based on analytical data.

Water samples were disinfected, and aliquots were systematically collected at different contact times and subjected to laboratory analyses employing Gas Chromatography-Mass Spectrometry (GC-MS) while UV/Visible spectra of the sample was automatically recorded with a UV-VIS sensor. GC-MS was utilized to identify and quantify the specific DBPs present in the samples, while UV/VIS spectroscopy provided essential data on the organic matter content, which is a precursor for DBP formation during water disinfection processes.

The integration of GC-MS and UV/VIS spectral data enabled the development of chemometric models. These models were trained and validated to predict the formation potential of THMs and HANs in the water samples. The chemometric approach, leveraging multivariate statistical techniques, demonstrated a high predictive capability, offering a reliable method for assessing DBP formation risk based on UV/VIS spectra.

The study's findings highlight significant variations in DBP formation potential among the different water sources, influenced by the distinct organic compositions of the waters from Ireland and Barcelona. Especially the presence of bromide in the water had a significant influence on the formation potential.

Overall, this research provides a valuable framework for water quality monitoring and management, promoting the use of advanced analytical and predictive tools to enhance public health protection against harmful disinfection by-products.

LIST OF TABLES

Table 1: Summary of sample locations and characteristics of waters 12
Table 2: RMSE (in µg/l) values for various chemometric models calculated from the full scale experiments outlined in chapter 3.1 & 3.2. 29
Table 3: RMSE (in µg/l) values for various chemometric models calculated for TTHMs from pre-chlorine spiking experiments. 30
Table 4:RMSE (in µg/l) values for various chemometric models calculated for THANs from pre-chlorine spiking experiments. 30

LIST OF FIGURES

Figure 1: 3D drawing of the reactor lid for DBP formation measurements including the spectrometer probe on the left. Photo of the experimental setup at the ICRA lab on the right. ...8
Figure 2: Evolution of temperature data at the Ter treatment plant. First three experimental runs from September 2023. 9
Figure 3: Spectral changes during a typical shock chlorination experiment at different points in time. 9
Figure 4: Evolution of the raw absorption at 290nm (picked randomly in the UV range) in blue and measured DBP concentrations from GC-MS measurements in orange. Dashed colored lines correspond to spectra shown in Figure 3. 10
Figure 5: Measured total THM concentrations in blue vs. the derived formation potential in orange. 11
Figure 6: Time series graph of the absorption at 290nm. Areas marked in red indicate measurements affected by air bubbles around the optical windows of the spectrometer. 12
Figure 7: Barcelona Ter plant data (based on raw water, clarified water and post-AC water). Left side: test set data. Right:all data points 15
Figure 8: Model developed for Ter plant applied to all measured spectra (dots) and overlayed with lab measurements (crosses) 16
Figure 9: Model developed on two different Irish raw water sources. Left side: test set data. Right:all data points 17
Figure 10: Model developed for Irish surface water applied to all measured spectra (dots) and overlayed with lab measurements (crosses). 18
Figure 11: Model built on water from the Abrera treatment plant including raw water, clarified water, EDR water and after activated charcoal. 19
Figure 12: Model developed for Abrera water applied to all measured spectra (dots) and overlayed with lab measurements (crosses). 20
Figure 13: PLS model built from all different water samples (left) vs. a Ridge Regression model based on the same PLS wavelengths. 21
Figure 14: Model built on a combination of Ter treatment plant water and Irish surface water. .22
Figure 15: Model based on Ter and Abrera data showing separation of data. 22
Figure 16: Example of the decrease of HAN concentration during one sample run. 23
Figure 17: All pre-chlorination spectra from all samples were used to build these models (left: with one latent variable, right: with two latent variables). 24
Figure 18: Model built on derivative spectra rather than absorption spectra from all samples. .24

Figure 19: Models from pre-chlorinated water samples combining Ter and Irish data with one latent variable (left) and two (right). 25
Figure 20: PLS models of HAN based on absorption spectra with one latent variable (left) and two (right). 26
Figure 21: PLS model for HAN based on derivate spectra with one latent variable. 26
Figure 22: HAN models built for Ter (top left), Abrera (top right) and Ter + Ireland (bottom) samples. 27
Figure 23: Models built on delta spectra between the first and last spectrum of an experimental run. 28
Figure 24: Applying source specific models to other water samples. Left: using the Abrera model (with high bromide concentrations) on Ter data and the other way around on the right side. 29

ABBREVIATIONS

Abbreviation	Description
BCAN	Bromochloroacetronitrile
BDCM	Bromodichloromethane
DBAN	Dibromoacetonitrile
DBCM	Dibromochloromethane
DBP	Disinfection By-Product
DCAN	Dichloroacetonitrile
FP	Formation Potential
HAN	Haloacetonitrile
PLS	Partial Least Squares
TBM	Tribromomethane
TCAN	Trichloroacetonitrile
TCM	Trichloromethane (chloroform)
THAN	Total Haloacetonitriles
THM	Trihalomethane
TTHM	Total Trihalomethanes
UV/VIS	Ultra-Violet / Visible

1.0 EXPERIMENTAL DESIGN

To be able to build prediction algorithms based on UV/VIS spectroscopy it is necessary to generate pairs of UV/VIS spectra together with reference values of the target value under evaluation. In our case this target value is the concentration of disinfection byproducts (DPBs). From the expected concentration range of DBPs and the capabilities of online UV/VIS spectrometers in terms of repeatability of the spectrum under real-life conditions (e.g. changing water temperature, changing flow rate, varying turbidity and other background variations) and difference in the spectra themselves from one DBP to another it was deemed realistic to only predict sums of DPB formation potential. The goal of this work was to predict the formation potential of water sample for groups of DBPs

such as Trihalomethanes (THMs) or Haloacetonitriles (HANs) as sum parameters and not for individual DPBs such as Bromodichloromethane (BDCM).

However, since the laboratory analysis capabilities to distinguish a list of DBPs are readily available at ICRA individual DBPs were measured and then sums of DBPs were calculated for the prediction algorithm development.

In this report we have defined formation potential of DBPs the concentration of DBPs that will be formed after 72h of contact time with chlorine, reaching a residual around 0.5 and 1 ppm after that time.

2.0 EXPERIMENTAL SETUP

2.1 Mechanical design

The main goal of the mechanical design was to have as little as possible plastic in contact with water to avoid the generation of DBPs from other sources than the water under investigation. Secondly, the total volume of water should be as low as possible as water samples from Irland had to be shipped to Spain for analysis.

Therefore, the measurement chamber consists of a glass cylinder specifically built for these experiments. The spectrometer probe for the analysis is inserted through a lid build from Teflon, which should be inert for DBP formation. A three-dimensional drawing of this lid can be found in Figure 1. The pink coloured O-rings have not been used in the final setup as the Teflon lid sealed very nicely directly with the glass cylinder. The small orange connector in the drawing was meant as a sample port to draw samples from the reactor for analysis by GC-MS.

A photo of the experimental design can also be seen in Figure 1. The spectrometer is held in place by a laboratory stand and is submersed in the water without a headspace above the aqueous phase. This is particularly important as DBPs can be volatile and would be lost for reference analysis by GC-MS.

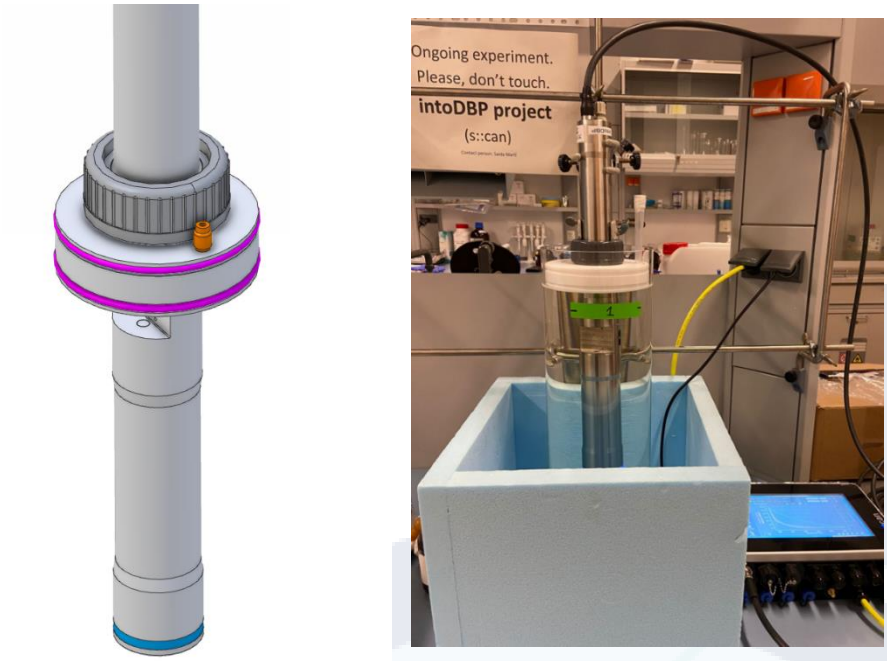


Figure 1: 3D drawing of the reactor lid for DBP formation measurements including the spectrometer probe on the left. Photo of the experimental setup at the ICRA lab on the right.

2.2 Experimental procedure

Two spectrometer probes (s::can, spectro::lyser V3 with serial numbers 23010228 and 23010236) where used to measure the UV/VIS absorption spectra during the reaction of sodium hypochlorite with the water samples.

Therefore, the water sample was filled into the measurement reactor and an initial spectrum of the water was measured. Then the sample was shock chlorinated with a solution of sodium hypochlorite and the lid was closed. The concentration of the hypochlorite solution was selected to match real conditions and therefore obtain between 0.5 and 1 ppm after the 72 hours that lasted the experiment. No buffer was used in these experiments to avoid precipitation of salts that would interfere in the UV-Vis measurement. Temperature was maintained at laboratory conditions around 21°C. From there on spectra were recorded continuously and automatically stored on the industrial PC connected to the spectrometer.

Initially the idea was to run two experiments side by side from the same water where the second experiment is started with a time delay of 16 hours from the first one so that sampling for GC-MS measurements must not been undertaken during night. The first reactor would be start at 5pm the first days whereas the second experiment starts at 9am the next morning and both running experiments are stopped at the same time. During post processing of the data the two experiments would be merged to form one full data set with a complete time series. Figure 2 shows time series data for such “double experiments” measured in September 2023.

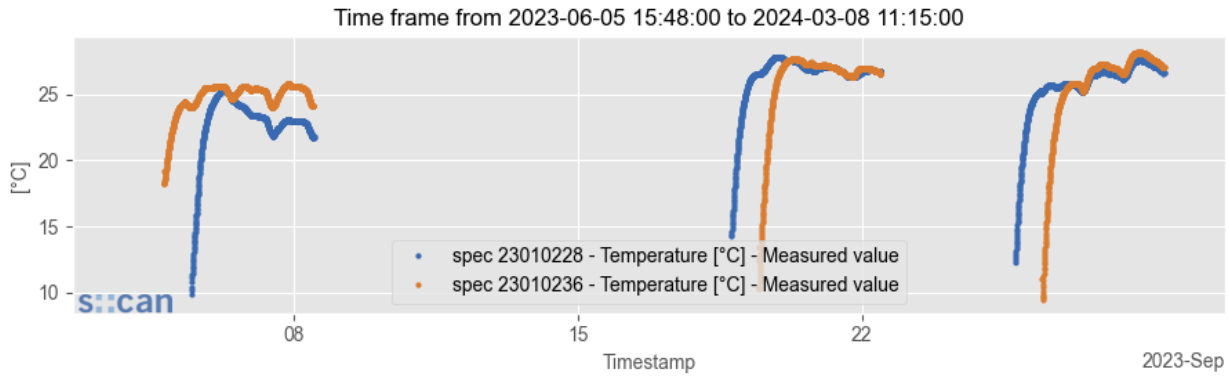


Figure 2: Evolution of temperature data at the Ter treatment plant. First three experimental runs from September 2023.

This procedure was used only once to analyse the three initial samples from the Ter treatment plant. After this experiments data acquisition was changed in such a way that only at the beginning of the experiments frequent samples have been taken and during the later hours of the runs only a handful lab samples have been analysed. This was done to safe cost on the GC-MS measurements and seemed appropriate as only at the beginning big changes in the formation of DBPs were observed and kinetics slowed down significantly after the initial period of the experiments. Additionally, to saving costs the lab work was also easier as two different samples could be analysed at the same time.

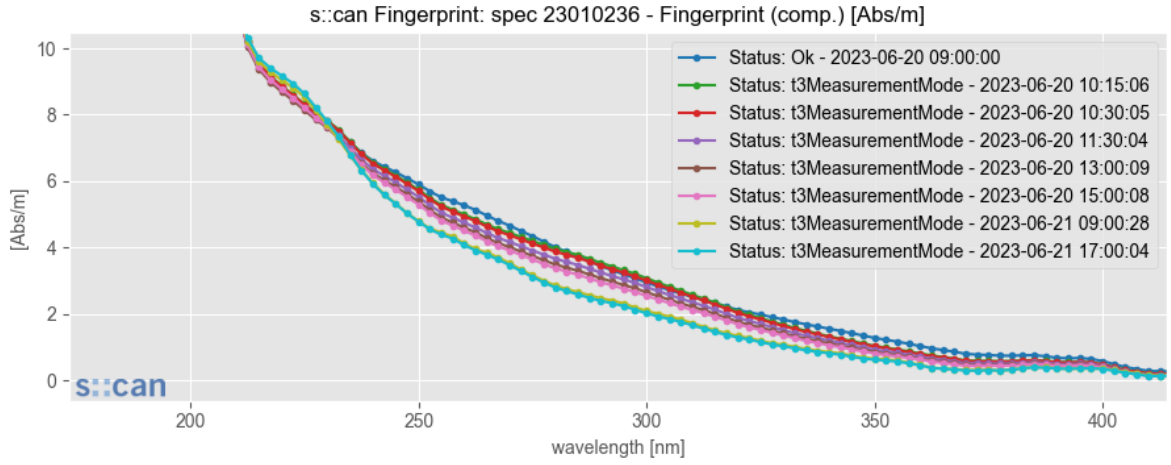


Figure 3: Spectral changes during a typical shock chlorination experiment at different points in time.

This procedure resulted in a data set for each experiment containing of the spectral data, exemplarily shown in Figure 3 together with an Excel spreadsheet containing the measured DBP concentrations that were summed for analysis. These values could then be linked to corresponding spectra as can be seen from Figure 4. The combination of spectra with reference values from the lab were further used to calculate the DBP formation potential and model it from the spectral data.

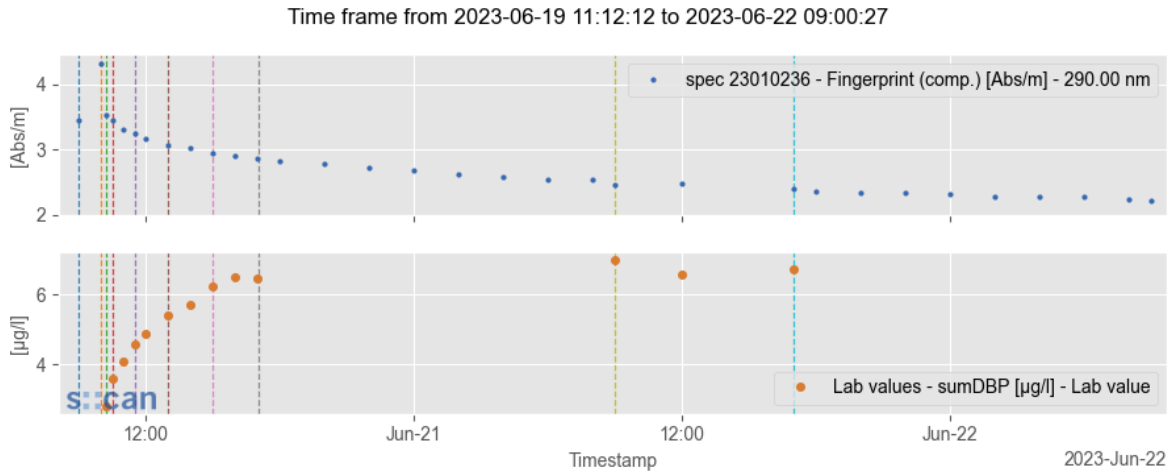


Figure 4: Evolution of the raw absorption at 290nm (picked randomly in the UV range) in blue and measured DBP concentrations from GC-MS measurements in orange. Dashed coloured lines correspond to spectra shown in Figure 3.

2.3 Calculating the Disinfection By-Product formation potential

The formation potential (FP) of a measurement run performed as outline in the previous chapter was defined as the final concentration of DBPs measured by GC-MS at the last sampled drawn (after 72 hours of contact time). This was taken as the maximum possible concentration of DBPs that could be generated. Therefore, the last measured spectrum had a formation potential of 0mg/l by default whereas the first spectrum prior to chlorination had this maximum measured concentration (of the last sample) as the formation potential. For samples with according to lab measurements of DBPs taken in between the formation potential was calculated according the formula: FP(t) = max(DBP) - DBP(t).

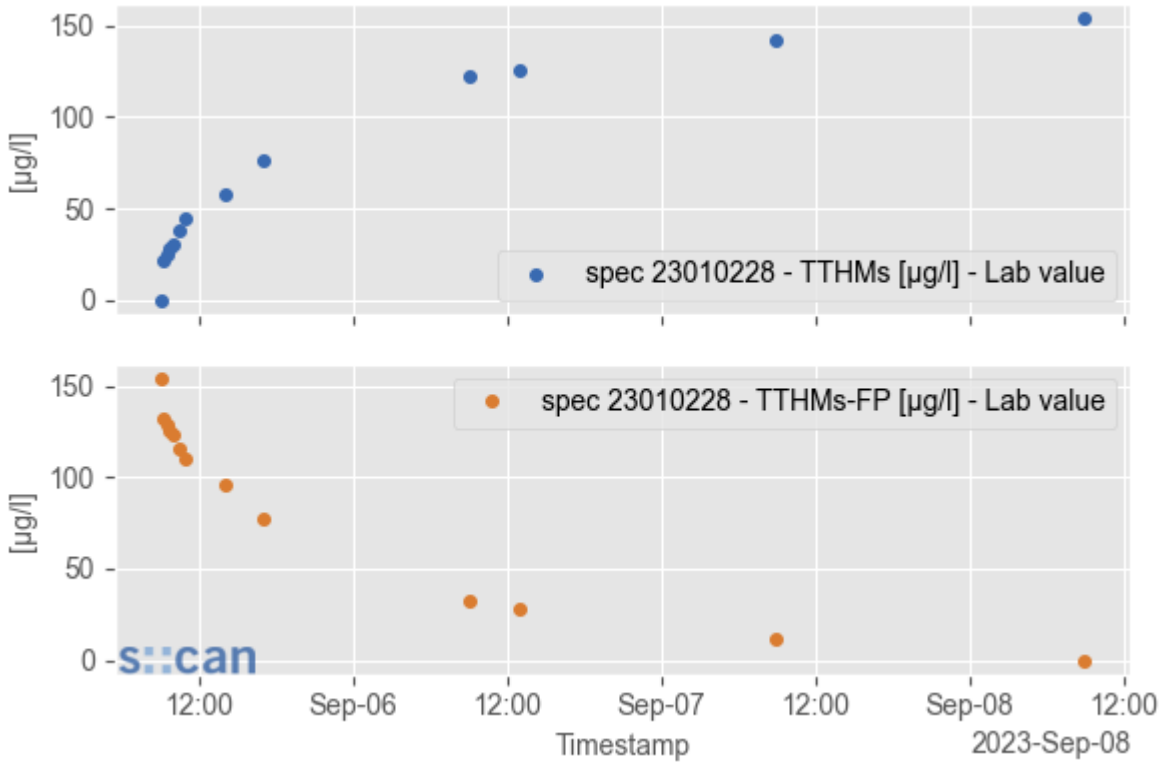


Figure 5: Measured total THM concentrations in blue vs. the derived formation potential in orange.

All further discussed models that predict the formation potential decay during the experiments were done with these FP values. All of these models predict only the total THM concentrations. Predicting the levels of THANs using this method wasn't considered viable, since these DBPs undergo hydrolysis and therefore the concentration of the compound undergoes an initial increase followed by a decay and therefore the concentration at the end of the experiment is lower than in the middle of the experiment. An exception to this observation was in the case of bromine rich samples, as brominated HANs are more stable than chlorinated HANs. Nevertheless, we decided to focus our efforts in predicting TTHMs.

2.4 Technical challenges

The main challenge regarding the measurements have been air bubbles formed after the start of the measurement. The bubbles tended to form around the optical windows of the spectrometer probe interfering with the path of light and therefore potentially interfered with the measurement. See a time series graph of one experiment that shows such time frames of interference in Figure 6.

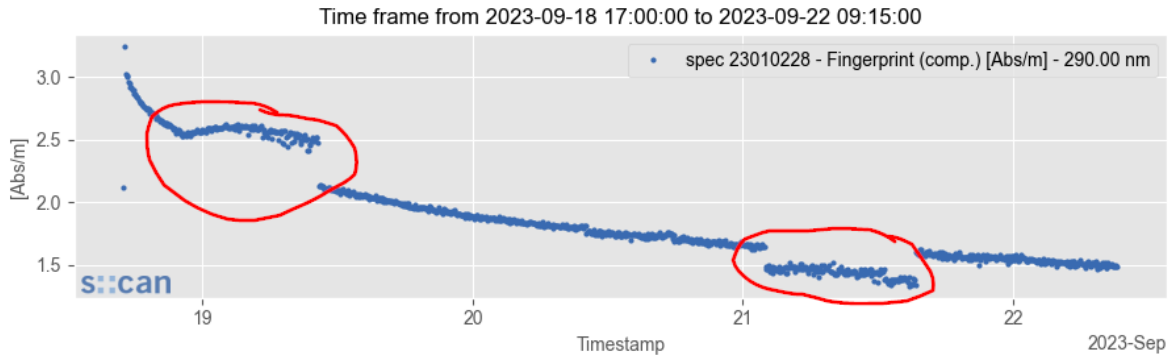


Figure 6: Time series graph of the absorption at 290nm. Areas marked in red indicate measurements affected by air bubbles around the optical windows of the spectrometer.

Several measures have been tested to reduce the amount of bubble formation. These included warming up the samples to room temperature before filling them into the measurement reactor. However, since the spectrometer is submersed in the water sample it is inevitable that the water will heat up during the time interval of the experiment. This effect can be clearly seen in Figure 3 and Figure 4. Secondly, samples were purged with argon to replace any dissolved air. The taken measures helped to reduce the formation but were not capable of completely avoiding the formation of bubbles. As long as the bubbles were small enough the turbidity compensation algorithm built into the software of the spectrometer was able to eliminate the influence on the recorded spectra. Larger bubbles distorted the spectra too much and had to be removed manually during a data pretreatment step before using the data for the development of a DBP formation potential algorithm.

2.5 Sample selection for algorithm development

Table 1: Summary of sample locations and characteristics of waters.

Sample Location	Description
Ter treatment plant of ATL	Raw surface water
Ter treatment plant of ATL	Post clarified water
Ter treatment plant of ATL	Post activated charcoal water
Abrera treatment plant Llobregat river	Raw surface water with high bromine content
Abrera treatment plant Llobregat river	Post clarified water
Abrera treatment plant Llobregat river	Post electrodialysis reversal water
Abrera treatment plant Llobregat river	Post activated charcoal water

Irish surface water from the Carrowmore Lough and Nephin Valley	High concentrations of humic acids

For a description of each of the case study site please refer to *deliverable 2.1* in chapter 2. The Abrera treatment plant is not part of the case study sites but samples from the Llobregat river were considered important for the algorithm development as it contains high concentrations of bromide which potentially changes the characteristics of the DBPs formed.

Samples from Ireland were different from the Spanish water samples in terms of the organic compounds dissolved which contain mainly humic acids which are bigger and heavier than the organic matter observed in Spanish rivers. Also, pH of Nephin Valley was lower than the rest of the samples (around 5).

2.6 Collected data

The used GC-MS method by ICRA is described in deliverable 2.1 in chapter 5.6 for the measurement of THMs and HANs. Samples were withdrawn from the reactor and transferred to a 60 millilitres glass vial containing ascorbic acid to stop the reaction. Samples were stored in the fridge and liquid-liquid extracted immediately after the experiment finished (72h). Extracts were injected in the GC-MS within 1 and 2 weeks.

In total 17 experiments were carried out.

3.0 ALGORITHM DEVELOPMENT

For the development of prediction algorithms, the measured UV/VIS absorbance spectra were used together with the measured DBP concentrations determined by the GC method described in deliverable 2.1. Due to the complexity of DBP formations predicting specific DBPs such a BDCM was not considered a realistic goal. Therefore, concentration values derived from the GC measurements were summed up and the sum of THM (TCM, BDCM. DBCM. TBM) or HAN (DCAN, TCAN, BCAN, DBAN) values were then used in the development of the prediction algorithms.

Additionally, to running the described experiments with water samples, blank measurements with milliQ water were carried out to determine background effects and to measure the pure contribution of dosing sodium hypochlorite / hypochlorous acid to the UV/VIS spectra.

Algorithms were developed for the case study sites in Barcelona as well as Ireland. First, one algorithm was developed per site and then different models were “cross-tested” to see if a model developed for location A would also work for location B etc. The developed models are summarized in the following sections.

All data points used for model development were split into two data sets. 1/3 of data was used as the test set and not used during developing the model. On the other hand, 2/3 of the data was used as training data set for calculating the model.

Root Mean Square Errors (RMSE) shown on the following correlation graphs all relate to the show data points (e.g. test set only, or all data points combined).

3.1 Per water source

We started with analysing data from each location separately but by combining water from different treatment steps (e.g. raw water, clarified water and post AC water) into one model. Different PLS models were built with varying wavelengths, different latent variables and with or without mean centering.

For all locations the best fitting models were based on wavelengths 272,5 and 342,5 nm with two latent variables (or components) and after applying mean centering to the spectra.

3.1.1 Ter Plant (Barcelona)

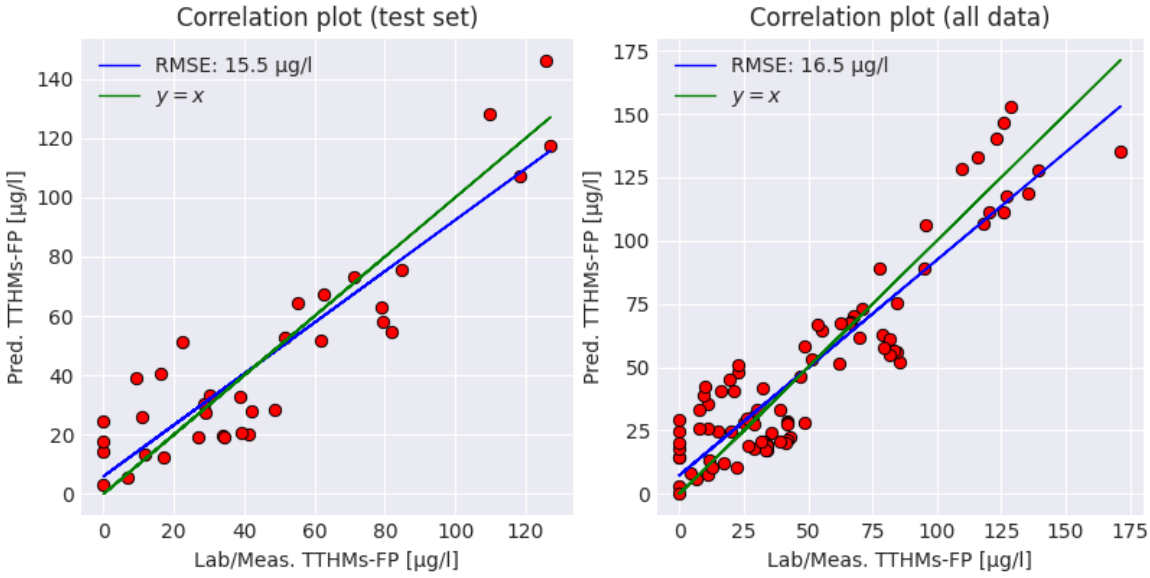


Figure 7: Barcelona Ter plant data (based on raw water, clarified water and post-AC water). Left side: test set data. Right:all data points.

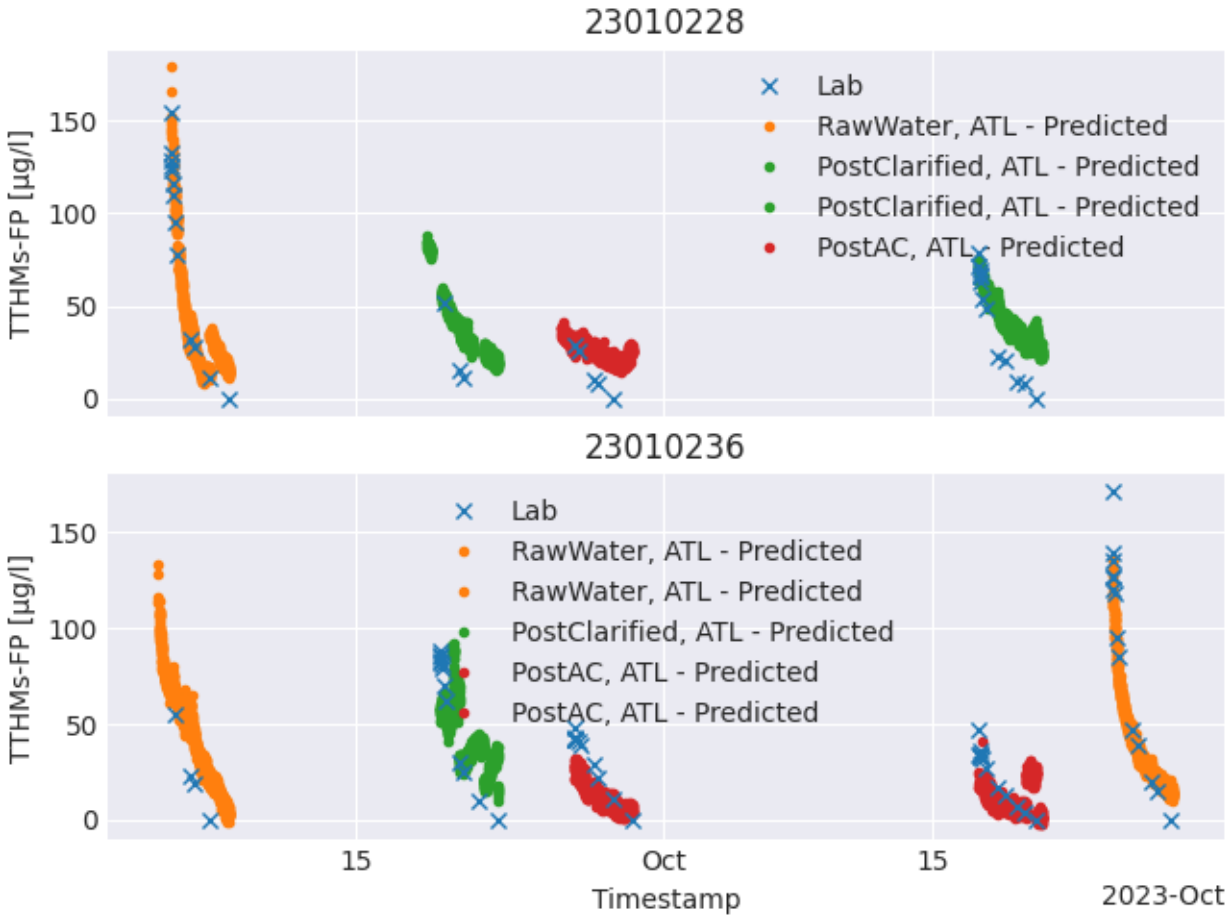


Figure 8: Model developed for Ter plant applied to all measured spectra (dots) and overlayed with lab measurements (crosses).

3.1.2 Irish Surface Water

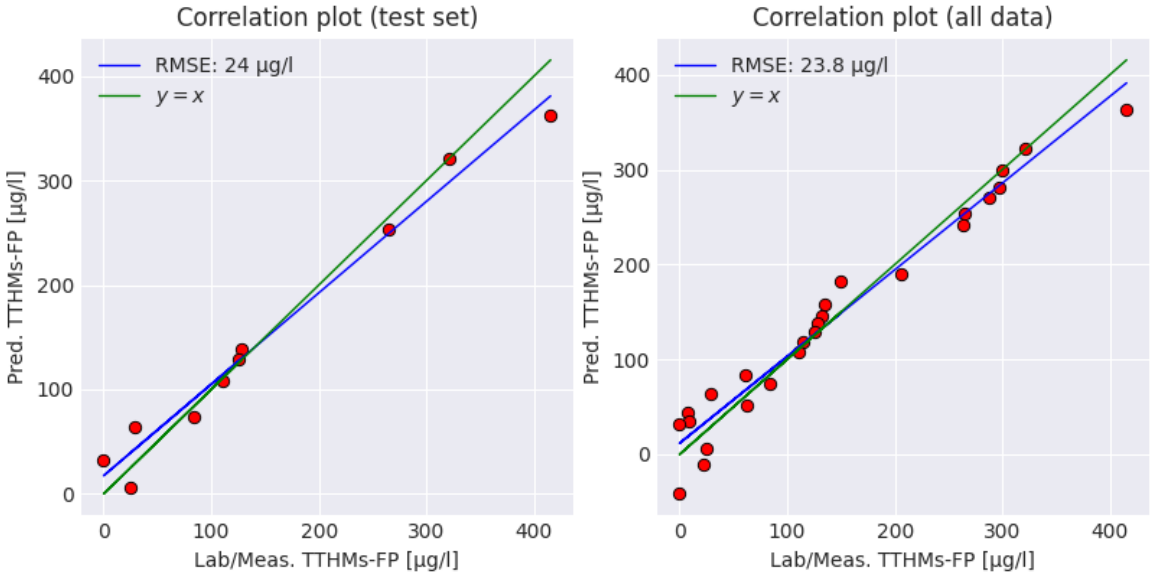


Figure 9: Model developed on two different Irish raw water sources. Left side: test set data. Right:all data points.

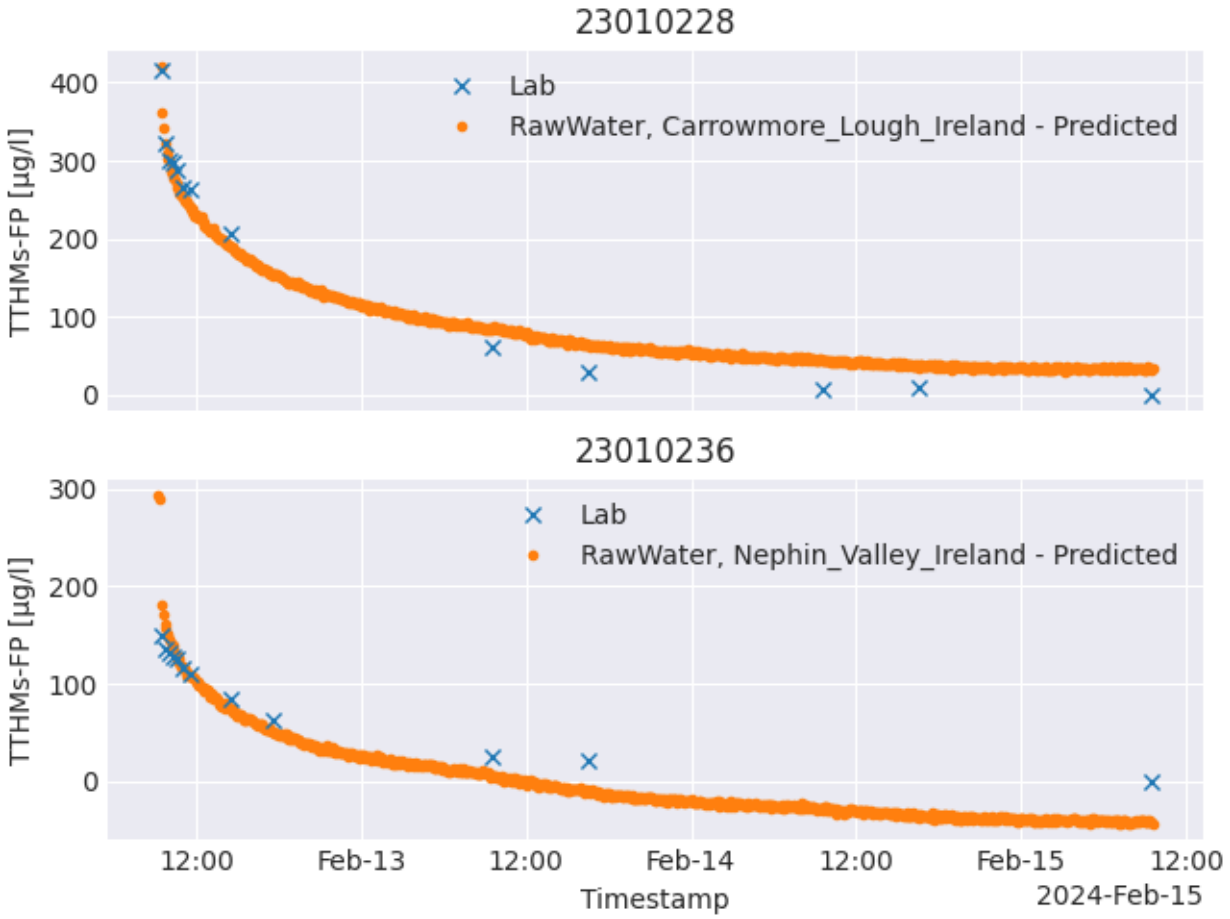


Figure 10: Model developed for Irish surface water applied to all measured spectra (dots) and overlayed with lab measurements (crosses).

3.1.3 Abrera Water

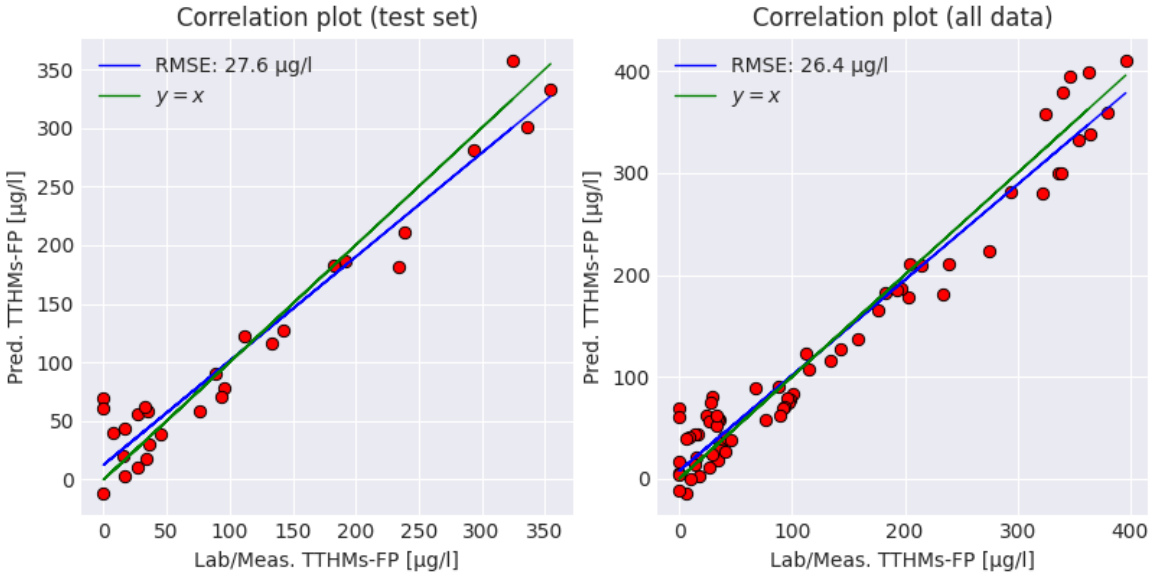


Figure 11: Model built on water from the Abrera treatment plant including raw water, clarified water, EDR water and after activated charcoal.

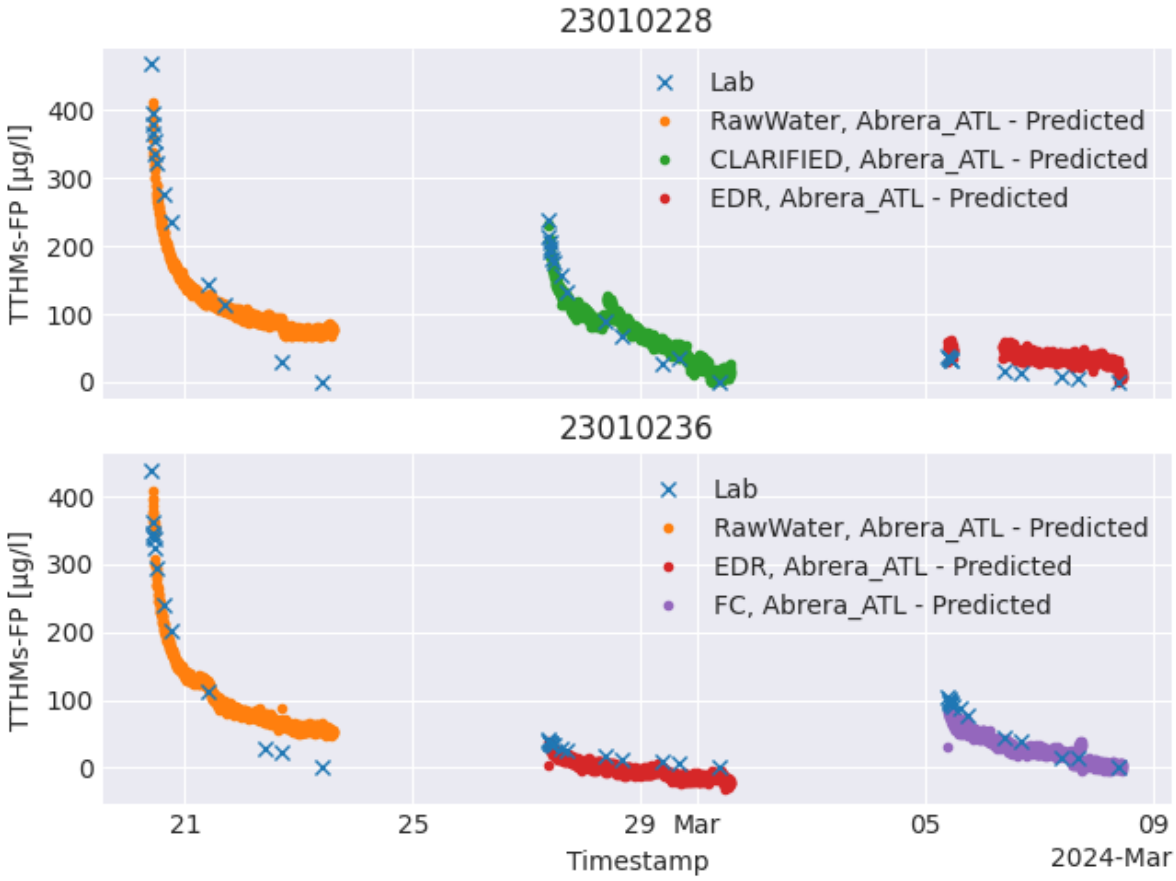


Figure 12: Model developed for Abrera water applied to all measured spectra (dots) and overlayed with lab measurements (crosses).

3.2 One combined model from all data points

Next step was to combine all data points from all different sources into one global model which would be the most beneficial from a commercial point of view. This would give a one-fits-all approach for potential customers.

The best potential model that could be achieved using PLS used 7 wavelengths {220, 240, 255, 272.5, 285, 310, 342.5} nm and 2 latent variables (Figure 13 left side) and Ridge-Regression model with the same wavelengths (right side).

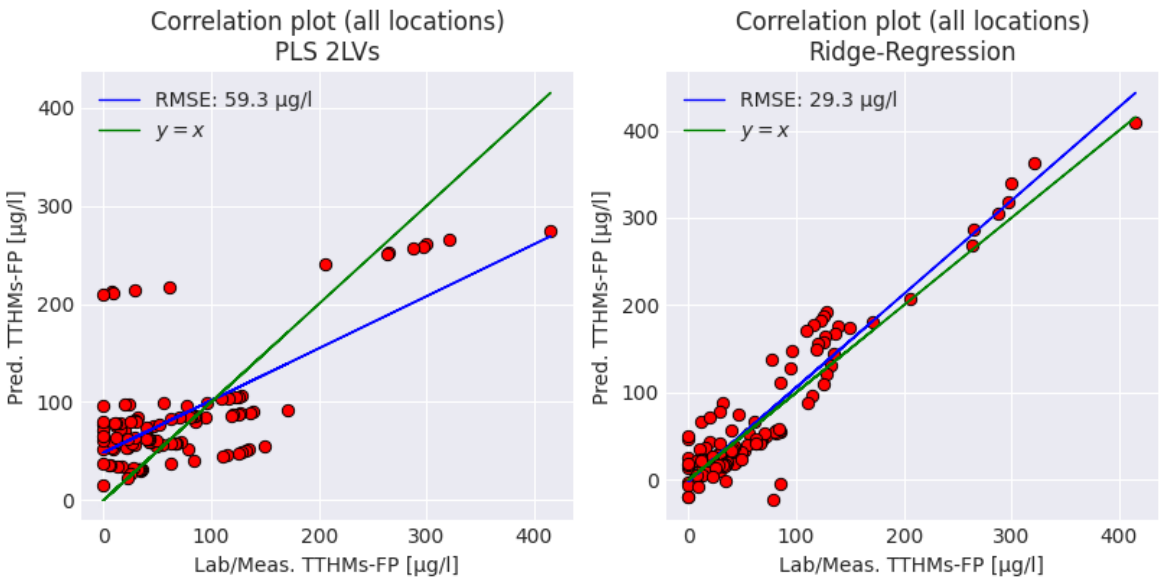


Figure 13: PLS model built from all different water samples (left) vs. a Ridge Regression model based on the same PLS wavelengths.

What can be clearly seen from the PLS model above is that one single model is not working. However, the Ridge model seems to fit much better. However, the ridge-regression model has a lot more degrees of freedom compared to a PLS model and can therefore easily lead to an overfitted model. To verify if overfitting occurred a completely independent set of data would be needed in this case.

Since the Abrera samples are different from the rest of the data sets by its high concentration of bromide another model was calculated based only on the samples from the Ter plant and Ireland. The wavelength selection algorithm selected three wavelengths with 2 latent variables: {272.5, 285, 342.5}

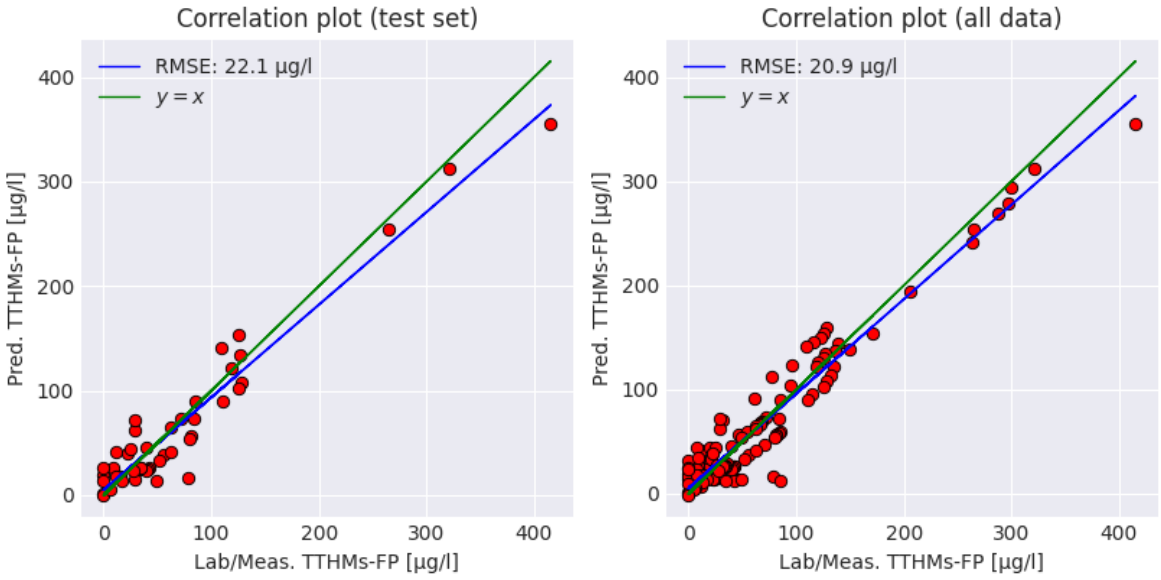


Figure 14: Model built on a combination of Ter treatment plant water and Irish surface water.

The combination of Spanish samples (Ter and Abrera data) was used as well to build a specific model based on the the same wavelengths {272.5, 285, 342.5} and two latent variables. It is plotted in Figure 15 and clearly shows that data is separated in two groups and therefore the data can't be combined successfully.

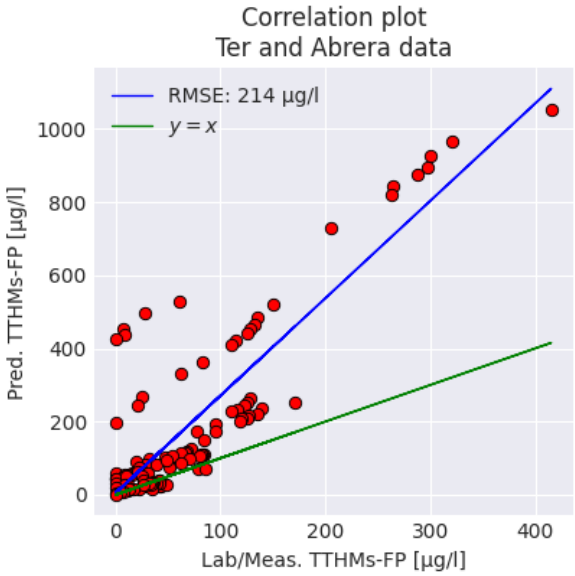


Figure 15: Model based on Ter and Abrera data showing separation of data.

Based on the above it is very likely that the bromide in the Abrera water samples sets these samples apart from the other samples and therefore can't be added to a combined model. The difference in organic matter between the Ter water samples and the Irish surface water samples (high concentration in humic acids) has no impact to the model and can be combined.

3.3 Model based only on unchlorinated water samples

All models presented above are based on spectra acquired during a full experimental run as described in chapter 2.2. However, such experiments are complex and expensive to run at customer locations and could be prohibitive to applying UV/VIS spectroscopy for DPB prediction. Therefore, an additional approach was tested: Only the spectrum of the unchlorinated water sample was used in combination with the maximum THM / HAN concentration measured during the chlorination experiments. HAN concentrations typically decrease during one sample run (see Figure 16) due to their hydrolysis constants . For the following models the maximum concentration of HANs was used and not the final concentration at the end of the measurement run.

Below are lab results from Ter raw water assays from September. The amount of the volatile HANs formed decreases at some point.

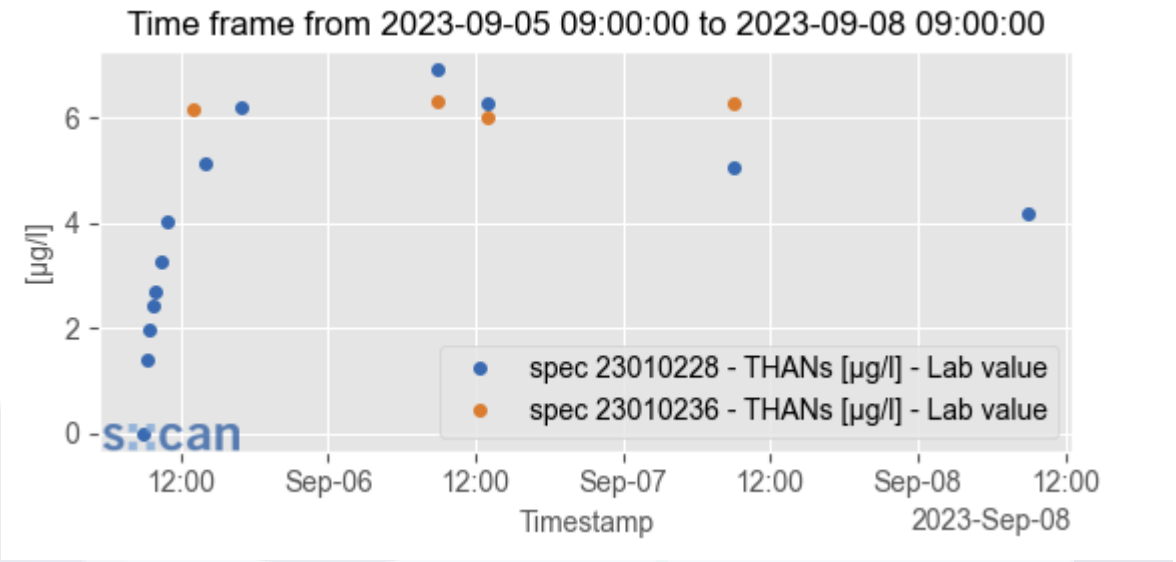


Figure 16: Example of the decrease of HAN concentration during one sample run.

3.4 THM prediction models

The first model that was built on all pre-chlorine spiking spectra from all different available water samples using the final THM concentrations of each sample run. Three wavelengths of the raw spectrum have been used: {240, 250, 280} nm. One latent variable alone was not enough to capture the complexity of the dataset. Therefore, again two latent variables had to be used (see Figure 17 for the comparison).

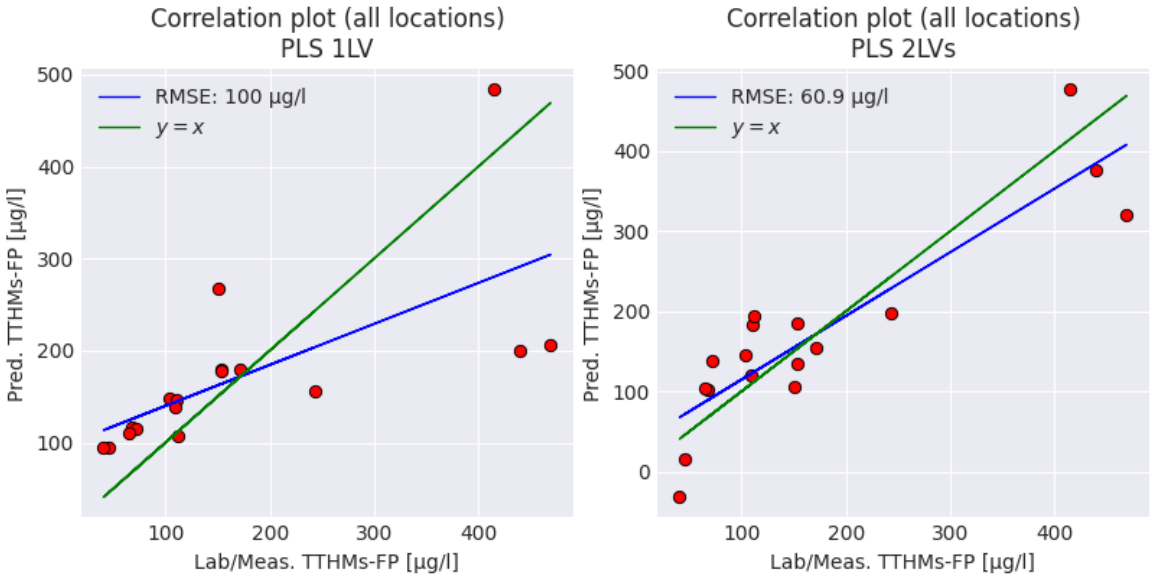


Figure 17: All pre-chlorination spectra from all samples were used to build these models (left: with one latent variable, right: with two latent variables).

Best results, however, have been achieved by using the first derivative spectra and not the absorption spectra directly using the same wavelengths as above. The result is shown in Figure 18. By only using one latent variable the RMSE could be even reduced to the 2 latent variable model. Obviously, the necessary information for the model is more apparent for the algorithm when using the derivative spectrum. The two samples with predicted values below 0 are the Abrera EDR samples.

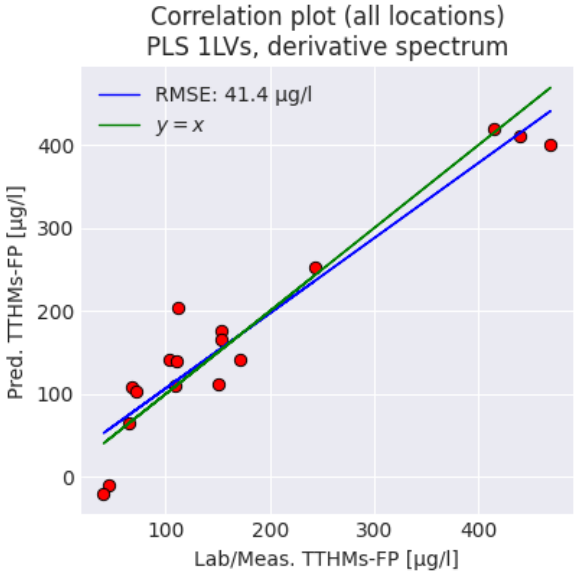


Figure 18: Model built on derivative spectra rather than absorption spectra from all samples.

Additionally, models that combine the Ter and the Irland water samples were calculated from the absorption spectra using wavelengths {240, 250, 280} nm and 1 and 2 latent

variables for comparison against the all-in-one model. From Figure 19 it can be seen that these two data sets can be combined. The one overestimated sample in the left graph is the raw water sample from Nephin valley. The underestimated samples around 150µg/l are three raw water samples from the Ter treatment plant.

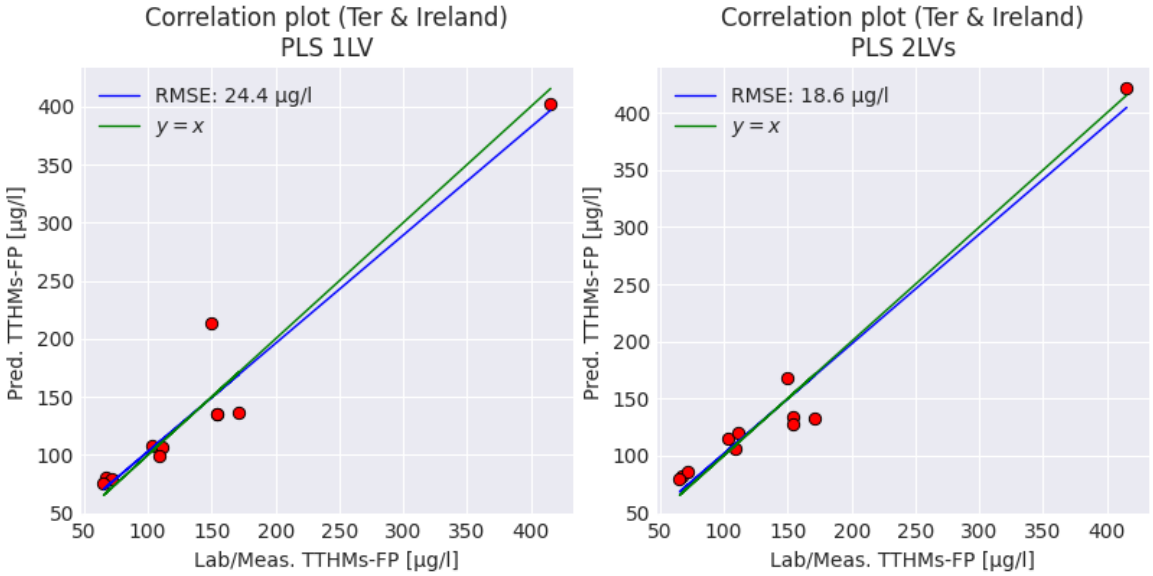


Figure 19: Models from pre-chlorinated water samples combining Ter and Irish data with one latent variable (left) and two (right).

Once again, the models suffer accuracy when brominated DBPs are forming due to bromide rich water.

3.5 HAN prediction models

The first developed model used all available pre-chlorinated spectra for the prediction of HAN formation. As can be seen from the below graphs no satisfactory model could be achieved compared to the THM prediction which gave acceptable correlations using the same wavelengths {240, 250, 280} nm.

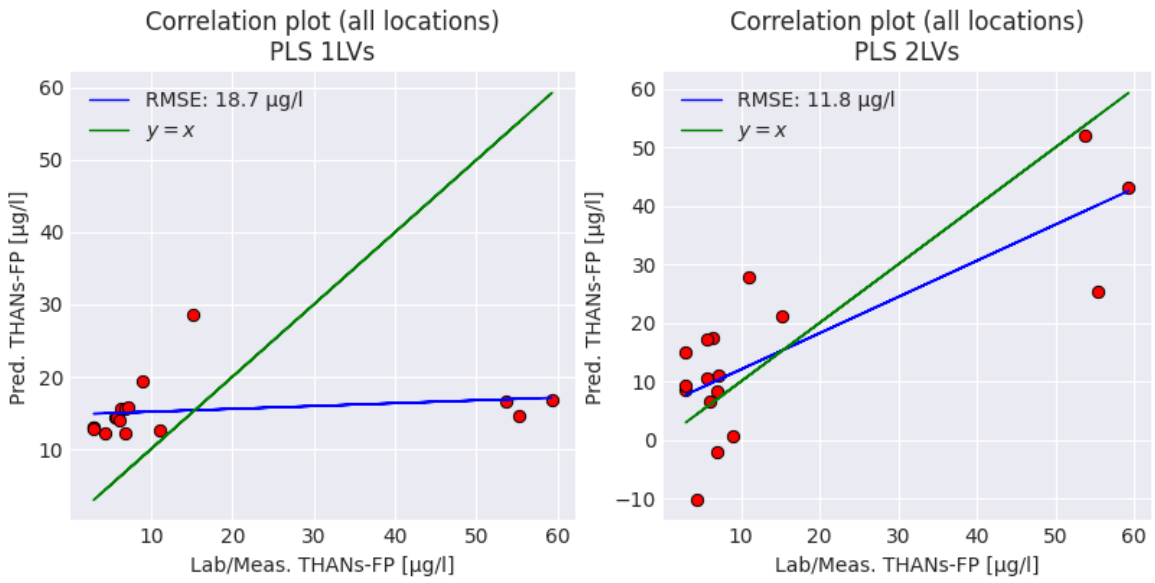


Figure 20: PLS models of HAN based on absorption spectra with one latent variable (left) and two (right).

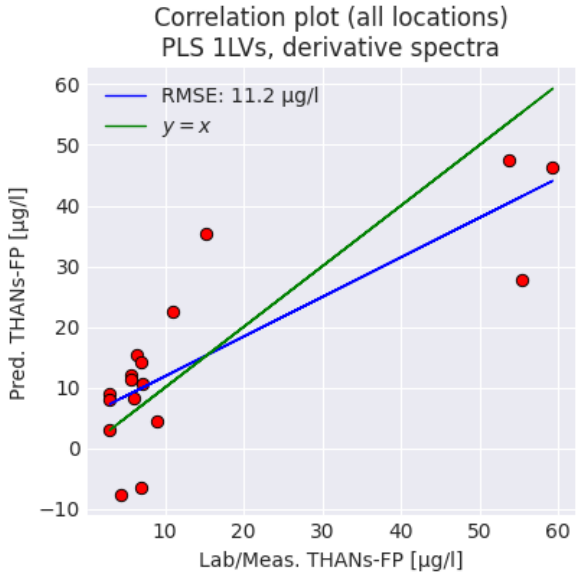


Figure 21: PLS model for HAN based on derivate spectra with one latent variable.

Summarizing a one-fits-all prediction model for HAN concentrations was not achieved. Therefore, site specific models have been developed for the Ter and Abrera treatment plant. Data from Irland didn't contain enough data points to build a model from that data alone. Therefore, the Irish data was combined with Ter data as this has proven successful for previous models already.

Each model is built from the derivative spectra on wavelengths {250, 265, 300} nm contains with one latent variable.

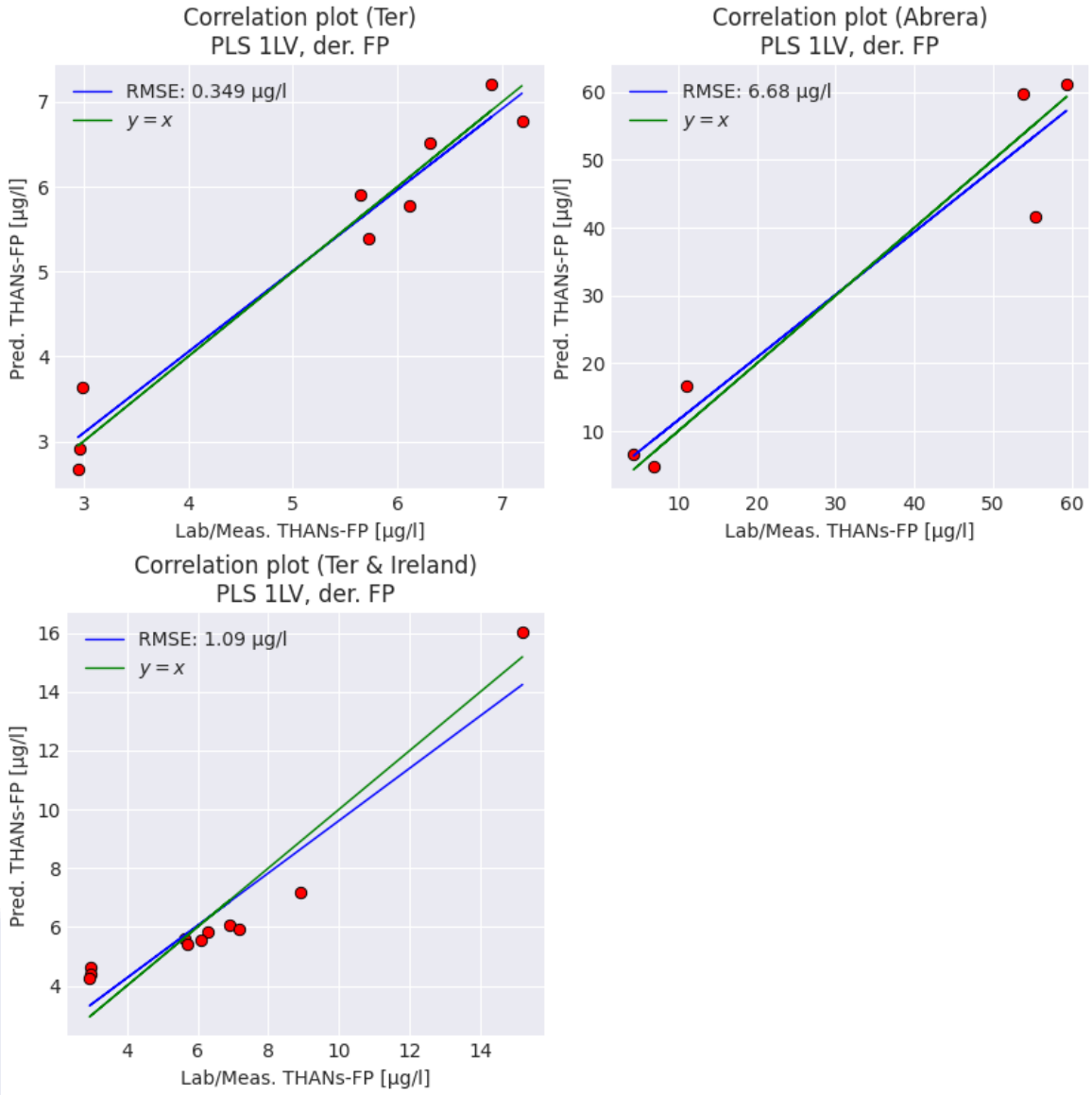


Figure 22: HAN models built for Ter (top left), Abrera (top right) and Ter + Ireland (bottom) samples.

Samples from Abrera showed the highest concentrations of HANs.

3.6 Delta spectrum of unchlorinated minus final spectrum

Another hypothesis that was examined was to use not the pre-chlorinated spectrum alone for the prediction of the final THM or HAN concentrations but use the mathematical difference of the raw water minus the final spectrum after all DBPs have formed as the basis of the model building. The idea here was that such a delta spectrum would contain more information on the actual DBPs formed during the chemical reaction itself.

However, this hypothesis didn't prove correct as can be seen in Figure 23. The achieved RMSE values are way above the values achieved from the pre-chlorinated spectrum

alone as outlined in the previous chapter. Models were not only built for THMs but also for HANs but are not shown here.

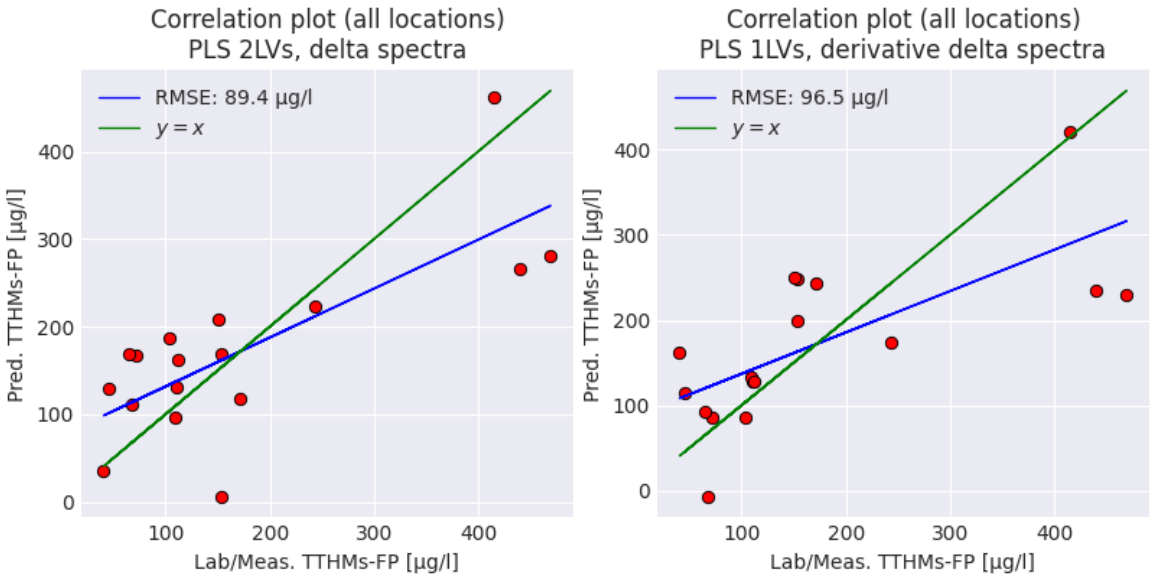


Figure 23: Models built on delta spectra between the first and last spectrum of an experimental run.

3.7 Difference between high and low bromide concentrations

As described in the chapter 3.2 a combination of one single model that fits all different water samples would be the most beneficial model from the business perspective. Models built so far have shown that a combination of samples containing higher concentrations of bromide in the water cannot be combined with samples that have low concentrations.

We also tested the possibility to cross use models. E.g. applying a model used from bromide rich water (Abrera) to water samples with low concentrations (Ter). Data is plotted in Figure 24. Shown models are PLS models with two latent variables, built on wavelengths {272.5, 342.5} nm with mean centering applied.

On the left side Ter water sample THM concentrations (raw water, clarified water and post AC water) were predicted using the model trained on Abrera data. Interestingly, the data seems to build two groups where the data points around the 1:1 line (green) are all data points from Ter post-AC water. The Abrera model (built on all Abrera samples) can predict the THM concentration of the final Ter water but not for the raw or clarified water.

Applying the Ter model to the Abrera data gives a linear response but highly underestimates the concentrations of the formed THMs. In such a situation the output of the model could be calibrated in a linear way to reduce the RMSE value.

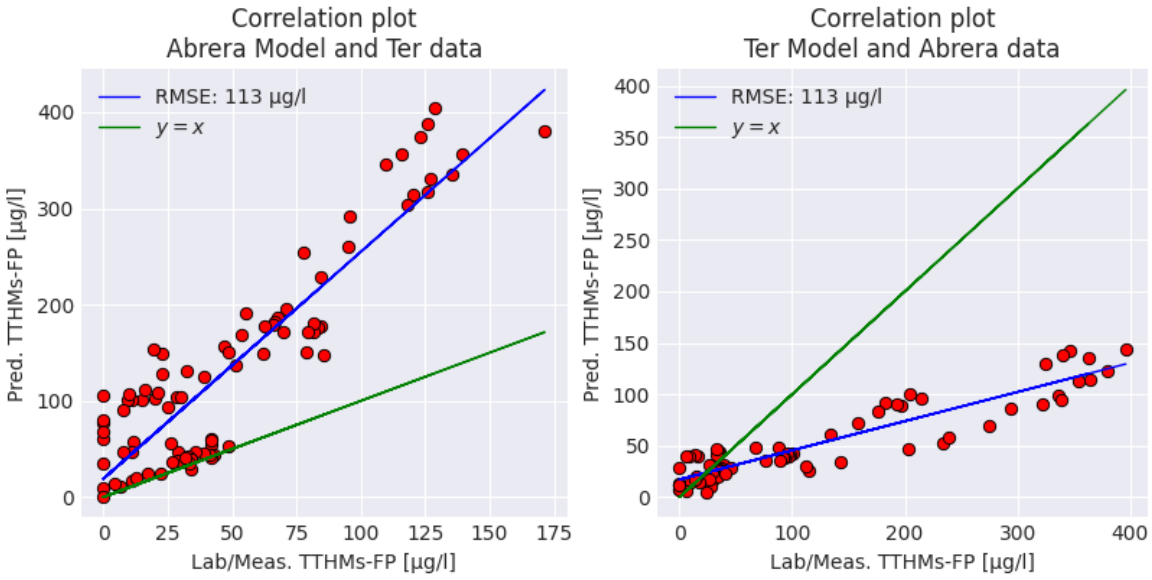


Figure 24: Applying source specific models to other water samples. Left: using the Abrera model (with high bromide concentrations) on Ter data and the other way around on the right side.

4.0 SUMMARY

At the beginning site-specific experiments were conducted and models for the formation potential have been derived. Models have been built around the experiments using all laboratory measurements along the time elapsed measurement of the shock chlorination experiments.

Table 2: RMSE (in µg/l) values for various chemometric models calculated from the full scale experiments outlined in chapter 3.1 & 3.2.

RMSE of TTHM models	Test Data Set	All Data	Comment
Ter Plant	15,5	16,5	Raw to final water
Irish Surface Water	24,0	23,8	two locations
Abrera	27,6	26,4	Raw to final water
Ter & Ireland	22,1	20,9	
Ter & Abrera	-	-	Not compatible
One model for all	-	-	Not working

The next step was to take only the spectra before the addition of chlorine and use this limited number of spectra for TTHM and THAN prediction which is closer to a potential application in the future. The obtained results are summarized in Table 3 and Table 4 below. Please note that using the derivative fingerprints allowed the reduction to one latent variable with RSME values comparable to values shown in Table 1 using the absorption spectra with two latent variables.

Table 3: RMSE (in µg/l) values for various chemometric models calculated for TTHMs from pre-chlorine spiking experiments.

RMSE of TTHM models	1 latent variable	2 latent variables	Comment
One model for all (absorbance spectra)	100,0	60,9	
One model for all (derivative spectra)	41,4	-	
Ter & Ireland (derivative spectra)	24,4	18,6	
Ter & Abrera (derivative spectra)	-	-	Not compatible

Table 4:RMSE (in µg/l) values for various chemometric models calculated for THANs from pre-chlorine spiking experiments.

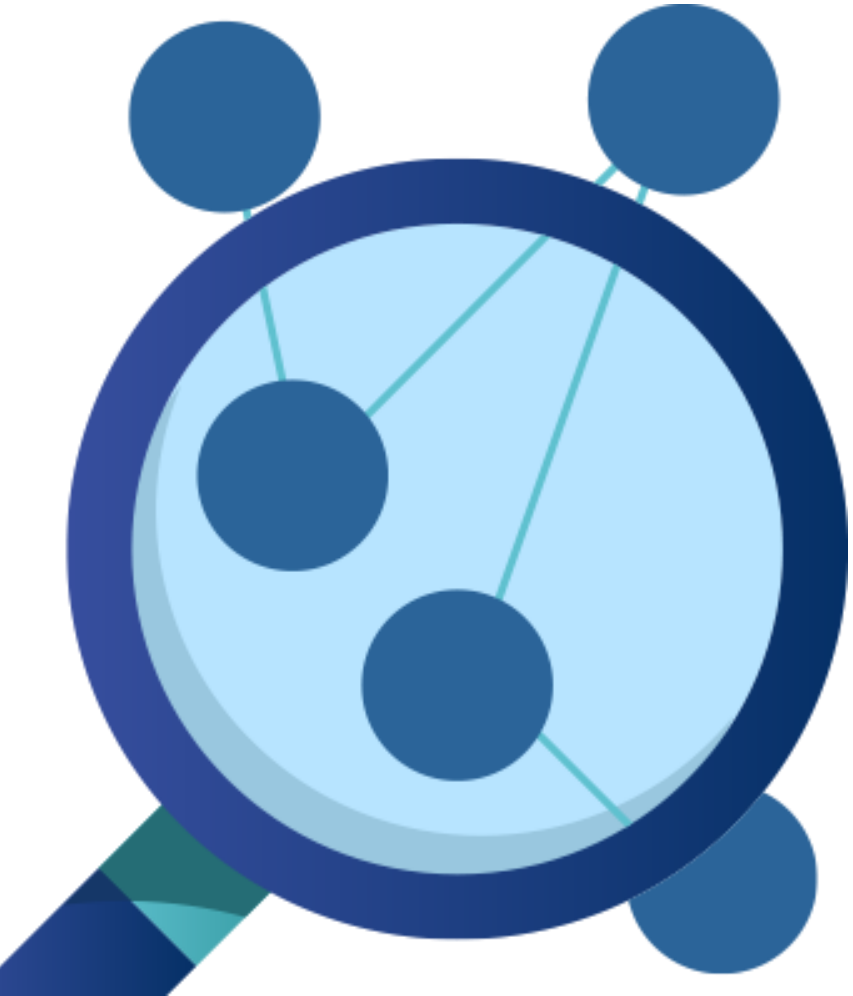
RMSE of THAN models	1 latent variable	2 latent variables	Comment
All samples (Absorption spectrum)	18,7	11,8	
All samples (Derivative spectrum)	11,2	-	
Ter (Derivative spectrum)	0,3	-	
Abrera (Derivative spectrum)	6,7	-	
Ter & Ireland (Derivative spectrum)	1,1	-	

Summarizing the tables above, it doesn't seem to be necessary to run full temporal experiments but taking sample of the unchlorinated waters and using them together with the maximum concentration of DBPs that form after a certain period of time (in our case 72h for THMs for example) is enough to achieve good site-specific correlations with the models. However, since the amount of data was limited for this type of experiment more data points would be beneficial to increase the quality of the models and making the model more global in the sense of applicable for a large variety of waters from different origins. S::can is already working with one customer in Australia to apply the gained knowledge in the field which will also increase the amount of available data.

Another very important take away from the described experiments is that models built with waters containing bromide are not compatible with models of waters without bromide. This is not too surprising but complicates the applicability in the real world. Again, the amount of data from bromide rich waters was limited and needs to be extended

in the future. On the other hand, waters with different characteristics of organic matter could be combined for the prediction algorithms.

After all the work conducted it seems possible to build a valuable tool for water treatment plant operators to optimize their treatment plants.



intoDBP

Funded by
the European Union