Journal of Hydrology 649 (2025) 132455

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Establishing correlations between time series of wastewater parameters under extreme and regular weather conditions

Ming Cheng [a], Margherita Evangelisti [a], Sacha Gobeyn [b], Francesco Avolio [c], Dario Frascari [a], Marco Maglionico [a], Valentina Ciriello [a,*], Vittorio Di Federico [a]

[a] Department of Civil, Chemical, Environmental, and Materials Engineering, University of Bologna, Viale Risorgimento 2, Bologna, 40136, Italy
[b] Fluves, Stropkaai 55, Ghent, 9000, Belgium
[c] Hera SPA - Water Direction, Modena, Italy

ARTICLE INFO

This manuscript was handled by Corrado Corradini, Editor-in-Chief.

Keywords:
Extreme rainfall events
Wastewater
Detrended cross-correlation analysis
ARDL
Turbidity
Electrical conductivity
Water level
Sewer systems
Climate change

ABSTRACT

This study investigates the correlations between key wastewater parameters – water level, turbidity, and electrical conductivity – under varying weather conditions, including extreme rainfall events such as the May 2023 flood event in Bologna, Italy. Data collected via IoT-based sensors are analyzed using Detrended Cross-Correlation Analysis and Autoregressive Distributed Lag (ARDL) models. The results highlight significant correlations between water level and other parameters, with distinct patterns emerging during extreme and regular weather periods. Notably, water level correlates negatively with electrical conductivity, particularly during flood events, due to the dilution effect of rainwater. Turbidity shows a complex relationship with water level, influenced by weather conditions and the opposing effects of different factors. ARDL models further demonstrate the potential to predict turbidity and electrical conductivity from water level data, offering valuable insights for wastewater management in urban areas.

1. Introduction

The analysis of extreme events is critical in hydrology for pattern recognition and risk assessments to quantify the implications for the environment and infrastructure (e.g., Focaccia et al., 2021; Felisa et al., 2022; Kohanpur et al., 2023; Hop et al., 2024; Xu et al., 2024). Extreme precipitation events impact sewer systems and urban flooding (e.g., Liu et al., 2024). While progress has been made in safeguarding freshwater resources, there are still significant threats to the quality of receiving water (e.g., Jalliffier-Verne et al., 2017).

Evaluation by the European Commission of the Urban Waste Water Treatment Directive identified stormwater overflows, including urban surface runoff and combined sewer overflows, as major sources of pollution for aquatic environments. Climate change is expected to worsen the impact of these emissions on water resources, with more frequent intense rainfall and severe droughts. In this context, innovative solutions are needed to efficiently and cost-effectively reduce pollution from urban runoff and sewer overflows (e.g., Rosenberger et al., 2021; Wang et al., 2023).

Here, we focus on the sewer system of Bologna in the Emilia-Romagna Region of Italy, and analyze the system response to regular weather conditions and the exceptional extreme rainfall event that happened in May 2023. Specifically, from May 16th to 18th, 2023, the Emilia-Romagna Region experienced its second severe flooding event in two weeks, following the first inundation on May 2nd. Both flooding events were caused by intense and continuous rainfall over the Region, including the city of Bologna. The impact on water levels in the southern-eastern part of the Region was significant, with some rivers reaching their record high levels (eGEOS, 2023).

In the context of the European Project StopUP - *Protecting the aquatic environment from urban runoff pollution*, financed by the EU under the Horizon Europe Programme, data loggers based on the Internet of Things (IoT) were installed on May 8th, 2023 and were able to record the impact of the event. These sensors monitor continuously, with a sampling frequency of fifteen minutes, three key wastewater parameters, namely water level, turbidity, and electrical conductivity at the outlet of the 800 km long combined sewer system of Bologna, which is also the inlet of the Bologna wastewater treatment plant (WWTP). The installation of the sensors, developed by the company FLUVES (a partner of the Project), follows up-to-date best practices in environmental monitoring. Moreover, continuous measurements of the

* Corresponding author.
E-mail address: v.ciriello@unibo.it (V. Ciriello).

https://doi.org/10.1016/j.jhydrol.2024.132455
Received 12 September 2024; Received in revised form 14 November 2024; Accepted 15 November 2024
Available online 5 December 2024

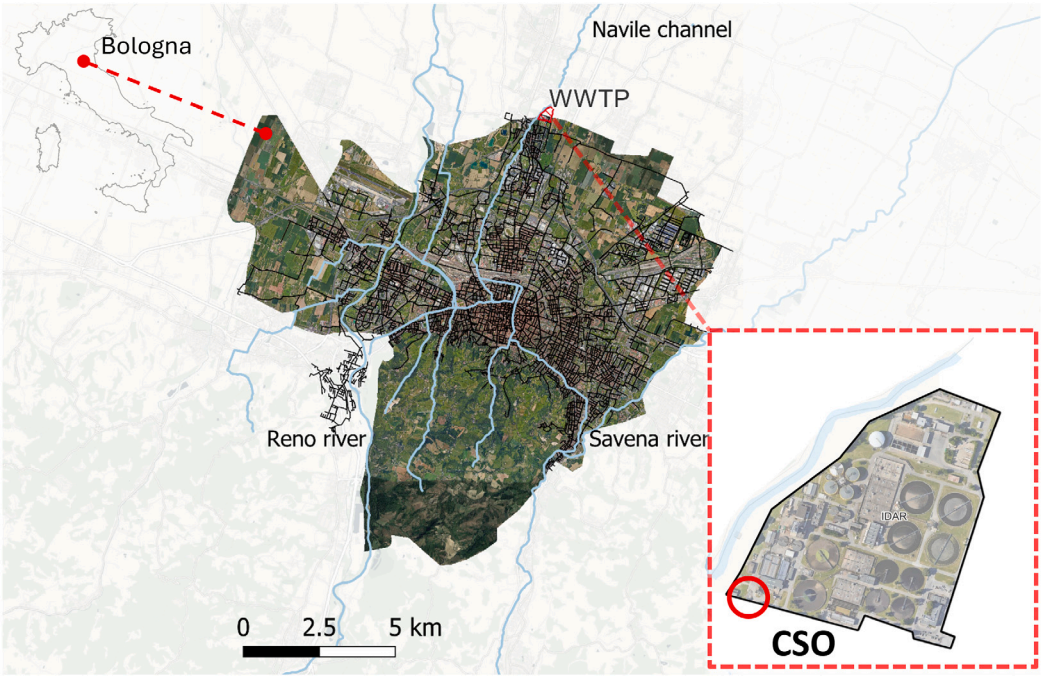

Fig. 1. Map of the sewer system of Bologna (Italy, latitude 44.495, longitude 11.343): main watercourses, combined sewer network and WWTP.

wastewater parameters allow the detection of the complete dynamics of hydrographs and pollutographs, which is the basis to evaluate the full variability of flow and pollutant concentrations (Gruber et al., 2006).

The availability of these data and the extent and importance of the rainfall event (and associated flooding) have suggested two avenues of investigation. These general goals are of importance for sewer system modeling in general: (i) assessing the impact of the rainfall event on the tributary basin of the main outlet of the sewer system in Bologna, to determine if such an event could pose a threat, as has been the case for larger river basins experiencing critical conditions; (ii) statistically evaluating, under extreme and regular weather conditions, the correlations among the wastewater parameters and between them and rainfall, with the additional scope of predicting some of these quantities, particularly turbidity and electrical conductivity, based on the water level.

Specifically, we use cross-correlations to examine the relationship between time series. Since non-stationary and noisy time series are ubiquitous in the real world, we adopt detrended cross-correlation analysis (DCCA) to reduce the impact of non-stationary features and noise (Podobnik and Stanley, 2008). DCCA helps us study the local and global fluctuations in the series and explore long-range cross-correlations using power laws. In contrast to standard DCCA, we use higher-order polynomials to approximate the local trend and identify the non-linearity inside the time series (Shadkhoo and Jafari, 2009; Horvatic et al., 2011; Marinho et al., 2013). Additionally, to quantify the lag effect, we analyze the cross-correlation between lag time series (Shen, 2015), and we develop a linear prediction by Autoregressive Distributed Lag (ARDL) models (Nkoro et al., 2016; Pesaran et al., 1995), in dry and wet seasons, of turbidity and electrical conductivity based on the water level.

To the best of our knowledge, there was a lack of studies in this field that employed these data-driven tools to interpret the relationships between hydrological variables and water quality parameters based on (continuous) time series data collected by IoT sensors. If a comprehensive literature review on the relationship between hydrological and water quality parameters at the sewer system level appears to be lacking – this gap seems to be present also at larger spatial scales like river basins – specific studies addressing this topic are abundant; relevant examples are Reisinger et al. (2019) and Mahaut and Andrieu (2019) for the cities of Baltimore and Nantes respectively. Related works for river systems are (Prathumratana et al., 2008) and Navratil et al. (2020). At both the sewer and river scales, the methodologies range from simple regression analysis to direct simulations of flow and transport equations, making it challenging to compare directly with the present results.

The structure of the paper is as follows: Section 2 presents the case study, the experimental set-up at the WWTP, and the methods, including the stationary Generalized Extreme Value (GEV) distribution and Cross-Correlation Analysis (CCA) with Detrended Fluctuation Analysis (DFA). Section 3 illustrates the results and discusses their physical meaning and implications, and a set of conclusions and perspectives closes the paper (Section 4).

2. Materials and methods

In this Section, we provide detailed information about the case study involving the sewer system serving the City of Bologna. This includes the experimental setup and the sensors used to collect data on water level, electrical conductivity, and turbidity at the inlet of the Bologna WWTP. We also describe how a database is created to store this data along with rainfall information. Additionally, we discuss the use of the stationary GEV distribution to study extreme events and the statistical tools employed for CCA among the different time series collected by the sensors.

2.1. Case study

The City of Bologna, located in the Emilia-Romagna Region of Italy, has an urbanized area of 4700 ha and is drained by a combined sewer system having a total length of 800 km, see Fig. 1. The WWTP serving the city, with a population equivalent of 600,000, is located on the northern boundary of the served basin. Both the network and the plant are operated by the multi-utility Hera S.p.A. (partner of the StopUP project). Overflows occurring in wet weather periods are received by three main watercourses (Reno River, Savena River, and Navile Channel) through 123 spills, or combined sewer overflows (CSOs). In particular, during rainfall events, the Navile Channel receives the overflow before the inlet to the WWTP. The case study domain includes the

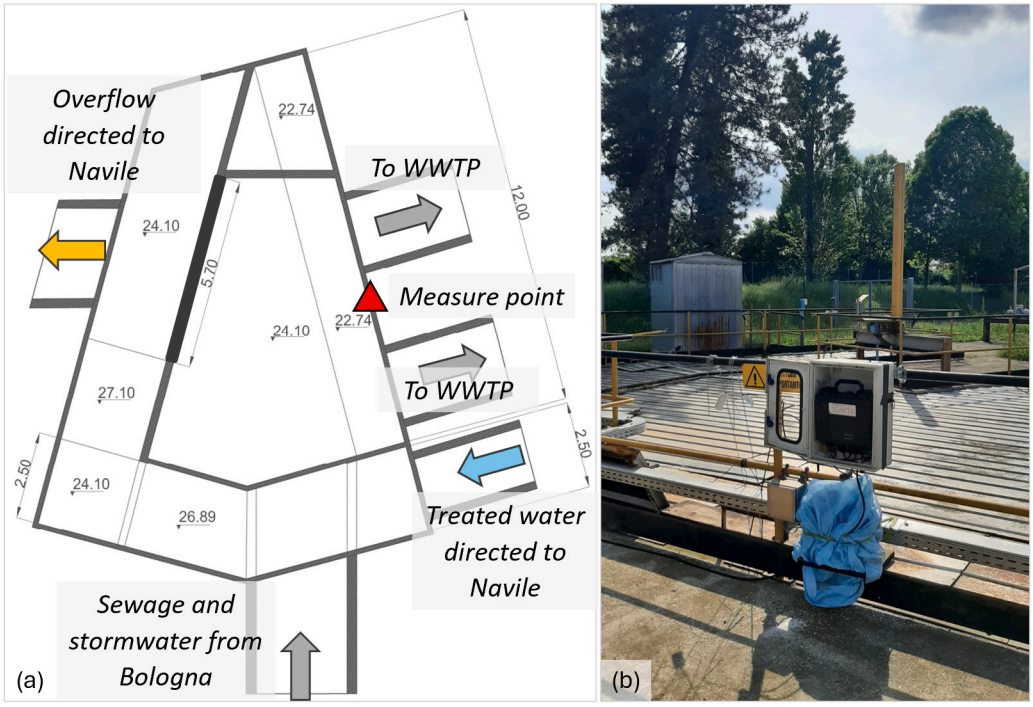

Fig. 2. (a) Top view of the manufact and location of sensors at the inlet of the WWTP (red point). (b) View of sensor's datalogger set-up. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

main CSO of Bologna. Indeed, based on long-term hydraulic simulations developed in Maglionico et al. (2009) on the whole sewer network, the CSO just upstream of the inlet to the WWTP is the main contributor in terms of the amount of mass and volumes discharged every year.

The measuring point is located in the pipe before the inlet to the WWTP which drains sewage and stormwater (Fig. 2). Installation in the CSO itself was impractical since the collector transports both spills during rainfall events and the treated water by the WWTP, see Fig. 2a. However, the continuous measure of the water level in the pipe provides information on verified overflows and their relative volume.

The data collection setup is based on the Internet of Things (IoT). It involves connecting turbidity (Aqualabo NTU, 0–4000 NTU, < 5% accuracy) and electric conductivity (Aqualabo C4E, 0–200 ms/cm, 1 < 5% accuracy) sensors to an IoT logger. Turbidity is measured using optical reflection at a 90° angle, while electric conductivity is measured using a titanium and graphite electrode. A relative pressure sensor (STS PTM/N, 0–5 m, < 0.1% accuracy) measures the water level. The data logger is responsible for processing the signals from each sensor and sending the data via LTE-M (fallback: NB-IoT) to a (MQTT-) endpoint. Measurements are taken every fifteen minutes. This setup was installed in early May 2023, and the data collected for the first 8 months, from mid-May to the end of December 2023, are analyzed in Section 3.

The pluviometric regime is analyzed using data recorded by the weather station "Bologna Idrografico", managed by the Regional Agency ARPAE. Table 1 lists the main characteristics of rainfall and temperature registered for the 8 months investigated. The cumulative rainfall depth ranges from 5.4 mm in December to 347.9 mm in May against an average of 75.8 mm.

Concurrently, high-time resolution data of electrical conductivity, turbidity, and water level collected by the sensors are subjected to a validation process and to the detection of outliers (Bertrand-Krajewski et al., 2021). Fig. 3 shows boxplots of turbidity and electrical conductivity for the eight months analyzed. In May 2023, a large number of outliers in the turbidity data highlights the exceptional nature of the events that occurred. As mentioned in Razguliaev et al. (2024), these measurements have been used as a "proxy" for monitoring various pollutants. Previous studies have shown that turbidity can be a reliable indicator for indirectly measuring other pollutants. During rainy periods, numerous pollutants in CSOs are attached to suspended particles in water (Rossi et al., 2005; Langeveld et al., 2005).

2.2. Stationary GEV distribution

Climate change favors extreme events that may impact sewer systems. Focusing on the precipitation depth, different methods can describe the precipitation distribution, such as the GEV distribution, and the threshold model (Coles et al., 2001; Katz et al., 2002; Cheng et al., 2014; De Paola et al., 2018). In this work, we have used the stationary threshold model based on the GEV distribution to estimate the probability density function (PDF) and the cumulative distribution function (CDF) of rain events featuring a precipitation depth d higher than a threshold u, during a rain sampling period that can range from 1 to 6 h. Thus, the conditional CDF is expressed as:

$$P_F(d';\mu,\tilde{\sigma},\gamma|d>u)=1-(1+\frac{\gamma d'}{\tilde{\sigma}})^{-\frac{1}{\gamma}}, \quad (1)$$

where $d' = d - u$ indicates the excess rain depth above the threshold u, μ is the location parameter of GEV distribution, $\tilde{\sigma}=\sigma+\gamma(u-\mu)$, σ is the scale parameter and γ decides the shape of distribution. The maximum likelihood method was used to find the optimal parameters μ,σ,γ.

According to the threshold model (Katz et al., 2002), the negative loglikelihood function can be derived from Eq. (1) as

$$-\log P_f(d;\tilde{\sigma},\gamma)=\log(\tilde{\sigma})+(1+\frac{1}{\gamma})\log(1+\frac{\gamma(d-u)}{\tilde{\sigma}}) \quad (2)$$

$$1+\frac{\gamma(d-u)}{\tilde{\sigma}}>0, \qquad \text{for all} \quad d, \quad (3)$$

where the sample is only taken when the rainfall depth exceeds the threshold $d > u$. The relationships between these parameters are presented below:

$$\tilde{\sigma}=\sigma+\gamma(d-\mu). \quad (4)$$

Table 1
Thermo-pluviometric regime at the "Bologna Idrografico" weather station in May-December 2023.

Month	Cumulative rainfall depth (mm)	Max Rainfall Intensity (mm/h)	Average Temperature (°C)	Max Temperature (°C)	Min Temperature (°C)
May	347.9	52.8	18.5	30.3	11.3
June	45.1	24.8	24.5	37.1	17.0
July	40.6	29.2	28.1	39,2	17.4
August	33.5	11.6	27.0	40.1	16.4
September	27.9	9.6	23.8	33.8	14.5
October	64.5	16.8	19.9	32.4	10.2
November	41.2	11.2	11.4	21.8	3.4
December	5.4	1.6	8.7	22.4	1.8

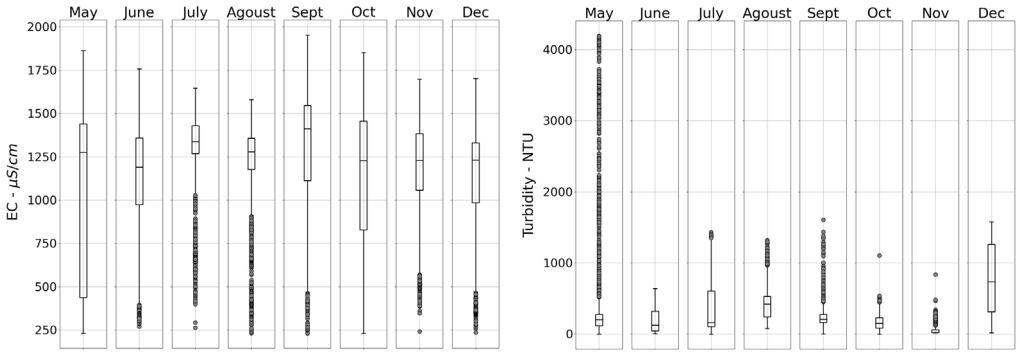

Fig. 3. Boxplots of electrical conductivity (left panel) and turbidity (right panel) data in May-December 2023.

2.3. Detrended cross-correlation analysis

To study the correlation between any two different time series among the four considered (i.e., rainfall, water level, electrical conductivity, and turbidity), we regard each time series as a random process. DFA (Jun et al., 2006) is a commonly used tool for analyzing the auto-correlation of time series. Standard cross-correlation methods assume the time series to be stationary and ignore non-stationary trends, which can lead to inaccurate correlation values. When studying the unknown correlation between time series with non-stationary features, we apply DCCA (Podobnik and Stanley, 2008), which can account for non-stationarity effects. To incorporate periodic features and capture higher-order terms for power series, high-order polynomials to approximate the local trend (Horvatic et al., 2011) are used. DCCA has been successfully applied in hydrology and proven to be effective in time series analysis (Adarsh et al., 2020; Nogueira, 2019).

The two time series of interest are denoted by $\{x_i\}_{i=1}^N$ and $\{y_i\}_{i=1}^N$. They can be any combination of turbidity (Nephelometric Turbidity Unit, NTU), rainfall depth (mm), electrical conductivity (µS/cm), and water level (mm). Before adopting DCCA methods, we first need to define the increment series R_x and R_y as

$$R_x^k = \sum_{i=1}^{k} x_i, \quad R_y^k = \sum_{i=1}^{k} y_i, \tag{5}$$

where $k = 1, 2, \dots, N$. Then, the whole time series is divided into $N - n$ overlapping boxes. Each box includes the $n + 1$ points sequence, R_x^i to R_x^{i+n}, and in each box the 'local trend' is computed by polynomials $P_l(R_x)$ or $P_l'(R_y)$ approximated by least squares fitting, i.e.

$$\arg\min_{P_l} \sqrt{\frac{1}{n+1}\sum_{k=i}^{i+n} \|R_x^k - P_{l,k,i}\|^2}, \tag{6}$$

where $P_{l,k,i}$ represents the polynomials at the kth point defined on the interval starting from the ith to the $i+n$-th points with the highest order l. Then, the residual for each box, f^2_{DCCA-l}, is defined as

$$f^2_{DCCA-l}(n,i) \equiv \frac{1}{n-1}\sum_{k=i}^{i+n}(R_x^k - P_{l,k,i})(R_y^k - P'_{l,k,i}), \tag{7}$$

and adding up the residual for all boxes, we obtain the cross-correlation F^2_{DCCA-l}, reading

$$F^2_{DCCA-l} = \sum_{i=1}^{N-n} f^2_{DCCA-l}(n,i). \tag{8}$$

Subsequently, the cross-correlation coefficient ρ is introduced as

$$\rho(n) = \frac{F^2_{DCCA-l}(n)}{F_{DFA}(n)F'_{DFA}(n)}, \tag{9}$$

where the definitions applied for the DFA are similar to those for the DCCA, e.g.

$$F^2_{DFA}(n) = \sum_{i=1}^{N-n} f^2_{DFA-l}(n,i), \tag{10}$$

$$f^2_{DFA-l}(n,i) \equiv \frac{1}{n-1}\sum_{k=i}^{i+n}(R_x^k - P_{l,k,i})^2. \tag{11}$$

By definition, $-1 \le \rho \le 1$. Its sign indicates whether there is a positive or negative correlation between the selected time series (rainfall, electrical conductivity, turbidity, and water level), while its absolute value indicates the strength of the relationship between the two time series.

3. Results and discussion

3.1. Preliminary analysis of extreme rainfall events

Here, the methodology described in Section 2.2 is applied to the rainfall time series. The considered time series consists of rainfall depth measured at the weather station "Bologna Idrografico", with a 15-minute time step resolution, from January 1990 to December 2022 (https://simc.arpae.it/dext3r/). These data are compared with those referred to the flood event occurred on May 9th–17th 2023.

The maximum rainfall depths in historical data are gathered at frequencies ranging from 1 h to 24 h. The average value of the time series, for the different frequencies, falls within the range of 0.5 to 1.4. The threshold, represented as u, is designated to 0.85 mm, which corresponds to the midpoint of that interval. The optimal parameters

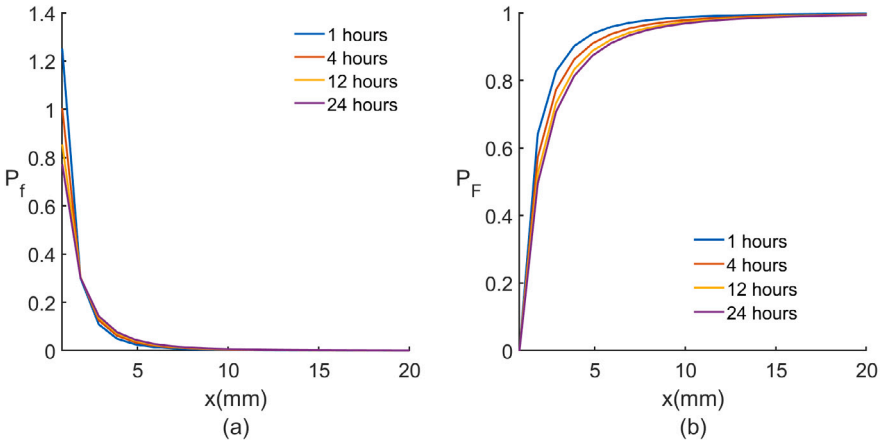

Fig. 4. (a) PDF and (b) CDF given by the GEV distribution, for sampling frequencies in the range 1–24 h.

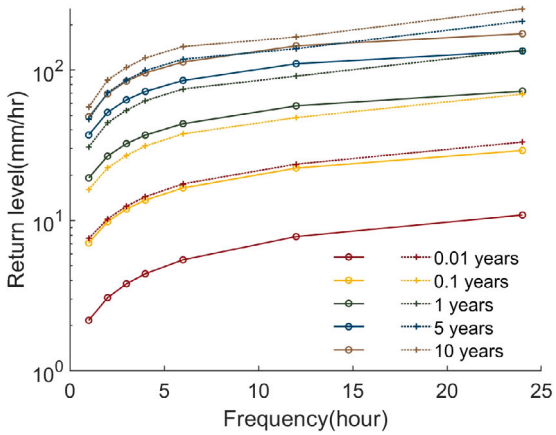

Fig. 5. Maximum rainfall intensity value (return level) against the sampling frequency for return periods in the range 0.01–10 years. Bold lines are referred to historical data, dashed lines to flood event data.

are determined by solving the maximum likelihood function using the Quasi-Newton method. Fig. 4 depicts the PDF and CDF for six different sampling frequencies. It is seen that the sampling frequency affects marginally the probabilistic behavior of the maximum rainfall depth.

Referring to the Poisson–GP model, the probability of occurrence of a maximum rainfall depth exceeding u can be approximated as

$$P(d > u) \approx \mathrm{Num}(d > u)/\mathrm{Num}(d), \tag{12}$$

while, for a given return period, the correspondent maximum rainfall depth value exceeding u is (see Coles et al., 2001)

$$d = u + \frac{\sigma}{\gamma}([N n_s P(x > u)]^{\gamma} - 1), \tag{13}$$

where n_s is the sampling frequency and N(year) is the return period.

Fig. 5 shows the maximum rainfall value (intensity) as a function of the sampling frequency for different return periods. It can be first observed that the intensity increases markedly with the return period as expected (the different curves). The intensity also increases with the sampling frequency, first in a nonlinear fashion and then, for values of sampling frequency around 5–6 h, with a nearly linear trend with a generally low rate of increase. Results are provided both for historical data (bold curves) and data associated with the flood event (dashed curves). The latter curves are always higher than their historical counterparts, highlighting the exceptional nature of the flood event; the difference between each pair of curves decreases with increasing return time.

3.2. Cross-correlation among time series

Here, the application of DCCA reveals the cross-correlation between the water level and the other parameters at the measurement station, namely the electrical conductivity and turbidity, and the rainfall depth. The rationale for this choice is that (i) a prediction of the water level based on the rainfall depth is of general hydrologic interest and is relevant for the management of the WWTP, and (ii) water level is much easier to measure than turbidity or electrical conductivity, and deriving the latter quantities from water level is of technical interest for the WWTP management.

First, we select three classical correlation models to eliminate the non-stationary trend in the time series, bringing them to a constant mean value by the detrending process (to enable auto-correlation computing). Let us denote with P_l a polynomial with the highest order equal to l. Hence, the constant polynomial, $P_0 = a_0$, linear polynomial, $P_1(Y) = a_0 + a_1 Y$, and quadratic polynomial, $P_2(Y) = a_0 + a_1 Y + a_2 Y^2$, are employed to approximate the local trend in (7) where a_0, a_1, and a_2 are constant coefficients and Y can be any accumulation times series of rainfall depth, electrical conductivity, turbidity and water level.

Fig. 6 illustrates the variation of the detrended cross-correlation coefficient between the water level and (a) electrical conductivity, (b) rainfall depth, (c) turbidity during the dry season, and (d) turbidity in the wet season, as a function of the time scale n. The average rainfall depth defines the dry/wet season in this context. For example, if the average rainfall over the whole period considered is 0.4 mm/day, the value associated with the flood event is 2.85 mm/day, while in June (wet season) is 0.96 mm/day. No lag effects are considered here. Results for different polynomials are not markedly different. For example, we observe minor variations for the electrical conductivity-water level correlation, where higher-order polynomials suggest a less negative correlation coefficient, and for the turbidity-water level correlation in the wet season, where higher-order polynomials indicate a less positive correlation.

The cross-correlation between water level and the other variables exhibits interesting results. In particular, the negative cross-correlation between electrical conductivity and water level becomes already significant at a time scale of 10^2 h. This negative cross-correlation, already detectable at relatively small time scales, is physically explained by considering that rainwater acts as a dilution factor (Bersinger et al., 2015). In this regard, Khamis et al. (2017) uses electrical conductivity to proxy the duration of hydrological events. While the cross-correlation between water level and precipitation is univocal, with values above 0.5 already for $n < 10^2$ h, thus denoting the impact of rainfall events on the selected hydraulic system at any time scale, the interrelation between turbidity and water level is not.

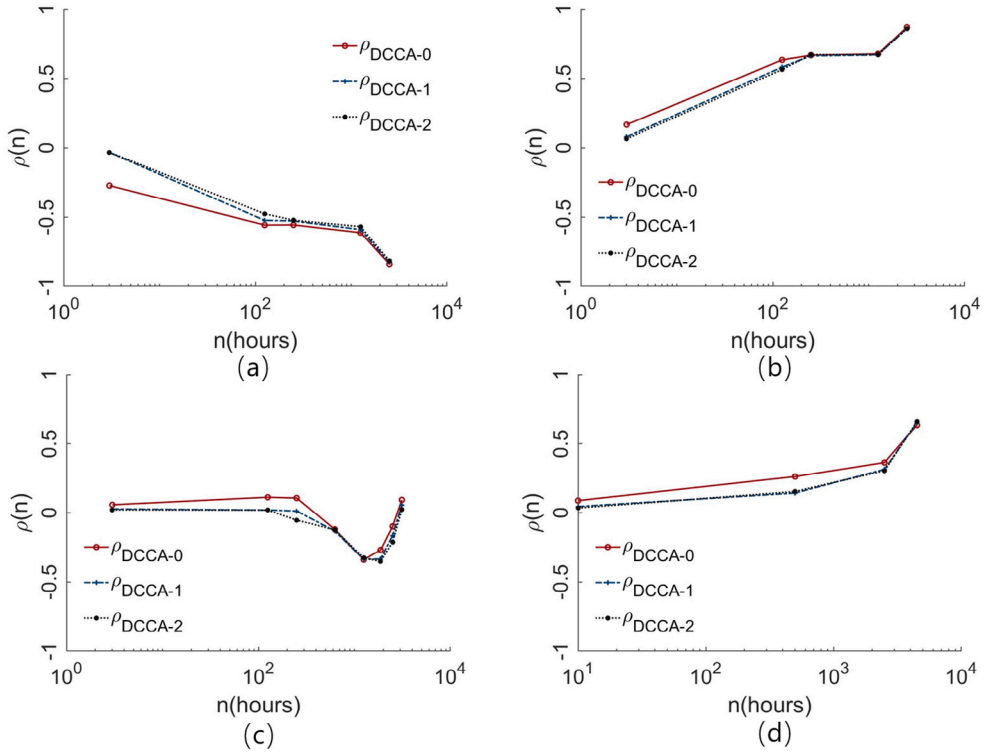

Fig. 6. Cross-correlation coefficients as a function of box size n (hours) for water level and (a) electrical conductivity, (b) rainfall depth, (c) turbidity in the dry season, and (d) turbidity in the wet season.

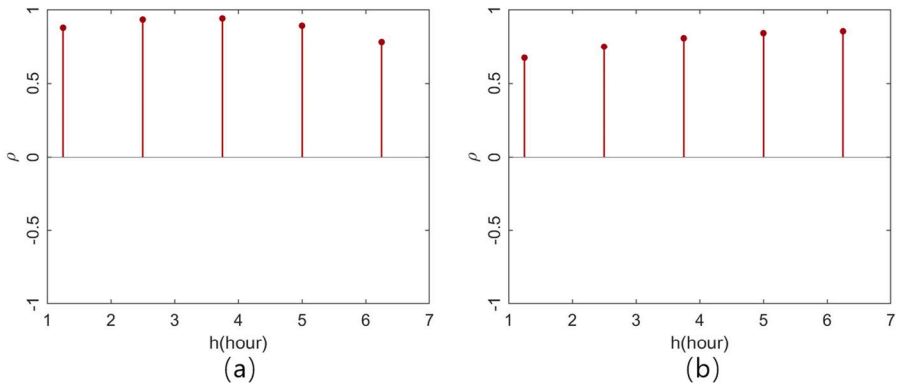

Fig. 7. Correlation coefficients between water level and rainfall depth vs lag h (hours) in (a) May 9th–13th (flood event) (b) June 1st–5th (regular wet season).

Turbidity, as defined in Anderson (2005), refers to the cloudiness or muddiness of water, and it is a measure of the optical property that causes light to scatter and be absorbed by the water/solids suspension. According to continuous monitoring conducted by a study in Pau, southern France (Bersinger et al., 2015), a peak in turbidity was observed in the sanitation network of an urban catchment during rainfall events, coinciding with an increase in flow. However, the magnitude of the peaks varied from one event to another. In this regard, according to Fig. 6, the cross-correlation between turbidity and water level presents two different behaviors over dry and wet seasons. During the wet season, turbidity has a weak positive association with water level, until the time scale is $n < 10^3$ hours, and then it increases rapidly with n above 0.5. During the dry season, there is no clear evidence of a cross-correlation between the time series.

This first part of our analysis highlights the importance of investigating the cross-correlation between time series with a multi-time scale approach to quantify scale-dependent correlations. Our results show that the correlation between water level and (a) electrical conductivity is negative at any time scale and increases with n, (b) rainfall depth is positive at any time scale and increases with n, (c) turbidity during dry season is sensibly zero until $n < 10^2$ h and then exhibits opposite trends with n with no evident characteristics, (d) turbidity during the wet season is always positive but not significant unless $n > 10^3$ h.

After obtaining these initial results, we further examined the correlations between the quantities of interest by analyzing different time lags (h) in the time series data. For this analysis, we used an average box size of 120 h for the DCCA, i.e., a fixed time scale. The investigation is developed for a five-day time frame, identified in three distinct periods: (a) May 9th–13th, 2023, representing the first flood event, (b) June 1st–5th, 2023, which served as a typical period of wet weather, and (c) August 15th–19th, 2023, representing dry weather. We assessed the correlation for time lags ranging from 1.25 to 6.25 h (multiples of 1.25 h).

Regarding the relationship between water levels and rainfall depth, Fig. 7a illustrates that the correlation coefficient varies slightly with h, exhibiting a maximum at 3.75 h. This nearly constant correlation

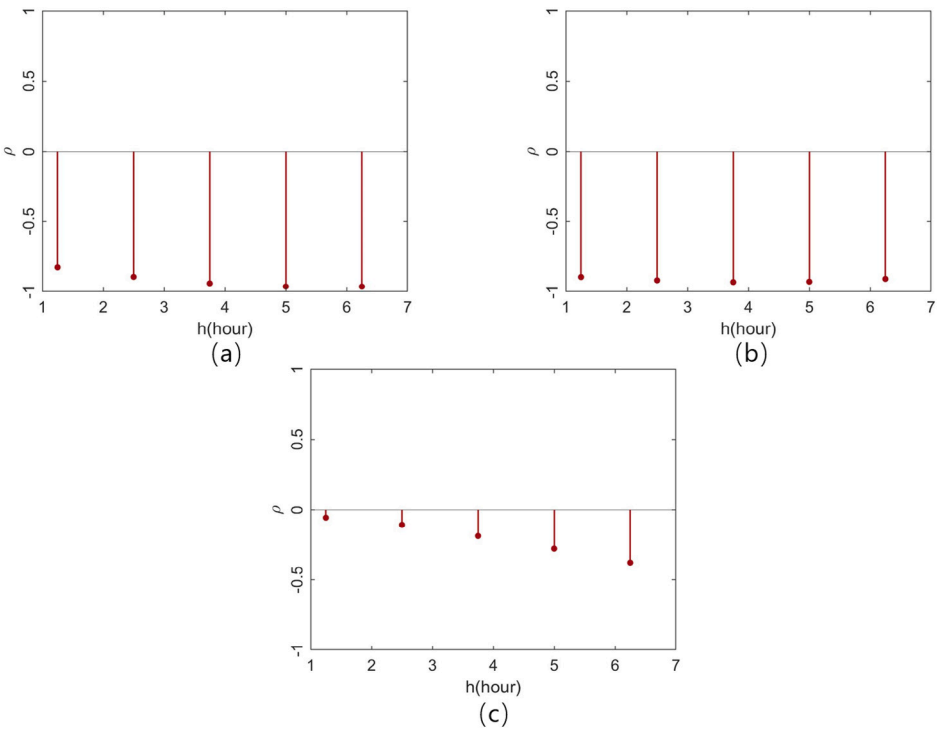

Fig. 8. Correlation coefficients between water level and electrical conductivity vs lag h (hours) in (a) May 9th–13th (flood event) (b) June 1st–5th (regular wet season) (c) August 15th–19th (regular dry season).

coefficient, close to one, is primarily due to our focus on the five days associated with the flooding event. During this period, the continuous rainfall enhances the correlation and reduces the impact of any delay. In contrast, when analyzing a regular wet period, as shown in Fig. 7b, the correlation increases with the lag, and the maximum occurs at 6.25 h. Here, the correlation coefficient exceeds 0.5 even at $h = 1$, which aligns with previous results for $n = 10^2$ h (see Fig. 6b). The continued increase in the correlation with h indicates the specific delay time for the selected basin relative to a single event. Indeed, the water level typically increases a few hours after the rain event has occurred in basins of limited size, such as the one considered here.

The analysis of electrical conductivity and water levels reveals a negative correlation, as illustrated in Fig. 8a–b. The correlation coefficients reach a peak at 5–6.25 and 3.75–5 h, respectively. However, the trend of ρ against h remains relatively constant, especially in the latter case. Thus suggesting that the decrease in electrical conductivity closely follows an increase in water level, even though the highest correlation occurs after a few hours. During a regular dry period, Fig. 8c, the correlation coefficient increases with h but is less negative than the previous cases, staying below 0.5 in absolute value. To interpret these results, we need to consider that an increase in water level reflects a rise in flow rate, which enhances sediment transport capacity. Conversely, a higher flow rate leads to greater dilution, resulting in lower electrical conductivity. The negative correlation observed in Fig. 8a–b indicates that the dilution effect predominates during the wet season, while its influence diminishes during the dry season.

Finally, regarding the correlation between turbidity and water level, Fig. 9a shows that ρ is negative, and remains consistently below 0.5 in absolute value. It does not vary significantly with h. Conversely, Fig. 9b shows that ρ is positive during regular wet periods, as expected, and increases with h, although it remains below 0.5. In Fig. 9c, ρ is negative and increases in absolute value with h, reaching a maximum of about −0.5. During the regular wet season shown in Fig. 9b, it is known that transport capacity increases with flow rate, resulting in higher turbidity levels. The value of ρ for $h = 1$ is positive, consistent with previous results (Fig. 6d), but the further increase in ρ with higher h is insufficient to establish a relevant correlation. During the dry season, depicted in Fig. 9c, transport capacity is restricted by low flow rates. In particular, ρ is nearly zero for $h = 1$, aligning with earlier results (Fig. 6c), and the decrease in ρ with increasing h is not sufficient to determine a high correlation in this scenario either. In the case of the flood event shown in Fig. 9a, the negative value of ρ is sensibly constant with h and below 0.5 in absolute value. This is likely due to a significant increase in dilution capacity, which remained relatively constant during the five days considered, reflecting the extreme nature of the event.

3.3. Autoregressive distributed lag models for the prediction of the time series

Here, we analyze short-term correlations between the time series. According to the results of DCCA, at least linear approximations exist between turbidity, electrical conductivity, and water level. As such, we develop a linear prediction by ARDL models in dry and wet seasons (Nkoro et al., 2016; Pesaran et al., 1995). While there are various techniques for time series analysis, such as recurrent neural networks (Connor et al., 1994; Hewamalage et al., 2021), ARMA neural networks (Hwarng and Ang, 2001; Hwarng, 2001), and Gaussian processes (Roberts et al., 2013; Seeger, 2004; Brahim-Belhouari and Bermak, 2004), these methods demand significant computational resources for training and are more sensitive to overfitting. In contrast, the ARDL model is straightforward to deploy and is well-suited for online prediction since its complexity is comparable to solving a least squares problem.

We assume different models for each month from May to August, which are characteristic of the flood events (May), regular rainy period (June), and regular dry period (July and August). The first half of each month is used as the training set to compute the coefficients for the linear model. Once the coefficients are derived, the model is applied to predict the electrical conductivity and turbidity for the rest of the month as functions of the water level. To avoid overfitting, we set the number of parameters within the ARDL model to no more than 10, with at least a thousand observations sampled for each 15 min.

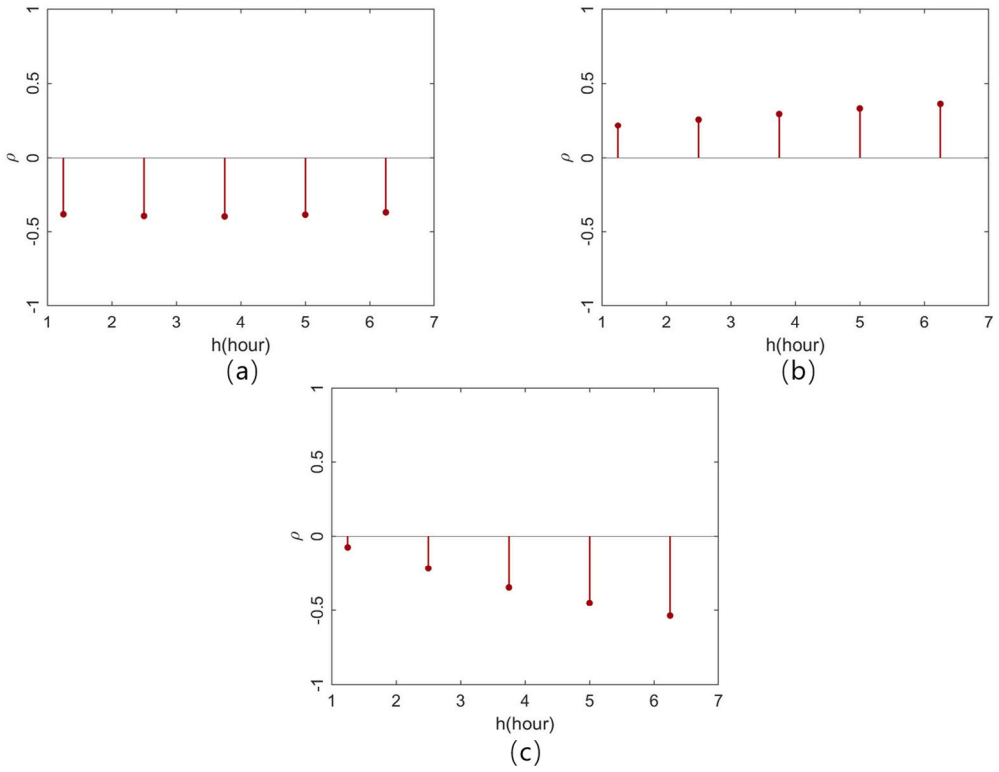

Fig. 9. Correlation coefficients between water level and turbidity vs lag h (hours) in (a) May 9th–13th (flood event) (b) June 1st–5th (regular wet season) (c) August 15th–19th (regular dry season).

Table 2
l_1 norm and R^2 of ARDL models for electrical conductivity (EC) and turbidity (Tur).

Month	l_1 EC	R^2 EC	l_1 Tur	R^2 Tur
May	28.89	0.9769	46.79	0.9783
June	25.28	0.9709	12.28	0.9774
July	18.32	0.9696	61.63	0.8901
August	24.21	0.9445	22.98	0.9699

The target time series to predict is denoted as Y, either turbidity or electrical conductivity. The variable time series is X which is the water level. The maximum lags for the target and variable are defined as p and q. The general ARDL model reads as

$$\Phi(L)Y_y = \phi + \Gamma(L)X_t + \epsilon_t \tag{14}$$

where $\Phi(L) = 1 - \Phi_1 L - \cdots - \Phi_p L^p$ and $L^p Y_t = Y_{t-p}$. $\Gamma(L) = \gamma_0 - \gamma_1 \Gamma^1 - \cdots - \gamma_q \Gamma^q$ and $\Gamma^q = X_{t-q}$. ϕ is the constant coefficient. ϵ_t is the i.i.d white noise $\epsilon_t \sim \mathcal{N}(0, 0.01)$.

With measurements taken every 15 min, we set for turbidity and electrical conductivity $p = 5$ and for water level $q = 5$ (i.e., 1.25 h), and compute the linear combination coefficients by the least square method.

To evaluate the prediction error, we introduce l_1 norm as $e_{l_1} = \frac{1}{n}\sum_{i=1}^{n} |Y_{pred} - Y_{true}|$ where n is the total length of time series. Table 2 collects l_1 and R^2 of ARDL models for electrical conductivity and turbidity.

The comparison between actual and predicted time series for the four months is illustrated in Figs. 10–13. The predictions from the ADRL models are generally accurate with R^2 values equal to or above 0.9 in all cases. By calculating the cross-correlation coefficient, as outlined in Eq. (9), we observed that the correlation between water level and electrical conductivity is consistently negative across the four selected months, which is expected. Furthermore, this negative correlation increases in absolute value from dry to wet conditions. In this short-term analysis, the cross-correlation coefficient for the relationship between water level and turbidity shows varying signs and magnitudes in absolute value, which does not provide a clear indication, as noted in previous analyses.

Finally, we use a fast Fourier transform to get the frequency spectrum of the water level time series in May, June, and July (see Fig. 14). The results indicate that the highest frequency observed in June and July is approximately the same and higher than in May. This increase in oscillation frequency suggests that the linear model may lose accuracy. Several factors may contribute to this phenomenon, including hydraulic operations at the outlet of the sewer system.

4. Conclusions

This study offers a comprehensive, data-driven analysis of wastewater parameters in response to both extreme and regular weather conditions in Bologna's sewer system. DCCA reveals significant correlations between water level and electrical conductivity. Notably, there is a negative correlation during flood events due to rainfall dilution. Turbidity displays a seasonally dependent correlation with water level, showing weak positive correlations during wet seasons and negative correlations during dry periods. The analysis of time lags highlights the temporal dynamics of these correlations, with peak values varying based on weather conditions. Additionally, the study employs ARDL models to demonstrate the potential for predicting turbidity and electrical conductivity from water level data, taking advantage of the fact that water level measurements are prone to less uncertainty, and can be collected remotely. In perspective, we plan to analyze a broader dataset collected by the sensor system over the project's duration to strengthen our findings. We want to mention that another extreme rainfall event affected the city of Bologna and surrounding areas on October 19th, 2024. Culverted streams, some eventually discharging into Bologna WWTP, overflowed, causing significant damage. Large volumes of urban runoff were released into the environment. Thus, yet another

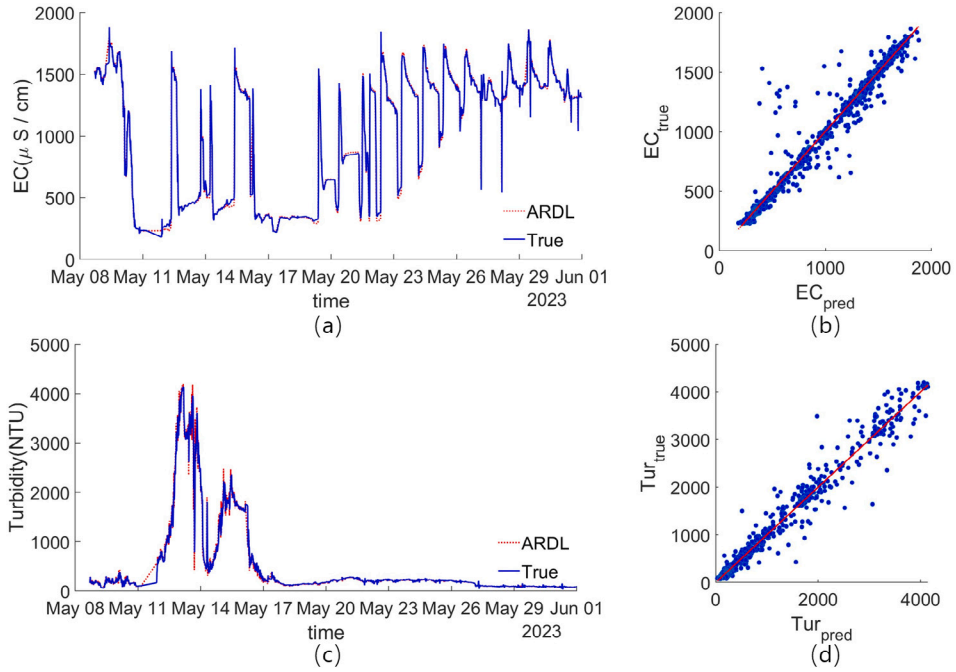

Fig. 10. ARDL model predictions in May (blue lines) and observed time series (red lines) for (a) electrical conductivity and (c) turbidity. Correspondent regression plots are in panels (b) and (d). The correlation coefficient (Eq. (9)) between electrical conductivity and water level is −0.9794, while between turbidity and water level is 0.4565. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

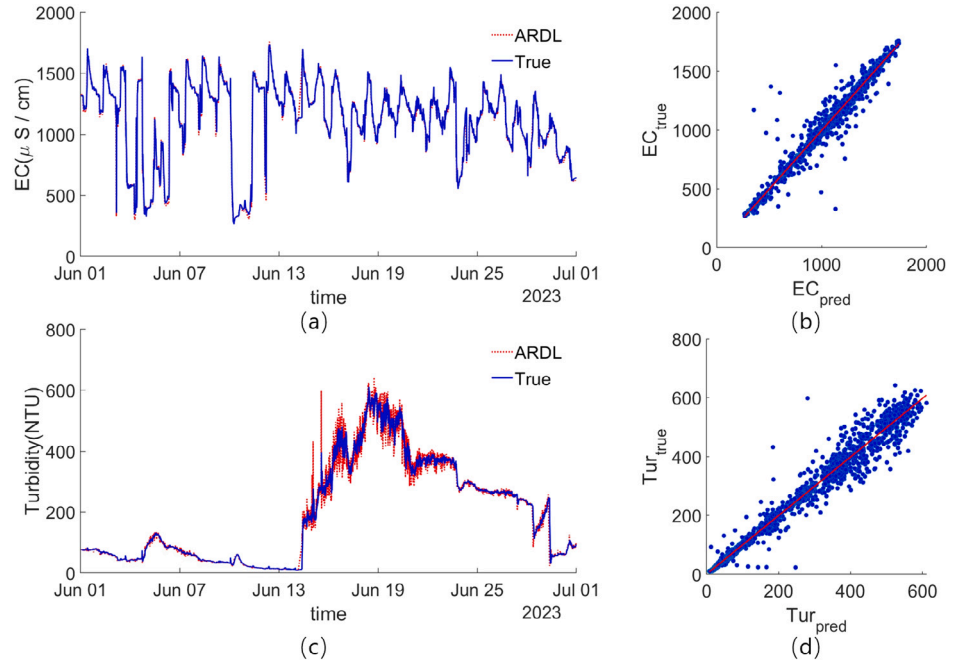

Fig. 11. ARDL model predictions in June (blue lines) and observed time series (red lines) for (a) electrical conductivity and (c) turbidity. Correspondent regression plots are in panels (b) and (d). The correlation coefficient (Eq. (9)) between electrical conductivity and water level is −0.4302, while between turbidity and water level is −0.8841. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

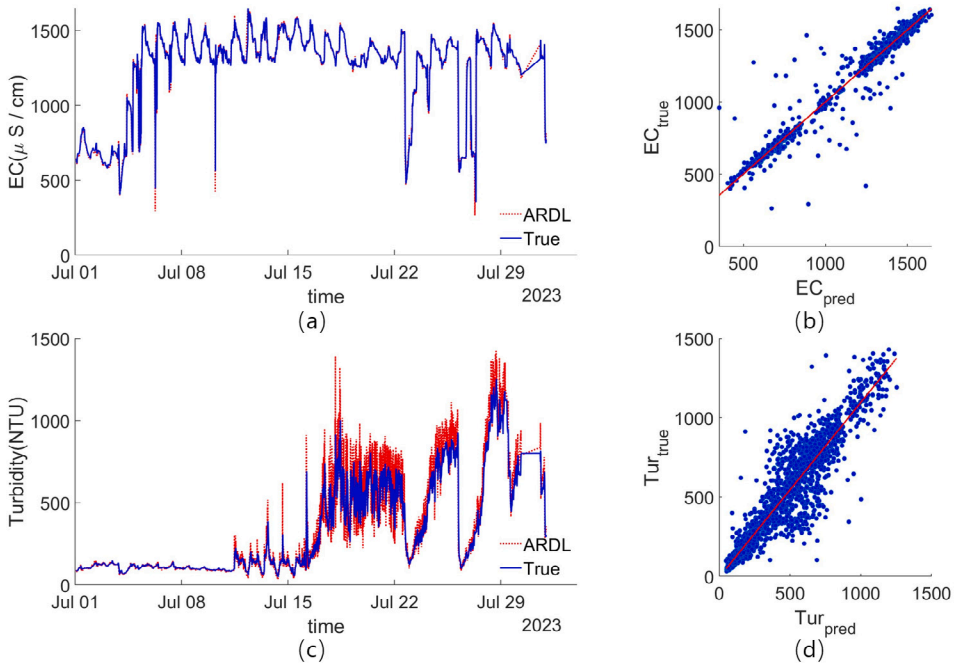

Fig. 12. ARDL model predictions in July (blue lines) and observed time series (red lines) for (a) electrical conductivity and (c) turbidity. Correspondent regression plots are in panels (b) and (d). The correlation coefficient (Eq. (9)) between electrical conductivity and water level is −0.3587, while between turbidity and water level is −0.5468. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

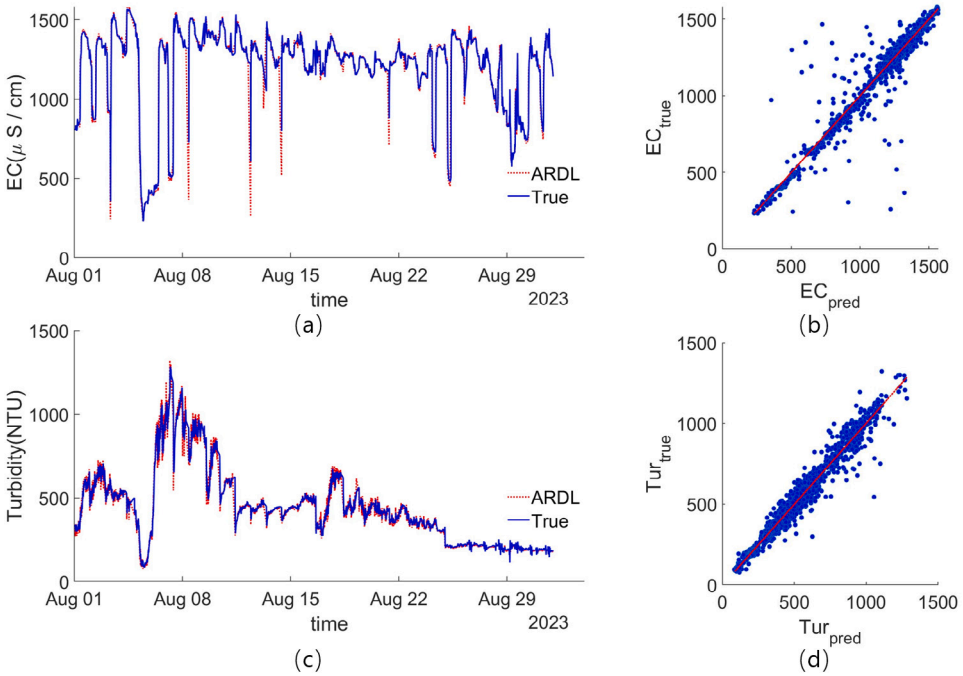

Fig. 13. ARDL model predictions in August (blue lines) and observed time series (red lines) for (a) electrical conductivity and (c) turbidity. Correspondent regression plots are in panels (b) and (d). The correlation coefficient (Eq. (9)) between electrical conductivity and water level is −0.9372, while between turbidity and water level is −0.4401. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

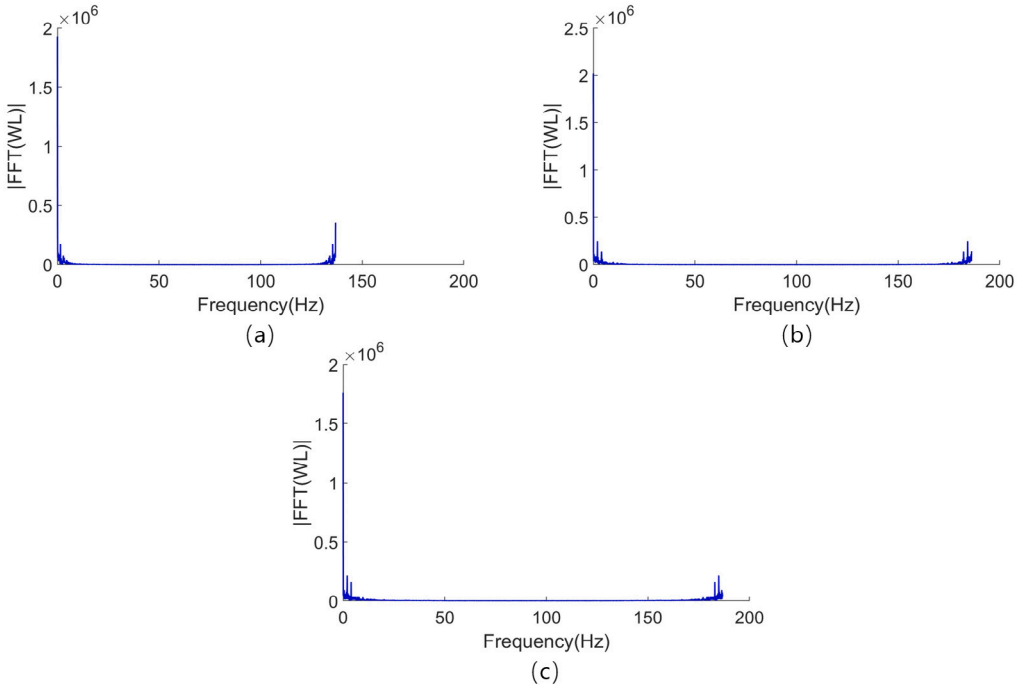

Fig. 14. Frequency spectrum of the water level time series in (a) May, (b) June, and (c) July.

extreme rainfall event reinforces the need to explore with a combination of tools (data-driven methods and physically-based models) the relationship between hydrologic and wastewater quality parameters.

CRediT authorship contribution statement

Ming Cheng: Writing – original draft, Software, Methodology, Investigation, Data curation, Conceptualization. **Margherita Evangelisti:** Writing – original draft, Investigation, Data curation. **Sacha Gobeyn:** Writing – original draft, Data curation. **Francesco Avolio:** Data curation. **Dario Frascari:** Writing – original draft, Investigation. **Marco Maglionico:** Writing – original draft, Investigation, Data curation. **Valentina Ciriello:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Vittorio Di Federico:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the European Union's Horizon Europe Programme for funding the StopUP project under Grant Agreement 101060428. Availability of research data: hydrological data are available at the repository https://simc.arpae.it/dext3r/, while data for electrical conductivity, rainfall depth, water level and turbidity registered at Bologna wastewater treatment plant from mid-May to December 2024 can be downloaded at https://doi.org/10.5281/zenodo.13741951.

Data availability

We have shared the link in the Acknowledgments.

References

Adarsh, S., Nourani, V., Archana, D., Dharan, D.S., 2020. Multifractal description of daily rainfall fields over India. J. Hydrol. 586, 124913. http://dx.doi.org/10.1016/j.jhydrol.2020.124913.

Anderson, C., 2005. Turbidity: US geological survey techniques of water-resources investigations, book 9, chap. A6. 7. In: National Field Manual for the Collection of Water-Quality Data. US Geological Survey Reston, VA.

Bersinger, T., Hécho, I.L., Bareille, G., Pigot, T., Lecomte, A., 2015. Continuous monitoring of turbidity and conductivity in wastewater networks: An easy tool to assess the pollution load discharged into receiving water. Revue des sciences de l'eau / J. Water Sci 28 (1), 9–17. http://dx.doi.org/10.7202/1030002ar.

Bertrand-Krajewski, J.-L., Clemens-Meyer, F., Lepot, M., 2021. Metrology in Urban Drainage and Stormwater Management: Plug and Pray. IWA Publishing, http://dx.doi.org/10.2166/9781789060119.

Brahim-Belhouari, S., Bermak, A., 2004. Gaussian process for nonstationary time series prediction. Comput. Statist. Data Anal. 47 (4), 705–712.

Cheng, L., AghaKouchak, A., Gilleland, E., Katz, R.W., 2014. Non-stationary extreme value analysis in a changing climate. Climatic Change 127, 353–369. http://dx.doi.org/10.1007/s10584-014-1254-5.

Coles, S., Bawa, J., Trenner, L., Dorazio, P., 2001. An Introduction to Statistical Modeling of Extreme Values, vol. 208. Springer.

Connor, J.T., Martin, R.D., Atlas, L.E., 1994. Recurrent neural networks and robust time series prediction. IEEE Trans. Neural Netw 5 (2), 240–254.

De Paola, F., Giugni, M., Pugliese, F., Annis, A., Nardi, F., 2018. GEV parameter estimation and stationary vs. non-stationary analysis of extreme rainfall in African test cities. Hydrology 5 (2), 28. http://dx.doi.org/10.3390/hydrology5020028.

eGEOS, 2023. Flood in Emilia-Romagna, Italy. In: Technical Report EMSN154. Copernicus - Emergency Management Service - Mapping, p. 32, downloaded on July, 2nd 2024 at https://emergency.copernicus.eu/mapping/list-of-components/EMSN154.

Felisa, G., Panini, G., Pedrazzoli, P., Di Federico, V., 2022. Combined management of groundwater resources and water supply systems at basin scale under climate change. Water Resour. Manag 36, 915–930. http://dx.doi.org/10.1007/s11269-022-03059-7.

Focaccia, S., Panini, G., Pedrazzoli, P., Ciriello, V., 2021. A meta-modeling approach for hydrological forecasting under uncertainty: Application to groundwater nitrate response to climate change. J. Hydrol. 603, 127173. http://dx.doi.org/10.1016/j.jhydrol.2021.127173.

Gruber, G., Bertrand-Krajewski, J.-L., Beneditis, J.D., Hochedlinger, M., Lettl, W., 2006. Practical aspects, experiences and strategies by using UV/VIS sensors for long-term sewer monitoring. Water Pract. Technol 1 (1), wpt2006020. http://dx.doi.org/10.2166/wpt.2006.020.

Hewamalage, H., Bergmeir, C., Bandara, K., 2021. Recurrent neural networks for time series forecasting: Current status and future directions. Int. J. Forecast. 37 (1), 388–427.

Hop, F.J., Linneman, R., Schnitzler, B., Bomers, A., Booij, M.J., 2024. Real time probabilistic inundation forecasts using a LSTM neural network. J. Hydrol. 635, 131082. http://dx.doi.org/10.1016/j.jhydrol.2024.131082.

Horvatic, D., Stanley, H.E., Podobnik, B., 2011. Detrended cross-correlation analysis for non-stationary time series with periodic trends. Europhys. Lett. 94 (1), 18007. http://dx.doi.org/10.1209/0295-5075/94/18007.

Hwarng, H.B., 2001. Insights into neural-network forecasting of time series corresponding to ARMA (p, q) structures. Omega 29 (3), 273–289.

Hwarng, H.B., Ang, H., 2001. A simple neural network for ARMA (p, q) time series. Omega 29 (4), 319–333.

Jalliffier-Verne, I., Leconte, R., Huaringa-Alvarez, U., Heniche, M., Madoux-Humery, A.-S., Autixier, L., Galarneau, M., Servais, P., Prévost, M., Dorner, S., 2017. Modelling the impacts of global change on concentrations of Escherichia coli in an urban river. Adv. Water Resour. 108, 450–460. http://dx.doi.org/10.1016/j.advwatres.2016.10.001.

Jun, W.C., Oh, G., Kim, S., 2006. Understanding volatility correlation behavior with a magnitude cross-correlation function. Phys. Rev. E 73 (6), 066128. http://dx.doi.org/10.1103/PhysRevE.73.066128.

Katz, R.W., Parlange, M.B., Naveau, P., 2002. Statistics of extremes in hydrology. Adv. Water Resour. 25 (8–12), 1287–1304. http://dx.doi.org/10.1016/S0309-1708(02)00056-8.

Khamis, K., Bradley, C., Stevens, R., Hannah, D.M., 2017. Continuous field estimation of dissolved organic carbon concentration and biochemical oxygen demand using dual-wavelength fluorescence, turbidity and temperature. Hydrol. Process. 31 (3), 540–555. http://dx.doi.org/10.1002/hyp.11040.

Kohanpur, A.H., Saksena, S., Dey, S., Johnson, J.M., Riasi, M.S., Yeghiazarian, L., Tartakovsky, A.M., 2023. Urban flood modeling: Uncertainty quantification and physics-informed Gaussian processes regression forecasting. Water Resour. Res. 59 (3), http://dx.doi.org/10.1029/2022wr033939.

Langeveld, J., Veldkamp, R., Clemens, F., 2005. Suspended solids transport: an analysis based on turbidity measurements and event based fully calibrated hydrodynamic models. Water Sci. Technol. 52 (3), 93–101. http://dx.doi.org/10.2166/wst.2005.0065.

Liu, K., Kinouchi, T., Tan, R., Heng, S., Chhuon, K., Zhao, W., 2024. Unraveling urban hydro-environmental response to climate change and MCDA-based area prioritization in a data-scarce developing city. Sci. Total Environ. 948, 174389. http://dx.doi.org/10.1016/j.scitotenv.2024.174389.

Maglionico, M., Bolognesi, A., Casadio, A., 2009. Evaluation of CSOs pollution impact and mitigation effects obtained with storage tanks. pp. 680–687, 8th International Conference on Urban Drainage Modelling.

Mahaut, V., Andrieu, H., 2019. Relative influence of urban-development strategies and water management on mixed (separated and combined) sewer overflows in the context of climate change and population growth: A case study in Nantes. Sustainable Cities Soc. 44, 171–182. http://dx.doi.org/10.1016/j.scs.2018.09.012.

Marinho, E., Sousa, A., Andrade, R.F.S., 2013. Using detrended cross-correlation analysis in geophysical data. Phys. A 392 (9), 2195–2201. http://dx.doi.org/10.1016/j.physa.2012.12.038.

Navratil, O., Boukerb, M.A., Perret, F., Breil, P., Caurel, C., Schmitt, L., Lejot, J., Petit, S., Marjolet, L., Cournoyer, B., 2020. Responses of streambed bacterial groups to cycles of low-flow and erosive floods in a small peri-urban stream. Ecohydrology 13 (4), http://dx.doi.org/10.1002/eco.2206.

Nkoro, E., Uko, A.K., et al., 2016. Autoregressive distributed lag (ARDL) cointegration technique: application and interpretation. J. Statist. Econom. Methods 5 (4), 63–91.

Nogueira, M., 2019. The sensitivity of the atmospheric branch of the global water cycle to temperature fluctuations at synoptic to decadal time-scales in different satellite-and model-based products. Clim. Dyn. 52 (1), 617–636. http://dx.doi.org/10.1007/s00382-018-4153-z.

Pesaran, M.H., Shin, Y., et al., 1995. An Autoregressive Distributed Lag Modelling Approach to Co-Integration Analysis, vol. 9514. Department of Applied Economics, University of Cambridge Cambridge, UK, http://dx.doi.org/10.1017/CCOL0521633230.011.

Podobnik, B., Stanley, H.E., 2008. Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series. Phys. Rev. Lett 100 (8), 084102. http://dx.doi.org/10.1103/PhysRevLett.100.08410.

Prathumratana, L., Sthiannopkao, S., Kim, K.W., 2008. The relationship of climatic and hydrological parameters to surface water quality in the lower Mekong river. Environ. Int. 34 (6), 860–866. http://dx.doi.org/10.1016/j.envint.2007.10.011.

Razguliaev, N., Flanagan, K., Muthanna, T., Viklander, M., 2024. Urban stormwater quality: A review of methods for continuous field monitoring. Water Res. 249, 120929. http://dx.doi.org/10.1016/j.watres.2023.120929.

Reisinger, A.J., Woytowitz, E., Majcher, E., Rosi, E.J., Belt, K.T., Duncan, J.M., Kaushal, S.S., Groffman, P.M., 2019. Changes in long-term water quality of baltimore streams are associated with both gray and green infrastructure. Limnol. Oceanogr 64 (1, SI), S60–S76. http://dx.doi.org/10.1002/lno.10947.

Roberts, S., Osborne, M., Ebden, M., Reece, S., Gibson, N., Aigrain, S., 2013. Gaussian processes for time-series modelling. Phil. Trans. R. Soc. A 371 (1984), 20110550.

Rosenberger, L., Leandro, J., Pauleit, S., Erlwein, S., 2021. Sustainable stormwater management under the impact of climate change and urban densification. J. Hydrol. 596, 126137. http://dx.doi.org/10.1016/j.jhydrol.2021.126137.

Rossi, L., Krejci, V., Rauch, W., Kreikenbaum, S., Fankhauser, R., Gujer, W., 2005. Stochastic modeling of total suspended solids (TSS) in urban areas during rain events. Water Res. 39 (17), 4188–4196. http://dx.doi.org/10.1016/j.watres.2005.07.041.

Seeger, M., 2004. Gaussian processes for machine learning. Int. J. Neural Syst 14 (02), 69–106.

Shadkhoo, S., Jafari, G., 2009. Multifractal detrended cross-correlation analysis of temporal and spatial seismic data. Eur. Phys. J. B 72, 679–683. http://dx.doi.org/10.1140/epjb/e2009-00402-2.

Shen, C., 2015. Analysis of detrended time-lagged cross-correlation between two nonstationary time series. Phys. Lett. A 379 (7), 680–687. http://dx.doi.org/10.1016/j.physleta.2014.12.036.

Wang, M., Liu, M., Zhang, D., Qi, J., Fu, W., Zhang, Y., Rao, Q., Bakhshipour, A.E., Tan, S.K., 2023. Assessing and optimizing the hydrological performance of grey-green infrastructure systems in response to climate change and non-stationary time series. Water Res. 232, 119720. http://dx.doi.org/10.1016/j.watres.2023.119720.

Xu, C., Zhong, P.-a., Zhu, F., Xu, B., Wang, Y., Yang, L., Wang, S., Xu, S., 2024. A hybrid model coupling process-driven and data-driven models for improved real-time flood forecasting. J. Hydrol. 638, 131494. http://dx.doi.org/10.1016/j.jhydrol.2024.131494.