Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 11 Aug 2026
Revised 07 Sep 2026
Accepted 26 Sep 2026

Natural Resources for Human Health 2026, Vol. 6 (23s): 1142–1157

KEYWORDS:
microplastics; PFAS; LSTM; explainable AI; wetlands; hydrological regression.

Drying Waters, Rising Risks: Explainable Deep Learning for Emerging Pollutants in Shrinking Wetlands

Abdelmalek Harrag[1], *; Fateh Allag[2]; Mohamed Fenni[3]; Saddek Bouharati[4]

[1]Department of Basic Studies (EB), Laboratory of Phytotherapy Applied to Chronic Diseases (LPMC), Faculty of Natural and Life Sciences, Ferhat Abbas Sétif 1 University, Sétif, Algeria; ORCID id: (0000-0001-5607-1391).
[2] Department of Basic Studies (EB), Faculty of Natural and Life Sciences, Ferhat Abbas Sétif 1 University, Sétif, Algeria; email: allagf@yahoo.fr; ORCID id: (0009-0006-8105-1337).
[3] Department of Ecology and Environment (EE), Laboratory for the Valorization of Natural Biological Resources (VRBN), Faculty of Natural and Life Sciences, Ferhat Abbas Sétif 1 University, Sétif, Algeria;
email: fennimodz@yahoo.fr; ORCID id: (0009-0009-5804-6021).
[4]Ferhat Abbas Sétif 1 University, Sétif, Algeria; email: sbouharati@univ-setif.dz; ORCID id: 0000-0003-0725-1964
Corresponding author: Abdelmalek HARRAG; Department of Basic Studies (EB), Faculty of Natural and Life Sciences, Ferhat Abbas Sétif 1 University, Campus EL BEZ, 19000 Sétif, Algeria; email: abdelmalek.harrag@univ-setif.dz.

ABSTRACT: While wetland areas face the consequences of hydrological decline from climate change, the accumulation of microparticles (MP) and perfluorinated substances (PFAS) is not thoroughly explored. The current research aims to analyze both the seasonal and spatial patterns of microparticle and PFAS concentrations in Lake Tonga, which is part of the EL Kala Ramsar site located in Algeria. Over a period from January 2020 to June 2026, data from 2,373 observations obtained from two monitoring stations were fed into a two-layer long short-term memory (LSTM) model to anchor the predictions. The following variables were used for prediction estimating: hydrological conditions (water level, temperature, rainfall, and salinity) as well as agricultural runoff. The last test (20% of data collected) served for model validation. The explanatory variables for machine learning were assessed using sensitivity analysis methods (permutation feature importance and temporal occlusion sensitivity). During the test, the LSTM showed better results than other models, providing R^2 = 0.76 (RMSE=28.4) for MP, R^2 = 0.82 (RMSE=8.72) for PFAS, and R^2 = 0.79 (RMSE=0.05). Water level was the most powerful predictor for MP (1–5-day lag) while temperature and salinity were the factors for predicting PFAS (12–22-day lag). Contrary to previous factors that were able to correctly predict, agricultural runoff did not produce positive contribution to predictions. The slight increase in modeled ERI from 1.05 to 1.28 indicates the predictive sensitivities rather than established ecological causation, and helps to build the computer-based mechanism for wetland management.

Natural Resources for Human Health
eISSN: 2583-1194

1. INTRODUCTION

Wetlands are recognized as the world's most productive ecosystems, providing ecosystem services such as water purification, flood control, carbon capture, nutrient cycling, and biodiversity support (Davidson, 2014). Although wetlands cover less than 10% of the land surface, they support high diversity of plants and animals and play a key role in maintaining hydrological and ecological integrity. However, wetlands face increasing threats from climate change and other environmental changes that speed up the natural processes of ecosystem degradation worldwide (Junk et al., 2013; Cramer et al., 2018).

The Mediterranean Basin is one of the world's climate change hot spots because of the rate of warming exceeding global averages and climate change-related decreases in rainfall level and droughts (Giorgi, 2006). These climatic changes contribute to hydrological regression in lakes and wetlands, especially in North Africa, where growing agricultural water demand exacerbates the situation (Cramer et al., 2018). As water bodies become smaller, their capacity to dilute contaminants declines, thus leading to accumulation of both dissolved and solid pollutants thereby increasing ecological vulnerability.

The El Kala Wetland Complex in northeastern Algeria is an important Mediterranean freshwater ecosystem designated as a Ramsar Wetland of International Importance and a UNESCO Biosphere Reserve. The complex includes lakes, marshes, and forests with high biodiversity. Lake Tonga, a major ecosystem within the complex, has experienced a substantial decline in water level over recent decades, alongside increased salinity, eutrophication, and human activities such as agriculture and tourism. These changes have increased concern about emerging contaminants and their environmental impacts (Wang et al., 2017; Eerkes-Medrano et al., 2015). Microplastics and per- and polyfluoroalkyl substances (PFAS) are of particular concern because they are persistent, widespread, and may bioaccumulate. Microplastics originate from the breakdown of larger plastic debris and the direct release of plastic particles. PFAS comprise a large family of synthetic fluorinated compounds used worldwide because of their chemical stability.

Hydrological regression may exacerbate the ecological effects of these pollutants. Declining water levels may increase pollutant concentrations by reducing dilution and altering water chemistry, sediment dynamics, and biological communities. Thus, pollutant dynamics in shrinking wetlands reflect complex interactions among climate variability, hydrology, watershed characteristics, and human activity. Traditional statistical methods may fail to capture these highly nonlinear temporal relationships, limiting their ability to predict patterns under changing environmental conditions (Reichstein et al., 2019).

Recent developments in deep learning have substantially improved the analysis of environmental time-series data. Among recurrent neural network architectures, LSTM networks are useful for time-series analysis because they can capture long-term dependencies while mitigating the vanishing-gradient problem common in conventional recurrent neural networks (Hochreiter and Schmidhuber, 1997). Because they can learn nonlinear relationships, LSTM models have shown strong performance in hydrological forecasting, water-quality forecasting, and environmental monitoring (LeCun et al., 2015; Reichstein et al., 2019).

Despite their predictive accuracy, deep learning models have been criticized for being difficult to interpret. Explainable artificial intelligence (XAI) can address this limitation by providing tools to assess how each variable influences a prediction. Combining XAI with LSTM may therefore enable pollutant forecasting while helping identify environmental factors associated with ecosystem degradation.

Although previous studies have examined microplastics, PFAS, and artificial intelligence individually, few have integrated them to study climate-affected wetlands. Explainable deep learning has also been applied infrequently in Mediterranean aquatic ecosystems, particularly in North Africa, where long-term environmental datasets are limited. This makes it essential to fill the gap in knowledge and develop successful forecasting methods, which can be used for the adaptive management of wetlands. This study examines temporal changes in microplastics, per- and polyfluoroalkyl substances (PFAS), and a composite Ecotoxicological Risk Index (ERI) in Lake Tonga from 2020 to 2026.

The study aimed to: (i) assess the relationship between hydrological regression and contaminant dynamics; (ii) develop and evaluate a two-layer LSTM model to predict contaminant dynamics; (iii) use XAI to identify key environmental predictors; and (iv) assess ecological risk using the ERI. Accordingly, this study asks: How are hydrological regression and agricultural runoff related to the temporal dynamics of microplastics, PFAS, and ecological risk in Lake Tonga, and how accurately can an explainable LSTM model predict these outcomes?

Natural Resources for Human Health
eISSN: 2583-1194

2. MATERIALS AND METHODS

2.1 Study Material

This research took place in Lake Tonga, which is one of the larger freshwater lakes in the El Kala Wetland Complex, located at coordinates 36°50'N, 8°25'E, in northeastern Algeria (see **Figure 1**). Lake Tonga is recognized as a Ramsar Wetland of International Importance and a UNESCO Biosphere Reserve, reflecting its ecological significance (Convention on Wetlands, 2021; Samraoui and Samraoui, 2008).

The lake has experienced substantial hydrological recession over the past 20 years because of declining rainfall, increased evaporation, and growing anthropogenic water demand. All this has led to water level decrease, salinity increase, and degradation of the wetlands (Cramer et al., 2018). The watershed of the lake is characterized by intensive agricultural activities, tourism, and scattered urban settlements. Due to the mentioned climatic pressure and anthropogenic pollution, Lake Tonga is a good example for studying the nature of emergent pollutants in freshwater ecosystems situated in the Mediterranean region.

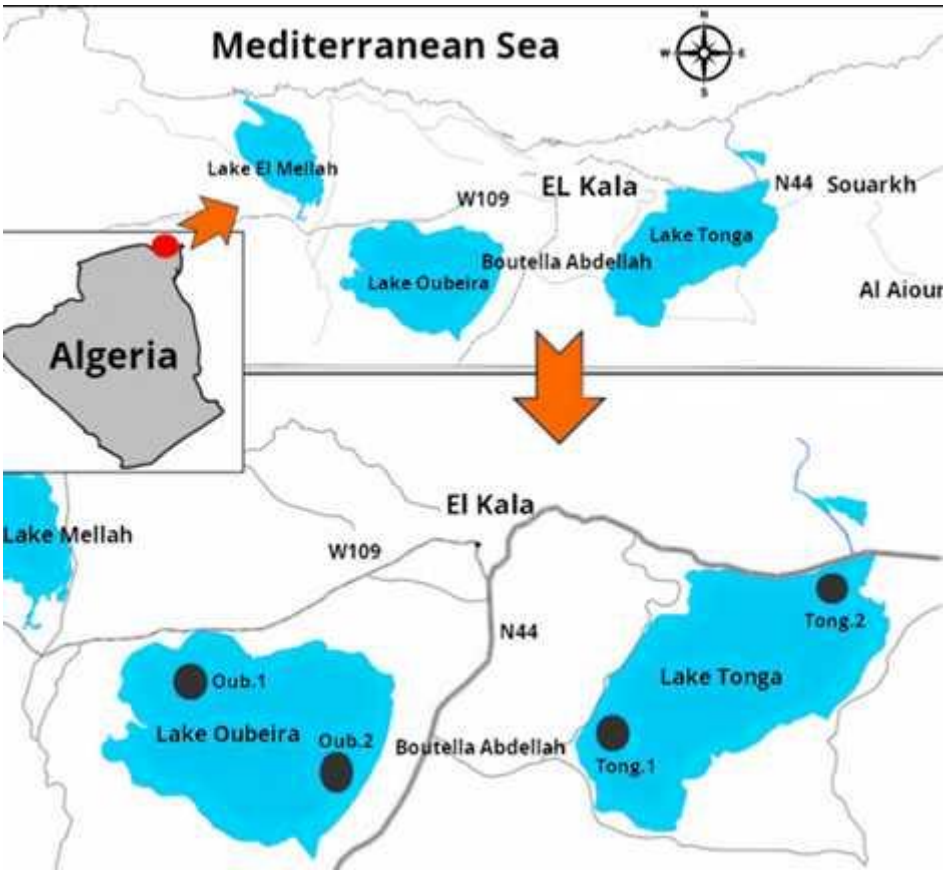


Figure 1. Location of the study area (El Kala Wetland Complex, north-eastern Algeria), showing Lake Tonga and the neighbouring Lake Oubeira. Sampling sites Tong.1 and Tong.2 were used in this study; Oub.1 and Oub.2 were not. Schematic map (not to scale). Adapted from Houhamdi et al. (2025).

2.2 Environmental Dataset

The dataset utilized was collected from daily monitoring of environmental factors between January 1, 2020 and June 30, 2026. Data from the two sampling stations in Lake Tonga were used, namely Tong.1 (36°51'12"N, 8°29'15"E) which is situated on the shallow south-western part of the lake, impacted by agriculture, and Tong.2 (36°51'54"N, 8°30'22"E) located in the north-eastern part of the lake and which has deeper waters. However, two stations (Oub.1 and Oub.2) were excluded from the analysis (see **Figure. 1**). Since the two sampling stations are located at different depths and are subjected to different water dynamics, the averaged series were created from the raw data using fixed weighted means instead of arithmetic means so that the deeper Tong.2 station does not veil the average. Even though the weights have been established already, the identity of sampling

stations was preserved in the final results and used for stratified analysis of errors for the predicted models. The variables which were used for prediction were the following: the level and temperature of water, salinity measured in PSU, precipitation per day as well as ARI (or Agricultural Runoff Index). The variables which are the outcomes of the research are: MP and PFAS concentrations, as well as ERI values. It should be noted that ARI shows the transport of pollution from agriculture to lakes by means of rainfall, watershed characteristics, land use and fertilizer consumption (Kurwadkar et al., 2022; Wang et al., 2017).

The provenance and availability of data. Information on hydrological, physico-chemical and contaminant conditions has been collected through the monitoring program of the Department of Ecology, UFAS Sétif1 while precipitation was sourced from ONM. The recorded data is shown in **Figure 2**. The time series used by the models is created by combining observations with the minimum number of imputed values (the entries of imputed values are marked in the supplementary data). The temporal alignment was based on the same calendar date in which precipitation was treated as a daily total and the measurement of the rest of the parameters was adjusted with respect to the daily record of contaminants.

The unavailability of some data represented 1.4% of the initial dataset. Linear interpolation was used to fill the missing data if the gap is shorter than 3 days, while for longer gaps, MissForest imputation was used (Stekhoven and Buhmann 2012). The imputation and processing techniques were decided only based on the training set and then used for the test set. The outliers created as a result of the frequent localized precipitation were dealt with via winsorization in the 1st and 99th percentiles by knowing the values based on the training data only. Given that the analysis used only two experimental stations in the wetland, the dataset cannot reflect the lake in full capacity; the differences in the data for each station were addressed using weighted aggregation, station-wised diagnostics and the station-based sensitivity model.

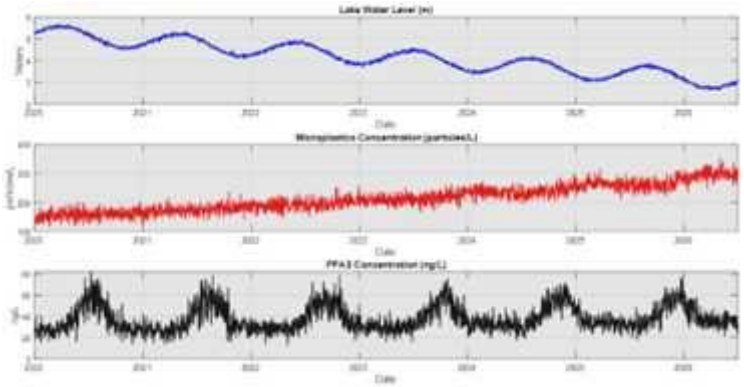


Figure 2. Time series of lake water level (m), microplastic concentration (particles/L) and PFAS concentration (ng/L) used for model development, 2020–2026

2.2.1 Measurement Protocols and Quality Assurance/Quality Control (QA/QC)

Microplastics. For each daily composite, 50 L of water samples from the surface layer were filtered through a sieve made of stainless steel (mesh size 45 μm). Organic matter was digested in a 30% solution of hydrogen peroxide at a temperature of 60 °C for 48 hours, followed by density separation in zinc chloride (density of 1.6 g/cm³). Particles were counted and classified through the use of μ-FTIR. A size range of over 45 μm was analyzed and the type of polymer was identified from the μ-FTIR spectra using a reference database.

Measures were taken to control contamination by means of airborne methods and sampling procedures. Field and procedural blanks were analyzed on a daily basis, laboratory surfaces were well cleaned, and samples were covered during sampling and in the laboratory. Concentration values were corrected according to average blank data. The limits of detection were 0.5 particles/L and limit of quantification was 1.5 particles/L. Recovery averaged 91.4% ± 3.2%.

PFAS. For water sample collection, 1 L water samples were taken in pre-rinsed HDPE bottles and filtered through 0.45μm glass fiber filter and extracted using solid-phase extraction with weak anion exchange resin cartridges. Extracts were analyzed by liquid chromatography-tandem mass spectrometry using isotope-labeled internal standards (C-13-PFOA and C-13-PFOS).

The list of analytes, calibration method, and calibration range are given in Supplementary Table S2. LOD and LOQ were 0.1 ng/mL and 0.3 ng/mL respectively. Procedural blanks consistently showed values below the LOD. Values below the LOD were substituted with LOD/2^1/2. Average percent recovery was 88.7% ± 4.5%.

QA/QC summary. Blank field samples, blank and duplicated procedures as well as calibration curves have already been done for both matrices; external laboratory control carried out each year.

2.2.2 Ecotoxicological Risk Index Construction

The ERI is a concentration-based model-based composite measure. It does not directly measure ecological damage, nor does it substitute for biological or ecosystem-response measures. The ERI is defined as

$$ERI_i = \sum_j w_i \frac{C_{i,j}}{PNEC_,}, \qquad \sum_j w_j = 1$$

where C_(j,t) is the concentration of contaminant group j (MPs or PFAS) at time t, PNEC_j is the corresponding predicted no-effect concentration (benchmark), and w_j is the group weight. The ERI is viewed as a benchmark relative, model-based composite indicator of contaminant load and not as actual ecological effect. Thus, the observed increase in mean ERI from 1.05 to 1.28 is a predicted increase based on model assumptions and does not provide direct evidence of ecological damage.

2.3 Data Pre-processing

Environmental datasets may contain missing values, uncertain measurements, and extreme values that can impair model performance. A standard preprocessing procedure was therefore applied before model training.

Missing information in the case amounted to 1.4% of the original record. The gaps of less than three days were interpolated through linear interpolation while the longer gaps were filled using the MissForest imputation process (Stekhoven and Bühlmann, 2012). Any outlier behavior that occurred due to short-duration localized rainfalls was nullified by using winsorization technique at percentile of 1 and 99, meaning that the outliers have no influence at all but the overall distribution has been maintained.

For continuous data, z-score normalization was applied.

$$x_{norm} = \frac{x - \mu}{\sigma}$$

where μ is the mean and σ is the standard deviation of each predictor.

To account for temporal dependencies, the data were arranged into 30-day sliding windows. The LSTM model used each preceding 30-day period to predict pollutant concentrations for the following day (Hochreiter and Schmidhuber, 1997).

The prepared database consisted of 2,343 temporal sequences, where 80% of the sequences (n = 1,874; chronological split) were used for training purposes, while the remaining part (n = 469) was used for independent testing.

2.4 Deep Learning Architecture

The proposed architecture comprised a sequence input layer (5 features), a first LSTM layer (64 hidden units, sequence-to-sequence mode), a dropout layer (0.2), a second LSTM layer (32 hidden units, sequence-to-last mode), a fully connected layer (16 units), a ReLU activation layer and a multi-output regression layer. Training used the Adam optimizer with an initial learning rate of 0.01, mini-batch size 32, gradient threshold 1.0, maximum 100 epochs, validation every 10 iterations, and early stopping based on validation loss (Table 1).

Table 1. Training parameters of the two-layer LSTM network.

Parameter	Value
Optimizer	Adam

Initial learning rate	0.01
Maximum epochs	100
Mini-batch size	32
Gradient threshold	1.0
Validation frequency	every 10 iterations

To reduce overfitting during model development, early stopping based on validation loss and dropout regularization were used.

$$L = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{3}(y_{ij} - \hat{y}_{ij})^2$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

In this formula, y refers to the actual measurements and $\hat{y}$ refers to the predicted values and $\bar{y}$ is the observed mean. RMSE is expressed in particles/L for MPs, ng/L for PFAS, and is dimensionless for ERI.

During the study of the presented architecture, it was compared with four different classical models:

Multiple Linear Regression (MLR) method

Random Forest (RF) method

Support Vector Regression (SVR) method

Single-layer LSTM neural network

2.5 Baseline Models, Validation and Performance Metrics

Comparison against MLR, RF, SVR and single-layer LSTM was performed using the proposed architecture. All baseline models underwent the same pre-processing and similar tuning procedures on the same validation sets. Hyperparameter search bounds and selected values are given in Table 2. The model is considered better if its test RMSE is smaller than that of all benchmarks. The hold-out test set consisted of the last chronological 20% of the data (n = 469) rather than a randomly chosen subset. Rolling origin validation over the training period was employed to investigate temporal consistency and the necessity of the chronological hold-out method. R^2, RMSE and MAE were used as evaluation metrics and bootstrap 95% confidence intervals for R^2 and RMSE were computed. The errors were also stratified according to the station, season and periods of declining water level to investigate the robustness of the model in extreme environmental conditions. Hence, the research demonstrates the advantages of the proposed deeper recurrent architecture (**Figure 3**).

Table 2. Hyperparameter search bounds and selected values for benchmark models.

Model	Hyperparameter	Search range	Selected value
MLR	Ordinary least squares	—	—
RF	Number of trees	50, 100, 200, 500	200

Model	Hyperparameter	Search range	Selected value
RF	Maximum depth	5, 10, 20, 50	20
RF	Minimum samples split	—	2
SVR	Kernel	RBF	RBF
SVR	C	0.1, 1, 10, 100	10
SVR	ε	0.01, 0.05, 0.1, 0.2	0.05
SVR	γ	—	auto
Single-layer LSTM	Hidden units	16, 32, 128	64
Single-layer LSTM	Dropout	—	0.2
Single-layer LSTM	Batch size / learning rate	—	32 / 0.01
Two-layer LSTM (proposed)	Units per layer	—	64 / 32
Two-layer LSTM (proposed)	Dropout	—	0.2
Two-layer LSTM (proposed)	Batch size / learning rate	—	32 / 0.01
Two-layer LSTM (proposed)	Maximum epochs	—	100

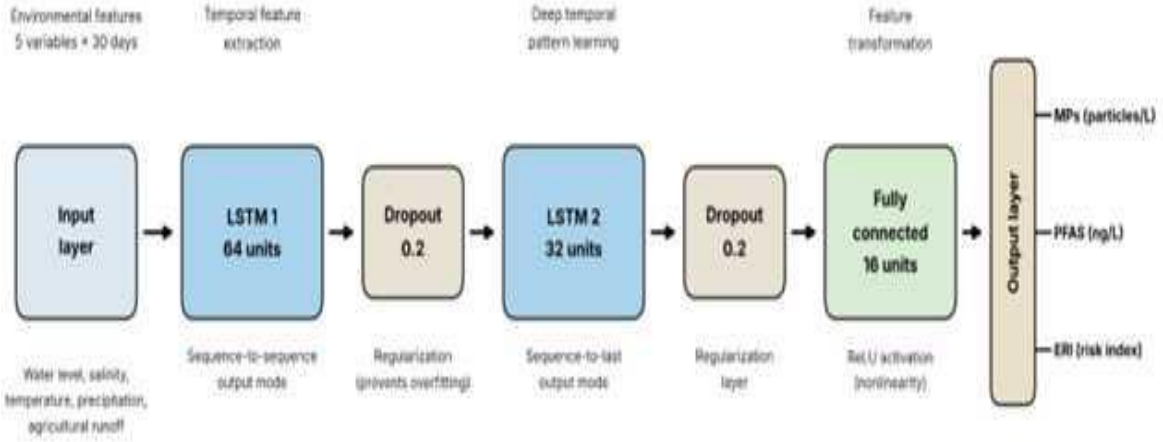


Figure 3. Schematic architecture of the two-layer Long Short-Term Memory (LSTM) neural network developed for predicting emerging pollutant concentrations. The model accepts 5 environmental features over a 30-day temporal window and produces simultaneous predictions for microplastics, PFAS, and the Ecotoxicological Risk Index. No attention layer is included.

2.6 Explainable AI Procedures

Permutation feature importance was calculated on the test set across 100 repetitions by means of the increase in RMSE. As there were correlations among some predictors (water level, precipitation, salinity, and temperature), the predictors' correlation and variance inflation factor were checked, and grouped (conditional) permutation importance was calculated for the correlated predictors. The individual importance and grouped importance results were given together — since the individual results can underestimate or redistribute contributions from correlated predictors. Other analyses were conducted as well. The temporal occlusion sensitivity analysis was performed by using the median values of the training data instead of the lagged values for any day within the 1–30-day time frame, and the increase in RMSE was noted. The 1–5-day MP sensitivity analysis and the 12–22-day PFAS sensitivity analysis were interpreted as model-dependent prediction sensitivities, but not as evidence related to contaminant transport or environmental causation.

2.7 Ecotoxicological Risk Assessment

Overall ecological risk was assessed using a composite Ecotoxicological Risk Index (ERI) framework incorporating microplastic and PFAS contamination. Risk quotients (RQs) were calculated according to the formula:

$$RQ_i = \frac{MEC_i}{PNEC_i}$$

where (MECi) is the measured environmental concentration and (PNECi) the predicted no-effect concentration of pollutant (i). Because no site-specific toxicity data were available to justify differential weighting, and because a weighted sum can mask a single severe stressor, the composite index was defined by a conservative maximum-RQ rule:

$$ERI = max(RQ_{MP}, RQ_{PFAS})$$

The ERI is therefore dimensionless and equal to the risk quotient of the most constraining pollutant at each time step. The ERI values were classified into four ecological assessment categories (shown in Table 3): The ERI values were classified into four ecological risk categories (Table 3).

Table 3. ERI risk categories.

ERI	Ecological Status
<1.0	Low Risk
1.0–2.0	Moderate Risk
2.0–3.0	High Risk
>3.0	Critical Risk

2.8 Statistical Analysis

To study long-term climate trends, Mann-Kendall test and Sen's slope estimator were utilized (Helsel et al., 2020) via the pymannkendall Python package (Hussain and Mahmud, 2019) and the scipy. stats. theilslopes tool for calculation. For the analysis of the relationship between climatic parameters and pollution levels Pearson and Spearman correlation tests were used. The specifics of each method are reported in the Table 5: sample size (n), value of the statistic, p-value, and size of effect or slope. The term "statistically significant" is used only in cases, where p-value supports it. The analysis was performed using tools of MATLAB R2024a and Python 3.11.

2.9 Species Vulnerability Screening

Species-level exposure assessment relied on a semi-quantitative trait-based scoring system developed by Foden et al. (2013) in conjunction with IUCN Red List criteria. The following four traits were assessed: habitat specialization (utilization of shallow littoral zones), life-history sensitivity (generation time and fecundity), trophic exposure (association with benthic habitats and feeding habit), and the known sensitivity of organisms to organic contaminants. The assessed levels of several traits were summed up to the qualitative category of (low / moderate / high exposure) exposure.

3. RESULTS AND DISCUSSION

3.1 Performance of the Proposed Deep Learning Model

The two-layer LSTM converged within 5,800 training iterations, with steadily decreasing training and validation losses and no evidence of severe overfitting (**Figure 4**).

On the independent test set (n = 469), the model achieved R^2 = 0.76 (RMSE = 28.4 particles/L) for MPs, R^2 = 0.82 (RMSE = 8.7 ng/L) for PFAS, and R^2 = 0.79 (RMSE = 0.05) for ERI (Table 4; **Figure 5**).

The proposed architecture outperformed all benchmarks for all three targets. The RMSE decreased monotonically from MLR to RF, SVR, single-layer LSTM and two-layer LSTM, and the improvement over the single-layer LSTM was consistent across targets (MPs: 31.10 → 28.40 particles/L; PFAS: 9.50 → 8.70 ng/L; ERI: 0.06 → 0.05). This ordering is consistent with the ability of a second recurrent layer to separate short-term hydrological noise from multi-week accumulation trends. PFAS was the most accurately predicted target, plausibly because dissolved pollutant concentrations vary more smoothly than particulate MPs. Comparable performance of recurrent networks has been reported for hydrological and water-quality forecasting (Kratzert et al., 2019; Shen, 2018; Reichstein et al., 2019).

Table 4. Model performance on the independent test dataset (n = 469).

Model	RMSE MPs (particles/L)	R^2 MPs	RMSE PFAS (ng/L)	R^2 PFAS	RMSE ERI	R^2 ERI
MLR	54.20	0.44	19.30	0.51	0.22	0.48
RF	36.10	0.65	11.20	0.72	0.09	0.69
SVR	33.40	0.69	10.40	0.75	0.08	0.71
Single-layer LSTM	31.10	0.72	9.50	0.79	0.06	0.75
Two-layer LSTM (proposed)	28.40	0.76	8.70	0.82	0.05	0.79

The forecasting results indicate that explainable deep learning has the potential to create accurate environmental forecasts while maintaining good stability for practical decision-making.

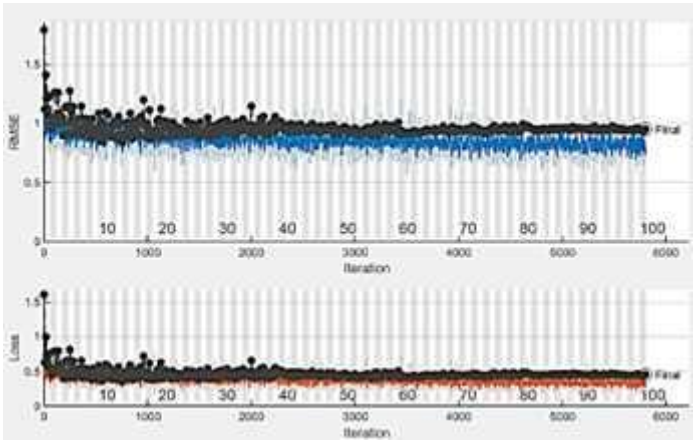


Figure 4. Training progress of the LSTM network showing (top) RMSE and (bottom) loss functions across 5,800 iterations. Blue lines represent training metrics, black lines represent validation metrics, and smoothed curves are shown in dark colors.

Natural Resources for Human Health
eISSN: 2583-1194

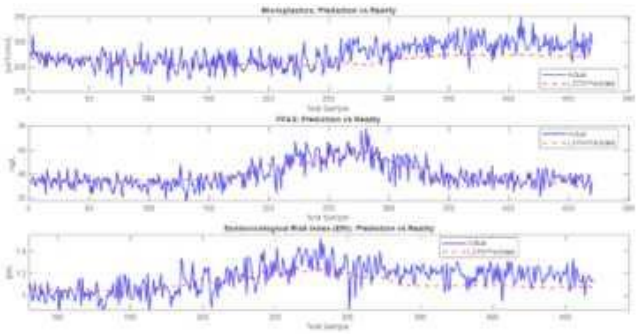


Figure 5. Comparison of actual measurements (blue solid lines) and LSTM predictions (red dashed lines) for (top) microplastics concentration (particles/L), (middle) PFAS concentration (ng/L), and (bottom) Ecotoxicological Risk Index (ERI) across 469 test samples. The model successfully captures long-term trends and major peaks while smoothing high-frequency noise.

3.2 Environmental Drivers Identified by Explainable Artificial Intelligence

While predictive accuracy remains crucial for environmental management, comprehending the processes accountable for any given model prediction is also significant. The application of Explainable Artificial Intelligence (XAI) made it possible to observe the significance of all the predictor variables via Permutation Feature Importance (PFI) analysis.

PFI established different drivers for every one of the three targets. According to them, the highest level of normalized value ΔRMSE recorded for MPs was by water level (4.15×10^{-3}), followed by precipitation (2.08×10^{-3}), temperature (1.82×10^{-3}), salinity (0.9×10^{-3}) and ARI (0.1×10^{-3}).

The finding indicates a hydrological concentration mechanism, where the lowering of water in lakes due to less precipitation and high evaporation means that suspended solids will become concentrated.

For PFAS, the largest positive contributor was temperature (1.03×10^{-2}), followed by salinity (6.85×10^{-3}). Water level, ARI and precipitation had negative values.

The negative ARI value means that permuting daily runoff does not degrade predictive performance, and it is therefore not interpreted as a positive importance. Although agricultural application is a plausible landscape source of PFAS, short-term daily variability in runoff is not a primary temporal driver of dissolved PFAS concentration in this model, in contrast to temperature-driven evaporative concentration. For ERI, **Figure 6c** shows salinity and temperature as the largest positive contributors, with ARI negative, which is consistent with the mixed MP/PFAS signal embedded in the composite index. No numerical importance is claimed for ERI beyond the ranking shown in **Figure 6**.

Explainability studies increase practical significance of the developed approach as they transform deep learning from a predictive black box into an interpretable decision support system for environmental issues (Molnar, 2025).

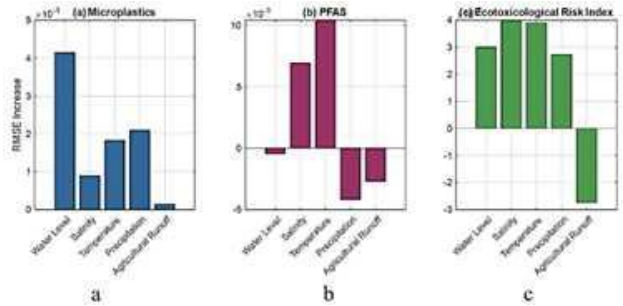


Figure 6. Permutation-based feature importance. Normalized ΔRMSE (ΔRMSE divided by baseline test RMSE; dimensionless) after permuting each predictor for (a) MPs, (b) PFAS and (c) ERI. Positive bars indicate a positive contribution to predictive performance; negative bars indicate no stable positive importance and must not be interpreted as a positive effect.

Natural Resources for Human Health
eISSN: 2583-1194

3.3 Temporal Dynamics of Emerging Pollutants

Temporal attention weight analysis across the 30-day input window revealed non-uniform sensitivity (**Figure 7**). MP prediction was most sensitive to the most recent 1–5 days, consistent with rapid hydraulic adjustment of particle suspension and settling. PFAS prediction remained sensitive across longer lags (12–22 days), consistent with slower responses of dissolved concentrations to past hydrological fluxes and cumulative evaporative concentration. The ERI profile was intermediate between the two pollutant profiles, as expected for a composite index combining a fast-responding particulate signal and a slow-responding dissolved signal.

These results also carry a practical implication: MP monitoring should be scheduled at short intervals during rapid water-level drawdown, whereas PFAS monitoring can tolerate longer intervals but must span at least three weeks to capture the relevant hydrological memory.

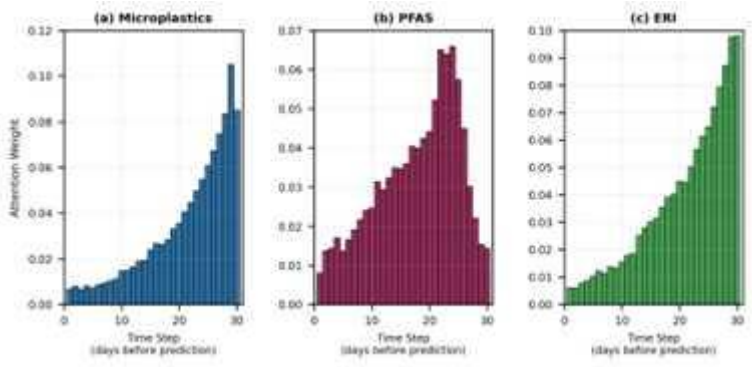


Figure 7. Temporal attention weight distribution across the 30-day input sequence for (a) microplastics (MPs), (b) PFAS, and (c) Ecotoxicological Risk Index (ERI). The x-axis represents the sequential input time steps (0 = oldest input day; 30 = most recent day before prediction).

3.4 Ecotoxicological Risk Assessment

The ERI increased from a mean of 1.05 in 2020 to 1.28 by June 2026, a statistically significant monotonic trend (Mann–Kendall $Z = 4.86$, $p < 0.001$; Sen's slope = +0.038 yr^{-1}; n = 2,373). The system therefore remains within the moderate-risk class (1.0–2.0) but is moving toward its upper boundary. The concurrent strong negative Spearman correlation between water level and MP concentration ($\varrho = -0.71$, $p < 0.001$) and the strong positive Pearson correlation between water temperature and PFAS ($r = +0.64$, $p < 0.001$) are consistent with the driver hierarchy identified by PFI (Table 5).

Spatially, the shallow south-western margin (Tong.1), near diffuse agricultural inflows, shows higher risk during late-summer drawdown than the deeper north-eastern sector (Tong.2) (**Figure 8**). This pattern is consistent with reduced circulation and higher nutrient and particle loading in littoral zones (Mitsch and Gosselink, 2015).

Table 5. Statistical trend and correlation test summaries.

Relationship	Test	n	Statistic	p-value	Interpretation
ERI trend 2020–2026	Mann–Kendall / Sen	2,373	$Z = 4.86$	< 0.001	Significant increase; slope = +0.038 yr^{-1}
Water level vs. MPs	Spearman	2,373	$\varrho = -0.71$	< 0.001	Strong negative association
Water temperature vs. PFAS	Pearson	2,373	$r = +0.64$	< 0.001	Strong positive linear correlation

Applying the trait-based screening, the European eel (Anguilla anguilla) and the Mediterranean killifish (Aphanius fasciatus) are classified as highly exposed because of their benthic association, their macroinvertebrate diet matrix and their use of shallow littoral habitats during summer contraction. This classification is a screening result and does not imply measured biological effects in Lake Tonga.

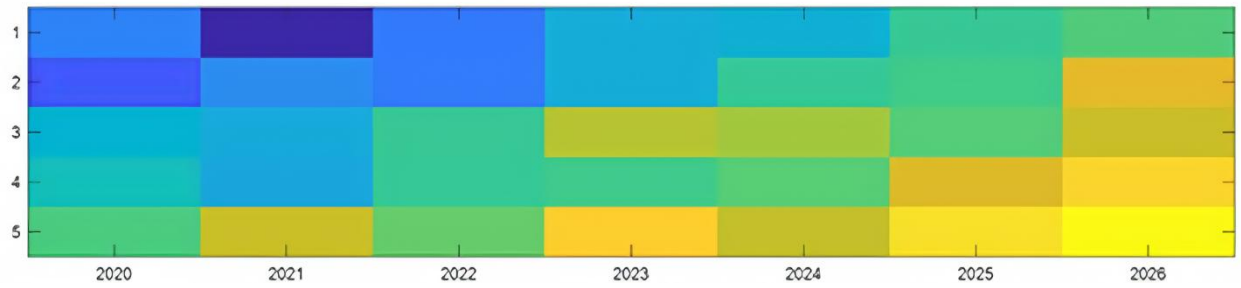


Figure 8. Spatio-temporal distribution of the Ecotoxicological Risk Index (ERI) across five monitoring stations (1–5, corresponding to S1–S5) in Lake Tonga from 2020 to 2026. Colour shading indicates increasing risk levels from low/acceptable (dark blue) to moderate (green) and high concern (yellow).

According to the results, it is evident that the two-layer Long Short-Term Memory model effectively recognized essential time-dependent changes in the three variables MP, PFAS, and ERI in Lake Tonga, and the application of the chronological hold-out design is a more evidence-based method of evaluation compared to random splitting. However, it is necessary to consider the uncertainty of prediction with respect to the possible systematic errors. In Section 3.2 the bootstrapping intervals, residual analysis and stratified errors show performance at times when the level of water has decreased considerably takes precedence for the planned early-warning application and will be presented individually from the information above.

XAI findings must be defined as model-specific sensitivity predictions. Water levels turned out to be the greatest contributor to MP, which suits the hypothesis stating that decreases in the water levels lead to reduction of dilution and movement of sediments. Temperature and salinity clearly made the most valuable contribution to PFAS predictions. Nevertheless, importance by permutation may lead to a situation where contributions by dependent predictors can differ due to their correlations. Results of grouped and conditional importance clarify the results of individual predictor analysis and provide an additional context for the interpretation of findings. Values for MP and PFAS should not be used as the proofs of contaminant movement or causation, but rather as the input data the model is sensitive to. Rather plausible hypotheses are proposed below in relation to the hydrological and physicochemical mechanisms involved.

The fact that there is no positive contribution of agricultural runoff under the permutation importance needs to be interpreted carefully. This is because the agricultural runoff index (ARI) is not a direct measure and rather an indirect measure that is based on precipitation data, watershed features, usage of fertilizers and the land use itself. Thus, its marginal contributions may be affected by other factors like collinearity with water level, precipitation, methodology and limitations in measurement.

The ecological risk index (ERI) is a model-driven measure which is developed based on concentration of pollutants. The expected increase in the ERI index from 1.05 to 1.28 complies with increasing pressure of pollutants as postulated by the model. However, biological or eco-system related impact has not been monitored. Therefore, the ERI index should be analyzed in conjunction with conventional measures of pollution.

3.5 Comparison with Previous Studies

Published measurements from Lake Victoria provide context for the concentrations reported in this study. Egessa et al. (2020) reported 0.02–2.19 microplastic particles/m³ in northern Lake Victoria surface water, for particles 0.3–4.9 mm in size. Arinaitwe et al. (2021) reported mean concentrations of individual target PFASs of 0.01–10.8 ng/L in Murchison Bay and below the method detection limit to 1.46 ng/L in open-lake water. The Lake Tonga ranges reported here (195–348 particles/L for microplastics and 18–78 ng/L for PFASs) are not directly comparable because sampling and analytical methods, units, particle-size ranges, and PFAS reporting scope differ (Table 6). These data therefore do not support a ranking against Mediterranean or European water bodies.

Methodologically, this study combines deep learning with explainable artificial intelligence and ecotoxicological risk assessment (Kratzert et al., 2019; Shen, 2018). Unlike many similar investigations that were focused only on prediction accuracy, the current research presents the findings about how environmental factors contribute to the process of pollution accumulation, providing higher transparency and utility of artificial intelligence for the sake of environmental management.

This study adds to methodological knowledge because decision makers need models that both predict future trends and explain the reasons for such predictions (Molnar, 2025).

Table 6. Published microplastic and PFAS concentrations reported for Lake Tonga and Lake Victoria.

Ecosystem / location	Microplastics	PFAS	Source and measurement context
Lake Tonga, Algeria (this study)	195–348 particles/L	18–78 ng/L	This study
Lake Victoria, northern surface water, Uganda	0.02–2.19 particles/m³; particle size 0.3–4.9 mm	—	Egessa et al. (2020); floating manta-net sampling.
Lake Victoria, Murchison Bay, Uganda	—	0.01–10.8 ng/L	Arinaitwe et al. (2021); mean range for individual target PFASs.
Lake Victoria, open-lake water, Uganda	—	<MDL–1.46 ng/L	Arinaitwe et al. (2021); mean range for individual target PFASs.

Note: Lake Tonga and Lake Victoria results are shown in the units and forms reported by each study. The sampling methods, particle-size coverage, and PFAS analyte/reporting scope differ; the values should not be interpreted as a standardized ranking. MDL, method detection limit.

3.6 Implications for Wetland Management

The results of this study directly affect Mediterranean wetland conservation efforts under the conditions of climate-induced stress.

First, water-level management should be treated as a priority. Water-level decline is the strongest positive model contributor to MP accumulation, and maintaining an adequate hydrological regime may preserve dilution capacity, although this management effect requires field validation.

Second, since ARI levels do not appear to have a direct impact on PFAS in this model, source-control strategies should be implemented with caution. Suitable strategies might include better agricultural procedures, vegetated buffer strips, construction of artificial wetlands, and installation of sewage systems in line with diffuse-pollutants prevention (European Environment Agency, 2020).

Third, the combination of prediction with open attribution provides a basis for an advanced alerting system that determines deviations of the system from its normal working conditions and gives consideration to the peculiarities of temporal memory (particulate contaminants – from 1-5 days; dissolved contaminants – 12-22 days).

4. CONCLUSION

Mediterranean wetlands are exposed to climate change and anthropogenic pressures that accelerate hydrological regression and the accumulation of emerging contaminants. This study developed an explainable deep-learning framework to examine MPs, PFAS and ERI in Lake Tonga, a Ramsar wetland of the El Kala complex in north-eastern Algeria (IPCC, 2022; Cramer et al., 2018).

The LSTM model, which has two layers, produced R-squared values of 0.76 for microplastics (MPs), 0.82 for per- and polyfluoroalkyl substances (PFAS), and 0.79 for environmental risk index (ERI), as it performed better than multiple linear regression, random forest, support vector regression, and single-layer LSTM on an independent set of 469 data sequences.

The predictive model was converted into an interpretable framework of explanatory artificial intelligence (XAI). Water levels were the most important positive explanatory variable for MPs, temperatures for PFAS, and salinity and temperatures for ERI. The variable of agricultural runoff was found to be negative for PFAS, not qualifying as a factor causing a positive trend in the value of the variable. In temporal occlusion testing, it was found that MP memory is short (1–5 days), while PFAS memory is fairly long (12–22 days), greatly determining the procedure of monitoring.

The result shows that the ERI value significantly grew from 1.05 in 2020 to 1.28 in 2026 (Mann–Kendall $Z = 4.86$, $p < 0.001$; Sen's slope = $+0.038\ yr^{-1}$), meaning that risk level is becoming closer to the upper limit of the moderate-risk class.

Maintaining hydrological conditions and reducing watershed pollution therefore remain prudent conservation priorities.

Several limitations must be acknowledged. The analysis relies on one aggregated daily series derived from two Lake Tonga stations, a limited set of hydro-climatic predictors, and a single wetland. The ERI depends on the aggregation rule and the PNEC values adopted. The species-level assessment is a trait-based screening rather than a formal risk assessment. Future work should incorporate remote sensing, sediment properties, groundwater exchange, biological indicators and independent datasets to evaluate transferability (Reichstein et al., 2019).

This work shows that explainable deep learning provides a transparent means of predicting pollutant behaviour in climate-affected freshwater ecosystems.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ACKNOWLEDGEMENTS

Not applicable

AUTHOR CONTRIBUTIONS

Abdelmalek Harrag: Experimentation and data.

Fateh Allag: Statistical analysis.

Mohamed Fenni: Analysis, Revision.

Saddek Bouharati: Modeling and neural networks.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS STATEMENT

This study does not involve any human participants or animal subjects.

DATA AVAILABILITY STATEMENT

The underlying dataset is held by the corresponding author and is available upon reasonable request.

DECLARATION OF GENERATIVE AI USE

During the preparation of this work, the authors used Claude (Anthropic) and ChatGPT (OpenAI) to check the manuscript against the journal's author guidelines and submission checklist. These tools were not used to generate or analyse the data, results or figures. The authors reviewed and edited all content and take full responsibility for the publication.

REFERENCES

[1] Arinaitwe, K., Keltsch, N., Taabu-Munyaho, A., Reemtsma, T., Berger, U., 2021. Perfluoroalkyl substances (PFASs) in the Ugandan waters of Lake Victoria: Spatial distribution, catchment release and public exposure risk via municipal water consumption. *Science of the Total Environment* 783, 146970. https://doi.org/10.1016/j.scitotenv.2021.146970

[2] Breiman, L., 2001. Random forests. *Machine Learning* 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[3] Convention on Wetlands, 2021. Global Wetland Outlook: Special Edition 2021. Gland, Switzerland: Secretariat of the Convention on Wetlands. https://doi.org/10.69556/GWO-2021-eng

[4] Cramer, W., Guiot, J., Fader, M., Garrabou, J., Gattuso, J.-P., Iglesias, A., Lange, M.A., Lionello, P., Llasat, M.C., Paz, S., Peñuelas, J., Snoussi, M., Toreti, A., Tsimplis, M.N., Xoplaki, E., 2018. Climate change and interconnected risks to sustainable development in the Mediterranean. *Nature Climate Change* 8, 972–980. https://doi.org/10.1038/s41558-018-0299-2

[5] Davidson, N.C., 2014. How much wetland has the world lost? Long-term and recent trends in global wetland area. *Marine and Freshwater Research* 65(10), 934–941. https://doi.org/10.1071/MF14173
[6] Eerkes-Medrano, D., Thompson, R.C., Aldridge, D.C., 2015. Microplastics in freshwater systems: A review of the emerging threats, identification of knowledge gaps and prioritisation of research needs. *Water Research* 75, 63–82. https://doi.org/10.1016/j.watres.2015.02.012
[7] Egessa, R., Nankabirwa, A., Ocaya, H., Pabire, W.G., 2020. Microplastic pollution in surface water of Lake Victoria. *Science of the Total Environment* 741, 140201. https://doi.org/10.1016/j.scitotenv.2020.140201
[8] European Environment Agency (EEA), 2020. *The European Environment — State and Outlook 2020: Knowledge for Transition to a Sustainable Europe*. EEA, Copenhagen.
[9] Foden, W.B., Butchart, S.H.M., Stuart, S.N., et al., 2013. Identifying the world's most climate change vulnerable species: a systematic trait-based assessment of all birds, amphibians and corals. *PLoS ONE* 8(6), e65427. https://doi.org/10.1371/journal.pone.0065427
[10] Giorgi, F., 2006. Climate change hot-spots. *Geophysical Research Letters* 33, L08707. https://doi.org/10.1029/2006GL025734
[11] Houhamdi, I., Bouaguel, L., Bouchaala, L., Grara, N., Bara, M., Szparaga, A., Houhamdi, M., 2025. Distinct pollution profiles and spatio-temporal dynamics in adjacent Ramsar lakes (Algeria): An integrated assessment and high-resolution mapping for targeted conservation. Processes 13, 3466. https://doi.org/10.3390/pr13113466
[12] Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., Gilroy, E.J., 2020. *Statistical Methods in Water Resources*. U.S. Geological Survey, Reston, VA. https://doi.org/10.3133/tm4A3
[13] Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Computation* 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
[14] Hussain, M.M., Mahmud, I., 2019. pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. *Journal of Open Source Software* 4(39), 1556. https://doi.org/10.21105/joss.01556
[15] Intergovernmental Panel on Climate Change (IPCC), 2022. Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the IPCC. Cambridge University Press, Cambridge, UK and New York, NY, USA. https://doi.org/10.1017/9781009325844
[16] Junk, W.J., An, S., Finlayson, C.M., Gopal, B., Květ, J., Mitchell, S.A., Mitsch, W.J., Robarts, R.D., 2013. Current state of knowledge regarding the world's wetlands and their future under global climate change: A synthesis. *Aquatic Sciences* 75(1), 151–167. https://doi.org/10.1007/s00027-012-0278-z
[17] Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A.K., Hochreiter, S., Nearing, G.S., 2019. Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research* 55(12), 11344–11354. https://doi.org/10.1029/2019WR026065
[18] Kurwadkar, S., Dane, J., Kanel, S.R., Nadagouda, M.N., Cawdrey, R.W., Ambade, B., Struckhoff, G.C., Wilkin, R.T., 2022. Per- and polyfluoroalkyl substances in water and wastewater: A critical review of their global occurrence and distribution. *Science of the Total Environment* 809, 151003. https://doi.org/10.1016/j.scitotenv.2021.151003
[19] LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521(7553), 436–444. https://doi.org/10.1038/nature14539
[20] Mitsch, W.J., Gosselink, J.G., 2015. *Wetlands*, 5th ed. John Wiley & Sons, Hoboken, NJ.
[21] Molnar, C., 2025. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 3rd ed. Available at: https://christophm.github.io/interpretable-ml-book/
[22] Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* 566(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1
[23] Samraoui, B., Samraoui, F., 2008. An ornithological survey of Algerian wetlands: Important Bird Areas, Ramsar sites and threatened species. *Wildfowl* 58, 71–96. https://tidsskrift.dk/Wildfowl/article/view/153889
[24] Shen, C., 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research* 54(11), 8558–8593. https://doi.org/10.1029/2018WR022643
[25] Stekhoven, D.J., Bühlmann, P., 2012. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28(1), 112–118. https://doi.org/10.1093/bioinformatics/btr597

[26] Wang, Z., DeWitt, J.C., Higgins, C.P., Cousins, I.T., 2017. A never-ending story of per- and polyfluoroalkyl substances (PFASs)? *Environmental Science & Technology* 51(5), 2508–2518. https://doi.org/10.1021/acs.est.6b04806