Disinfection by-product prediction models

Deliverable D3.2, WP3

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 39, 31.01.2026
Final version date	Month 39, 31.01.2026
Updated version date	26.06.2026
Dissemination Level	PU (Public)
Deliverable No.	D3.2
Work Package	WP3
Task	T3.3
Lead Beneficiary	Eurecat (EUT)
Contributing Beneficiaries	CAT
Report Author	Laura Vinardell (EUT) Danillo Lange Souza Borges Dos Santos (EUT)
Reviewed by	Irene Jubany (EUT), Mathias Ernst (DVGW-TUHH)
Approved by	Anissa Grieb (DVGW-TUHH)
Updated by	Irene Jubany (EUT)
Update approved by	Anissa Grieb (DVGW-TUHH)

History of versions		
Version	**Publication Date**	**Change**
1	30.01.2026	Sent to Commission
1.1	05.06.2026	Updated Figure 24 and eq 16 since an error was detected. Text of section 4.5.2 modified accordingly.
2	26.06.2026	Sent to Commission

Part. No.	**Participant organization name**	**Short name**
1 (Coordinator)	German Association for Gas and Water, local research branch at Hamburg University of Technology	DVGW-TUHH
2	Politecnico di Milano	POLIMI
3	Berlin Centre of Competence for Water	KWB
4	BioDetection Systems B.V.	BDS
5	Centre Tecnológic de Catalunya (EURECAT)	EUT
6	German Environment Agency	UBA
7	Helmholtz-Centre for Environmental Research	UFZ
8	Consorci Concessionari d'Aigües per als Ajuntaments i Industries de Tarragona	CAT
9	Tutech Innovation GmbH	TUTECH
10	Metropolitana Milanese SPA	MM
11	Multisensor Systems Ltd.	MSS
12	The National University of Water and Environmental Engineering	NUWEE

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

Abstract

The formation of disinfection by-products (DBPs) in drinking water distribution networks (DWDNs) represents a major challenge for water utilities, as it requires balancing effective microbiological protection with the minimization of chemical risks to public health. DBP occurrence is governed by a complex interplay of disinfectant type and dose, water quality, hydraulic residence time, operational conditions, and secondary disinfection practices. While numerous DBP models have been proposed in the literature, most are based on laboratory-scale studies or simplified assumptions that do not fully capture the spatial and temporal dynamics of full-scale distribution systems.

This deliverable presents a comprehensive modelling framework for predicting the formation and evolution of DBPs in a real DWDNs within the SafeCREW project. The framework integrates hydraulic modelling, mechanistic chemical kinetics, laboratory-scale experiments, field monitoring data, and data-driven approaches to support operational decision-making under realistic conditions. An EPANET-based hydraulic model was developed and calibrated to accurately estimate water residence times across the network, forming the backbone for subsequent water quality simulations.

A mechanistic chlorine decay model was implemented and calibrated using online sensor data, accounting for spatial variability across different network regions. In parallel, complementary data-driven models based on machine learning were developed to estimate chlorine decay parameters and to identify key drivers influencing disinfectant consumption. Although predictive performance for the decay constant itself was moderate, the data-driven approach achieved comparable downstream chlorine residual predictions and provided valuable process insights through feature importance analysis.

Building on the chlorine decay model, mechanistic formation models for regulated DBPs, namely trihalomethanes (THMs) and haloacetic acids (HAAs), were developed using network-scale simulations. The THM models showed good agreement with field measurements in most network locations, while HAAs proved more difficult to model due to weaker correlations with available water quality parameters. In addition, a kinetic model for chlorate formation associated with hypochlorite degradation was developed and validated for both concentrated (18%) and diluted (0.8%) hypochlorite solutions, highlighting the strong influence of temperature and storage conditions. Finally, the formation and decay of nitrogenous DBPs, specifically haloacetonitriles (HANs), were modelled based on laboratory experiments and successfully validated against distribution network measurements.

Overall, the results demonstrate that the proposed integrated modelling framework can reliably predict DBP formation and evolution in a full-scale DWDN. The models are suitable for near real-time implementation and provide a valuable tool for optimizing disinfection strategies, improving regulatory compliance, and reducing potential health risks associated with DBPs under climate-resilient water management scenarios.

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

Table of contents

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

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

Abbreviations

Acronym	Description
ABS	Absorbance
BCAA	Bromochloroacetic Acid
BDCAA	Bromodichloroacetic acid
BDCM	Bromodichloromethane
BCAN	Bromochloroacetonitrile
BDS	BioDetection Systems B.V.
CDBAA	chlorodibromoacetic acid
CS	Case Study
DBAN	Dibromoacetonitrile
DBAA	Dibromoacetic Acid
DBCM	Dibromochloromethane
DBP	Disinfection by-product
DCAA	Dichloroacetic acid
DCAN	Dichloroacetonitrile
DOC	Dissolved Organic Carbon
DWDN	Drinking Water Distribution Network
DWTP	Drinking Water Treatment Plant
EB	Pumping Station (from the Catalan acronym)
EC	European Commission
EPANET	Water distribution system modelling software
FCV	Flow Control Valve
GC	Gas Chromatography
HAN	Haloacetonitrile
HAAs	Haloacetic Acids
HS-SPME	Headspace Solid-Phase Microextraction
LC-QqQ	Liquid Chromatography Triple Quadrupole
LOD	Limit of Detection
MAE	Mean Absolute Error
MBAA	Bromoacetic acid
MBAN	Monobromoacetonitrile
MCAA	Monochloroacetic Acid
MCAN	Monochloroacetonitrile
MIAN	Iodoacetonitrile
ML	Machine Learning
MSS	Multisensor Systems Ltd.
NOM	Natural Organic Matter
NTU	Nephelometric Turbidity Unit
PCA	Principal Component Analysis
SHAP	SHapley Additive exPlanations
SUVA	Specific Ultraviolet Absorbance
TBAA	Tribromoacetic acid
TCAA	Trichloroacetic Acid
TCAN	Trichloroacetonitrile

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

TCM	Trichloromethane (Chloroform)
THAA	Total Haloacetic Acids
THAN	Trihaloacetonitrile
THM	Trihalomethane
TOC	Total Organic Carbon
TSS	Total Suspended Solids
TTHM	Total Trihalomethanes
UV	Ultraviolet
UV254	Ultraviolet absorbance at 254 nm
WHO	World Health Organization

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. Disinfection by-products

Access to safe drinking water is a fundamental prerequisite for public health and societal well-being. While drinking water treatment effectively prevents the transmission of waterborne diseases, the use of chemical disinfectants can unintentionally lead to the formation of disinfection by-products (DBPs). Disinfectants such as chlorine, chloramines, chlorine dioxide, and ozone are widely applied worldwide, primarily because they provide a residual disinfectant that preserves microbiological safety throughout drinking water distribution networks (DWDNs). However, reactions between these disinfectants and naturally occurring organic and inorganic constituents in water can result in the formation of DBPs, some of which pose potential risks to human health.

Since their first identification in the 1970s, several thousand DBPs have been reported in disinfected drinking water. Among these, trihalomethanes (THMs) and haloacetic acids (HAAs) are the most prevalent and extensively regulated compounds. Other DBP families, such as nitrogenous and aromatic DBPs, are typically detected at lower concentrations but often exhibit higher toxicity, including mutagenic, genotoxic, or carcinogenic effects. As a result, regulatory frameworks worldwide impose increasingly stringent limits on selected DBPs, placing significant pressure on water utilities to ensure compliance while maintaining microbiological safety.

The formation and fate of DBPs in DWDNs depend on a complex interplay of factors, including the type and dose of disinfectant, source water quality, natural organic matter (NOM) characteristics, presence of inorganic species such as bromide, hydraulic residence time, and operational conditions within the distribution system. Unlike treatment plants, distribution networks are dynamic and spatially heterogeneous systems, where water quality continues to evolve due to prolonged disinfectant contact times, variable flows, storage tanks, and secondary disinfection practices. Consequently, DBP formation may persist or even increase after water leaves the treatment plant, making their control particularly challenging.

Several strategies have been proposed to mitigate DBP formation, including precursor removal during treatment, optimization of disinfectant type and dose, reduction of water residence time, and implementation of additional treatment processes prior to consumption. However, these measures often involve trade-offs between microbiological protection, chemical risk, operational complexity, and cost. In this context, the management of DBPs represents a critical scientific, technical, and regulatory challenge with direct implications for public health and water utility operations.

Mathematical modeling has emerged as a key tool to support DBP management, enabling the assessment of formation mechanisms, prediction of concentration levels, and evaluation of alternative operational strategies. By providing quantitative insight into DBP dynamics, models can support proactive decision-making and reduce reliance on extensive monitoring campaigns. Nevertheless, the development of reliable and transferable DBP models remains a major challenge, particularly at the scale of real drinking water distribution networks.

1.2. Disinfection by-products modelling

DBP models play an increasingly important role in drinking water management, as they enable utilities and regulators to improve control strategies, minimize health risks, and support compliance with regulatory standards when combined with monitoring data. Over the past decades, a large number of

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

DBP modeling studies have been published, significantly advancing the understanding of DBP formation mechanisms and key influencing parameters. However, most of these models have been developed at the laboratory scale, under controlled conditions that do not fully represent the complexity of real drinking water DWDNs.

Only a limited number of studies have attempted to model DBP formation using data from full-scale distribution systems, and even fewer have incorporated accurate hydraulic residence time estimations through network simulation tools such as EPANET. Water residence time is a critical determinant of DBP formation, as it governs the duration of contact between disinfectants and DBP precursors. Simplified or estimated residence times can therefore lead to significant uncertainties in model predictions. Moreover, key operational features commonly present in real systems, such as chlorine booster stations, have been systematically neglected in existing modeling approaches, despite their potential to strongly influence DBP formation patterns.

Another major limitation in current DBP models relates to the characterization of natural organic matter. Most studies rely on bulk parameters such as total organic carbon (TOC) or dissolved organic carbon (DOC) as surrogates for DBP precursors, even though these parameters include fractions that do not actively participate in DBP formation. Only a small number of studies have incorporated more selective indicators, such as UV absorbance at 254 nm (UV254) or specific ultraviolet absorbance (SUVA), which better reflect the reactive fraction of organic matter. Similarly, most modeling efforts have focused on total trihalomethanes (TTHMs), while relatively few studies have addressed individual DBP species or other DBP families. This represents a significant gap, as individual DBPs can differ substantially in toxicity and regulatory relevance. In addition, bromide concentration—one of the key drivers of DBP speciation and toxicity—has frequently been overlooked.

With respect to chlorate, modeling efforts have so far been limited to systems disinfected with chlorine dioxide. Although the formation of chlorate from hypochlorite degradation is well documented, no studies have explicitly modeled this pathway within DWDNs, despite its growing relevance in systems relying on sodium hypochlorite.

Overall, DBP modeling in real DWDNs is constrained by data scarcity, low spatial and temporal resolution of monitoring programs, insufficient characterization of key chemical parameters, simplified hydraulic representations, and the omission of secondary disinfection processes. These limitations significantly reduce model reliability and hinder their transferability to operational conditions. There is therefore a clear need for comprehensive and scalable modeling frameworks that integrate laboratory-derived knowledge, real network monitoring data, hydraulic simulations, chemical kinetics, and advanced data-driven or hybrid approaches. Addressing these gaps is essential to enable reliable prediction of DBP formation in full-scale drinking water distribution systems and to support more effective, evidence-based management strategies.

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

2. Objectives

The objective of this deliverable is to develop, calibrate, and validate a comprehensive framework for predicting the formation and evolution of DBPs in a full-scale DWDN (Case study number 3, CS#3). The proposed framework integrates hydraulic modelling, mechanistic chemical kinetics, laboratory experiments, and data-driven approaches to support improved operational control and risk management of DBPs under real distribution system conditions.

Specifically, the objectives of this work are to:

1. Develop and validate a hydraulic model of the selected case study distribution network using EPANET, enabling accurate estimation of water residence times and operational conditions throughout the system.
2. Implement and calibrate a chlorine decay model based on mechanistic principles and supported by online sensor data and hydraulic simulations.
3. Develop complementary data-driven models to estimate chlorine decay parameters, with the aim of improving prediction accuracy and gaining insight into the key drivers of chlorine consumption.
4. Establish mechanistic prediction models for regulated DBPs, including THMs and HAAs, using network-scale simulations informed by monitoring data, laboratory-derived formation potentials, and hydraulic residence times.
5. Develop and calibrate kinetic models for nitrogenous DBPs, particularly HANs, based on laboratory-scale experiments and validate their applicability at distribution network scale.
6. Develop and validate a chlorate formation model associated with hypochlorite degradation, considering different disinfectant concentrations and temperature conditions relevant to primary disinfection and booster stations.
7. Evaluate the performance and limitations of the proposed models through comparison with field measurements across multiple locations within the distribution network.
8. Provide a modelling framework that can be implemented for near real-time DBP prediction and operational decision support, contributing to improved disinfection control while minimizing DBP formation and associated health risks.

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

3. Materials and methods

3.1. Drinking water distribution network hydraulic modelling

The case study used for the development of models is the DWDN of CAT (Tarragona). The EPANET hydraulic representation model is shown in Figure 1.

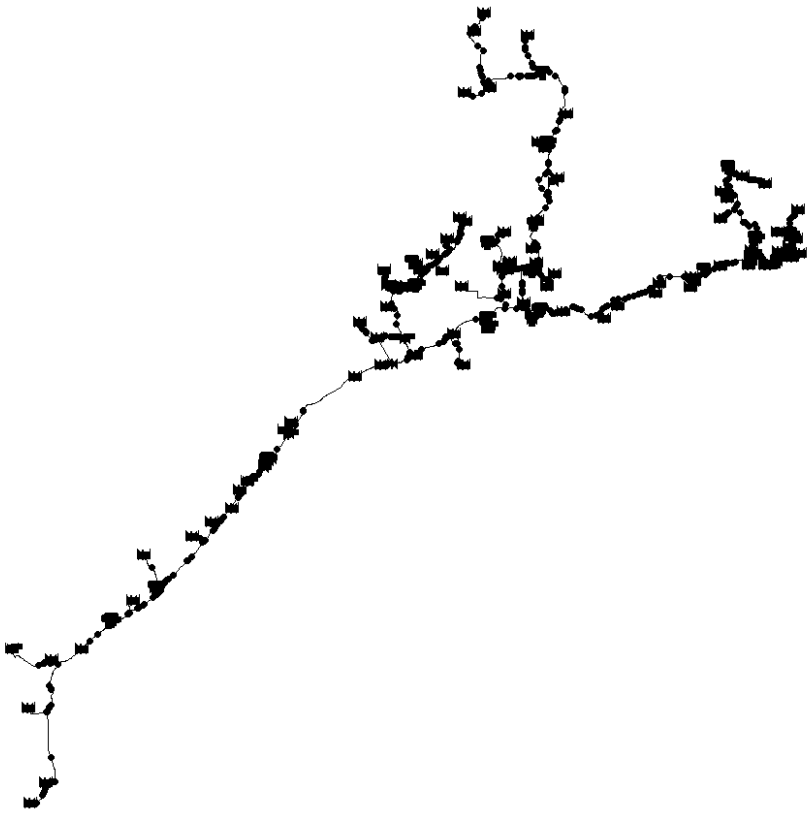

Figure 1: CS#3 global hydraulic model (EPANET representation)

The hydraulic model was calibrated using flowmeter measurements (used to define node demands in the DWDN), tank level sensors and pump curves. The hydraulic modelling was limited to three zones, illustrated in Figure 2, where EB indicates pumping station (from the Catalan acronym):

- Zone 1: from EB1 to Alcanar. Being EB1 the effluent of the drinking water treatment plant (DWTP) and Alcanar an end point of the network
- Zone 2: from EB1 to EB3 and EB10 inlet. Being EB1 the effluent of the DWTP and EB3 a tank in the middle of the DWDN.
- Zone 3: from EB10 to Masies Catalanes Albiol. Being EB10, a chlorine pumping station, and Masies Catalanes Albiol, an endpoint of the network.

The following simplifications were made to the real network to facilitate the simulations:

- Multiple tanks were simulated with as a unique tank with an equivalent volume (Σ tanks = 1 tank with equivalent volume)
- Connected reservoirs were simulated as a node with demand
- Tank storage variations were set according to the simulated water level

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

- Pipes with bidirectional flow (inlet–outlet) were simulated with two flow controlling valves (FCV)
- Increase of the maximum tank level to prevent overflow

The model validation was done through tank levels (comparing modelled and measured values) and considering the experience of CS#3 operators (CAT).

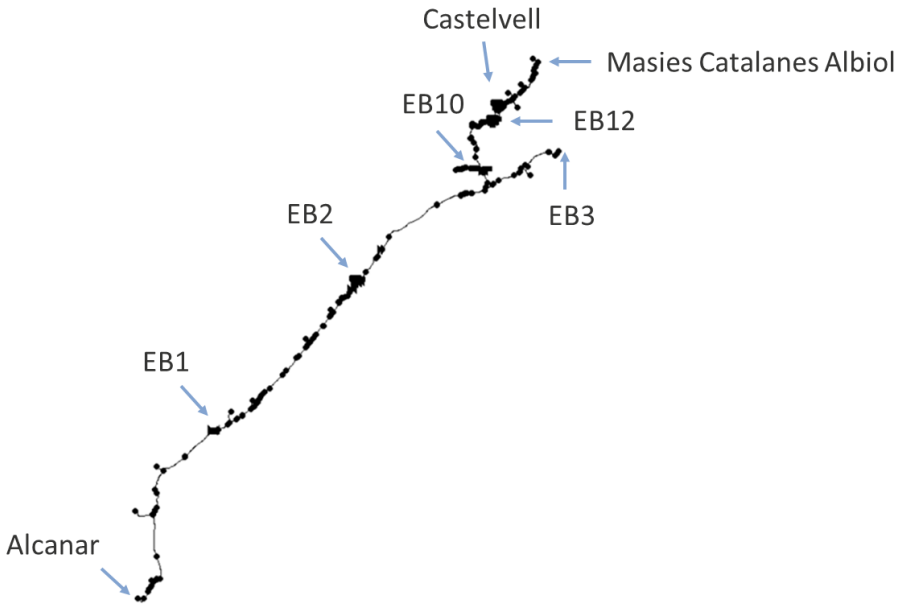

Figure 2: Hydraulic model of CS#3 section used in this study (EPANET representation). Storage tanks are indicated.

3.2. Chlorine decay model

3.2.1. Mechanistic model

Free chlorine reacts with NOM, inorganic compounds present in water, and pipe material, and decays with residence time. This phenomenon is generally expressed as a 1st order decay model (see Eq. 1).

$$Cl_f = Cl_0 \cdot \exp(-k \cdot t) \qquad \textit{Eq. 1}$$

Parameter *k* can be estimated in regions where Cl_f (final free chlorine concentration), Cl_0 (initial free chlorine concentration), and residence times are known. Using residence time from EPANET, hydraulic simulations and free chlorine measurements from online sensors, *k* was estimated using Eq. 1.

It is well known that chlorine decay in the wall and in the bulk is different: biofilm and/or pipe material react differently than at the bulk water. Thus, the *k* parameter is generally divided into k_{wall} (k_w) and k_{bulk} (k_b) (see Eq. 2).

$$k = k_w + k_b \qquad \textit{Eq. 2}$$

Chlorine decay was simulated in the network pipes and in tanks (DWDN sections shown in Figure 2). For tanks, *k* was assumed to be mainly k_b. Thus, tanks were calibrated to estimate k_b using Eq. 1 and

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

assuming $k_w \sim 0$. For pipes, k_b calibrated in tanks was used and a global k was determined using Eq. 1. Additionally, a global k was also calibrated using Eq. 1.

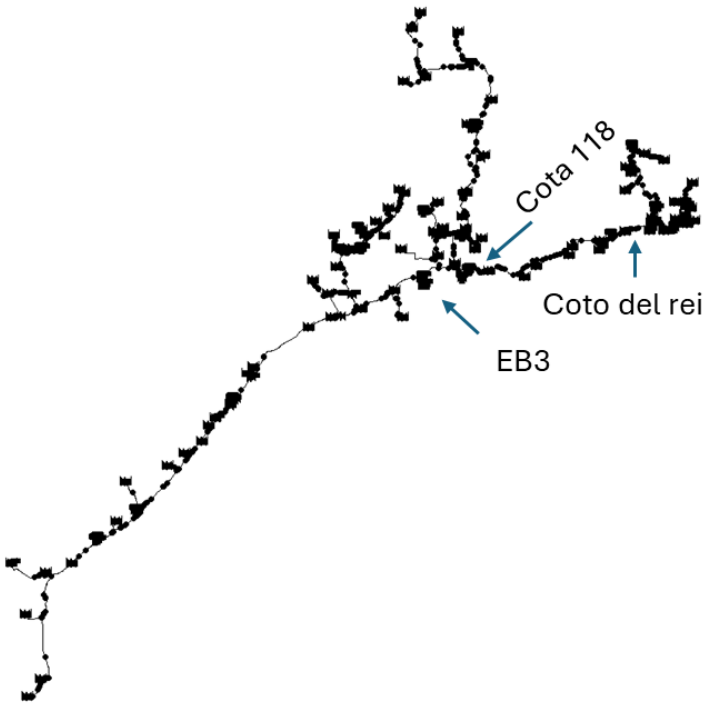

Figure 3: CS3 drinking water distribution network and the modelled tanks for the free chlorine decay model

3.2.2. Data-driven model for improving chlorine decay parameter

The model described in the previous section provides one estimate of k, based on mechanistic assumptions and parameterizations, while a machine learning (ML) approach can infer k from empirical data and interactions among covariates.

The primary objective was to compare the mechanical model against data-driven algorithms that estimate k, then propagate both estimates through the chlorine decay equation to evaluate which approach yields more accurate predictions of Cl_f. A secondary objective was to assess feature importance of data-driven models to extract process insights into drivers of chlorine decay.

Data and target

The working dataset comprised process and water-quality variables of the case study. The description of all used variables is detailed in Table 1. Each parameter was aggregated at hourly intervals to ensure temporal consistency and reduce noise from short-term fluctuations. The target variable was the apparent first-order decay constant k, derived from historical chlorine decay observations.

Table 1: Water-quality variables used in the data-driven model for chlorine decay rate estimation

Variable	Description	Unit of Measure	Frequency
ABS	Absorbance (254 nm)	1/cm	Hourly
CRE	Color	mg/L	
CTR	Conductivity	m^3/h	
THT	Temperature	°C	
TOC	Total Organic Carbon	mg/L	

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

TTR	Turbidity	NTU	
TSS	Total suspended solids	mg/L	

The residence time (*t*) values were available and used only in the final chlorine calculation. The ML models did not predict *t*, but *k* directly. Also, chlorine observations were not included in the training dataset of the models, to prevent knowledge leakage and overfitting.

Preprocessing and feature engineering

The following methods were followed while developing the data-driven models:

- Basic data cleaning: missing-value imputation (where applicable), outlier inspection, type casting, and removal of clearly erroneous records to prevent distortion of regression models.
- Scaling/normalization: considered where required by algorithmic assumptions (e.g., for linear models and distance-based learners).
- Derived features:
  - Lagged features to capture short-term memory (example $x^{t^{-1}}$) for ABS, CTR, and more variables.
  - Temporal encodings (e.g., time-of-day, day-of-week) to reflect operational cycles based on *t*.
  - Interactions via multiplicative terms (e.g., TOC × TSS, CTR × ABS) to expose synergistic or antagonistic effects on *k*.

The preprocessing and feature engineering augmentations aimed to map nonlinearities and cross-dependencies that the mechanical model may not explicitly represent. Also, the objective was to maximize the information that the data could provide, exploring unseen relationships between variables.

Model class and tuning

After training a portfolio of regression models with extensive grid search over hyperparameters and algorithm families, including *HistGradientBoostingRegressor* (*HistGB*) and *Random Forest*, the tuning strategy used nested parameter grids and cross-validation to mitigate overfitting and select robust hyperparameters.

Let $f(\mathbf{x};\theta)$ denote a regressor mapping input features $\mathbf{x}$ to $\hat{k}$. The loss minimized is the mean squared error (MSE):

$$L(\theta) = \left(\frac{1}{N}\right) \cdot \Sigma_{\{i=1\}}^{\{N\}} \left(k_i - f(x_i;\theta)\right)^2 \qquad \textit{Eq. 3}$$

With K-fold cross-validation to estimate generalization error. Performance was summarized via:

- *R2* (coefficient of determination) for *k*
- *MAE/MAPE* for *k* (robust to outliers)

$$MAE_{Cl} = (1/M) \cdot \Sigma_{\{j=1\}}^{\{M\}} \left| C_{f,obs}{}^{j} - C_f^j \right| \qquad \textit{Eq. 4}$$

- Downstream chlorine metrics, evaluating Cl_f produced by each *k* against measured chlorine.

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

Although *R2* on *k* was not high (0.10 in region 1 and 0.15 in region 2, reflecting the difficulty of predicting an effective rate constant that aggregates multiple reaction pathways and transport effects), the downstream chlorine predictions were notably accurate when passed through exponential decay.

Mechanical model baseline

The mechanical model provided a baseline estimate *k* derived from domain equations and calibrated constants, its output is similarly propagated and was used to compare with the data-driven model performance while estimating Cl_f.

3.3. THMs and HAAs mechanistic model

An intensive sampling campaign was carried out by CAT from April 2024 to June 2025, as described in D3.1 (Database and infographics from one year monitoring of the drinking water distribution networks of two case studies: Milan city and Tarragona). During this period, several water quality parameters were measured, including THMs, HAAs, turbidity, bromide, TOC, temperature, conductivity, and free chlorine.

A Principal Component Analysis (PCA) was performed to identify potential relationships between DBPs and the measured water quality parameters. The variables studied were:

- ΔDBPs: DBPs formation (in concentration) from the DWTP outlet and the evaluated tank (THMs and HAAs, total and individuals)
- ΔCl: Chlorine decay (in concentration) from the DWTP outlet and the evaluated tank
- TOC
- Free chlorine concentration
- Temperature
- pH
- Bromide concentration
- Fluoride concentration
- Electrical conductivity (EC) at 20ºC

PCA results were used to select the variables included in the THMs and HAAs estimation model. Only samples taken from tanks before secondary chlorination (i.e., after booster chlorination) were included, to capture the effect of free chlorine evolution without interactions from an additional chlorination step.

EPANET software was used to model THMs and HAAs formation along the DWDN. EPANET uses Eq. 5 to estimate DBPs kinetics:

$$\frac{dDBP}{dt} = K_{DBP} \cdot DBP \cdot (DBP_L - DBP) \qquad \text{Eq. 5}$$

Where K_{DBP} is the formation coefficient for the specific DBP, DBP is the DBP concentration (total THMs/HAAs or individual compounds), and DBP_L is the limiting concentration (the maximum generated, linked to formation potential). Simulations were run with the same constraints as the residence time simulations (indications node demands, pump, and valve management) including the following input information:

- The limiting concentration of each compound (total or individual): DBP_L

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

- The formation coefficient: K_{DBP}
- The initial DBP concentration (measured at EB1, the DWTP outlet)

THMs and HAAs were simulated at the different end-points of the region studied: Alcanar, EB3 and Masies Catalanes Albiol (see Figure 2).

3.4. HANs and HAAs modeling based on lab-scale experiments

3.4.1. Lab-scale experiments for HANs and HAAs

Lab-scale experiments were conducted to calibrate the kinetic models for HANs and HAAs. Water from CS#3 was collected at the DWTP just before the final chlorine disinfection.

Samples were chlorinated in the lab to obtain an initial free chlorine concentration around 2 mg/L. A total of 57 bottles (60 mL) were prepared by filling them with the source water and adding 15 µL of 12.8 g/L sodium hypochlorite solution and were stored at different temperatures (14 bottles at 10 °C, 14 bottles at 21 °C, and 28 bottles at 30 °C). Experiments at 30 °C were done in triplicate. Samples were collected at t = 0 h, 1 h, 2 h, 4 h, 24 h, 7 days, 9 days, and 14 days, and the chlorine reaction was quenched using 0.1 M thiosulfate. Samples were stored at 4ºC until analysis.

3.4.2. HANs and HAAs modelling

HANs hydrolysis kinetics was based on a modification of the equation described in (Yu & Reckhow, 2015). In Yu and Rechkhow (2015) work, HANs degradation is expressed as a function of HANs concentration, pH, and free chlorine and no formation is considered. Therefore, a factor accounting for the formation was including. This factor was considered dependent on free chlorine concentration. Therefore, the kinetic for the HANs dinamics in the network were defined as:

$$\frac{d[HAN]}{dt} = -k_1 \cdot [HAN] - k_2 \cdot [HAN] \cdot [OCl^-] + k_3 \cdot [OCl^-] \qquad \textit{Eq. 6}$$

This was solved together with the following expression that describes free chlorine kinetics:

$$\frac{d[OCl^-]}{dt} = -k_4 \cdot [OCl^-] \qquad \textit{Eq. 7}$$

Results from lab-scale experiments were used to estimate the four parameters k_1, k_2, and k_3 for each HAN and k_4 for hypochlorite. K1, k2 ad k3 estimation was done using R code and the function *optim* from the package *stats* and *ode* function from the package *deSolve*. K4 was estimated by solving the 1st order equation.

3.4.1. HANs sampling campaigns in the DWDN

Five tanks were sampled during four sampling campaigns: July 2023, November 2023, March 2024, and May 2024 (one campaign per season). Sampling points included the water treatment plant outlet and four tanks: WTP outlet, EB12, Coto del rei, Calafell les Brises and Sant Jaume dels Domenys (see Figure 4).

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

Additionally, another sampling campaign was conducted on 02/06/2025 at: WTP outlet, EB3, Cota 118, L'Arboç and Sant Jaume dels Domenys. These data were compared with the model developed in the previous section.

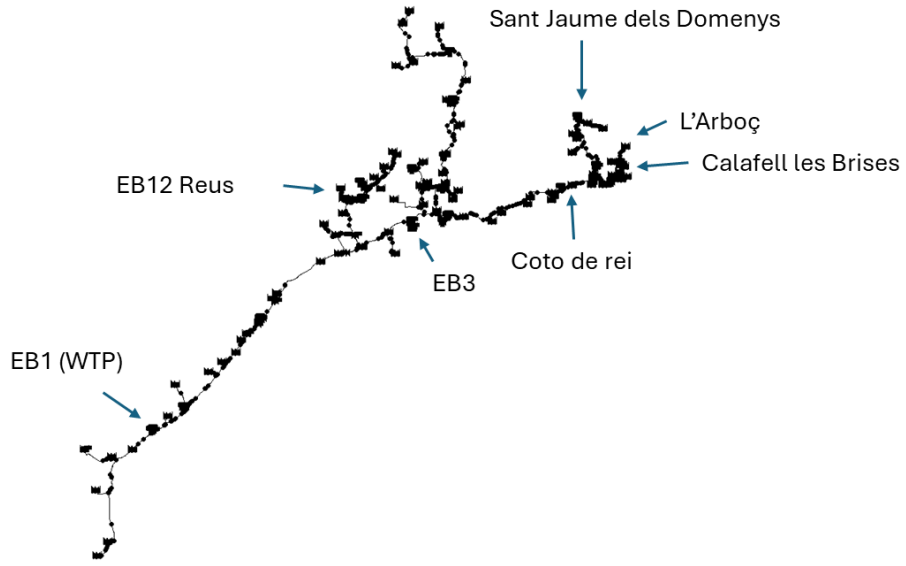

Figure 4: Distribution network of the CS3 (from EPAET) showing the sampling locations where HANs were monitored.

3.5. Chlorate

Chlorate is generally found in DWDNs that use hypochlorite as disinfectant due to the degradation of the stored concentrated hypochlorite. Hypochlorite degradation is accelerated by temperature, storage time, pH conditions, and disinfectant concentration. In this work, chlorate formation was studied for two concentrations, 18% and 0.8%, corresponding to the primary disinfection and the chlorine booster stations (having on-site hypochlorite generation), respectively.

3.5.1. Chlorate formation from hypochlorite at 18%

Chlorate formation from hypochlorite mainly occurs during hypochlorite storage, particularly when the solution is diluted in water. The chemical degradation pathways leading to chlorate formation are described by Eq. 8 to Eq. 10 (Boylard, Michele & Snyder Fair, Patricia, 1992):

$$NaClO \rightarrow Na^{+} + ClO^{-} \quad \textit{Eq. 8}$$

$$H_2O + ClO^{-} \rightleftarrows HClO + OH^{-} \quad \textit{Eq. 9}$$

$$2HClO + ClO^{-} \rightarrow ClO_3^{-} + 2Cl^{-} + 2H^{+} \quad \textit{Eq. 10}$$

When hypochlorite is stored under basic pH conditions, Eq. 10 is replaced by Eq 11. In this case, hypochlorite degradation follows second-order kinetics, as described by Eq. 12. This situation typically occurs when highly concentrated hypochlorite solutions are used, since both pH and ionic strength increase, leading to degradation kinetics that differ from those observed at neutral pH. Adam and Gordon (Adam & Gordon, 1999) investigated these kinetics at laboratory scale over a temperature range of 25 to 55 °C and for different initial hypochlorite concentrations.

$$3ClO^{-}_{(aq)} \rightarrow ClO^{-}_{3\,(aq)} + 2Cl^{-}_{(aq)} + 2H^{+}_{(aq)} \quad \textit{Eq. 11}$$

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

$$\frac{d([ClO^-])}{3dt} = -k_c[ClO^-]^2 \quad \text{Eq. 12}$$

For hypochlorite storage tanks, the evolution of hypochlorite concentration was simulated using a discrete time approach with 1-hour time steps. Hypochlorite degradation kinetics were solved to determine the variation in hypochlorite concentration between consecutive time steps, and the corresponding chlorate formation was calculated based on reaction stoichiometry.

The analytical solution of the hypochlorite degradation kinetics under basic pH conditions (Eq. 12) is given in Eq. 13. It was solved discretely over the simulated period. The initial hypochlorite concentration was set to 18% (equivalent to 2.54 M), while the initial chlorate concentration was assumed to be zero.

$$[ClO^-]_{(i+1)}{}^{basic\ pH} = \frac{1}{3k_c{}^{basic\ pH} \cdot \Delta t + \frac{1}{[ClO^-]_i}} \quad \text{Eq. 13}$$

Kinetics are temperature dependent: hypochlorite degradation is faster at high temperatures. The relation between temperature and kinetic constant is defined by the Arrhenius equation (Eq. 14, (IUPAC Compendium of Chemical Terminology, 1997)).

$$k_c = Ae^{\frac{-E_a}{RT}} \quad \text{Eq. 14}$$

Where the pre-exponential factor *A* and activation energy E_a are temperature-independent, *R* is the universal gas constant, and *T* is the absolute temperature expressed in Kelvin. For basic pH, Adam et al. (1999) presented different kinetic constants for different initial hypochlorite concentrations and different ionic strengths. The parameters *A* and *Ea/R* from Eq. 14 were estimated based on (1999) previous work and are presented in Table 2.

Table 2: Hypochlorite degradation kinetic parameters used in this study

	$\frac{E_a}{R}\left[\frac{J}{mol}\right]$	$A\left[\frac{M}{s}\right]$
Concentrated hypochlorite conditions	-12150	$1.51 \cdot 10^{10}$

To validate these parameters, a test was done using hypochlorite at 18% and storing it for 4 days at 26ºC. The observed results were compared with the theoretical estimation.

3.5.2. Chlorate formation from hypochlorite at 0.8%

The kinetics of chlorate formation from hypochlorite were studied at three different temperatures (10 °C, 20 °C, and 30 °C). For each temperature, twelve 30 mL amber bottles were prepared. At each sampling time, two bottles were used: one bottle was used for the determination of free chlorine, while the second bottle was used to prepare a 1:100 dilution of the 0.8% hypochlorite solution in a 50 mL Falcon tube containing sodium thiosulfate, for subsequent chlorate analysis.

Samples were collected at different times: t = 0 h, 1 day, 3 days, 1 week, 2 weeks, 3 weeks, 4 weeks, 5, 6, 7 and 8 and frozen until analysis.

Experimental results were used to estimate the kinetic parameters estimation for Eq. 14.

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

3.6. Analytical methods

THMs were analyzed with Agilent 6890, HS-GC/µECD. Detected THMs were chloroform (trichloromethane, TCM), bromodichloromethane (BDCM), dibromochloromethane (DBCM) and bromoform (tribromomethane, TBM).

HAAs were analyzed by liquid chromatography coupled to tandem mass spectrometry (LC-QqQ), using an Agilent 1290 Infinity II UHPLC system coupled to a 6470 triple quadrupole mass spectrometer. Specific HAAs studied were dichloroacetic acid (DCAA), chloroacetic acid (MCAA), bromochloroacetic acid (BCAA), bromoacetic acid (MBAA), dibromoacetic acid (DBAA), trichloroacetic acid (TCAA), bromodichloroacetic acid (BDCAA), chlorodibromoacetic acid (CDBAA) and tribromoacetic acid (TBAA).

HANs were analyzed by headspace solid-phase microextraction (HS-SPME) followed by gas chromatography (Agilent 8890) coupled to tandem mass spectrometry (GC-QqQ, 7010B triple quadrupole mass spectrometer from Agilent Technologies). Specific HANs analyzed were chloroacetonitrile (MCAN), bromoacetonitrile (MBAN), iodoacetonitrile (MIAN), dichloroacetonitrile (DCAN), dibromoacetonitrile (DBAN), tricloroacetonitrile (TCAN) and bromochloroacetonitrile (BCAN).

Chlorate was analyzed using ionic chromatography using a DIONEX ICS-2100 instrument.

4. Results and discussion

4.1. Hydraulic model

The hydraulic model was used to obtain water residence time at each point of the network section. Simulation results were validated by comparing the predicted tank levels with the sensor measurements. Figure 5 shows the results for tank levels in Castellvell site as an example of the validation of the hydraulic model. Water residence time results from the simulation of five months (April to August 2025) are graphed in Figure 6 for different site locations. The average residence times for these tanks are summarized in Table 3. These results validated by CS#3 operators (CAT). The water residence time is increasing along the network with episodes of up to 300 h (12.5 days) and averaged values of 188 h (7.8 days) in the farthest point from the DWTP.

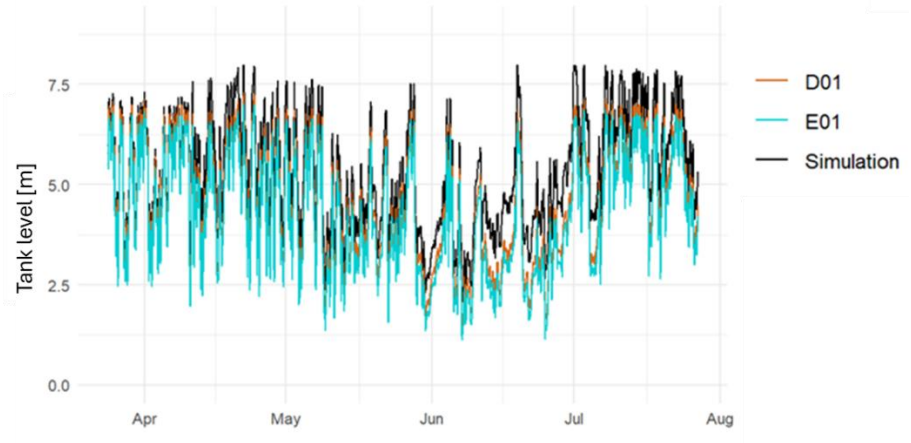

Figure 5: Tank level (Castellbell tank) obtained with the hydraulic model and comparison with level sensors. D01 and E01 are the two tanks shaping Castellvell. Data from 2025.

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

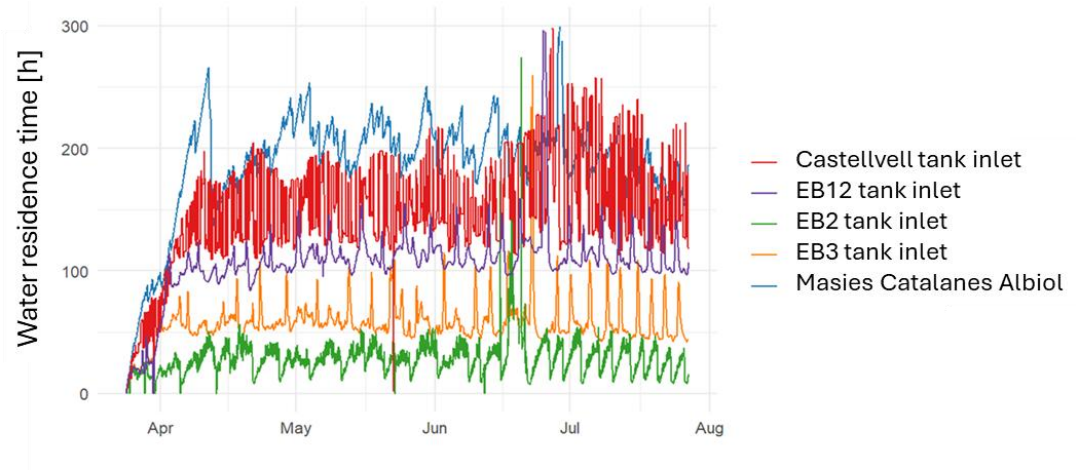

Figure 6: Water residence time at different points of the network (Castellvell, EB12, EB12 and EB3 tanks, and Masies Albiol). Data from 2025.

Table 3: Mean water residence time for different points of the network for the period April-August 2025.

Network point	Mean water residence time [h]
Tank EB2 inlet	19
Tank EB12 inlet	12
Tank Castellvell inlet	46
Tank EB2 inlet	36
Tank EB12 inlet	106
Tank Castellvell inlet	145
Tank EB3 inlet	58
Masies Catalanes Albiol	188
Alcanar	58

4.2. Chlorine decay model

4.2.1. Chlorine decay model in tanks (k_w)

Free chlorine decay in bulk water was calibrated in tank EB3. The expression for k_b is presented in Eq. 15. The results from the simulation with the calibrated values are illustrated in Figure 7 where it can be observed that the free chlorine concentration simulated for the outlet of the tank fits well the measurements.

$$k_b = 8{,}2 \cdot 10^7 \cdot e^{\frac{-6.6 \cdot 10^4}{T}} [h^{-1}] \quad \text{Eq. 15}$$

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

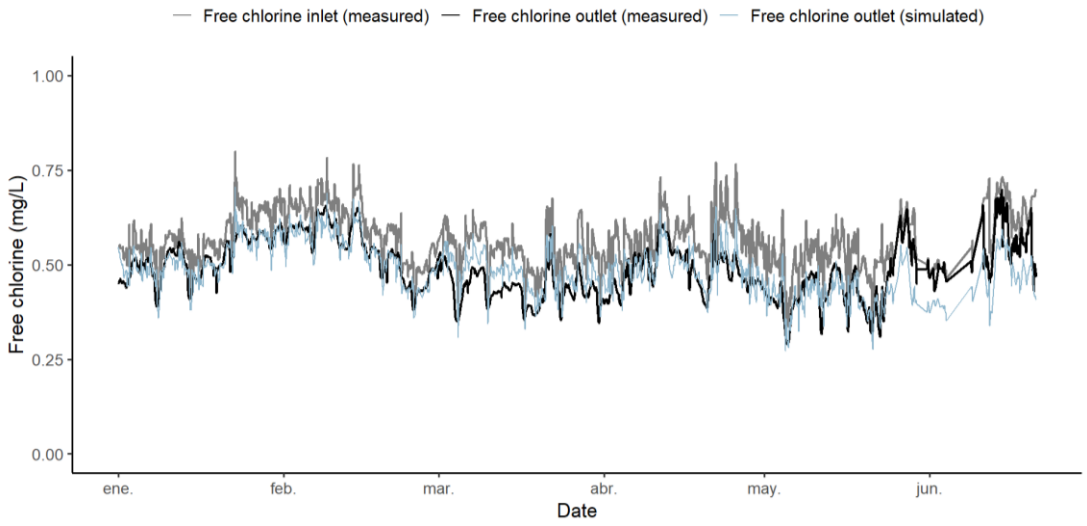

Figure 7: Free chlorine in Tank EB3. Comparison of measured and predicted concentration. Data from 2025.

The calibrated model was applied to two other tanks: one water storage tank (coto del rei, see Figure 8) and one chlorine booster tank (Cota 118, see Figure 9). For the storage tank, the model can predict the decrease in the chlorine concentration due to chlorine decay, and in the chlorine booster station, the model predicts well the increase in the chlorine concentration due to the addition of disinfectant. The errors between the model estimations and the values measured by the online sensors are summarized in Table 4.

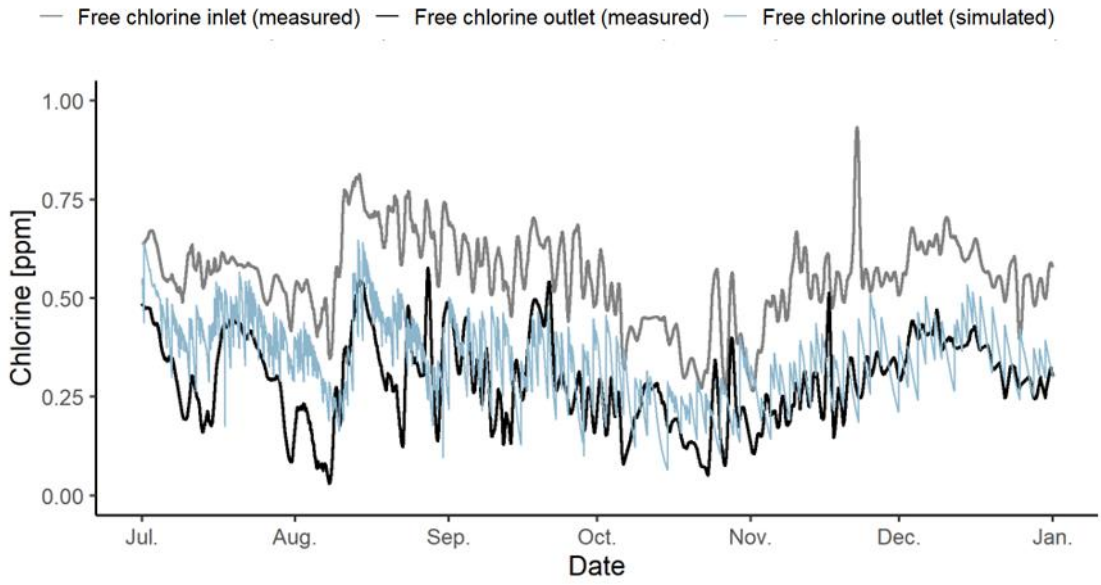

Figure 8: Chlorine decay model validation in a water storage tank (Coto del Rei). Data from 2025.

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

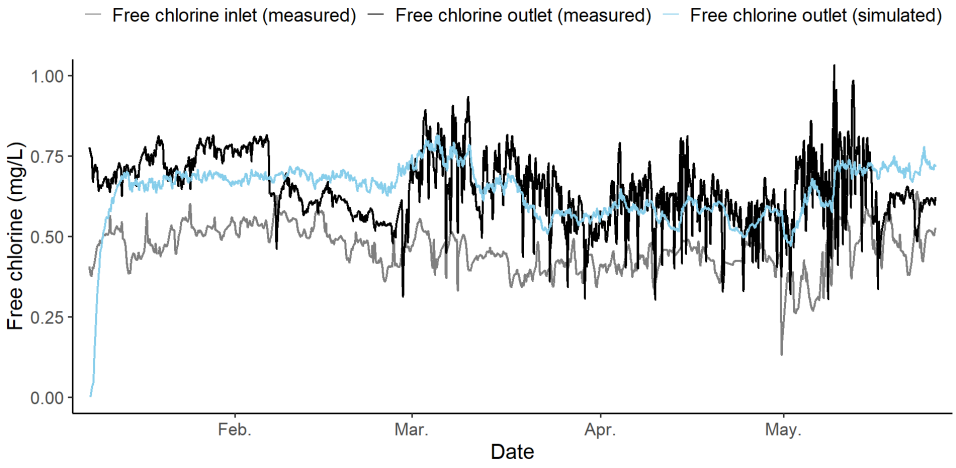

Figure 9: Chlorine decay model validation in a chlorine booster station (Cota 118). Data from 2025.

Table 4: Free chlorine estimation error in the chlorine decay model validation

Error estimator	EB3	Coto del rei	Cota 118
MAE [mg/L]	0.03	0.10	0.09

4.2.2. Chlorine decay model in pipes

The three zones described in Section 3.2.1 were calibrated individually to estimate a representative global decay constant k. This resulted in three distinct models, one for each region (model 1 for region 1, model 2 for region 2, and model 3 for region 3; see Figure 10). The corresponding results are shown in Figure 11. k-parameter is not a unique value for each region since this decay parameter depends on temperature. That is why these figures show a vector of k-values for each region.

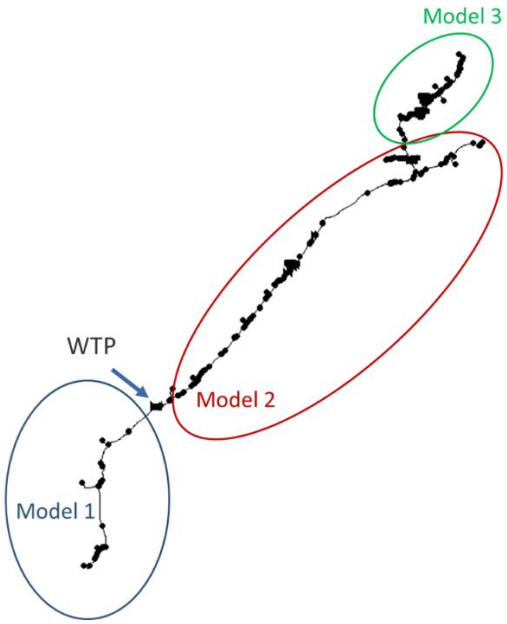

Figure 10: Sections of the network used for calibration of chlorine decay model for pipes.

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

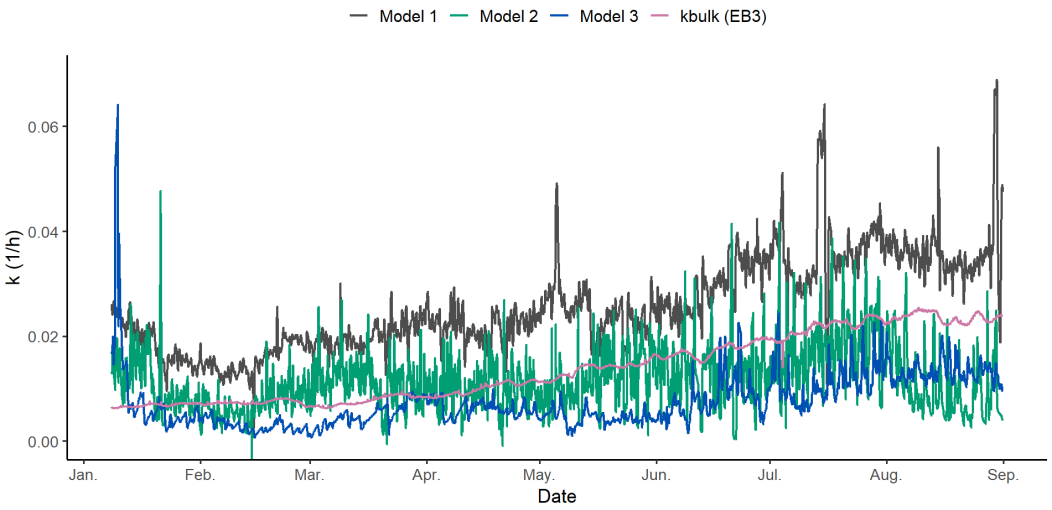

Figure 11: Chlorine decay parameter (k) obtained for the different studied regions and comparison with k_b calculated in tank EB3. Data from 2025.

The highest decay constant values were observed in region 1, followed by region 2 and region 3. In region 1, flow velocity is low due to reduced water demand, and a high presence of biofilm on pipe walls is suspected. This may explain the higher k-values compared to region 2, which corresponds to the main pipeline which is characterized by high water demand and large pipe diameters. Chlorine decay in region 3 is lower than in the other regions, likely because the water receives an additional chlorine dose at a booster station. When water undergoes secondary chlorination, chlorine may decay more slowly because less NOM is available, even if chlorine concentrations are high.

The bulk decay constant (k_b) calibrated using tank data (see section 4.2.1) is shown in red in Figure 11. For region 1, k_b is lower than the total decay constant k (k_b + k_w for region 1), which is consistent with decay being partially driven by bulk reactions and partially by wall reactions. In region 2, where pipe diameters are large (up to 1600 mm), k and k_b are very similar, suggesting that wall decay k_w is negligible. However, in region 3, k_b is higher than the total k obtained from network calibration. This indicates that chlorine decay may behave differently in areas receiving only primary disinfection compared to areas supplied with additional chlorination from booster stations. Consequently, it was not possible to define a single pair of k_b and k_w values applicable to the entire network.

Since CS#3 is equipped with multiple online free-chlorine sensors, region-specific decay models can be calibrated for each monitored section. For this reason, the analysis was simplified by working exclusively with the global decay constant k, without separating it into k_b and k_w, and calibrating a specific k value for each region. As a result, the network was divided into three regions corresponding to three distinct models, as illustrated in Figure 10.

4.3. Data-driven chlorine decay model

4.3.1. Results and comparison

The data-driven regressors on region 1 and 2 (e.g., HistGB) exhibited modest *R2* on the target *k*, consistent with the complexity of chlorine decay kinetics and the role of latent, unmeasured factors (e.g., micro-scale reactivity, speciation, and temperature-dependent pathways).

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

Despite the modest *R2*, the distribution of k^{ML} was physically plausible and free of pathological predictions after cleaning and hyperparameter tuning.

After predicting the *k* values, the Cl_f was then calculated using previously stated equations, and then the results were compared against the real chlorine values and the mechanistical model, which can be observed in Figure 12 and Figure 13.

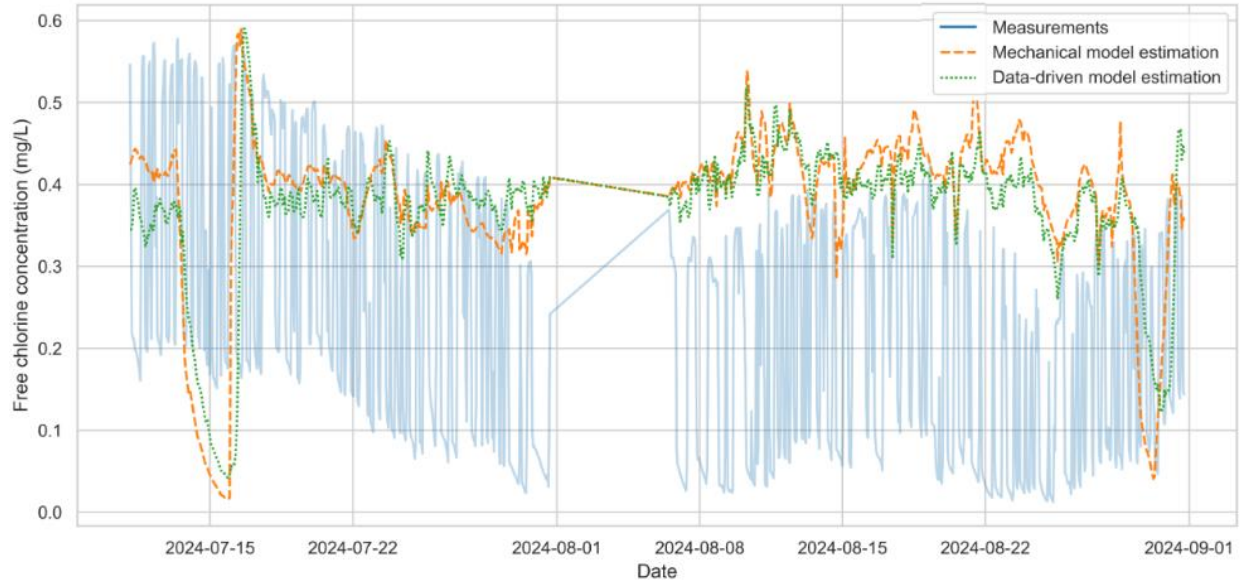

Figure 12: Comparison of chlorine concentration between measured values, mechanistical and data-driven models in region 1

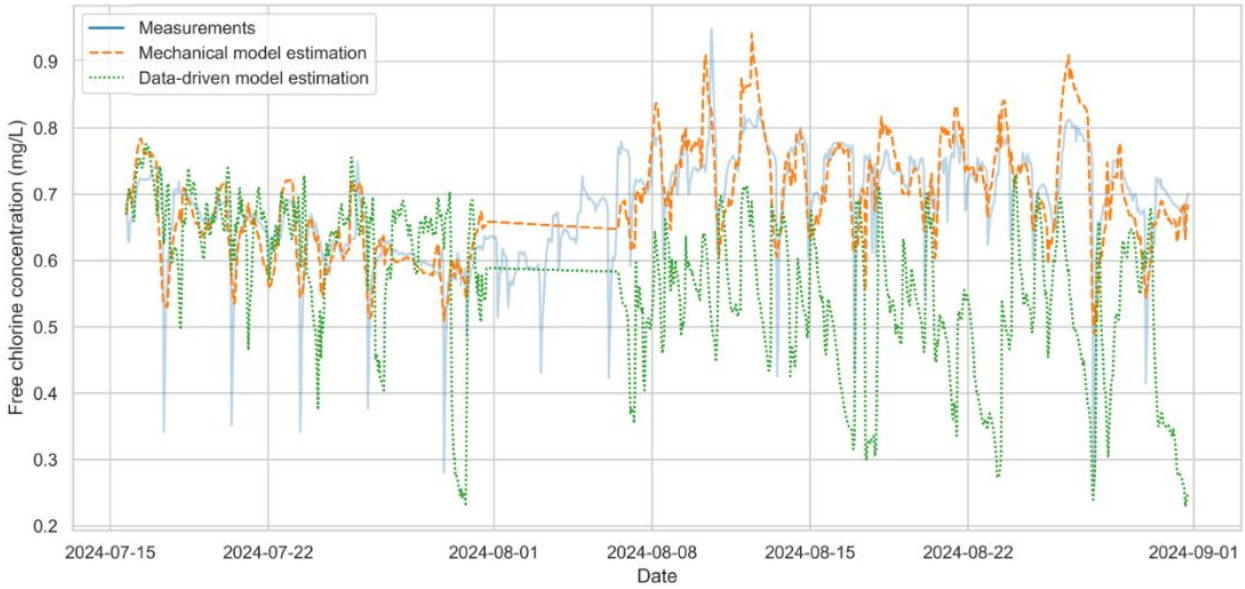

Figure 13: Comparison of chlorine concentration between measured values, mechanistical and data-driven models in region 2.

Downstream chlorine accuracy

When propagated through the decay equation, chlorine concentration derived with ML was very similar to both measured chlorine and mechanistical-model chlorine. Also, the absolute error was

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

acceptable across typical operating ranges of *t*. Moreover, in regimes with moderate residence times, the data-driven predictions showed parity with the mechanistical model, and in certain subsets (identified via stratification on TOC/TSS or Free chlorine concentration), the data-driven approach matched the measurements.

This outcome underscores a practical point: optimizing the intermediate variable *k* is relevant but not sufficient. The use of metric Cl_f is the operational KPI, and the data-driven model can produce competitive results even with imperfect k estimation.

4.3.2. Feature importance and Process Insight

Tree-based learners (e.g., *HistGB* and *Random Forest*) provide permutation or gain-based feature importances. This analysis was critical for identifying operational drivers of chlorine decay and guiding process optimization. In this experiment:

- Variables such as TSS and TOC, which are proxies for reactive organic matter, tended to rank highly, indicating strong associations with chlorine decay.
- Operational variables such as conductivity and absorbance contributed fairly, showing control and mixing-related effects
- Lags and interactions improved the explanatory power of the models, the multiplicative terms frequently surfaced in the top features, which is consistent with coupled reaction and adsorption dynamics.

All these importances provide diagnostic value beyond a single scalar *k*: they highlight which covariates most influence decay, supporting process optimization (for example, targeting TOC reduction or managing solids to stabilize chlorine residuals).

As observed in Figure 14, the SHAP (SHapley Additive exPlanations) (Lundberg, 2017) summary plots indicate that time-related features and engineered interactions play a dominant role in shaping the model's prediction of k. For example, the feature *day_of_year* appears as the strongest contributor, suggesting a marked seasonal or operational periodicity influencing chlorine decay. Several short-term operational indicators such as THT_1h (related to temperature) and lagged variables also show substantial impact, indicating that both immediate and delayed process conditions affect reactivity. A notable pattern is the prominence of interaction terms, reflecting that chlorine decay depends not only on individual patterns but on their combined behavior, which other kinds of models typically do not capture explicitly. The color gradients show that both high and low feature values can shift predictions in different directions depending on context, reinforcing the nonlinear, interaction-driven nature of the process. Overall, the analysis highlights how the data-driven model uncovers complex temporal and cross-feature relationships that provide deeper operational insight into the drivers of chlorine decay.

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

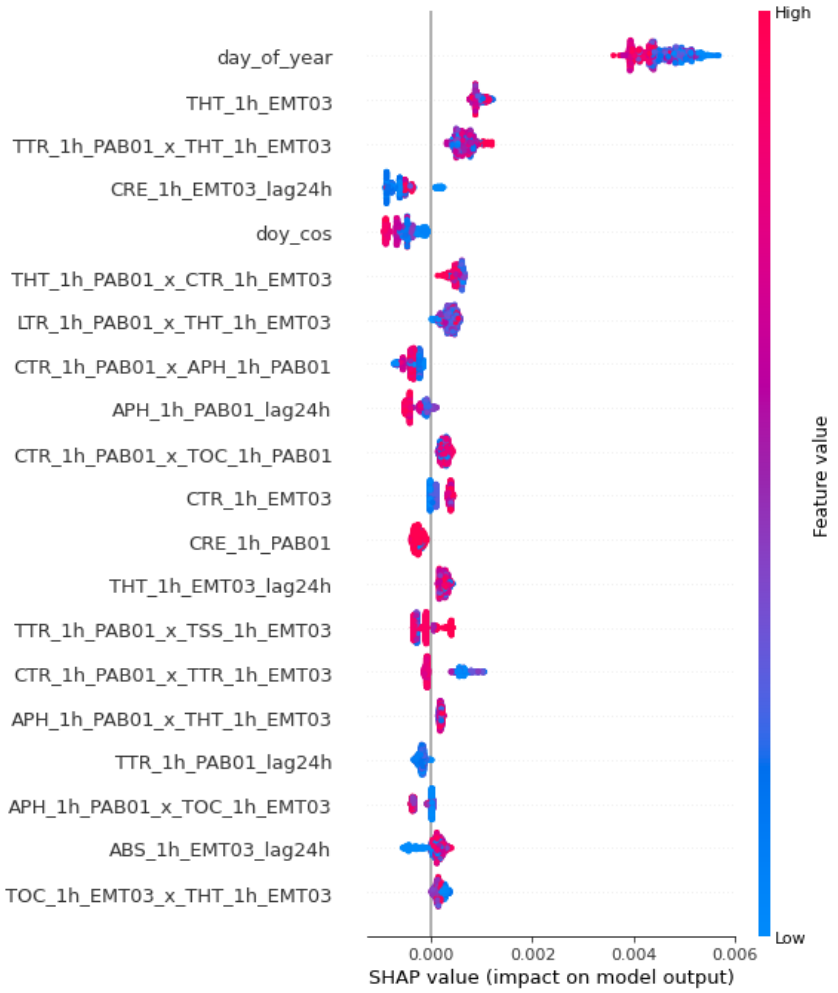

Figure 14: Feature importance values for region 1 data-driven model

While analyzing the region 2 model in Figure 15, the plot highlights *hour_cos* as the most influential feature, indicating that short-term diurnal cycles strongly affect the model's estimation of *k.* Several interaction terms also show substantial contribution, suggesting that the combined behavior of water flow and water quality-related is more informative than any single parameter alone. Lagged features such as *THT_1h_lag24h* (related to temperature) and *TTR_1h_lag7H* (related to turbidity) further indicate that past operational conditions influence current chlorine decay dynamics. Overall, the model also relies heavily on temporal patterns and cross-feature interactions, reinforcing the concept that chlorine decay is governed by nonlinear and time-dependent processes.

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

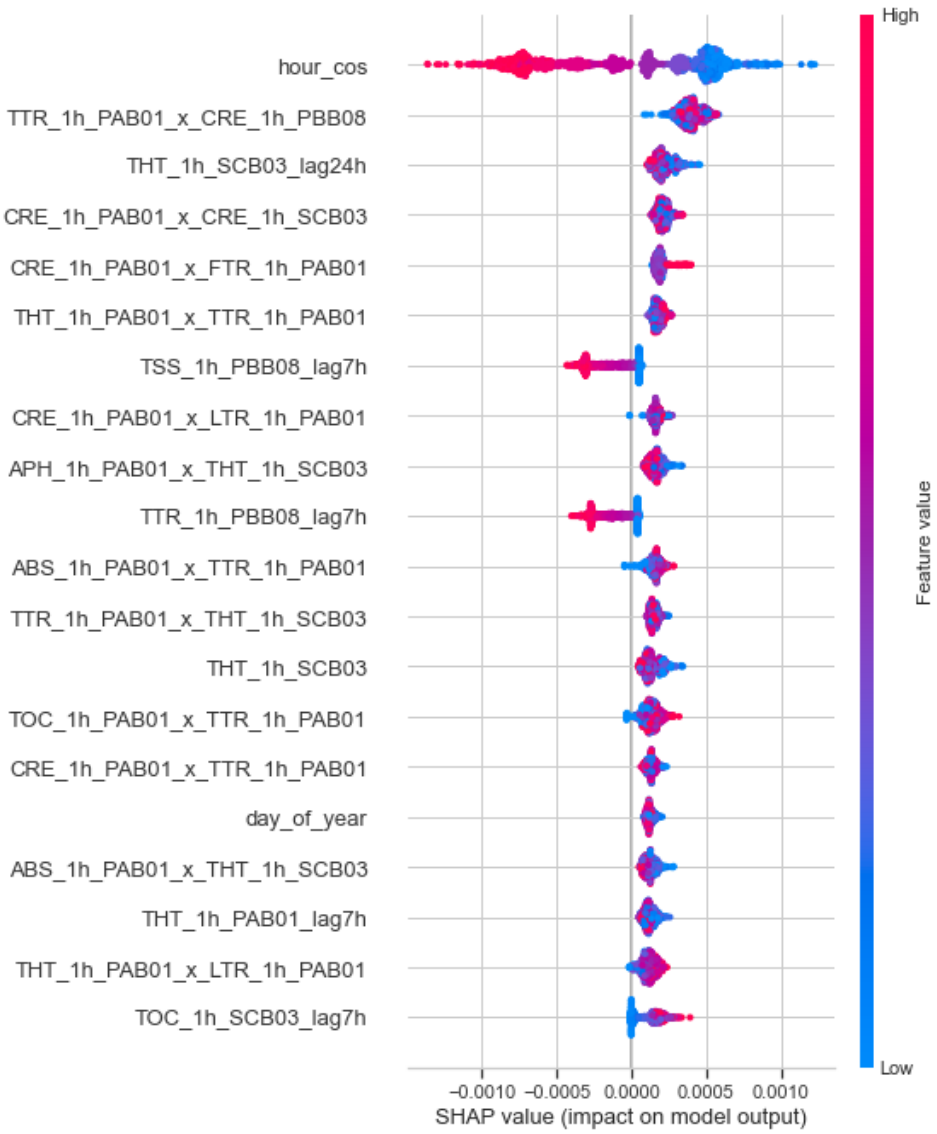

Figure 15: Feature importance values for region 2 data-driven model

Comparative summary

The data-driven model shows a competitive downstream chlorine accuracy with modest *R2* on k, it can capture nonlinearities and interactions, offer feature-importance insights for operations, requiring careful data governance and validation.

Among the enhancements provided by the data-driven model are the feature discovery profile which highlight actionable levers such as TOC and TSS, alongside interaction mechanics, guiding pretreatment and mixing strategies to reduce chlorine demand.

Moreover, the data-driven model is useful for naturally modeling nonlinear dependencies and cross-terms that otherwise might be difficult to include in closed-form traditional mechanistical models.

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

As for the scenario calibration, it would be needed in the future to retrain transfer learning for fast local calibration for new zones or conditions.

Finally, the data-driven model provides residual learning which can be used in the future for hybrid modelling, improving generalization and reducing bias.

4.3.3. Next steps and conclusions

Among the next steps, the metric focus is to continue prioritizing downstream chlorine accuracy, not only the performance metrics on *k*. Also, the inclusion of cross-site validations and temporal holdouts would be good for improving stability. Additionally, the plan is to explore physics-informed features and residual learning atop the mechanistical model, and by performing an operational integration with the mechanistical model.

The data-driven models achieved chlorine residual predictions that were very similar to both measured values and the mechanical model outputs. Also, the data-driven models add diagnostics and operational value via feature-importance analysis, making it a compelling complement to the mechanical model.

4.4. THMs & HAAs formation model

4.4.1. Potential relationship between DBPs and water quality measurements

The PCA results considering THMs, HAAs, and water quality parameters are graphed in Figure 16.

For THMs, the most significant relationships found were:

- TTHMs formed and on-site measured free chlorine (inversely)
- TTHMs formed and chlorine consumed (positively)
- TBM formed and conductivity (positively)

Secondary relationships found were:

- TTHMs and temperature
- TTHMs and TOC
- TTHMs and turbidity
- TBM and bromide

Therefore, it was decided to consider free chlorine decay and water temperature in the THMs model.

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

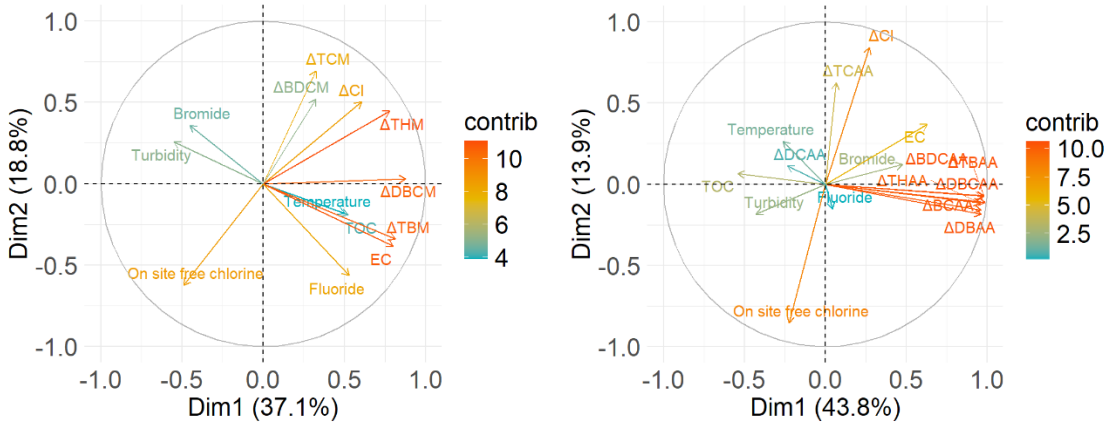

Figure 16: PCA results of the DBPs and water quality parameters interaction. THMs (left) and HAAs (right).

Regarding HAAs, the highest relationships found were:

- Among the HAAs (all the HAAs, except the TCAA) (positive relationship): ΔTBAA, ΔTHAA (total haloacetic acids), ΔDCAA, ΔBCAA and ΔDBAA
- ΔTCAA and chlorine decay (positively)
- ΔTCAA and onsite measured free chlorine (inversely)

Due to the low correlations found between HAAs and water quality parameters, and the fact that TCAA was correlated with chlorine decay, it was decided to use the same equation as for THMs for HAAs estimation.

4.4.2. THMs and HAAs simulation results

CS3 conducted several tests to determine the potential formation of THMs and HAAs. The results from 24/02/2025 and 02/06/2026 are presented in Table 5 for THMs (total and individual compounds) and in Table 6 for HAAs (total and individual compounds). These values were used to define the limiting concentration (DBP_L) for each compound.

Table 5: THMs formation potential from two sampling campaigns (24th of February and 2nd of June). LOD is the limit of detection.

DBP	BDCM	TCM	TBM	DBCM	TTHM
Unit	**µg/L**	**µg/L**	**µg/L**	**µg/L**	**µg/L**
LOD	**2,5**	**2,5**	**2,5**	**2,5**	**20**
20250224	17,6	29,4	4,0	16,0	66,9
20250602	18,3	29,1	5,2	17,9	70

Table 6: HAAs formation potential from two sampling campaigns (24th of February and 2nd of June). LOD is the limit of detection.

DBP	BCAA	DBAA	DCAA	MBAA	MCAA	TBAA	TCAA	THAA
Unit	**µg/L**	**µg/L**	**µg/L**	**µg/L**	**µg/L**	**µg/L**	**µg/L**	**µg/L**
LOD	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**12**
20250224	8,5	5,8	13,0	<LOD	1,4	7,8	8,6	29,0

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

20250602	3,3	2,5	7,3	<LOD	<LOD	1,7	2,1	9

The limiting concentration and the formation coefficient for each DBP are summarized in Table 7 for THMs and in Table 8 for HAAs. MBAA and MCAA were not modelled because the formation potential was below the limit of detection (LOD). TBAA concentrations were below the LOD in more than 50% of the observations during the evaluated period.

Table 7: parameters used for the THMs formation potential

	TTHMs	**TCM**	**BDCM**	**DBCM**	**TBM**
C_L [ppb]	70	30	20	20	10
K_{DBP}	13	4	10	13	13

Table 8: parameters used for the HAAs formation potential

	THAAs	**BCAA**	**DBAA**	**DCAA**	**TCAA**
C_L [ppb]	30	10	6	15	10
K_{DBP}	2	2	2	1	2

The THMs simulation results are plotted in Figure 17 for EB3, in Figure 18 for Alcanar and in Figure 19 for Masies Catalanes Albiol. At Masies Catalanes Albiol, the simulation underestimates the measured THM concentrations. This section of the network is the most complex to model, as it includes a rechlorination point and is the farthest from the WTP. The assumption that the THM formation coefficient is uniform across the entire network may therefore not be valid, and further analysis using alternative coefficients could lead to a better model fit.

It should therefore be assessed whether the limiting concentration can be assumed to be constant throughout the distribution network and whether the formation coefficient remains unchanged in the presence of rechlorination. In addition, given that limiting concentration is a key parameter in EPANET simulations, further work should evaluate the possibility of assigning it as a variable parameter (depending on raw water quality and seasonal conditions) rather than assuming it to be constant, as done in the present study.

At the other two study locations (EB3 and Alcanar), the simulations show very good agreement with the measured THM concentrations for all compounds, except for DBCM, which is slightly underestimated at Alcanar.

The results from the HAAs simulation are in Figure 20 for EB3, Figure 21 for Alcanar and Figure 22 for Masies Catalanes Albiol.

The HAA models perform worse than the THM models: concentrations are lower (generally below 5µg/L), and PCA results showed that chlorine dependence was only evident for TCAA.

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

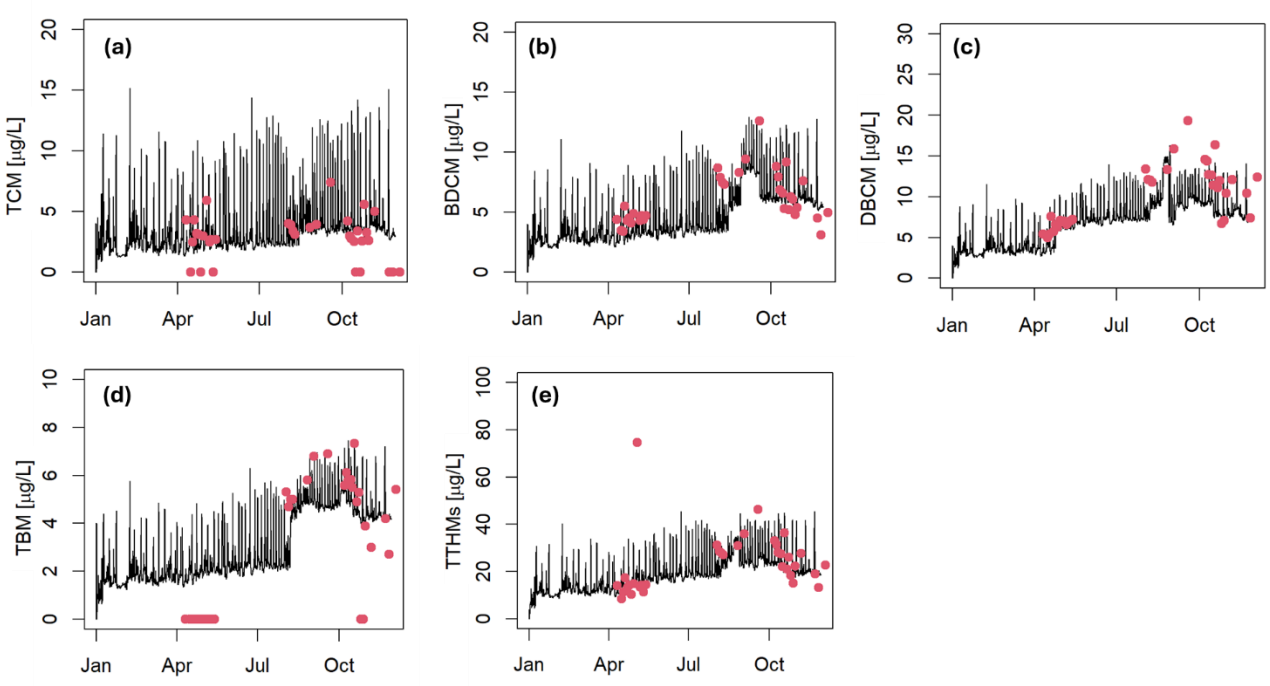

Figure 17: THMs simulated (black line) and measured (red dots) at EB3 (a) TCM, (b) BDCM, (c) DBCM, (d) TBM, and (e) TTHMs. Data from 2025.

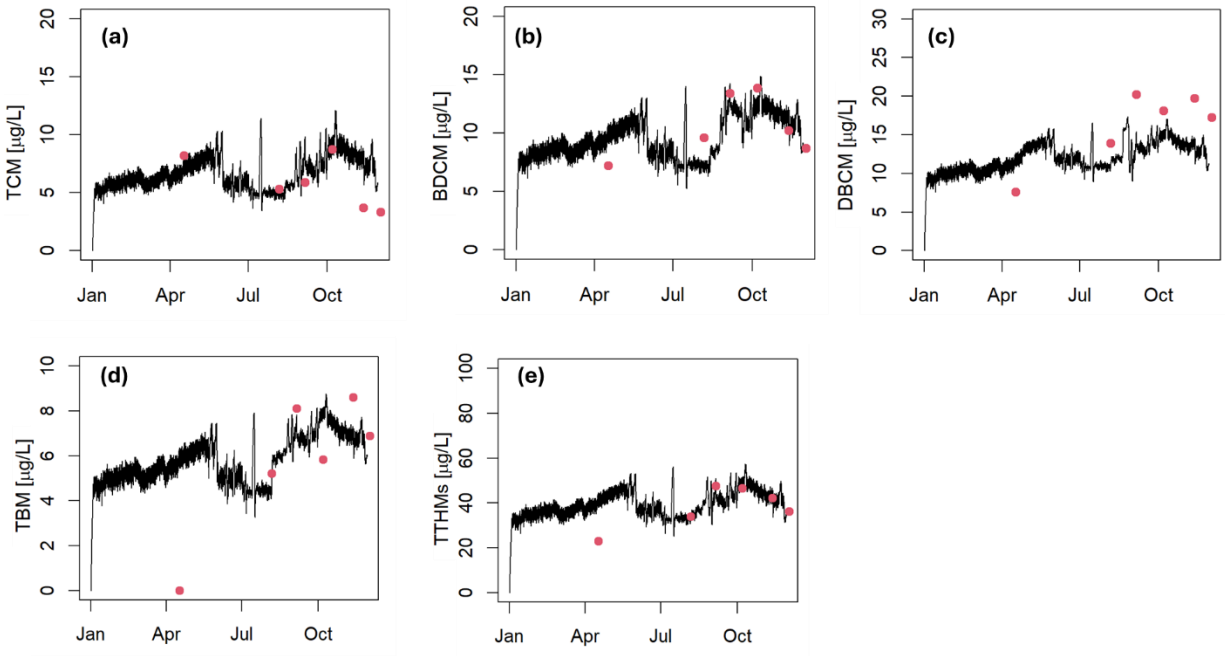

Figure 18: THMs simulated (black line) and measured (red dots) at Alcanar (a) TCM, (b) BDCM, (c) DBCM, (d) TBM, and (e) TTHMs. Data from 2025.

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

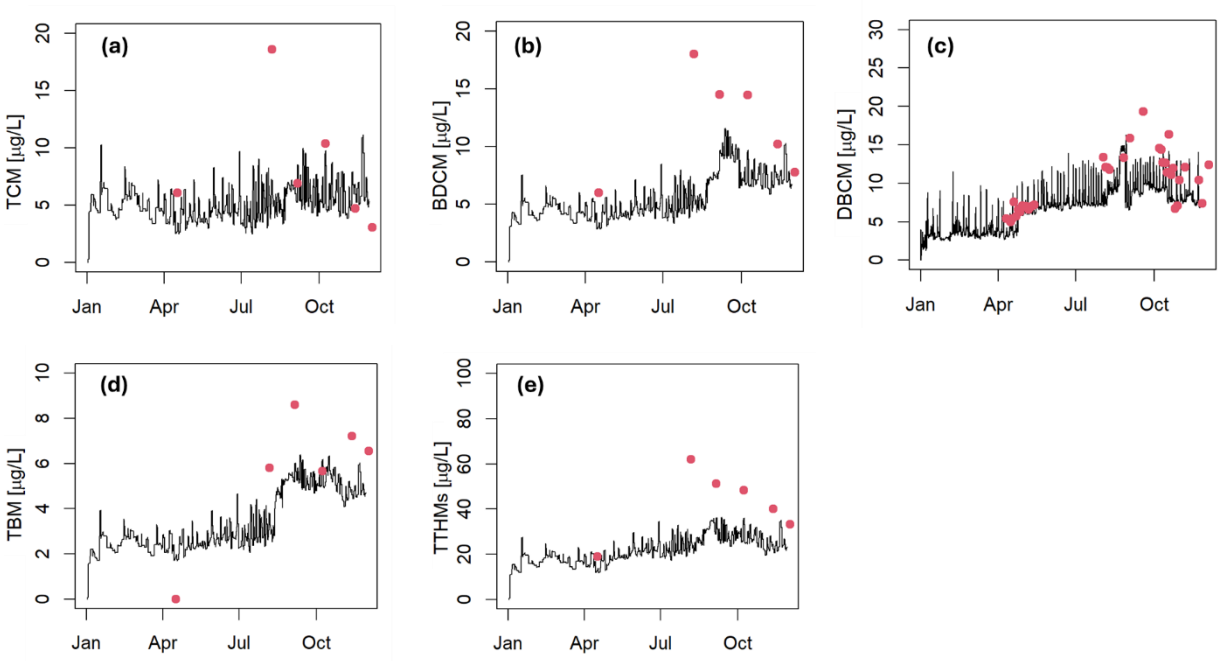

Figure 19: THMs simulated (black line) and measured (red dots) at Masies Catalanes Albiol (a) TCM, (b) BDCM, (c) DBCM, (d) TBM, and (e) TTHMs. Data from 2025.

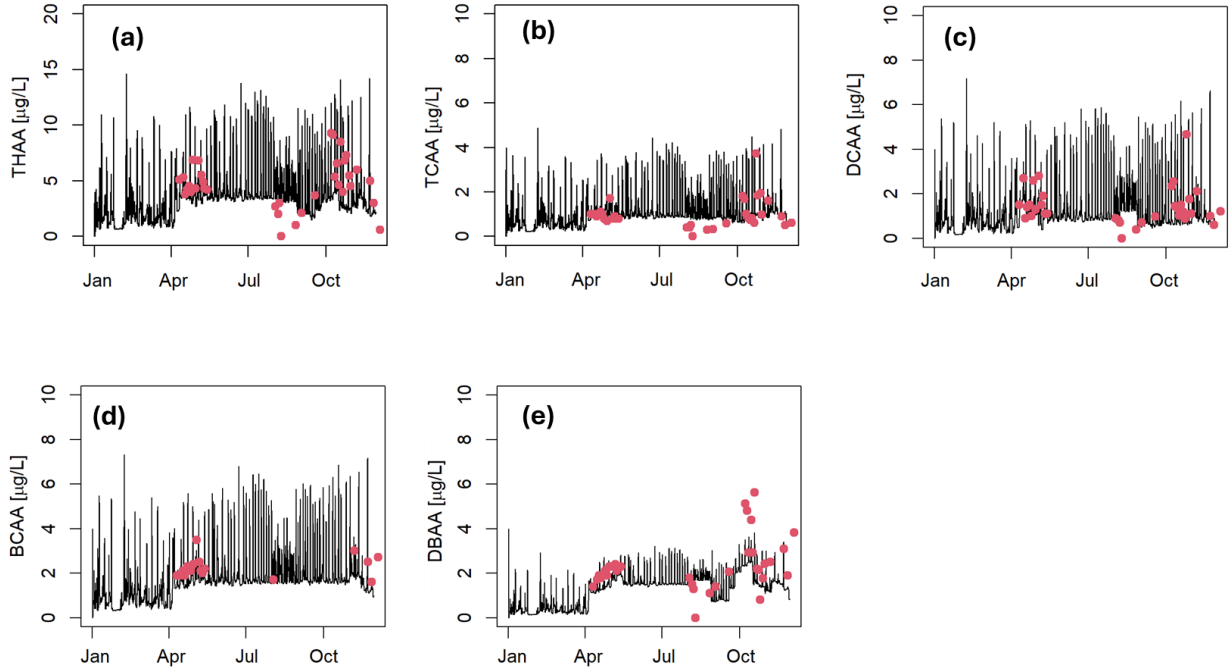

Figure 20: HAAs simulated (black line) and measured (red dots) at EB3 (a) TCM, (b) BDCM, (c) DBCM, (d) TBM, and (e) TTHMs. Data from 2025.

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

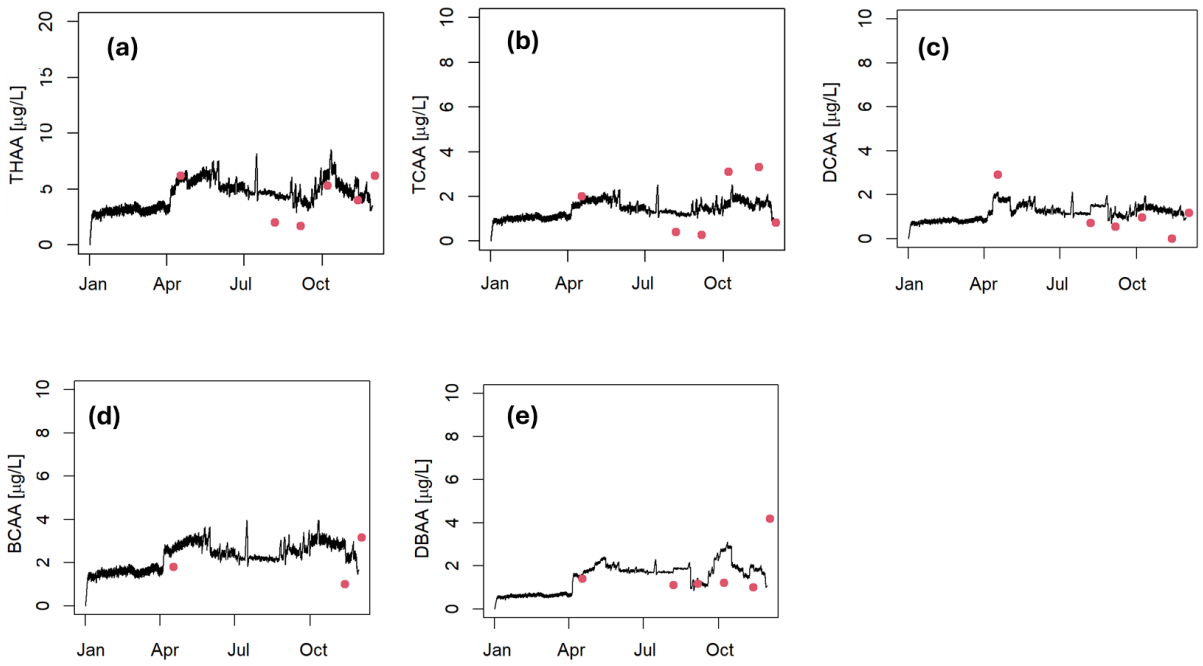

Figure 21: HAAs simulated (black line) and measured (red dots) at Alcanar (a) THAA, (b) TCAA, (c) DCAA, (d) BCAA, and (e) DBAA. Data from 2025.

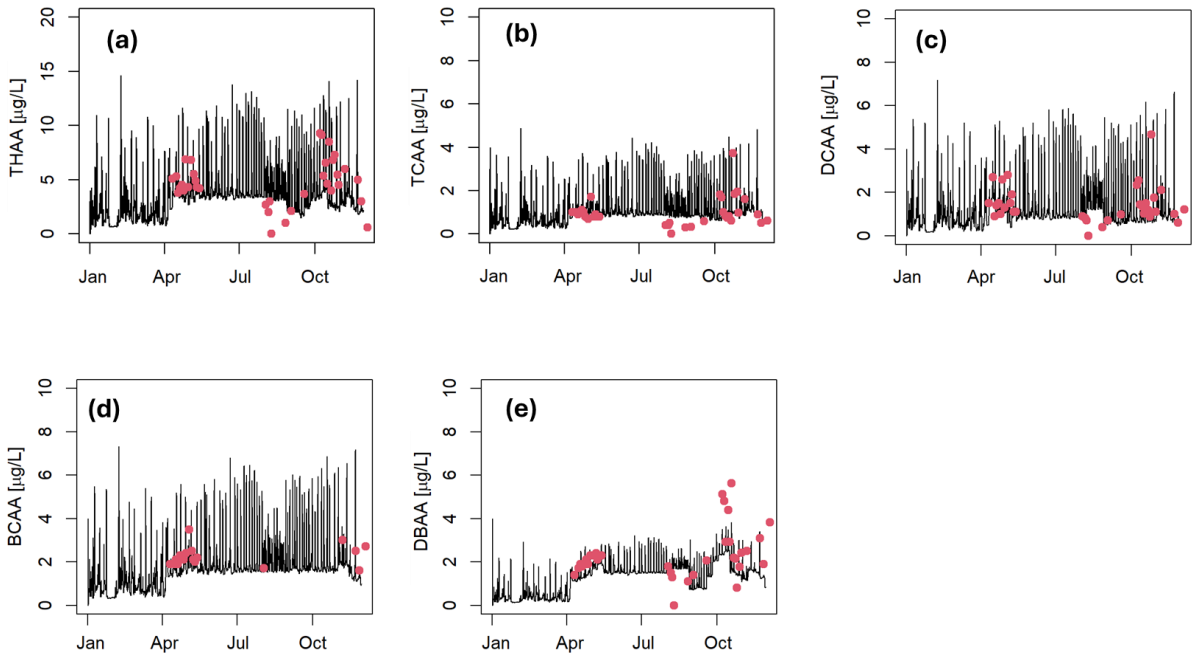

Figure 22: HAAs simulated (black line) and measured (red dots) at Masies Catalanes Albiol (a) TCM, (b) BDCM, (c) DBCM, (d) TBM, and (e) TTHMs. Data from 2025.

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

4.5. Chlorate formation model

4.5.1. Chlorate formation model from hypochlorite 18%

Chlorate formation kinetics was already validated in a previous work (Laura Vinardell Magre, 2026). In the context of this project, a lab-test was carried out for confirmation. Results are presented Figure 23 where the model adequately reproduces the experimental data.

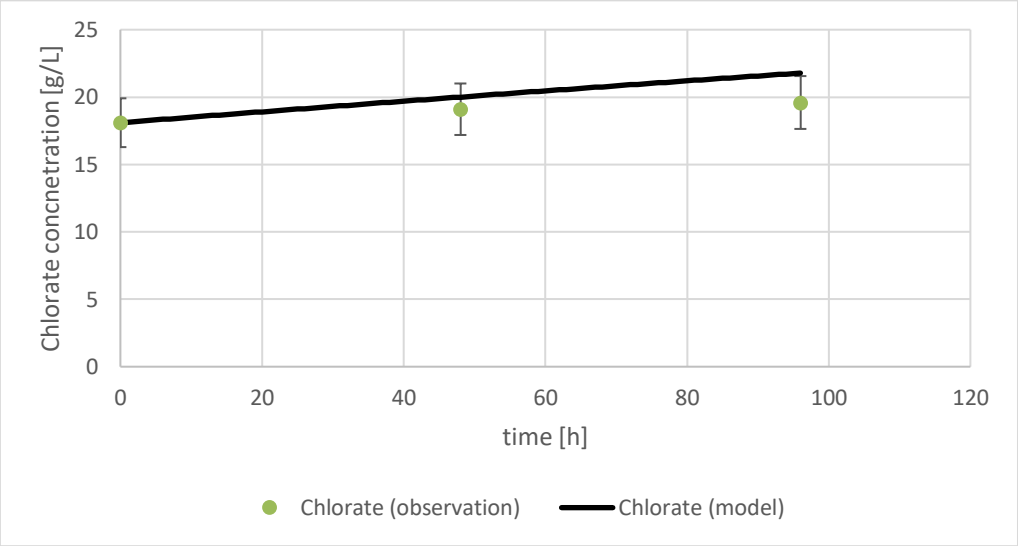

Figure 23: Chlorate formation (from hypochlorite 18%) model validation with experimental data.

4.5.2. Chlorate formation model from hypochlorite 0.8%

Results are presented in Figure 24. It indicates the chlorate concentration inside the hypochlorite tank at different temperatures (8, 21 – triplicate samples- and 30ºC) during two months. The SD was calculated for the triplicate samples (at 21ºC): the maximum divergence was found at time = 1h (17%), but generally the SD was around 5%.

The results indicate that 0.8% hypochlorite also generates chlorate significantly, although with slower kinetics. Chlorate concentration stored during two months at 8ºC increased 50 mg/L more that the initial concentration. In contrast, at temperature 21ºC, chlorate was almost triplicated (from 157 mg/L at the beginning to more than 400 mg/L after 2 months) and five times at temperature 30ºC. Results are plotted in Figure 24.

The kinetics of the degradation of hypochlorite 0.8% were calibrated with experimental data. The expression found is presented in Eq. 16. The model is presented in Figure 24 as lines for each temperature.

$$k_c = 7{,}11 \cdot 10^7 \cdot e^{\frac{-1{,}04 \cdot 10^4}{T}} \ [s^{-1}] \qquad \text{Eq. 16}$$

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

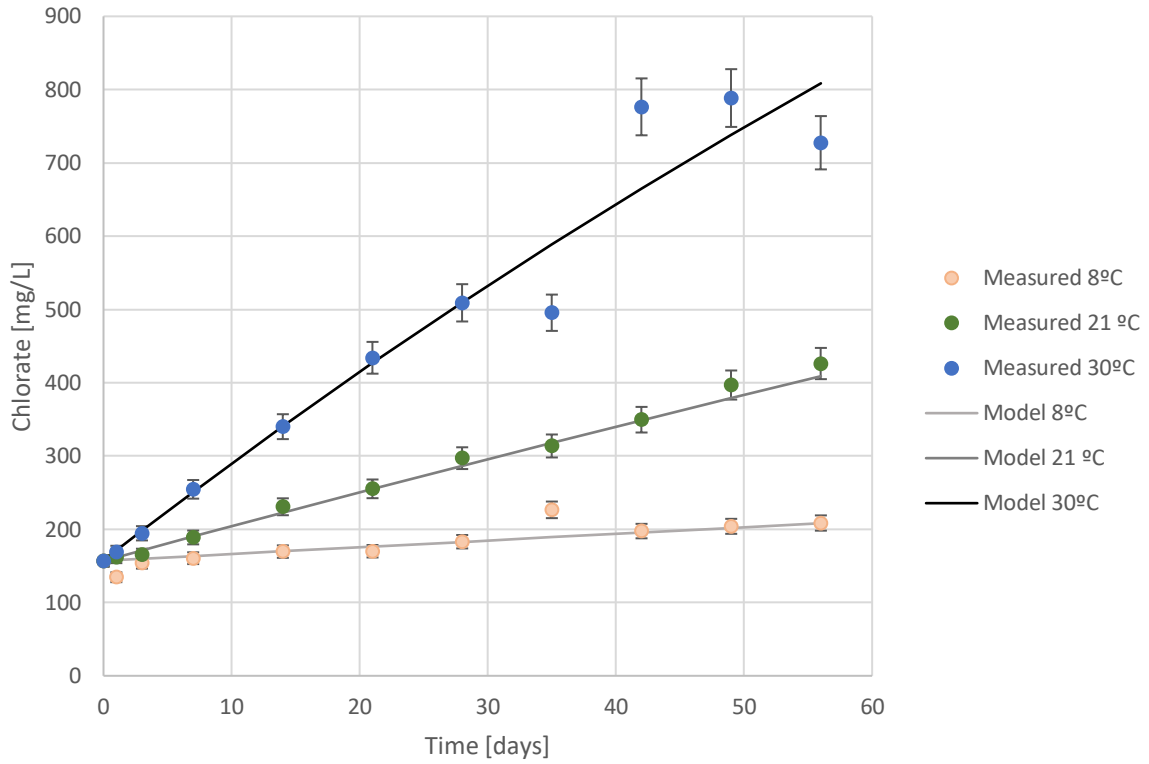

Figure 24: Observed and modelled chlorate from hypochlorite at 0.8% at three different temperatures (8ºC, 21ºC and 30ºC).

Overall, these findings open a potential avenue for further research. The data also suggests that the degradation may still follow second-order kinetics, likely because the solution conditions remain basic. Additional studies would be required to validate the proposed interpretation: namely, that the same kinetic model applies, but with a decay rate approximately 200 times smaller.

4.6. HANs formation model

4.6.1. HANs detected in CS#3

The HANs detected in the CS3 are presented in Figure 25. The concentration of non-regulated HANs appeared to be low (except for DBAA, below 2 ppb), considering that World Health Organization (WHO) suggests a maximum concentration of 20 ppb for DCAN and 70 ppb for DBAN (Organization, 2004). All measured values were approximately one order of magnitude below these guideline levels. TCAN and MIAN could not be detected (LOD=0,01 and 0,11 µ/L respectively).

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

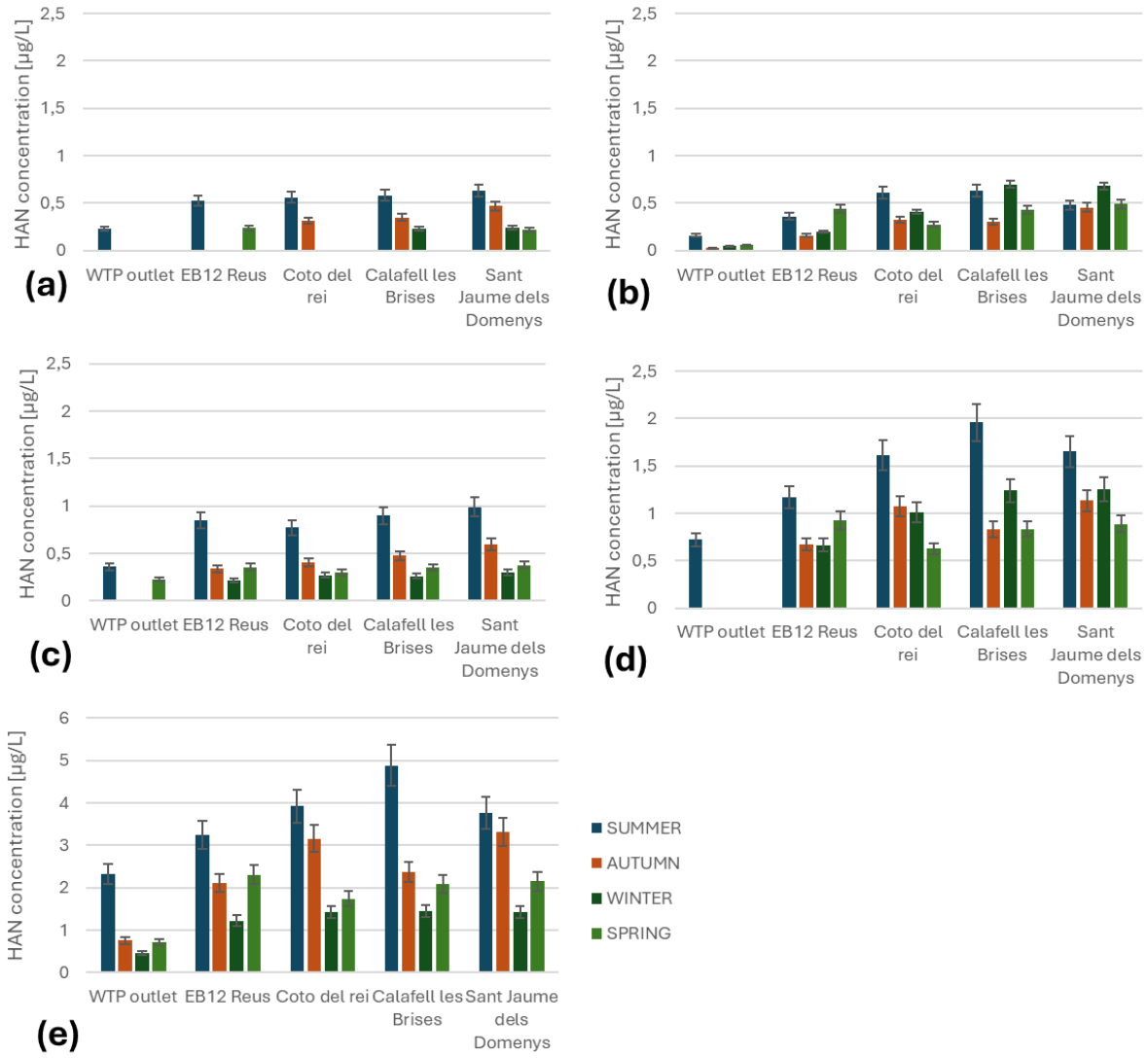

Figure 25: Measured HANs in the CS3 distribution network: (a) MCAA, (b) DCAA, (c) MBAA, (d) BCAA and (e) DBAA.

4.6.2. HANs formation lab-scale experiments

The results from the lab-scale experiments described in section 3.4.1 are plotted in Figure 26. Compounds with two halogens were detected at the highest concentrations, followed by mono-halogenated compounds and, finally, tri-halogenated compounds. This is consistent with the literature, which reports that trihaloacetonitriles (THANs) undergo very rapid hydrolysis and are therefore rarely detected. In addition, mono-haloacetonitriles (MHANs) are described as the most persistent species, which agrees with the observed data.

Figure 26 shows that after two weeks (t = 312 h), the most abundant HANs were observed at 9 °C, probably because the slower reaction kinetics prevented complete hydrolysis. The concentration peak of HANs occurred at temperatures of 21 °C or 9 °C for all compounds, except for MCAN. The peak concentration of the remaining HANs at 30 °C was likely not observed due to faster hydrolysis kinetics.

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

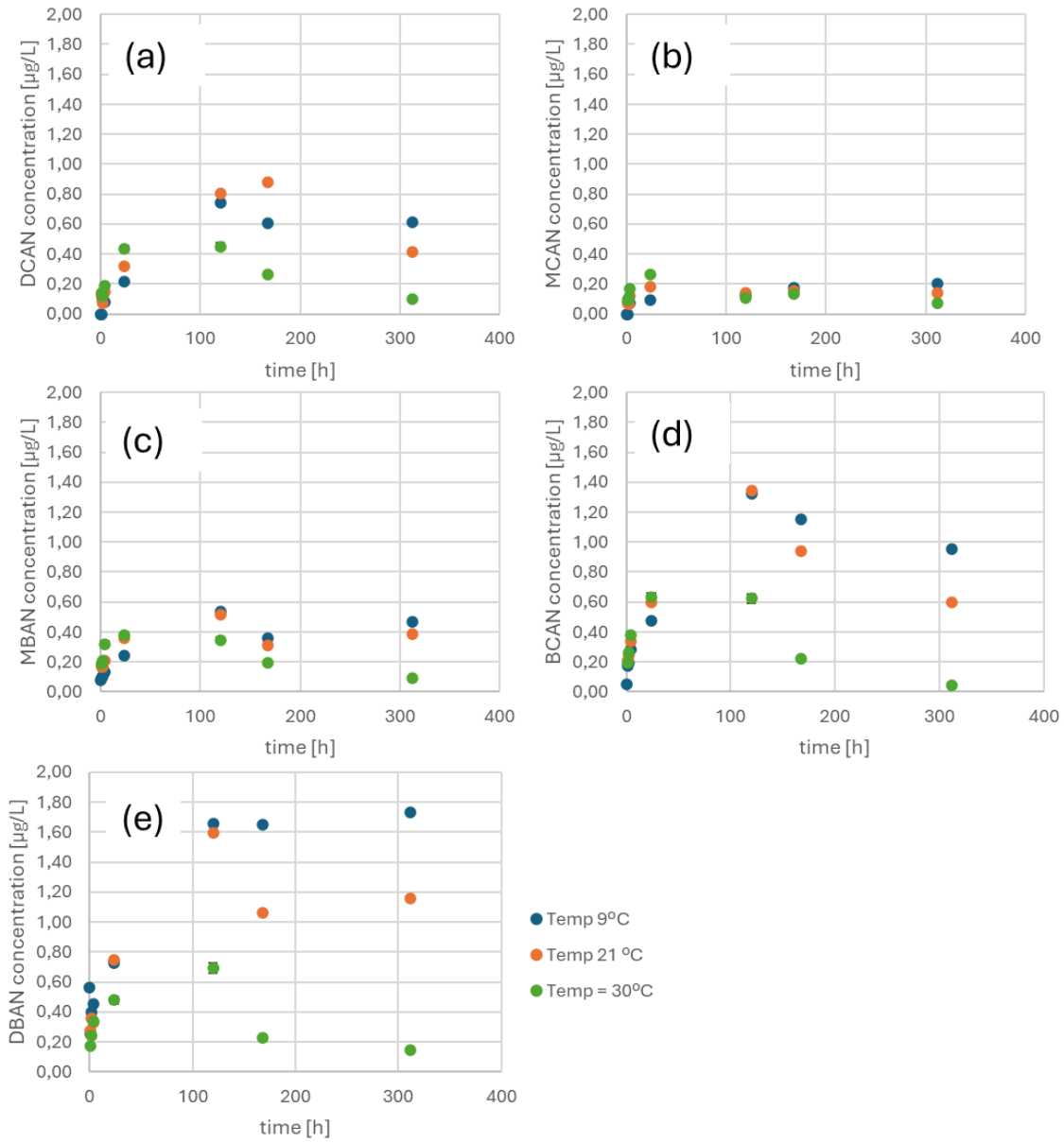

Figure 26: HANs concentration in lab experiments for kinetics calibration. Error bars for 30 ºC data indicate the standard deviation of triplicates.

4.6.3. HANs formation model calibration

The results from the parameter estimation for the calibration of the kinetics of HANs (Eq. 6 and Eq. 7) are summarized in Table 9. Figure 27 shows the fit for 9 and Figure 28 for 30 ºC. Figure 29 shows the fit for 21 ºC.

Some of the estimated parameters resulted in negative values (k_1 for MCAN and DCAN at T = 9ºC; and k_1 for MCAN at T = 21ºC), which do not have a clear physical interpretation under the assumption that the kinetic model is valid. Moreover, for a given compound, certain parameters showed limited or no variation with temperature. These observations suggest that the available dataset may be insufficient to fully capture the temperature dependence of the kinetic behaviour. Consequently, a more

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

comprehensive study including additional data points for each compound and temperature would likely lead to more reliable parameter estimates.

Table 9: Calibrated parameters for HANs kinetics

Temperature	Parameter	MCAN	DCAN	MBAN	BCAN	DBAN
9 ºC	K_1 [h^{-1}]	-6,53E+01	-5,01E-03	1,23E-02	1,76E-02	1,19E-02
	K_2 [$L \cdot mg^{-1} \cdot h^{-1}$]	2,09E+06	6,85E+02	4,00E+00	1,50E+02	3,08E+02
	K_3 [h^{-1}]	6,94E-04	3,35E-06	1,76E-06	4,50E-06	5,26E-06
	K_4 [h^{-1}]	8,21E-04				
21 ºC	K_1 [h^{-1}]	-3,45E+00	6,12E-03	1,23E-02	5,80E-03	2,12E-03
	K_2 [$L \cdot mg^{-1} \cdot h^{-1}$]	2,09E+06	6,85E+02	4,00E+00	1,50E+02	3,08E+02
	K_3 [h^{-1}]	3,12E-03	4,85E-06	1,76E-06	2,75E-06	2,72E-06
	K_4 [h^{-1}]	6,51E-03				
30 ºC	K_1 [h^{-1}]	3,07E+01	6,12E-03	1,68E-02	1,15E-02	9,65E-03
	K_2 [$L \cdot mg^{-1} \cdot h^{-1}$]	2,09E+06	6,85E+02	3,63E+02	1,50E+02	3,08E+02
	K_3 [h^{-1}]	5,02E-03	4,85E-06	3,66E-06	3,34E-06	3,19E-06
	K_4 [h^{-1}]	9,89E-03				

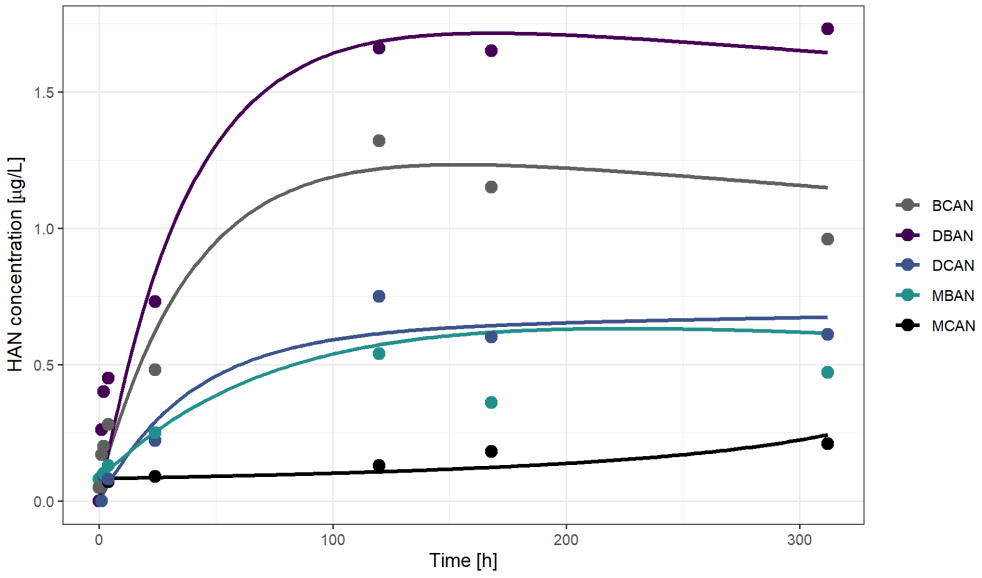

Figure 27: HANs observed (dots) and simulated (lines) at 9ºC

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

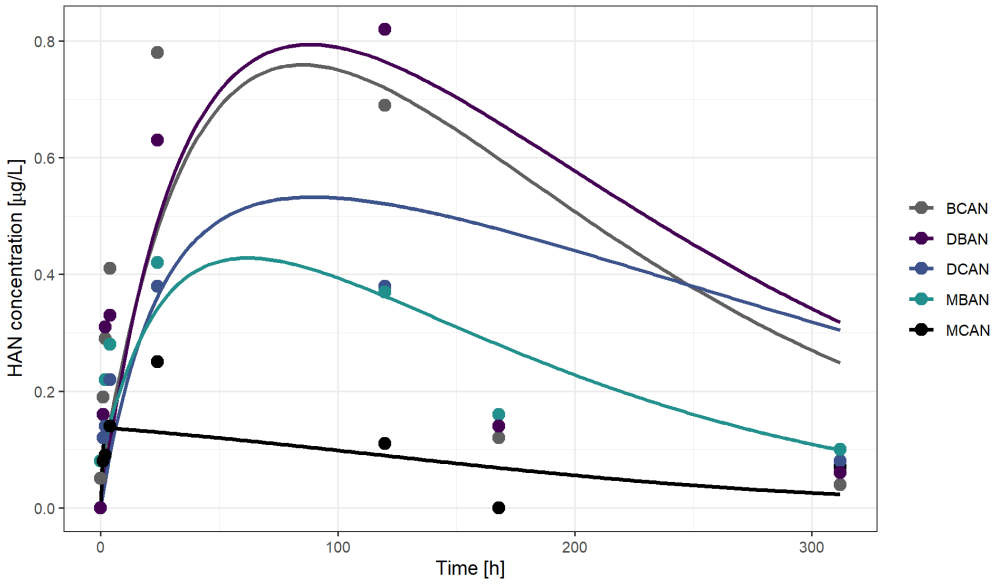

Figure 28: HANs observed (dots) and simulated (lines) at 30ºC

4.6.4. HANs measured in the distribution network

The HANs estimation model was validated by comparison with HAN concentrations measured in the distribution network. Data from the sampling campaign conducted in May 2024 was used and compared with the models developed for T = 21 °C. The temperature measured during the May 2024 campaign was 21.3 °C.

The results are plotted in Figure 29. Overall, the concentrations measured in the distribution network are in reasonable agreement with the model predictions, which were calibrated exclusively using laboratory data. For DCAN, the model shows good agreement with the network measurements, suggesting that it is a suitable candidate for describing DCAN behavior in the distribution system. In contrast, for MBAN and BCAN the model tends to overestimate the concentrations observed in the network, while for DBAN the model underestimates the measured values, failing to fully reproduce the observed concentrations.

It should be noted that although both laboratory and network waters originate from the same site, they were collected in different years (2025 for laboratory experiments and 2024 for the sampling campaign); therefore, differences in water quality and HAN precursor composition were expected. Additionally, residence times at the network sampling points are approximate, whereas residence time in laboratory experiments is a well-controlled and accurately defined parameter, which may further contribute to the observed discrepancies.

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

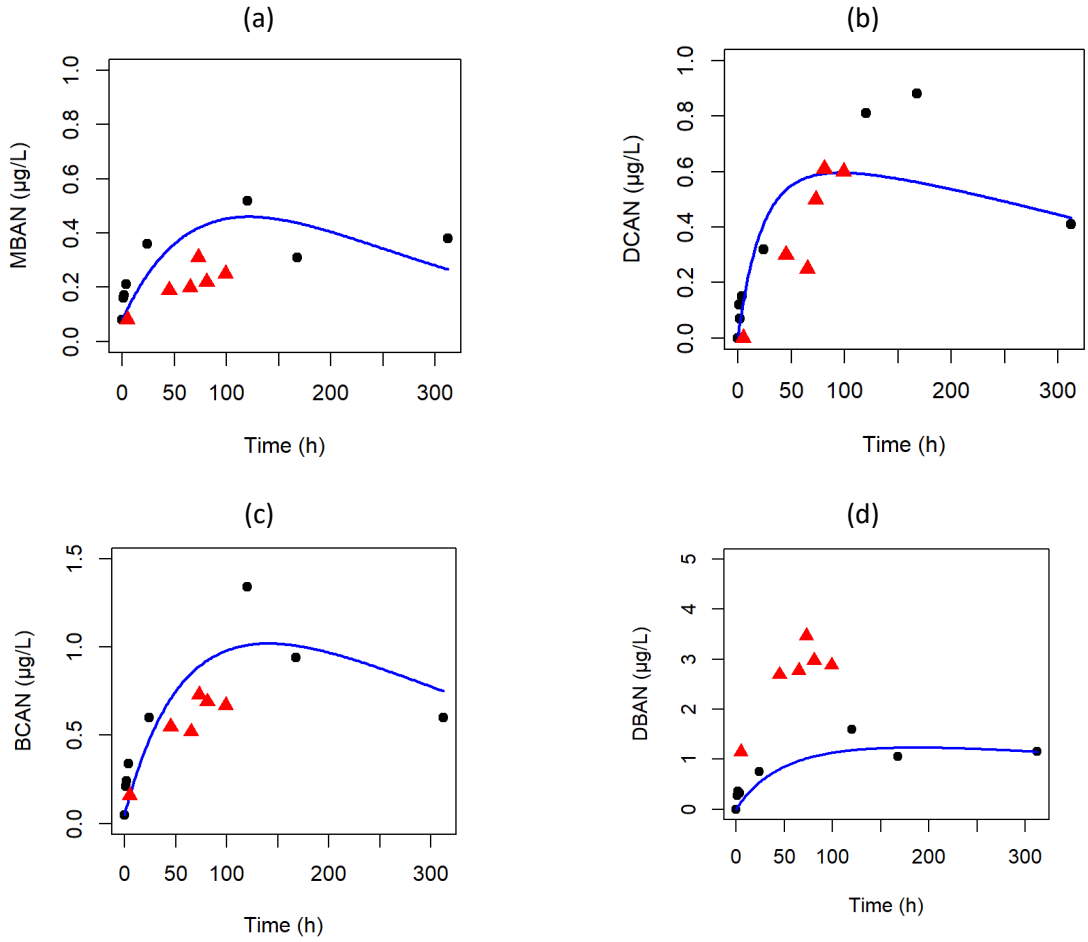

Figure 29: HANs concentration at 21ºC. Experimental results (black dot), simulation (blue line) and observed values in the network (red triangles) for (a) MBAN, (b) DCAN, (c) BCAN and (d) DBAN.

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

5. Conclusions

A mechanistic chlorine decay model was developed based on water residence time and initial chlorine concentration. The model provides a good fit to the data; however, a single set of parameters applicable to the entire distribution network could not be identified. Considering that water utilities typically operate multiple chlorine sensors, a regionalized modelling approach, with different parameter sets for distinct network zones, may be sufficient and operationally feasible.

The chlorine decay model was subsequently used as the basis for the THM formation model, as chlorine concentration was the only variable showing a statistically significant relationship with THM levels. The THM model showed good performance in Regions 1 and 2, while further refinement is required for Region 3, where rechlorination occurs.

For HAAs, only weak or non-significant relationships were found with the variables analyzed, which considerably complicates their modelling and limits the robustness of the proposed approach.

The chlorate formation model available in the literature for 18% hypochlorite solutions was successfully validated. In contrast, a kinetics for chlorate formation from 0.8% hypochlorite solutions was developed. The results indicated that 0.8% hypochlorite does generate chlorate, particularly at higher temperatures.

A kinetic model for HAN formation was also calibrated based on laboratory experiments. The resulting model is consistent with the concentration observed in the distribution network. Although HANs are currently unregulated DBPs, their presence in the network was found to be relatively low (generally below 2µg/L) and accordingly well below the guideline values proposed by the WHO.

The developed models are now ready to be implemented as online simulations in CS#3 for THMs, HAAs, HANs and chlorate prediction. In the future they will be very useful for risk mitigation and the optimal management of the water distribution network.

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

6. References

Adam, L. C., & Gordon, G. (1999). Hypochlorite Ion Decomposition: Effects of Temperature, Ionic Strength, and Chloride Ion. *Inorganic Chemistry*, *38*(6), 1299-1304. https://doi.org/10.1021/ic980020q

Boylard, Michele & Snyder Fair, Patricia. (1992). Occurrence of chlorate in hypochlorate solutions used for drinking water disinfection. *Environmental Science & Technology*, *26*(8), 1663-1665.

Laura Vinardell Magre. (2026). *Full-scale water distribution netwroks modelling for desinfection by-products estimation*. Universitat Politècnica de Catalunya.

Organization, W. H. (2004). *Halogenated Acetonitriles in Drinking-water. Background document for development of WHO Guidelines for Drinking-water Quality* (WHO/SDE/WSH/03.04/98). World Health Organization.

Yu, Y., & Reckhow, D. A. (2015). Kinetic Analysis of Haloacetonitrile Stability in Drinking Waters. *Environmental Science & Technology*, *49*(18), 11028-11036. https://doi.org/10.1021/acs.est.5b02772

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