Environmental Science
Water Research & Technology

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: A. Pedregal-Montes, E. Jennings, R. Marcé and M. J. Farré, *Environ. Sci.: Water Res. Technol.*, 2026, DOI: 10.1039/D6EW00128A.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/es-water

View Article Online
DOI: 10.1039/D6EW00128A

Water Impact Statement: This study shows that the ability to provide safe drinking water is strongly influenced by upstream climate variability, reservoir dynamics, and operational decisions that shape disinfection by-product formation risk. The findings highlight the need to manage source waters and reservoirs as active control points, supporting more resilient and integrated strategies to safeguard water supplies under increasing hydroclimatic uncertainty.

Integrating catchment, climate and reservoir drivers to estimate the risk of THM formation at a drinking water treatment plant inlet

Angela Pedregal-Montes[a,b*], Eleanor Jennings[c], Rafael Marcé[d], Maria José Farré[a]

[a]Catalan Institute for Water Research (ICRA), Carrer Emili Grahit 101, Parc Científic i Tecnològic de la Universitat de Girona, 17003 Girona, Spain

[b]University of Girona, Plaça de Sant Domènec 3, 17004 Girona, Spain

[c]Centre for Freshwater and Environmental Studies, Dundalk Institute of Technology, A91 K584 Dundalk, Ireland

[d]Centre for Advanced Studies of Blanes (CEAB), Spanish National Research Council (CSIC), 17300 Blanes, Spain

*Corresponding author.

Email address: apedregal@icra.cat (A. Pedregal-Montes)

Abstract

Using long-term monitoring and machine learning, this study links upstream hydrometeorology, reservoir processes, and operations to source water conditions relevant to trihalomethane (THM) formation risk at the Mediterranean Ter drinking water treatment plant (DWTP) in Spain, supplied by a three-reservoir cascade (Sau-Susqueda-Pasteral). Based on exploratory analyses, three target variables were selected as indicators of THM formation risk:

View Article Online
DOI: 10.1039/D6EW00128A

dissolved organic carbon (DOC) and water temperature (WT) at the DWTP inlet and fluorescent dissolved organic matter (fDOM) at the Susqueda withdrawal depth. Permutation importance results using Random Forest and LSTM models indicated that withdrawal-layer conditions at Susqueda dominate downstream variability: DOC was most strongly associated with extracted fDOM and other withdrawal water quality variables, whereas inlet WT was primarily controlled by Susqueda withdrawal temperature. For fDOM at Susqueda, reservoir storage volume emerged as a major driver, highlighting the influence of water availability, retention time, and stratification on DOM dynamics. Optimized LSTM models predicted the three target variables with strong validation skill (R^2 and KGE > 0.8). Scenario simulations identified seasonal windows of opportunity for THM risk reduction, with selective withdrawal targeting low-fDOM or cooler layers reducing indicator-based THM formation risk at the DWTP inlet, particularly during warm stratified periods and post-summer rainfall transitions. The effectiveness of this strategy was event-dependent and constrained by reservoir levels and gate accessibility. These results highlight opportunities to reduce DBP formation risk through upstream management, supporting a shift from end-of-pipe control to multi-barrier strategies, particularly in regions facing increasing hydroclimatic stress.

Keywords: source water quality, source water management, disinfection by-products, machine learning, Mediterranean

View Article Online
DOI: 10.1039/D6EW00128A

1. Introduction

Surface water supplies for drinking water are experiencing increasing pressure from various factors, including climate change, escalating water demand, and diverse human activities, with direct consequences for both water availability and raw water quality [1–3]. In chlorinated drinking water treatment plants (DWTP), utilities must balance effective disinfection with the need to limit disinfection by-products (DBPs), which form when disinfectants react with natural organic matter and inorganic constituents (e.g., halides) present in source waters [4]. Due to their potential long-term health effects [5,6], more DBP species are being regulated in drinking water (Directive (EU) 2020/2184). Among DBPs, trihalomethanes (THMs) remain the most widely monitored and well-characterized group, yet because many other (often unregulated) DBPs may pose equal or greater toxicological concern [7], treatment plants are increasingly shifting from reactive compliance responses to proactive management of precursor conditions [8–10].

Operational control therefore often focuses on monitoring source water precursor indicators at the DWTP inlet using dissolved organic matter (DOM) surrogates such as dissolved organic carbon (DOC), ultraviolet absorbance at 254 nm (UV_{254}), or fluorescent DOM (fDOM), together with variables that influence formation kinetics, such as water temperature. These indicators are commonly incorporated into site-specific DBP risk tools to support operational decisions [11]. However, rapid shifts in source water conditions such as temperature spikes, hydrologic events, or precursor pulses can reduce the response time available to operators and compromise water safety [12–14]. A key limitation of many DBP predictive tools is that they are primarily DWTP centered and may not explicitly account for upstream drivers of DOM and water temperature, including catchment forcing, reservoir processes, and water source management [15]. This gap is particularly relevant in Mediterranean regions, where droughts and intense rainfall are projected to intensify, with strong implications for DOM dynamics in rivers and reservoirs [16–18]. Understanding the link between upstream controls and inlet indicators is therefore essential for more anticipatory and climate-adaptive management from source to tap.

Linking upstream forcing to source water conditions at a DWTP inlet requires methods that can represent both catchment driven inputs and managed reservoir transformations at relevant time scales

View Article Online
DOI: 10.1039/D6EW00128A

[19]. This is challenging because DOM dynamics reflect interacting hydrological, physical, and biogeochemical processes [20,21], while catchment DOM monitoring is typically low frequency (often monthly). Process-based hydrological and biogeochemical models can help bridge this gap by providing temporally continuous estimates of catchment DOM and discharge [22], but downstream water quality reaching the DWTP additionally reflects nonlinear interactions among meteorology, reservoir stratification and internal processing, and selective withdrawal operations that are difficult to parameterize explicitly in complex, highly managed reservoir systems [23]. In this context, machine learning (ML) approaches offer a practical alternative for forecasting operationally relevant source water indicators, as they can learn empirical, potentially lagged and nonlinear relationships directly from multi-source monitoring and operational datasets without requiring explicit representation of all underlying processes [24].

Previous work at the Ter DWTP showed that DOC and raw water temperature measured at the plant inlet can be used as practical indicators to classify THM formation risk using an empirical risk matrix, as raw waters are predominantly influenced by DOM and typically show low contributions from inorganic precursors (e.g., bromide) [25]. This framework reflects pre-treatment THM formation risk, defined as the potential for DBP formation during subsequent disinfection based on source water conditions. Building on this approach, this study examines the upstream controls, predictability, and operational leverage points governing these indicators in the Mediterranean Ter river-reservoir-DWTP continuum supplying the Barcelona metropolitan area. We compiled and analyzed an extensive spatio-temporal dataset spanning catchment forcing, reservoir water quality and stratification, reservoir storage and gate operations, and DWTP inlet monitoring. The dataset was complemented by daily upstream DOC and discharge simulations from a previously validated catchment model to represent inflow variability at appropriate temporal resolution. We first evaluated relationships among DOM proxies and THM observations to support the use of DOC as a consistent indicator, then assessed longitudinal DOC patterns to identify the most influential upstream control points along the continuum. Subsequently, we applied ML-based driver attribution and prediction to quantify the dominant hydrometeorological, water quality, and operational drivers of DOC and water temperature at the DWTP inlet, and finally

tested alternative selective-withdrawal strategies to assess how operational choices could shift indicator-based THM risk classes. By focusing on source water indicators and their upstream drivers, the study provides a basis for more anticipatory, climate-adaptive source water management that can support THM risk mitigation without attempting to directly model DBP formation within the treatment process.

2. Methods

2.1. Study area and river-reservoir-DWTP continuum

This study focuses on the Ter DWTP, located in Cardedeu, Catalonia (northeastern Spain), operated by the *Ens d'Abastament d'Aigua Ter-Llobregat* (ATL), which supplies drinking water to the Barcelona metropolitan area (≈4.5 million inhabitants). The DWTP sources raw water from the Ter River Basin, which drains into a cascade of three reservoirs, Sau, Susqueda and Pasteral, with storage capacities of 166 hm^3, 233 hm^3, and 1.5 hm^3, respectively (Figure 1). A detailed map of the Ter River catchment, including elevation, river network, and reservoirs, is provided in Figure S1.

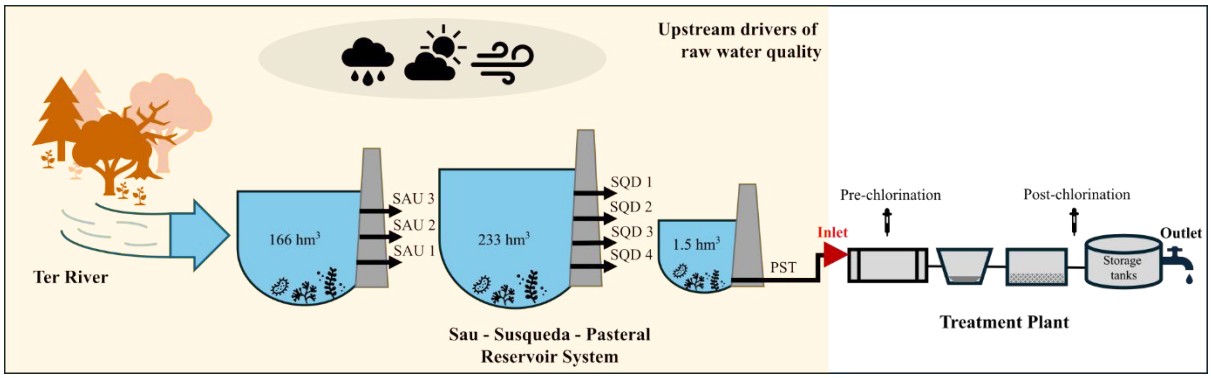


Figure 1. Schematic of the Ter River-reservoir-DWTP continuum showing upstream drivers, reservoir storage and selective withdrawal at Sau and Susqueda (SQD), bottom withdrawal at Pasteral (PST), and conventional treatment with pre-chlorination (chlorine dioxide and sodium hypochlorite) and post-chlorination (sodium hypochlorite) at the DWTP.

Reservoir water levels and releases are managed by regional water authorities to accommodate multiple uses, including hydroelectric generation, ecological flow maintenance, and recreational activities. In contrast, water quality management is primarily conducted by ATL through selective withdrawal

View Article Online
DOI: 10.1039/D6EW00128A

operations at the Sau and Susqueda reservoirs, which are equipped with three and four intake levels, respectively (Figure 1; Figures S2-S3). These intake structures allow operators to select withdrawal depth based on water quality conditions; however, the set of available intake levels varies with reservoir water level, which determines the accessibility of individual gates. Water is subsequently withdrawn from the bottom outlet of the Pasteral reservoir, where discharge rates can be regulated, and transported to the DWTP via a pipeline, with an approximate travel time of 12 hours to the DWTP inlet. Additional characteristics of the reservoirs are provided in Table S1. Although the three-reservoir configuration provides substantial buffering of hydrological variability and raw water quality, the system remains sensitive to extreme meteorological conditions that can alter reservoir stratification and organic matter dynamics, thereby influencing raw water quality at the DWTP inlet [26].

The Ter DWTP employs a conventional treatment train that includes both pre-chlorination (combined dosing of chlorine dioxide (ClO_2) and sodium hypochlorite (NaClO) to increase oxidation and disinfection capacity) and post-chlorination (NaClO). Pre-chlorination is the stage most susceptible to DBP formation, as disinfectants are applied when organic precursor concentrations in raw water are highest. Consequently, operational control focuses on limiting DBP formation by adjusting disinfectant dosage and treatment conditions in response to raw water characteristics. Previous studies have shown that, at the Ter DWTP, raw water DOM concentration and water temperature are particularly relevant factors influencing THM formation risk, as they control precursor availability and reaction kinetics under local treatment conditions [25]. Due to the spatial extent of the supply network and variability in water demand, hydraulic retention times (HRT) within the distribution system range from several hours to multiple days. Therefore, to ensure compliance with drinking water regulations, which establish a maximum allowable concentration of 100 μg L^{-1} for total THMs at the consumer tap, ATL applies a more conservative internal operational limit of 50 μg L^{-1} for total THMs at the DWTP outlet.

2.2. Study workflow

This study followed a structured, multi-step workflow to investigate the upstream controls, predictability, and operational sensitivity of water quality variables associated with THM formation risk at the Ter DWTP inlet.

First, exploratory analyses were conducted using long-term monitoring data to characterize the temporal variability of organic matter and water temperature at the DWTP inlet, their relationships with THM concentrations at the outlet, and their connection to upstream hydrometeorological conditions along the river-reservoir-DWTP continuum. These analyses were used to support the selection of candidate water quality indicators and to inform predictor preselection for subsequent modeling. The exploratory analysis leveraged the full monitoring record (2015-2023) to capture long-term variability and hydroclimatic extremes, whereas subsequent ML modeling was constrained to a shorter period defined by the availability of high-frequency reservoir data required to represent key upstream drivers.

Second, ML models were implemented to quantify the relative importance of upstream drivers influencing the selected indicators. Random Forest (RF) and Long Short-Term Memory (LSTM) models were trained using preselected hydrometeorological, operational, and water quality predictors, and permutation importance (PI) was applied to attribute the contribution of individual drivers. Third, LSTM models were refined for each indicator by reducing predictor sets based on driver attribution results and predictive performance. The optimized LSTM models were then used to evaluate predictability and temporal dynamics under observed conditions.

Finally, the calibrated LSTM models were applied to simulate alternative reservoir operation scenarios designed to reduce THM formation risk at the DWTP inlet. Model outputs under baseline and scenario conditions were subsequently translated into THM formation risk classes using published empirical relationships.

2.3. Data collection and preprocessing

Data were collected from multiple locations along the river-reservoir-DWTP continuum to characterize meteorological, hydrological, operational, and water quality variability. Table 1 summarizes the datasets used in the study and their main characteristics.

Meteorological forcing was characterized using daily air temperature, total precipitation, and solar radiation obtained from the ERA5 reanalysis product of the European Centre for Medium-Range Weather Forecasts [27]. ERA5 provides global atmospheric data at a spatial resolution of 0.25°. Data were

View Article Online
DOI: 10.1039/D6EW00128A

extracted for the grid cell encompassing the Ter reservoir system and used to represent atmospheric drivers influencing catchment processes, reservoir stratification, and water temperature dynamics.

Hydrological inputs from the Ter River were represented using daily discharge and DOC concentrations simulated by the Precipitation, Evapotranspiration, and Runoff Simulator for Solute Transport (PERSiST; Futter et al., 2014) coupled with the Integrated Catchments Model for Carbon (INCA-C, (INCA-C; Futter et al., 2007), previously validated for the Ter basin[30]. These simulations were used to represent upstream catchment inputs and to provide daily values for ML analysis. In addition, observed DOC concentrations were available at monthly resolution and were used exclusively for exploratory, continuum-scale analyses.

Table 1. Overview of datasets used in the study, including location, variables, temporal resolution, data source, and application in the analysis.

Location	Variable	Temporal resolution	Source	Use in study
Meteorology	Air temperature, total precipitation, solar radiation	Daily	ERA5 reanalysis	Exploratory and ML
Ter river	DOC	Monthly	Monitoring records	Exploratory
	Discharge (simulated)	Daily	Process-based catchment model	Exploratory and ML
	DOC (simulated)	Daily	Process-based catchment model	ML
Reservoirs	fDOM, WT, DO, turbidity, Chl-a	Daily (aggregated from 2-min profiler data)	In situ profilers	ML
	DOC	Monthly	Monitoring records	Exploratory
	Gate operation	Event-based (gate changes)	Reservoir operational records	Exploratory and ML
	Stored volumes	Daily	Reservoir operational records	ML
DWTP inlet	DOC, UV254, SUVA, WT	Daily	DWTP records	Exploratory and ML
DWTP outlet	Total THM concentration	Weekly	DWTP records	Exploratory

Note: Exploratory analyses used the full monitoring period (1 January 2015-31 December 2023). Machine learning (ML) models were trained and evaluated over the period constrained by reservoir profiler availability (4 February 2017-30 November 2020). Abbreviations: DWTP, drinking water treatment plant; DOC, dissolved organic carbon; fDOM, fluorescent dissolved organic matter; WT, water temperature; UV254, ultraviolet absorbance at 254 nm; SUVA, specific ultraviolet absorbance; DO, dissolved oxygen; Chl-a, chlorophyll-a; THM, trihalomethane.

Reservoir dynamics were characterized using a combination of monitoring records, high-frequency profiling data, and operational information. Monthly DOC concentrations at the surface and at extracted depths were available for all reservoirs and were used for exploratory analyses. For Sau and Susqueda reservoirs, high-frequency water quality data (2-minute resolution) from profiling buoys installed near the dams were aggregated to daily resolution at the surface (0-5m) and at extraction depths (2.5 m above and below the active gate). At each site, a YSI EXO2 multiprobe recorded turbidity (NTU), chlorophyll-a (Chl-a, mg L^{-1}), dissolved oxygen (DO, mg L^{-1}) water temperature (ºC) and fDOM (QSU) throughout the water column. EXO fDOM can be used as a surrogate for colored DOM (CDOM) (excitation 365 ± 5 nm; emission 480 ± 40 nm), and raw fDOM data were water temperature corrected following [31]. Additionally, reservoir operational data, including daily stored volume and gate operation, were used to represent storage dynamics and selective withdrawal. No profiler data were available for the Pasteral reservoir due to its small volume and short HRT (~1 day) relative to Sau and Susqueda. Profiler data availability differed between reservoirs. Susqueda profiler data were available from 4 February 2017 to 30 November 2020, whereas Sau profiler data were available from 4 February 2017 to 1 March 2020, after which the instrument failed due to damage sustained during Storm Gloria (19-24 January 2020; Berdalet et al., 2020). Historical time series of selected reservoir variables are available in the Supplementary Material (Figure S2-S3).

At the DWTP inlet, historical datasets of raw water quality variables were available daily, including DOC, UV254, specific ultraviolet absorbance (SUVA), and water temperature. At the DWTP outlet, total THM concentrations were measured at weekly resolution and used exclusively for exploratory analyses. Further methodological details for these variables are provided in the Supplementary Material (Text S1).

Exploratory analyses were conducted using the full available monitoring period (1 January 2015 – 31 December 2023). ML models were trained and evaluated over 4 February 2017 to 30 November 2020, with some gaps, defined by Susqueda profiler availability to retain the Storm Gloria period in the dataset; missing Sau profiler periods within this window were treated as extended gaps in the corresponding predictors. This period was selected to ensure the inclusion of high-frequency reservoir

View Article Online
DOI: 10.1039/D6EW00128A

variables required to represent withdrawal-depth water quality and operational dynamics. While it captured substantial hydroclimatic variability, including extreme events, it did not fully encompass the longer-term variability observed in the full monitoring record (2015-2023). Therefore, model results should be interpreted within the range of conditions represented during the model training and evaluation period. All datasets were subjected to quality control and harmonized prior to analysis. Variables used in ML were aligned on a common daily time step; interpolation was applied only to profiler-derived variables to support daily alignment, while all other predictors were already available at daily resolution. Detailed preprocessing procedures are provided in Text S1.

2.4. Exploratory analysis to support variable selection

At the DWTP inlet, DOM was characterized using multiple proxies to represent both organic carbon quantity and quality. DOC was used as a measure of concentration, while UV254 and SUVA were included to capture variations in DOM optical properties. Pearson correlation analysis was applied to assess relationships among DOM proxies, water temperature, and total THM concentrations measured at the DWTP outlet. Hydrometeorological variables, including air temperature, precipitation, river discharge, and reservoir storage volume, were analyzed in parallel with raw water quality variables to provide context for upstream influences on temporal variability at the DWTP inlet.

To evaluate spatial patterns in organic matter variability along the continuum, DOC was used as the sole DOM proxy, as it was the only organic matter variable consistently available across all locations. DOC concentrations were compared along the river-reservoir-DWTP continuum using seasonal summaries and correlation analysis; for the Sau and Susqueda reservoirs, DOC values at both surface and extracted depths were considered.

2.5. Machine learning framework

Machine learning models were used to simulate selected water quality variables relevant to THM formation risk assessment and to investigate their upstream drivers. Data-driven approaches are well suited to highly managed aquatic systems, as they can exploit high-frequency monitoring and long-term datasets to learn nonlinear relationships under varying environmental and operational conditions [24,33].

Two ML approaches were implemented in this study: RF and LSTM neural networks. These models were selected to provide complementary perspectives on driver attribution rather than to perform exhaustive model benchmarking. RF was used as a robust and interpretable baseline method, widely applied in environmental prediction tasks due to its ability to capture nonlinear relationships and handle correlated predictors. In contrast, LSTM networks were selected to represent state-of-the-art sequence modeling approaches, capable of learning temporal dependencies and lagged relationships in time series data.

RF is a non-parametric ensemble method that constructs multiple decision trees using bootstrap samples and aggregates their predictions, and has demonstrated strong predictive performance across a range of environmental forecasting applications [34–37]. LSTM networks were employed to represent temporal dependencies and lagged relationships in time-series data, and have shown strong performance in recent hydrological and water quality forecasting studies [37–41].

Driver attribution analyses were performed using both RF and LSTM models, while LSTM models were used for forecasting and scenario simulations. Predictor sets were tailored to each modeled variable to reduce redundancy and model complexity. All predictors were aligned to a common daily time step and harmonized prior to modeling (see Text S1). LSTM inputs were constructed using a fixed 14-day lookback window. Different window lengths were tested, and a 14-day window yielded the best predictive performance across all target variables. This window length was consistent with the short- to medium-term dynamics of the system. Models were trained and evaluated using a chronological split, with the first 80% of the time series used for training and the remaining 20% reserved as a held-out test period. Model performance was assessed using the coefficient of determination (R^2), root mean squared error (RMSE), mean absolute error (MAE), and Kling-Gupta efficiency (KGE). Reproducibility was ensured by fixing random seeds and enforcing deterministic settings. Additional details on model configuration, hyperparameter selection, and implementation settings are provided in Text S2.

2.6. Driver attribution using permutation importance

View Article Online
DOI: 10.1039/D6EW00128A

Candidate predictors were preselected for each modeled variable based on exploratory analyses, data availability, and system-specific process understanding, prior to driver attribution. All candidate predictors considered are listed in Table 1; the preselection step retained variables most directly connected to the target locations and dominant processes identified along the continuum.

Predictor relevance was quantified using PI, which evaluates the contribution of each predictor by randomly permuting its values and quantifying the resulting change in model performance [34]. This approach avoids biases associated with split-based importance measures in the presence of correlated predictors [42]. Importance was expressed as the percentage change in the coefficient of determination relative to the unpermuted model (ΔR^2, %).

PI was computed on the chronologically held-out validation period for both RF and LSTM models. For RF, each predictor was permuted repeatedly and ΔR^2 values were averaged across repetitions. For LSTM, predictors were permuted repeatedly across samples in the test set while keeping the remaining predictors unchanged; permutations were applied at the sequence level to preserve within-sequence temporal structure of non-permuted inputs. Attribution results from RF and LSTM were compared to assess the robustness of driver rankings across static and sequential learning frameworks.

Positive ΔR^2 values indicate a loss of predictive skill when a predictor is permuted and therefore denote an important contributor to model performance, whereas values near zero or negative indicate negligible or unstable contributions that may arise from sampling variability and collinearity [43]. Predictors were interpreted based on relative ranking and consistency across repetitions, consistent with recommendations for correlated predictors in predictive models [43]. As PI can be sensitive to multicollinearity among predictors, importance values may be distributed across correlated variables. Therefore, results were interpreted primarily in terms of relative ranking rather than absolute importance values.

2.7. Scenario simulations and THM formation risk estimation

View Article Online
DOI: 10.1039/D6EW00128A

Optimized LSTM models were used to simulate alternative reservoir operation scenarios designed to reduce THM formation risk at the DWTP inlet. Scenario simulations were conducted exclusively with LSTM models, given their ability to represent temporal dependencies in predictor time series.

Two operational scenarios were defined at the Susqueda reservoir, the last major operational control point upstream of the DWTP, to represent plausible selective-withdrawal strategies designed to influence THM risk indicators (DOC and water temperature) at the DWTP inlet: (i) extraction of water layers characterized by lower organic matter values, and (ii) extraction of the coolest available water layer. These scenarios were implemented by modifying the input time series of reservoir extraction depth and associated profiler-derived water quality variables while keeping all other predictors identical to baseline conditions. Scenario definition was constrained to physically accessible withdrawal layers based on observed profiler data and gate availability under prevailing reservoir water levels. Figure S3 illustrates the Susqueda reservoir stratification, water quality, water levels, and gate operation data used to define the scenarios.

Scenario predictions were computed only for dates where the required scenario inputs were available for the preceding 14 days (LSTM lookback window); therefore, gaps in scenario trajectories reflect incomplete profiler input data rather than model instability. For each scenario, the optimized LSTM models were used to simulate DOC and water temperature at the DWTP inlet for all eligible dates. Baseline simulations corresponding to observed operational conditions were also produced for comparison. Simulated DOC and water temperature time series were subsequently translated into THM formation risk classes using empirical, expert-based relationships developed previously for the Ter DWTP [25]. In this framework, DOC and water temperature were first classified into discrete levels based on predefined concentration and temperature ranges (Table S2), and combined THM formation risk classes were then assigned using a rule-based matrix linking these categories (Table S3). These risk classes represent pre-treatment THM formation risk at the DWTP inlet, defined as the potential for DBP formation during subsequent disinfection. Changes in THM formation risk under alternative scenarios were assessed by comparing simulated risk classes against baseline conditions over the simulated period.

View Article Online
DOI: 10.1039/D6EW00128A

3. Results and discussion

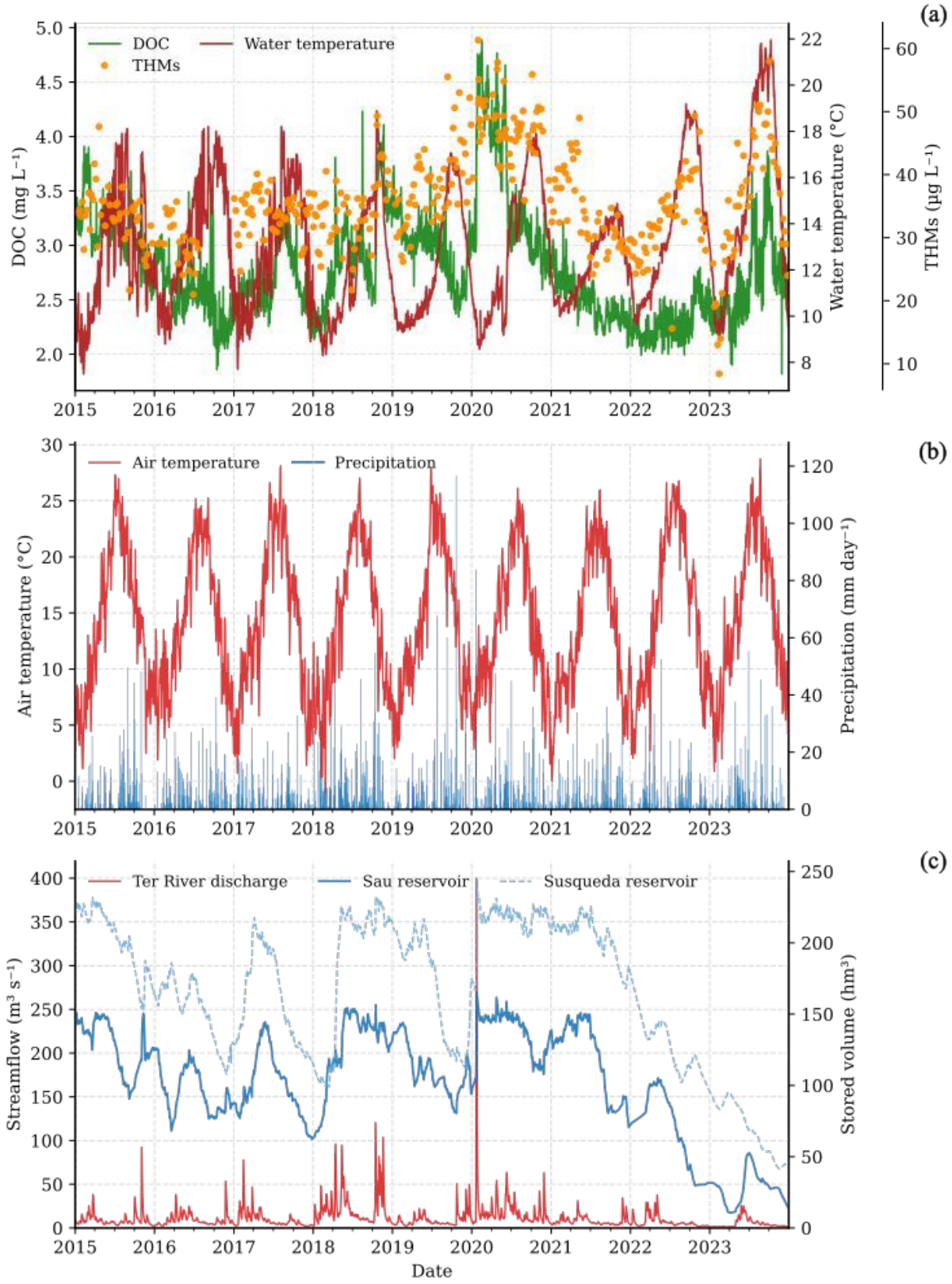
3.1. Relationships between DWTP inlet conditions and THM concentrations under hydro-meteorological variability

The three DOM proxies measured at the DWTP inlet, DOC, UV254 and SUVA, were strongly correlated with each other (r = 0.63-0.94; Figures S4-S5), indicating that they conveyed consistent information on DOM dynamics over the study period. Given this strong covariation, and because DOC was the only DOM proxy available consistently along the river-reservoir-DWTP continuum, subsequent analyses focused on DOC as the primary indicator of DOM variability at the DWTP inlet.

The temporal evolution of raw water DOC and water temperature exhibited seasonal variability, while departures from typical cycles were apparent during periods of hydro-meteorological extremes (Figure 2a). To place these patterns in context, atmospheric forcing (air temperature and precipitation) and hydrological and operational conditions are shown in Figures 2b-c. Total THM concentrations measured at the DWTP outlet displayed marked temporal variability (Figure 2a) and tended to be higher during periods of elevated DOC and/or higher raw water temperature, although the correspondence was not systematic. Pearson correlation analysis indicated positive but moderate associations between THMs and DOC (r = 0.39) and between THMs and water temperature (r = 0.25). While these correlations alone do not imply causality, they indicate that DOC and water temperature capture part of the variability associated with THM formation. In this study, these variables were used as operational indicators of pre-treatment THM formation risk at the Ter DWTP inlet, rather than direct predictors of THM concentrations.

Several periods highlighted the influence of extreme conditions on inlet water quality and potential THM formation. During Storm Gloria in early 2020, DOC increased abruptly (Figure 2a), concurrent with intense precipitation and hydrological disturbance (Figure 2b-c), likely reflecting enhanced mobilization and transport of catchment-derived organic matter [44]. In contrast, the prolonged drought from 2021 onwards coincided with elevated water temperatures and altered DOC dynamics (Figure 2a),

with DOC increases often occurring after rainfall events after extended dry periods, a response commonly reported in Mediterranean catchments [45].



View Article Online
DOI: 10.1039/D6EW00128A

Figure 2. Time series for the period January 2015-December 2023 showing (a) daily DOC concentrations and water temperature at the DWTP inlet together with weekly total THM concentrations at the DWTP outlet; (b) daily mean air temperature and total precipitation; and (c) daily Ter River discharge and stored volumes of the Sau and Susqueda reservoirs.

Finally, higher THM concentrations observed during the later part of the record (2022-2023; Figure 2a) suggested that THM formation risk during prolonged droughts was influenced not only by upstream hydrological conditions but also by source water management decisions at the system scale. During periods of reduced inflow and elevated temperatures, the Ter DWTP relies on operational adjustments within the reservoir cascade to manage raw water quality, while overall supply reliability is supported through blending of treated surface water with desalinated water prior to storage and distribution. While blending does not alter THM formation during treatment, it may increase bromide in distributed water, promoting the formation of more toxic brominated THMs [46]. Although bromide concentrations and THM speciation were not evaluated in this study, this highlights an important operational consideration for managing DBP precursors in source waters.

3.2. Longitudinal patterns of organic matter along the river-reservoir-DWTP continuum

DOC distributions revealed systematic spatial and seasonal differences along the continuum (Figure 3). DOC concentrations in the Ter River were generally lower and more variable than in downstream reservoirs, particularly during spring and early summer, reflecting the influence of catchment hydrology and episodic inputs [12,47]. In contrast, the Sau and Susqueda reservoirs exhibited higher DOC levels, particularly during summer and autumn, consistent with enhanced in-reservoir processing under stratified conditions [48,49]. At the Pasteral intake and the DWTP inlet, DOC values reflected an integrated downstream signal with reduced variability, suggesting the combined influence of upstream inputs and reservoir operations.

Correlation analyses further revealed clear spatial connectivity patterns along the continuum (Figure S6). DOC at Ter River showed weak or negative correlations with downstream locations (Sau-extracted: r = 0.38; Sqd-extracted: r = -0.39; PST: r = -0.41; DWTP-inlet: r = -0.30), indicating limited direct propagation of upstream riverine variability. Sau exhibited only modest correlations with downstream

sites (Sqd-extracted: r = 0.15; PST: r = 0.14; DWTP-inlet: r = 0.17). In contrast, Susqueda displayed strong correlations with Pasteral and the DWTP inlet (r = 0.94 and r = 0.91, respectively), consistent with PST-DWTP inlet coupling (r = 0.95) and indicating that DOC variability reaching the treatment plant is primarily controlled by the lower reservoirs.

Comparison with surface DOC values at Sau and Susqueda reservoirs (Figure S7) showed consistently higher concentrations and greater seasonal amplitudes in the reservoir epilimnion relative to extracted waters, particularly during stratified periods. This contrast highlights the role of selective withdrawal in modulating the quantity and character of organic matter delivered downstream [50]. Similar patterns were observed for surface DOC correlations (Figure S8), although relationships were generally weaker, reinforcing the relevance of withdrawal conditions for downstream water quality.

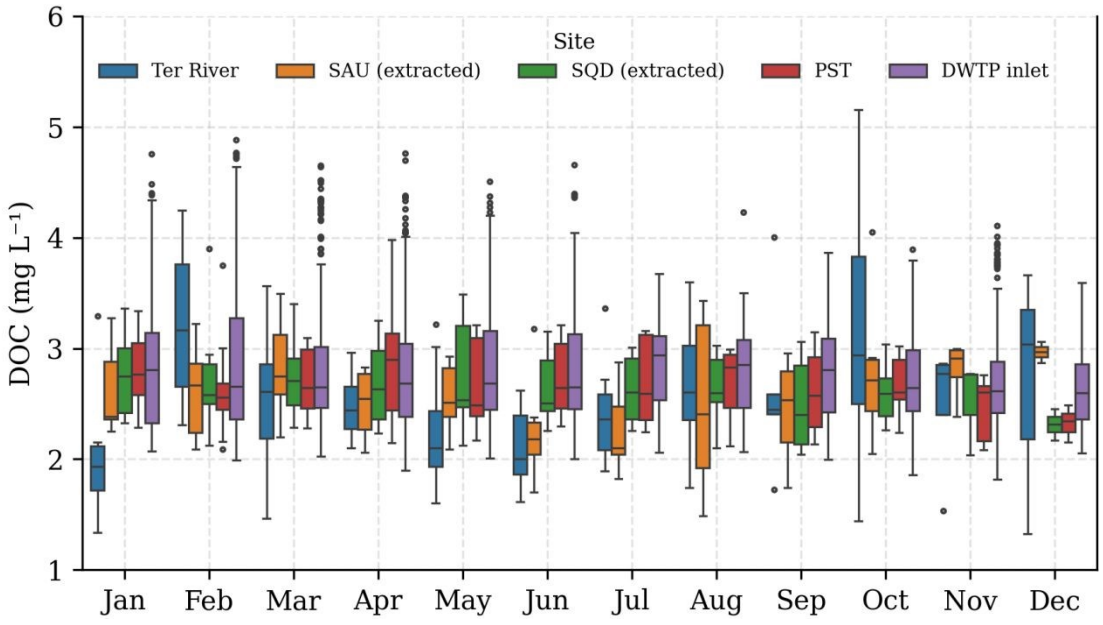


Figure 3. Seasonal distribution of DOC concentrations along the river-reservoir-DWTP continuum for the period January 2015-December 2023. Monthly boxplots are shown for the Ter River, Sau (extracted withdrawal depth), Susqueda (extracted withdrawal depth, SQD), Pasteral intake (PST), and the DWTP inlet (n_{Ter} = 69, n_{Sau} = 74, $n_{Susqueda}$ = 70, $n_{Pasteral}$ = 81, n_{DWTP} = 108).

Together, these analyses indicate that DOC variability at the DWTP inlet is strongly linked to conditions at Susqueda. This longitudinal connectivity identifies Susqueda as the key upstream control point for organic matter dynamics affecting the treatment plant. Given the availability of high-frequency monitoring at this location, fDOM measured at the extraction depth was retained as a proxy to

investigate short-term organic matter dynamics. In combination with DOC and water temperature measured at the DWTP inlet, these results define the three variables selected for subsequent analysis of upstream drivers, predictability and THM formation risk.

3.3. Driver attribution of water quality indicators

Figure 4 presents the PI results obtained using RF and LSTM models for the three selected target variables: DOC, water temperature, and fDOM. Although absolute importance values differed between models, both approaches revealed consistent patterns in the dominant drivers and their relative influence.

View Article Online
DOI: 10.1039/D6EW00128A

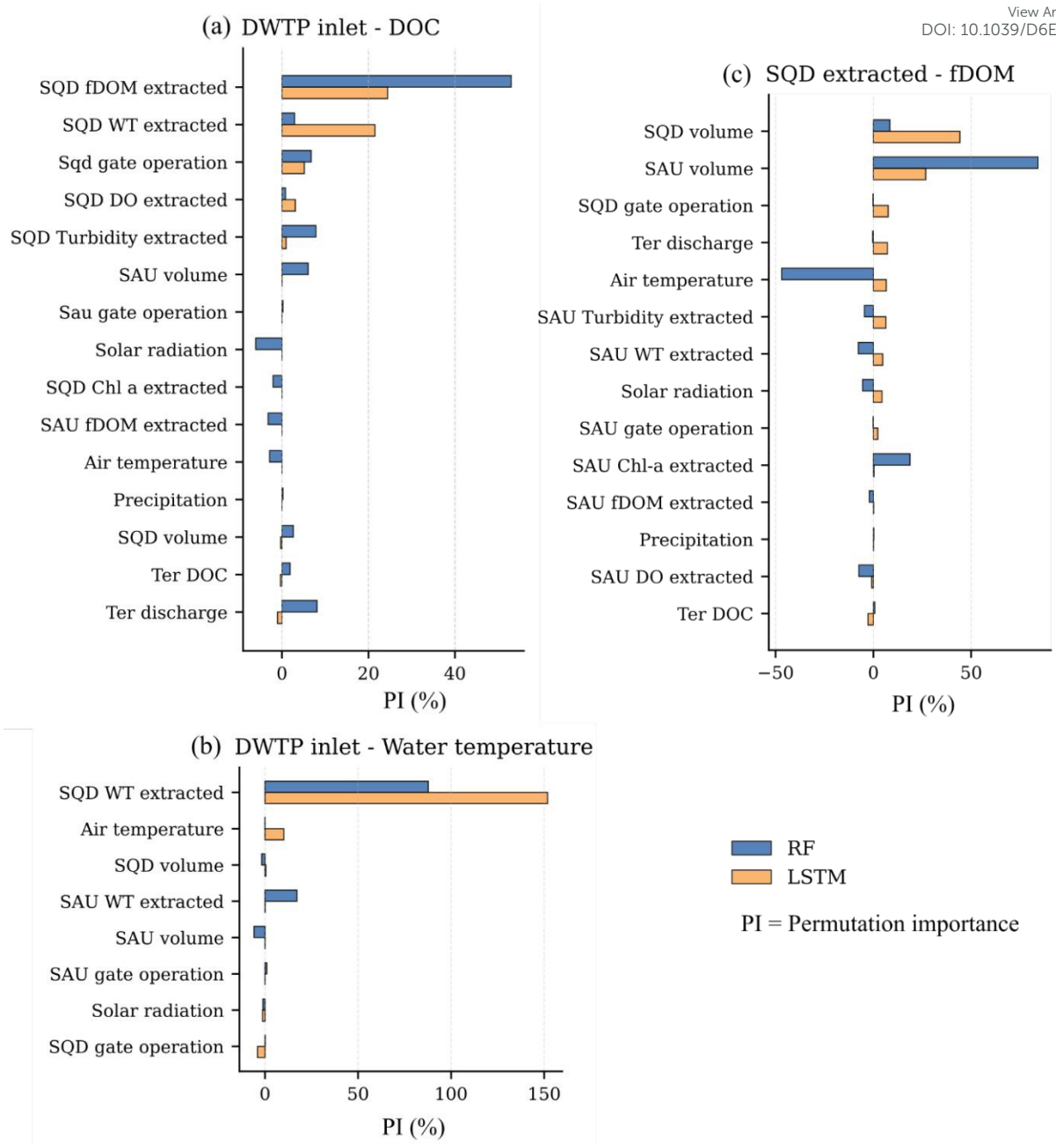


Figure 4. Feature contributions to predict (a) DOC at the DWTP inlet, (b) Water temperature (WT) at the DWTP inlet, and (c) fDOM at Susqueda extracted withdrawal depth. Permutation importance (ΔR^2,%) was computed on the validation period for both Random Forest (RF) and Long Short-Term Memory (LSTM) models; positive values indicate important predictors, whereas negative values indicate negligible or unstable contributions.

At the DWTP inlet, DOC variability was primarily associated with water quality conditions at the Susqueda withdrawal depth, with extracted fDOM at Susqueda emerging as the most influential predictor in both RF and LSTM models (Figure 4a). Additional contributions from DO, turbidity, water temperature and gate operation at Susqueda suggested that DOC reaching the DWTP reflects organic

View Article Online
DOI: 10.1039/D6EW00128A

matter characteristics shaped within the reservoir and transmitted downstream through selective withdrawal operations. Together, these variables represent optical properties, particulate inputs, and temperature-dependent processes, which may provide complementary information beyond DOC alone, as DOM responses can be linked to internal biological and physical controls that vary with stratification and residence time in reservoirs [51,52]. The higher DOC concentrations observed in the reservoirs relative to the upstream Ter River (Section 3.2) further indicate a substantial autochthonous component to organic matter dynamics, consistent with the elevated importance of in-reservoir water quality variables in the LSTM model [53]. In contrast, Ter River DOC and discharge showed weak or negligible contributions in the LSTM but slightly higher relevance in RF, reflecting methodological differences whereby RF captures static cross-sectional associations while LSTM emphasizes predictors that improve temporal forecasts across lag structures [54].

For water temperature at the DWTP inlet (Figure 4b), PI results clearly identified water temperature at the Susqueda extraction depth as the dominant driver, particularly in the LSTM model. Air temperature contributed to a lesser extent, while other upstream hydrometeorological variables had minimal explanatory power. This pattern indicated that thermal conditions at the DWTP were largely controlled by selective withdrawal at Susqueda, with meteorological forcing indirectly embedded in the reservoir thermal structure rather than acting as a direct driver at the inlet [55].

At Susqueda reservoir, fDOM was modeled as an independent target variable to better understand controls on organic matter quality at this key upstream location (Figure 4c). PI analysis highlighted reservoir storage volume as a major driver of fDOM variability, alongside withdrawal-related variables. This suggests that fDOM dynamics at Susqueda are closely linked to hydrodynamic and stratification conditions that regulate internal organic matter production, residence time, and vertical distribution [56]. The importance of storage volume was consistent with findings by Mercado-Bettín et al., 2025, who reported a similar role of volume in controlling fDOM dynamics at the Sau reservoir, where water volume acted as a surrogate for in-reservoir DOM production, with lower volumes associated with reduced fDOM and higher volumes corresponding to increased and more stable values. The agreement

View Article Online
DOI: 10.1039/D6EW00128A

between studies indicates that, in large managed reservoirs, water availability and storage conditions may shift DOM control from catchment-derived inputs toward internal biogeochemical processing.

Overall, RF and LSTM models yielded coherent driver rankings. RF provided an interpretable baseline of predictor relevance in a highly correlated system, while LSTM emphasized predictors that consistently improve temporal forecasts, reducing the apparent role of weaker or collinear variables. Across all target variables, Susqueda reservoir operational and withdrawal-related variables emerged as the dominant controls, underscoring the central role of reservoir management in shaping organic matter and thermal conditions that propagate to the DWTP inlet and influence THM formation risk.

It should be noted that PI estimates may be affected by multicollinearity among predictors, which can lead to shared or redistributed importance across correlated variables. Therefore, results are interpreted in terms of consistent patterns across predictors and modeling approaches rather than as precise quantitative measures of individual variable influence. This is particularly relevant in environmental systems where many drivers are interdependent [34].

3.4. LSTM prediction performance

Time series comparisons between observations and LSTM predictions (Figure 5) illustrate the predictive skill of the optimized models for DOC, water temperature, and fDOM across both training and validation periods. Overall, the LSTM models closely reproduced the timing and magnitude of variability, capturing both seasonal patterns and short-term fluctuations, and maintained strong performance during the validation period despite increased variability and the occurrence of extreme conditions.

View Article Online
DOI: 10.1039/D6EW00128A

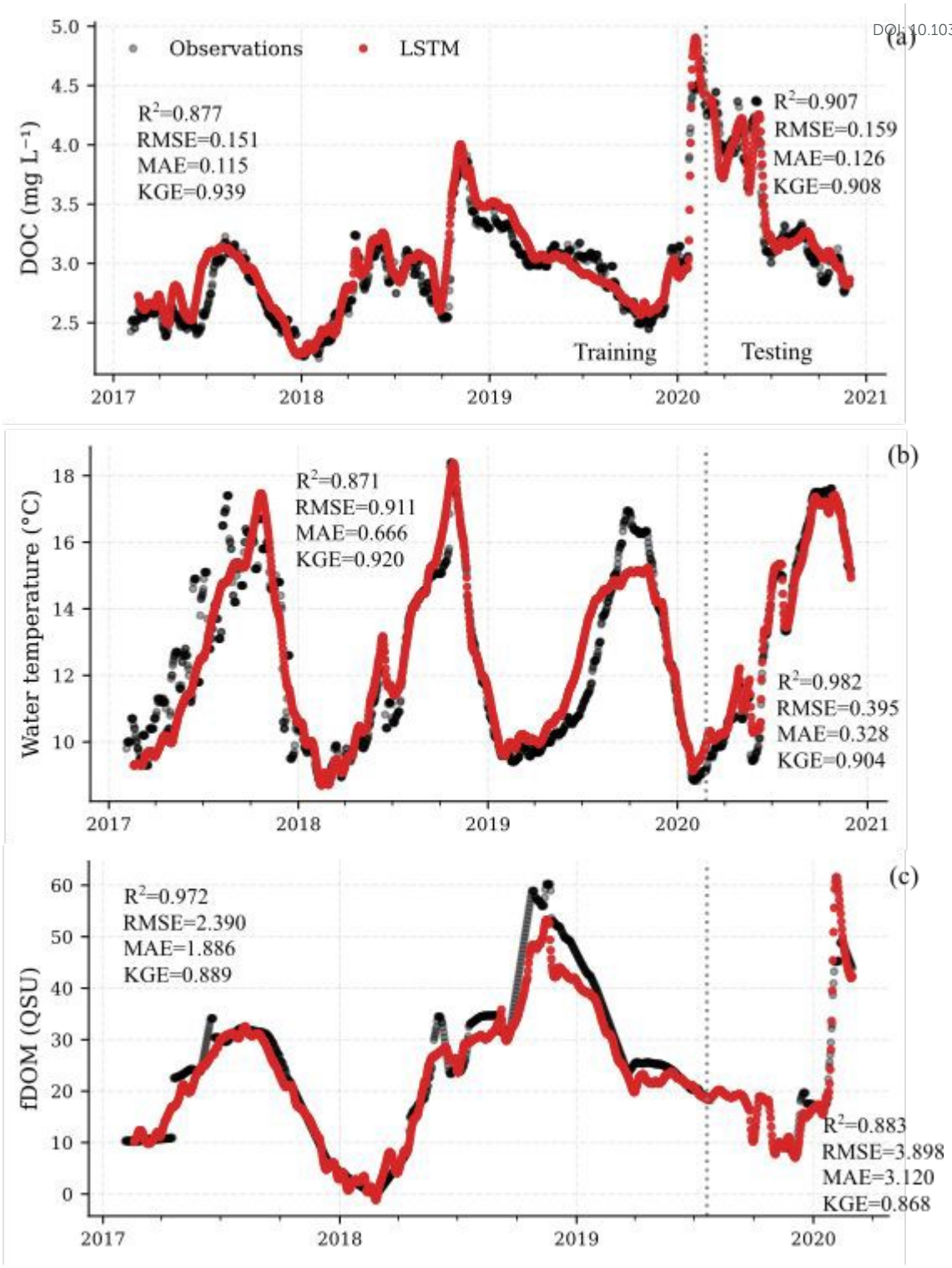


Figure 5. Observed and predicted daily time series for the three modeled water quality indicators using the optimized LSTM models: (a) DOC at the DWTP inlet, (b) water temperature at the DWTP inlet, and (c) fDOM at the Susqueda reservoir withdrawal depth. Black points show observations and the red line shows LSTM predictions. The vertical dashed line separates the training and testing periods. Performance statistics (R^2, RMSE, MAE, KGE) are reported for each period within each panel. Note that fDOM (panel c) covers a shorter record with observational gaps due to profiler availability, whereas DOC and water temperature are continuous over February 2017-November 2020.

View Article Online
DOI: 10.1039/D6EW00128A

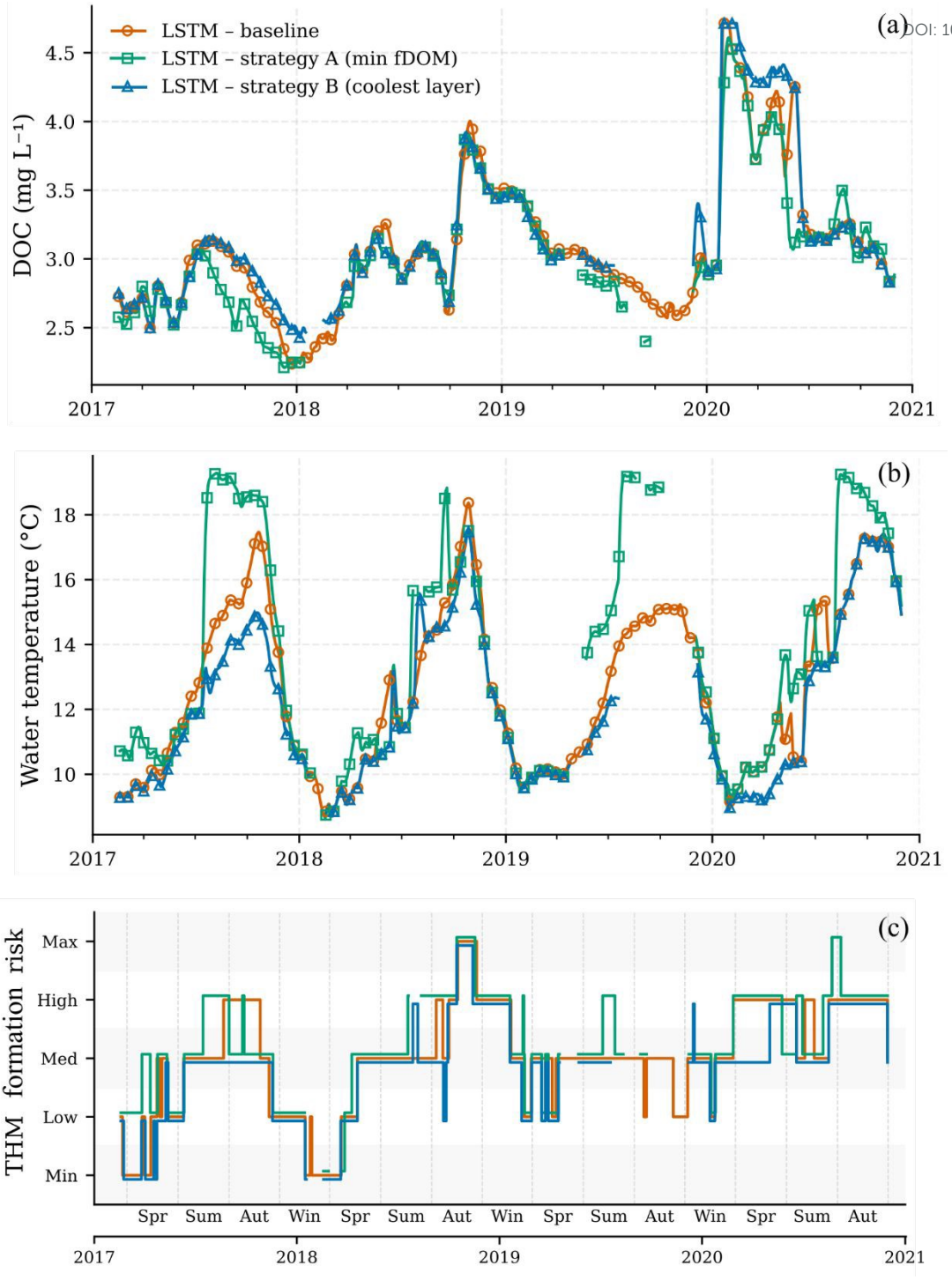
For DOC at the DWTP inlet, the highest predictability was achieved using a multivariate configuration comprising ten predictors dominated by Susqueda withdrawal water quality variables (fDOM, water temperature, DO, turbidity and Chl-a), together with extracted fDOM at Sau, reservoir storage volumes, and upstream river inputs (DOC and discharge). Water temperature at the DWTP inlet predictions required a simpler configuration, with Susqueda withdrawal temperature as the dominant predictor and air temperature providing secondary information. The strong performance obtained with this minimal configuration highlights the deterministic nature of downstream thermal dynamics once withdrawal-layer temperature was accounted for. In contrast, fDOM predictions at the Susqueda withdrawal depth required a broader predictor set to achieve optimal performance. The best results were obtained using all predictors retained from the PI analysis, with the exception of precipitation, which did not improve predictive skill, likely due to collinearity with river discharge and storage dynamics [43]. This result reflects the greater complexity of organic matter quality dynamics within the reservoir, which are influenced by interacting processes.

These results demonstrated that the LSTM models effectively translated the dominant drivers identified in Section 3.3 into robust generalizable predictions, while revealing clear contrasts in the number and type of predictors required to optimally predict each target variable.

3.5. Scenario-based simulations of reservoir operation and impacts on THM formation risk

Scenario-based simulations were conducted to explore how alternative selective withdrawal strategies at the Susqueda reservoir could influence DOC and water temperature at the DWTP inlet, and consequently lead to shifts in THM formation risk classes (Figure 6). The baseline record reflected that historical gate selection was not governed by a single formal decision rule, but rather by expert judgment and system conditions. In particular, the accessible withdrawal layers were constrained by reservoir water level and discrete gate availability (Figure S3), indicating that storage dynamics condition when (and whether) certain operational strategies can be implemented.

View Article Online
DOI: 10.1039/D6EW00128A



497

498 **Figure 6.** Evaluation of two alternative operational withdrawal strategies using the trained LSTM
499 models. Strategy A (organic matter-oriented) selects the Susqueda withdrawal depth associated with
500 the minimum fDOM, whereas Strategy B (temperature-oriented) selects the coolest available layer. The
501 strategies were propagated through the (a) DOC and (b) water temperature forecasting models to assess
502 potential trade-offs, and the resulting DOC and water temperature simulations were combined to
503 estimate (c) the total trihalomethane (THM) formation risk class at the DWTP inlet using the DOC-
504 temperature risk classes proposed by Godo-Pla et al., 2021. Strategy A (green) and Strategy B (blue)
505 are compared against the baseline scenario (red), which represents simulated historical operating
506 conditions. Scenario trajectories include gaps for dates when the required withdrawal-depth and profiler
507 inputs were unavailable.

View Article Online
DOI: 10.1039/D6EW00128A

Across the simulation period, the two operational strategies did not produce uniform benefits and did not always concur. The "minimum fDOM" strategy (Strategy A; Figure 6a) tended to reduce predicted DOC relative to baseline during selected periods, whereas the "coolest layer" strategy (Strategy B; Figure 6b) more consistently reduced predicted water temperature during stratified seasons. Because the THM formation risk class depends on both DOC and water temperature, a strategy only reduced risk when it lowered the variable that mattered most at that time. As a result, some periods showed a clear benefit from one strategy but little change from the other, while in other periods both strategies produced similar outcomes, especially when baseline conditions were already close to a risk-class threshold (Figure 6c).

The results indicate that both strategies can reduce THM formation risk relative to baseline conditions, but their effectiveness varies seasonally and depends on prevailing reservoir stratification. Two windows illustrate the potential for operational mitigation in this system. During summer-autumn 2017 and summer-autumn 2020 the simulations indicate that THM formation risk could have been reduced relative to baseline (Figure 6c). These windows align with conditions where late summer stratification and subsequent hydro-meteorological transitions can elevate risk; DOC can increase sharply after first post-summer rainfall events, while water temperature may remain relatively high, together shifting the system into higher risk categories. This pattern was consistent with seasonal behavior previously reported for DOC dynamics in the Ter catchment and with the sensitivity of Mediterranean systems to "first flush" events following dry periods [45,57]. In these periods, strategy A was more effective when DOC reductions were sufficient to shift the DOC class downward, while strategy B was more effective when reducing temperature shifted the temperature class or prevented transitions to the highest-risk combination.

A key insight from the scenario analysis was that optimizing DOC and temperature simultaneously was not always possible because the importance of individual drivers can diverge seasonally. Although DOC and water temperature at the DWTP inlet were positively associated with THM concentrations at the plant and are therefore useful practical indicators of formation risk [25], their dominant upstream control

View Article Online
DOI: 10.1039/D6EW00128A

at Susqueda may oppose each other during stratification. For example, cooler withdrawal layers can at times be associated with different organic matter quality signals than surface waters, and periods that minimize fDOM at the withdrawal depth may not coincide with the coolest available layer [58,59]. This helped explain why the two strategies diverged in some seasons (Figure 6a-b) and highlights the value of the predictive framework for decision support. Rather than relying on a single "rule-of-thumb", managers can evaluate the trade-off between lowering DOC-related risk versus lowering temperature-related risk in real time, conditional on current stratification and gate availability [60].

From an operational perspective, the results suggest a pragmatic approach. When the objective is short-term reduction of THM formation risk categories, managers could prioritize the strategy that targets the limiting component of risk at that moment. For example, strategy A during periods when DOC is near a class threshold and likely to increase (e.g., post-summer rainfall transitions), and strategy B during periods when temperature dominates risk (e.g., warm stratified conditions when DOC is relatively stable). However, the baseline record (Figure S3), emphasizes that these decisions are sometimes constrained by reservoir storage dynamics, which determine the set of withdrawal options available at any given time and are influenced by system-wide release requirements [60]. Therefore, the most actionable implication may be that source water quality management should be coordinated with water quantity governance [1]. Maintaining storage conditions that preserve selective withdrawal flexibility (when feasible) may increase the capacity to mitigate DBP precursor export during high-risk seasons [59,61]. Under increasing drought pressure and more variable inflows, integrating reservoir operations for both supply reliability and water quality may become essential to sustain risk reduction opportunities [62,63].

The simulated withdrawal strategies represent operationally feasible alternatives within the Ter reservoir system, as they were based on observed profiler data and reflect physically accessible withdrawal conditions under given reservoir water levels. Their feasibility is therefore primarily determined by reservoir storage conditions, which control the availability of intake gates, while broader water management objectives influence these strategies indirectly through their effect on reservoir storage.

View Article Online
DOI: 10.1039/D6EW00128A

3.6. Study limitations and transferability

A key limitation of the modeling framework is the restricted temporal coverage of the dataset used for model training and evaluation (2017-2020), imposed by the availability of high-frequency reservoir profiler data required to represent withdrawal-depth conditions. Although this period captures a range of hydroclimatic conditions, including extreme events such as Storm Gloria, it does not fully represent longer-term variability, particularly the prolonged drought conditions observed after 2020. As a result, model predictions and scenario simulations should be interpreted within the range of observed conditions, and caution is required when extrapolating beyond this domain. Furthermore, the estimated risk reductions depend on the predictive accuracy of the LSTM models and should therefore be interpreted as relative, scenario-based outcomes rather than exact forecasts.

In addition, the use of DOC and water temperature as indicators represents a simplified description of potential DBP formation processes and does not explicitly account for other influencing factors such as halides, pH, or disinfectant conditions which are typically included in DBP formation models applied within treatment plants and distribution systems [64,65]. The relevance of these indicators is therefore system-specific and should be assessed for each case study, particularly in systems with higher halide concentrations or other relevant DBP precursors.

A more comprehensive assessment of uncertainty, including contributions from input data, initial conditions, model parameters, and model structure, was beyond the scope of this study but represents an important direction for future work.

Despite these limitations, the consistency between long-term exploratory analyses and model-derived drivers supports the robustness of the identified control points, highlighting the value of the framework for operational decision support under observed system conditions. This is particularly relevant given that only a limited number of recent studies have applied ML approaches to model DOM dynamics (e.g., fDOM) in reservoirs [36,66], which typically focus on surface conditions and meteorological forcing due to data limitations and the complexity of representing internal reservoir processes.

4. Conclusions

View Article Online
DOI: 10.1039/D6EW00128A

This study demonstrates that pre-treatment DBP formation risk in a Mediterranean drinking water system is strongly shaped by upstream hydroclimatic conditions, reservoir processes, and operational decisions, highlighting the limitations of treatment-plant-focused approaches when applied in isolation. By integrating long-term monitoring data with machine learning, the results showed that conditions in key upstream reservoirs can dominate downstream variability in DBP-relevant indicators, highlighting reservoirs as active control points rather than passive buffers in drinking water supply systems.

In the Ter system, this control was exerted by the Susqueda reservoir, where longitudinal connectivity and operational choices primarily governed DOM dynamics at the drinking water treatment plant inlet. Strong covariation of DOM proxies supported the use of DOC as an operationally relevant indicator, while withdrawal-depth fDOM and temperature emerged as dominant drivers, emphasizing the importance of vertical reservoir structure and selective abstraction. The high predictive skill achieved for DOC, water temperature, and fDOM using LSTM models further indicates that data-driven approaches can effectively capture the combined influence of climate variability, storage dynamics, and operations in complex, managed catchment systems.

Scenario simulations revealed that upstream operational strategies, such as selective withdrawal, can reduce indicator-based THM formation risk at the DWTP inlet, but only within specific seasonal and hydroclimatic windows. The effectiveness of these interventions was constrained by reservoir levels, stratification state, and infrastructure limitations, emphasizing that water quality objectives must be coordinated with water quantity governance under increasing hydroclimatic stress.

Beyond the Ter system, this work highlights a broader and transferable opportunity for drinking water utilities. While the empirical risk relationships and operational constraints are site-specific, the overall framework (combining multi-source monitoring data, indicator-based risk metrics, machine learning, and scenario analysis) is broadly applicable to other reservoir systems. Many treatment plants already collect long-term and increasingly high-frequency data on hydrology, reservoir conditions, and raw water quality, and the integration of these datasets with globally available climate data products enables site-specific, data-driven analyses to support anticipatory DBP formation risk management. When combined with local knowledge of infrastructure and operating constraints, such approaches offer a

practical pathway to move from reactive end-of-pipe mitigation toward proactive, multi-barrier strategies that begin at the source. As climate change continues to intensify hydroclimatic variability, leveraging existing data and integrated modeling and forecasting frameworks will be essential for safeguarding drinking water quality across diverse regions.

Author contributions

Angela Pedregal-Montes: Writing-original draft, Methodology, Investigation, Formal analysis, Data curation. Eleanor Jennings: Writing-review & editing, Supervision, Conceptualization, Investigation, visualization. Rafael Marcé: Writing-review & editing, Conceptualization, visualization, Funding acquisition. Maria José Farré: Writing-review & editing, Supervision.

Conflicts of interest

There are no conflicts to declare.

Data availability

This study used publicly available meteorological reanalysis data from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/) and hydrological data for the catchment provided by the Catalan Water Agency (ACA) through the SDIM platform. Water quality data were supplied by the ATL water company and are not publicly available due to access restrictions. Data analysis and machine-learning modeling were performed in Python; the specific libraries used are listed in the Supplementary Information (SI). References cited in the SI are included in the article's reference list[67–70].

Acknowledgements

This project was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 956623 (MSCA-ITN-ETN-European Training Network, inventWater).

View Article Online
DOI: 10.1039/D6EW00128A

References

1 I. Delpla, A. V. Jung, E. Baures, M. Clement and O. Thomas, Impacts of climate change on surface water quality in relation to drinking water production, *Elsevier Ltd*, 2009, preprint, DOI: 10.1016/j.envint.2009.07.001.

2 B. Ma, C. Hu, J. Zhang, M. Ulbricht and S. Panglisch, Impact of Climate Change on Drinking Water Safety, *ACS Environmental Science and Technology Water*, 2022, **2**, 259–261.

3 R. I. Woolway, Y. Zhang, E. Jennings, T. Zohary, S. F. Jane, J. Jansen, G. A. Weyhenmeyer, D. Long, A. Fleischmann, L. Feng, B. Qin, K. Shi, H. Shi, W. Wang, Y. Tong, G. Zhang, J. Zscheischler, Z. Ren and E. Jeppesen, Extreme and compound events in lakes, *Springer Nature*, 2025, preprint, DOI: 10.1038/s43017-025-00710-w.

4 X. Liu, L. Chen, M. Yang, C. Tan and W. Chu, The occurrence, characteristics, transformation and control of aromatic disinfection by-products: A review, *Elsevier Ltd*, 2020, preprint, DOI: 10.1016/j.watres.2020.116076.

5 I. Evlampidou, L. Font-Ribera, D. Rojas-Rueda, E. Gracia-Lavedan, N. Costet, N. Pearce, P. Vineis, J. J. K. Jaakkola, F. Delloye, K. C. Makris, E. G. Stephanou, S. Kargaki, F. Kozisek, T. Sigsgaard, B. Hansen, J. Schullehner, R. Nahkur, C. Galey, C. Zwiener, M. Vargha, E. Righi, G. Aggazzotti, G. Kalnina, R. Grazuleviciene, K. Polanska, D. Gubkova, K. Bitenc, E. H. Goslan, M. Kogevinas and C. M. Villanueva, Trihalomethanes in drinking water and bladder cancer burden in the European Union, *Environ. Health Perspect.*, DOI:10.1289/EHP4495.

6 I. Kalita, A. Kamilaris, P. Havinga and I. Reva, Assessing the Health Impact of Disinfection Byproducts in Drinking Water, *ACS ES and T Water*, 2024, **4**, 1564–1578.

7 S. D. Richardson, M. J. Plewa, E. D. Wagner, R. Schoeny and D. M. DeMarini, Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: A review and roadmap for research, 2007, preprint, DOI: 10.1016/j.mrrev.2007.09.001.

8 S. E. Hrudey, B. Conant, I. P. Douglas, J. Fawell, T. Gillespie, D. Hill, W. Leiss, J. B. Rose and M. Sinclair, Managing uncertainty in the provision of safe drinking water, *Water Sci. Technol. Water Supply*, 2011, **11**, 675–681.

9 Y. Zhang, J. Deng, B. Qin, G. Zhu, Y. Zhang, E. Jeppesen and Y. Tong, Importance and vulnerability of lakes and reservoirs supporting drinking water in China, *Fundamental Research*, 2023, **3**, 265–273.

10 R. Xiao, Y. Deng, Z. Xu and W. Chu, Disinfection Byproducts and Their Precursors in Drinking Water Sources: Origins, Influencing Factors, and Environmental Insights, *Engineering*, DOI:10.1016/j.eng.2023.08.017.

11 S. Chowdhury, P. Champagne and P. J. McLellan, Models for predicting disinfection byproduct (DBP) formation in drinking waters: A chronological review, 2009, preprint, DOI: 10.1016/j.scitotenv.2009.04.006.

12 A. Pedregal-Montes, E. Jennings, D. Kothawala, K. Jones, J. Sjöstedt, S. Langenheder, R. Marcé and M. J. Farré, Disinfection by-product formation potential in response to variability in

dissolved organic matter and nutrient inputs: Insights from a mesocosm study, *Water Res.*, DOI:10.1016/j.watres.2024.121791.

13 A. Kozari and D. Voutsa, Impact of climate change on formation of nitrogenous disinfection by products. Part I: Sea level rise and flooding events, *Science of the Total Environment*, DOI:10.1016/j.scitotenv.2023.166041.

14 A. Kozari, S. Gkellis and D. Voutsa, Impact of climate change on formation of nitrogenous disinfection by-products. Part II: water blooming and enrichment by humic substances, *Environmental Science and Pollution Research*, DOI:10.1007/s11356-024-32960-4.

15 S. Chowdhury, Adaptation of water treatment processes for controlling disinfection byproducts in supply waters to compensate the effects of climate change, *Journal of Water Process Engineering*, DOI:10.1016/j.jwpe.2024.105081.

16 IPCC, *Climate Change*, 2023.

17 R. Swinamer, L. E. Anderson, D. Redden, P. Bjorndahl, J. Campbell, W. H. Krkošek and G. A. Gagnon, Climate-Driven Increases in Source Water Natural Organic Matter: Implications for the Sustainability of Drinking Water Treatment, *Environ. Sci. Technol.*, 2024, **58**, 11958–11969.

18 D. Y. Dorado-Guerra, J. Paredes-Arquiola, M. Á. Pérez-Martín, G. Corzo-Pérez and L. Ríos-Rojas, Effect of climate change on the water quality of Mediterranean rivers and alternatives to improve its status, *J. Environ. Manage.*, DOI:10.1016/j.jenvman.2023.119069.

19 R. Bhattacharya, J. R. Jones, J. L. Graham, D. V. Obrecht, A. P. Thorpe, J. D. Harlan and R. L. North, Nonlinear multidecadal trends in organic matter dynamics in Midwest reservoirs are a function of variable hydroclimate, *Limnol. Oceanogr.*, 2022, **67**, 2531–2546.

20 A. Senatore, G. A. Corrente, E. L. Argento, J. Castagna, M. Micieli, G. Mendicino, A. Beneduci and G. Botter, Seasonal and Storm Event-Based Dynamics of Dissolved Organic Carbon (DOC) Concentration in a Mediterranean Headwater Catchment, *Water Resour. Res.*, DOI:10.1029/2022WR034397.

21 G. A. Weyhenmeyer and J. Karlsson, Nonlinear response of dissolved organic carbon concentrations in boreal lakes to increasing temperatures, *Limnol. Oceanogr.*, 2009, **54**, 2513–2519.

22 V. Krysanova and J. G. Arnold, Advances in ecohydrological modelling with SWAT - A review, 2008, preprint, DOI: 10.1623/hysj.53.5.939.

23 F. Alizadeh, M. H. Niksokhan, M. R. Nikoo, A. Mishra, M. Al-Wardy and G. Al-Rawas, Enhancing water security through integrated decision-making and selective withdrawal for sustainable reservoir management, *Sci. Rep.*, DOI:10.1038/s41598-025-18027-5.

24 X. C. Nguyen, V. K. H. Bui, K. H. Cho and J. Hur, Practical application of machine learning for organic matter and harmful algal blooms in freshwater systems: A review, *Taylor and Francis Ltd.*, 2024, preprint, DOI: 10.1080/10643389.2023.2285691.

25 L. Godo-Pla, J. J. Rodríguez, J. Suquet, P. Emiliano, F. Valero, M. Poch and H. Monclús, Control of primary disinfection in a drinking water treatment plant based on a fuzzy inference system, *Process Safety and Environmental Protection*, 2021, **145**, 63–70.

26 J. Suquet, L. Godo-Pla, M. Valentí, L. Ferràndez, M. Verdaguer, M. Poch, M. J. Martín and H. Monclús, Assessing the effect of catchment characteristics to enhanced coagulation in drinking water treatment: RSM models and sensitivity analysis, *Science of the Total Environment*, DOI:10.1016/j.scitotenv.2021.149398.

27 H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume and J. Thépaut, The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 2020, **146**, 1999–2049.

28 M. N. Futter, M. A. Erlandsson, D. Butterfield, P. G. Whitehead, S. K. Oni and A. J. Wade, PERSiST: A flexible rainfall-runoff modelling toolkit for use with the INCA family of models, *Hydrol. Earth Syst. Sci.*, 2014, **18**, 855–873.

29 M. N. Futter, D. Butterfield, B. J. Cosby, P. J. Dillon, A. J. Wade and P. G. Whitehead, Modeling the mechanisms that control in-stream dissolved organic carbon dynamics in upland and forested catchments, *Water Resour. Res.*, DOI:10.1029/2006WR004960.

30 A. Pedregal-Montes, D. Mercado-Bettín, M. Futter, J. L. J. Ledesma, M. J. Farré, R. Marcé and E. Jennings, Seasonal forecasting of dissolved organic carbon in a Mediterranean catchment: Enhancing upstream control of disinfection by-product precursors, *Environ. Monit. Assess.*Submitted.

31 E. Ryder, E. Jennings, E. de Eyto, M. Dillane, C. NicAonghusa, D. C. Pierson, K. Moore, M. Rouen and R. Poole, Temperature quenching of CDOM fluorescence sensors: temporal and spatial variability in the temperature response and a recommended temperature correction equation, *Limnol. Oceanogr. Methods*, 2012, **10**, 1004–1010.

32 E. Berdalet, C. Marrasé and J. L. Pelegrí, Resumen sobre la Formación y Consecuencias de la Borrasca Gloria (19-24 enero 2020), 2020, preprint, DOI: 10.20350/digitalCSIC/12496.

33 M. Zhu, J. Wang, X. Yang, Y. Zhang, L. Zhang, H. Ren, B. Wu and L. Ye, A review of the application of machine learning in water quality evaluation, *Elsevier B.V.*, 2022, preprint, DOI: 10.1016/j.eehl.2022.06.001.

34 L. Breiman, in *Random Forest. Machine Learning*, 2001, vol. 45, pp. 5–32.

35 T. D. Harris and J. L. Graham, Predicting cyanobacterial abundance, microcystin, and geosmin in a eutrophic drinking-water reservoir using a 14-year dataset, *Lake Reserv. Manag.*, 2017, **33**, 32–48.

36 D. Mercado-Bettín, R. Paíz, V. McCarthy, E. Jennings, E. de Eyto, A. M. Gallegos, M. Dillanee, J. C. Garcia, J. J. Rodríguez and R. Marcé, A machine learning approach to driver attribution of dissolved organic matter dynamics in two contrasting freshwater systems, 2025, preprint, DOI: 10.5194/egusphere-2025-4049.

37 C. Fournier, R. Fernandez-Fernandez, S. Cirés, J. A. López-Orozco, E. Besada-Portas and A. Quesada, LSTM networks provide efficient cyanobacterial blooms forecasting even with incomplete spatio-temporal data, *Water Res.*, 2024, **267**, 122553.

38 J. C. Pyo, Y. Pachepsky, S. Kim, A. Abbas, M. Kim, Y. S. Kwon, M. Ligaray and K. H. Cho, Long short-term memory models of water quality in inland water environments, *Elsevier Ltd*, 2023, preprint, DOI: 10.1016/j.wroa.2023.100207.

39 S. Hochreiter and J. Schmidhuber, Long Short-Term Memory, *Neural Comput.*, 1997, **9**, 1735–1780.

40 D. Wang, C. Zhang, A. Li, Y. Guo, H. Zhang and C. Tan, Spatio-temporal analysis and prediction for raw water quality of drinking water source by improved RNN algorithm, *Journal of Water Process Engineering*, 2025, **71**, 107164.

41 J. Ruan, Y. Cui, Y. Song and Y. Mao, A novel RF-CEEMD-LSTM model for predicting water pollution, *Sci. Rep.*, 2023, **13**, 20901.

42 C. Strobl, A. L. Boulesteix, A. Zeileis and T. Hothorn, Bias in random forest variable importance measures: Illustrations, sources and a solution, *BMC Bioinformatics*, DOI:10.1186/1471-2105-8-25.

43 A. Fisher, C. Rudin and F. Dominici, *All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously*, 2019, vol. 20.

44 I. Caballero, M. Roca, M. B. Dunbar and G. Navarro, Water Quality and Flooding Impact of the Record-Breaking Storm Gloria in the Ebro Delta (Western Mediterranean), *Remote Sens. (Basel).*, 2023, **16**, 41.

45 J. L. J. Ledesma, A. Lupon, E. Martí and S. Bernal, Hydrology and riparian forests drive carbon and nitrogen supply and DOCg :g NO3-stoichiometry along a headwater Mediterranean stream, *Hydrol. Earth Syst. Sci.*, 2022, **26**, 4209–4232.

46 S. D. Richardson, Disinfection by-products and other emerging contaminants in drinking water, *TrAC - Trends in Analytical Chemistry*, 2003, **22**, 666–684.

47 P. A. Raymond and R. G. M. Spencer, in *Biogeochemistry of Marine Dissolved Organic Matter*, Elsevier, 2015, pp. 509–533.

48 W. Zheng, Y. Chen, Y. Niu, P. Xu, H. Hao and B. Dong, Disinfection by-product formation potential in response to seasonal variations in lake water sources: Dependency on fluorescent and molecular weight characteristics, *Science of the Total Environment*, DOI:10.1016/j.scitotenv.2024.177891.

49 E. Munthali, R. Marcé and M. J. Farré, Drivers of variability in disinfection by-product formation potential in a chain of thermally stratified drinking water reservoirs, *Environ. Sci. (Camb).*, 2022, **8**, 968–980.

50 L. Cáceres, D. Méndez, J. Fernández and R. Marcé, From End-of-Pipe to Nature Based Solutions: a Simple Statistical Tool for Maximizing the Ecosystem Services Provided by Reservoirs for Drinking Water Treatment, *Water Resources Management*, 2018, **32**, 1307–1323.

51 M. Abbasi, M. Peacock, S. Drakare, J. Hawkes, E. Jakobsson and D. Kothawala, Water residence time is an important predictor of dissolved organic matter composition and drinking water treatability, *Water Res.*, DOI:10.1016/j.watres.2024.121910.

52 D. W. Howard, A. G. Hounshell, M. E. Lofton, W. M. Woelmer, P. C. Hanson and C. C. Carey, Variability in fluorescent dissolved organic matter concentrations across diel to seasonal time scales is driven by water temperature and meteorology in a eutrophic reservoir, *Aquat. Sci.*, 2021, **83**, 30.

53 Y. Wu, H. Fang, L. Huang, C. He, Q. Shi, Y. Yi, D. He and K. Wang, Reservoir operation regulates the dynamics of dissolved organic matter in sediments, *J. Environ. Manage.*, 2025, **392**, 126850.

54 S. Hochreiter and J. Schmidhuber, Long Short-Term Memory, *Neural Comput.*, 1997, **9**, 1735–1780.

55 G. Hu, Z. Yang, Y. Yue, F. Bai and Y. Ren, Joint thermal regulation by selective withdrawal in serial cascade reservoir systems effectively improves reservoir and downstream ecological health, *Water Res.*, 2025, **281**, 123659.

56 X. Wang, H. Zhang, E. Bertone, R. A. Stewart and S. P. Hughes, Hybrid three-dimensional modelling for reservoir fluorescent dissolved organic matter risk assessment, *Inland Waters*, 2022, **12**, 463–476.

57 S. Bernal, A. Butturini and F. Sabater, Variability of DOC and nitrate responses to storms in a small Mediterranean forested catchment, *Hydrol. Earth Syst. Sci.*, 2002, **6**, 1031–1041.

58 G. Hu, Z. Yang, Y. Yue, F. Bai and Y. Ren, Joint thermal regulation by selective withdrawal in serial cascade reservoir systems effectively improves reservoir and downstream ecological health, *Water Res.*, 2025, **281**, 123659.

59 B. Zouabi-Aloui, S. M. Adelana and M. Gueddari, Effects of selective withdrawal on hydrodynamics and water quality of a thermally stratified reservoir in the southern side of the Mediterranean Sea: a simulation approach, *Environ. Monit. Assess.*, 2015, **187**, 292.

60 E. Soyer, H. Bayram, N. Canıgeniş and O. Eren, Decision support system for selective withdrawal in water supply reservoirs: an approach based on thermal stratification, *Water Quality Research Journal*, 2023, **58**, 99–110.

61 C. A. Murphy, S. L. Johnson, W. Gerth, T. Pierce and G. Taylor, Unintended Consequences of Selective Water Withdrawals From Reservoirs Alter Downstream Macroinvertebrate Communities, *Water Resour. Res.*, DOI:10.1029/2020WR029169.

62 M. Nazari and R. Kerachian, Optimal Operation of Reservoirs Considering Water Quantity and Quality Aspects: A Systematic State-of-the-Art Review, *Water Resources Management*, 2024, **38**, 5911–5944.

63 M. Reza Nikoo, N. Bahrami, K. Madani, G. Al-Rawas, S. Vanda and R. Nazari, A robust decision-making framework to improve reservoir water quality using optimized selective withdrawal strategies, *J. Hydrol. (Amst).*, 2024, **635**, 131153.

64 L. Liang and P. C. Singer, Factors Influencing the Formation and Relative Distribution of Haloacetic Acids and Trihalomethanes in Drinking Water, *Environ. Sci. Technol.*, 2003, **37**, 2920–2928.

65 S. Chowdhury, P. Champagne and P. J. McLellan, Models for predicting disinfection byproduct (DBP) formation in drinking waters: A chronological review, 2009, preprint, DOI: 10.1016/j.scitotenv.2009.04.006.

66 D. W. Howard, M. E. Lofton, R. Q. Thomas, A. D. Delany, A. Breef-Pilz and C. C. Carey, Near-Term Forecasts of Dissolved Organic Matter Exhibit Consistent Patterns of Accuracy Across Multiple Freshwater Reservoirs, *J. Geophys. Res. Biogeosci.*, DOI:10.1029/2025JG009064.

67 M. Kuhn and K. Johnson, *Applied Predictive Modeling*, Springer New York, New York, NY, 2013.

68 APHA, *AWWA and WPCF Standard Methods for the Examination of Waters and Waste Waters*, 2005.

69 T. Niedzielski and M. Halicki, Improving Linear Interpolation of Missing Hydrological Data by Applying Integrated Autoregressive Models, *Water Resources Management*, 2023, **37**, 5707–5724.

70 Z. Che, S. Purushotham, K. Cho, D. Sontag and Y. Liu, Recurrent Neural Networks for Multivariate Time Series with Missing Values, *Sci. Rep.*, 2018, **8**, 6085.

View Article Online
DOI: 10.1039/D6EW00128A

Data availability

This study used publicly available meteorological reanalysis data from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/) and hydrological data for the catchment provided by the Catalan Water Agency (ACA) through the SDIM platform. Water quality data were supplied by the ATL water company and are not publicly available due to access restrictions. Data analysis and machine-learning modeling were performed in Python; the specific libraries used are listed in the Supplementary Information (SI). References cited in the SI are included in the article's reference list[60–63].