



Research papers

Ai-driven morphoclimatic regional frequency modelling of sub-daily rainfall-extremes

Andrea Magnini ^{a,*}, Michele Lombardi ^b, Taha B.M.J. Ouarda ^c, Attilio Castellarin ^a^a Department of Civil, Chemical, Environmental and Materials Engineering (DICAM), University of Bologna, Bologna, Italy^b Department of Computer Science and Engineering (DISI), University of Bologna, Bologna, Italy^c Institut National de Recherche Scientifique (INRS), Quebec, Canada

ARTICLE INFO

ABSTRACT

Keywords:

Neural networks
Extreme rainfall
Regional frequency analysis
Italy
Sub-daily time-interval

Common main limitations affect standard approaches to regional frequency analysis (RFA) of rainfall extremes. Our study focuses on three of them that are rather frequent: regional models address (a) a single duration, or (b) a single exceedance probability at a time, and/or (c) hold a small-to-medium homogeneous region only. We use unsupervised ensembles of artificial neural networks (ANNs) to set up four alternative RFA models of sub-daily rainfall extremes. These are fed with annual maximum series of rainfall depth of any length collected at 2238 raingauges in a large and climatically and morphologically heterogeneous region. Our models can predict parameters of a Gumbel distribution for any location within the study area and any duration in the 1–24 h range. Prediction is based on mean annual precipitation (MAP), or on twenty morphoclimatic covariates. Validation is performed over an independent set of 100 gauges, where locally fitted Gumbel distributions are used as reference. A common literature approach where Gumbel parameters are functions of MAP is used as benchmark. Our results show that multivariate ANNs remarkably improve the estimation of percentiles relative to the benchmark approach. Finally, we show that the very nature of the proposed ANN models makes them suitable for interpolating predicted sub-daily rainfall quantiles across time-aggregation intervals and space and can be adapted for considering more flexible target frequency distributions (e.g. 3-parameter models).

1. Introduction

Several hydrological applications, such as the design and management of stormwater drainage systems, combined sewer overflows and flood control systems require an accurate estimation of the design storm (e.g., Claps et al., 2022; Camorani et al., 2005). The latter can be defined as the rainfall depth associated with a given duration and non-exceedance probability (commonly expressed in terms of return period). To produce an accurate estimation of the design storm, modelling the frequency regime of rainfall extremes in the location of interest is needed (e.g., Koutsoyiannis, 2007, Persiano et al., 2020). Timeseries of observed rainfall extremes (e.g., annual sequences of maximum rainfall depths for given durations), when locally available, are in many cases too short to perform robust at-site frequency analysis. This limitation is often addressed by means of regional frequency analysis (RFA), that consists in transferring observed data from other gauged locations to the target site (see e.g., Di Baldassarre et al., 2006, Castellarin et al., 2009, Blöschl, 2011).

In general, RFA consists of two main phases: (i) the delineation of a homogenous pooling-group of sites (i.e., region), containing gauged sites that are similar to the target one, and (ii) the definition of a regional model to transfer the information from the homogeneous region to the target site (e.g., Grimaldi et al., 2011). The scientific literature reports on a large number of different methods for conducting RFA of rainfall extremes (see e.g., Svensson and Jones, 2010). The homogeneous region is defined based on some hydrological similarity criteria (see e.g. Castellarin et al., 2001), and can be considered as fixed (i.e., the gauged sites are divided into fixed clusters) or specific for any target site, as in the region-of-influence approach (Burn, 1990). The regional transfer function defined in the second phase of RFA is highly variable depending on the specific approach: on the one hand, the target variable could be a specific percentile (e.g., Ouali et al., 2016a), a parameter of a probability distribution (e.g., Soltani et al., 2017), a statistical moment or L-moment (e.g., Modarres and Sarhadi, 2011; Ngongondo et al., 2011), or the complete time series itself (e.g., Requena et al., 2017, 2018); on the other hand, the observations of the gauged sites could be used alone or

* Corresponding author.

E-mail address: andrea.magnini@unibo.it (A. Magnini).

with some covariates of the target variable. Literature also reports on several methods that do not require the definition of a homogeneous region (methods based on regression, e.g., Brath et al., 2003, or interpolation, e.g., Claps et al., 2022).

Regarding the application of RFA to the estimation of flood quantiles, the scientific community has proposed several approaches that make use of advanced artificial intelligence (AI) techniques. Linear and non-linear techniques have been discussed for the definition of a homogeneous region (e.g., Ouarda et al., 2001; Ouali et al., 2016a). Models for regional flood frequency analysis can consider many morphological and climatic covariates (Msilini et al., 2022), consider non-linearity of the input-output relations (Ouarda and Shu, 2009), and represent the interaction between the input variables (e.g., Msilini et al., 2020). On the contrary, the literature does not report on many AI-aided RFA methods for modelling the frequency regime of extreme precipitation.

The preference for one model or another strongly depends on the specific case. Moreover, since certain knowledge of the frequency regime of a highly stochastic event is not possible, incontrovertible evaluation of regional models cannot be obtained (Di Baldassarre et al., 2009; Velázquez et al., 2011). However, it is clear which are the characteristics that a good regionalization method should have. First, the ability to profitably use as much recorded information (i.e., observed data) as possible. This is due to the nature of available official gauging networks, that consist of unevenly distributed rain gauges, and timeseries that are often short or fragmented (Libertino et al., 2018; Kidd et al., 2017). As a result, very short sequences are often discarded in regional analysis (e.g., Di Baldassarre et al., 2006). Second, the model should be as flexible as possible. Due to the fast development of technology and science, the amount, location and type of data available is highly dynamic. Thus, a model that can be easily adapted to these changes has a great advantage.

In the present study, the potential of a new AI-based approach to RFA of rainfall extremes is investigated and discussed. It is based on ensembles of unsupervised artificial neural networks (ANNs), that are able to predict the parameters of a selected extreme value probability distribution of the dimensionless extreme rainfall for any duration between 1 and 24 h. Following the general framework of the widely adopted storm index method (e.g., Di Baldassarre et al., 2006), the frequency regime of the dimensionless extreme rainfall is regionalized. This is the extreme rainfall depth timeseries divided by its mean value at each site for a given duration. In the present study we focused on the Gumbel distribution, but the approach could consider different models. The proposed method is simple, flexible, and innovative thanks to some characteristics. First, no clustering or target-pooling of available rain gauges is performed: all available annual sequences of maximum rainfall depth are used jointly. Second, no filter on a minimum length of annual sequences is needed and very short sequences (even with two observations) can be used. Third, training is performed simultaneously on all available durations, which leads to advanced interpolation of time-aggregation intervals capability, a very useful feature for practical applications such as the construction of intensity-duration-frequency curves (see Koutsoyiannis et al., 1998; Brath et al., 2003). Fourth, the modelled extreme value distribution can be predicted at ungauged locations within the study region, based on available morphoclimatic information.

The proposed approach is tested in the present study by implementing it through four kinds of ANNs with increasing complexity. The first makes use of the mean annual precipitation (MAP) alone (MAP-ANN), which is a classical proxy for frequency regime of rainfall annual maxima (see e.g., Schaefer, 1990; Alila, 1999; Castellarin et al., 2009). The second relies on an extended set of twenty morphoclimatic characteristics of the site of interest, including MAP (EXT-ANN). The third (EXT-PCA-ANN) and the fourth (EXT-CCA-ANN) models are fed on pre-processed versions of the same twenty input descriptors, that are obtained through principal component analysis (PCA), and canonical correlation analysis (CCA, see e.g., Di Prinzio et al., 2011).

We make use of a large dataset of gauged stations located in northern and central Italy. For each station, rainfall annual maximum series (AMS) for five different time-aggregation intervals, or durations for the sake of brevity, are available (i.e., 1 h, 3 h, 6 h, 12 h and 24 h). The maximum length of the AMS series is 90 years. In particular, 2238 stations representing a wide range of morphological and climatic conditions are used to train the four models. Validation of the four regional models is performed using data collected at 100 independent rain-gauges. The validation considers a traditional RFA method based on L-moments and MAP (see e.g. Di Baldassarre et al., 2006) as the baseline regional approach (hereafter also referred to as MAP-Lm), as well as the newly proposed models (i.e., MAP-ANN, EXT-ANN, EXT-PCA-ANN, EXT-CCA-ANN), and compare their predictions with those resulting from at-site frequency analyses (i.e. locally estimated Gumbel distributions).

Finally, the study shows a preliminary application of the proposed EXT-ANN model that adopts the 3-parameter Generalized Extreme Value (GEV) distribution (Jenkinson, 1955). In this preliminary application, the parameter that controls the skewness of the GEV distribution (i.e., the shape parameter) is regionalized through geostatistical interpolation procedure (e.g., Hengl, 2007), while the remaining two parameters are derived from the prediction of the EXT-ANN model. Testing the proposed approach for a 3-parameter distribution is important. The scientific literature clearly indicates that in some geographical and climatic contexts the flexibility of a 2-parameter Gumbel distribution, even though widely adopted in previous works (see e.g., Grieser et al., n.d; Svensson and Jones, 2010; Van den Besselaar et al., 2013; Piper et al., 2016; Maity, 2018; Caldas-Alvarez et al., 2022), is not enough for producing an accurate representation of the frequency of rainfall extremes (e.g., Koutsoyiannis and Baloutsos, 2000; Koutsoyiannis, 2004, Papalexiou et al., 2018).

Two main research questions are addressed within this research: (1) are ANNs useful and effective in RFA of rainfall extremes? (2) Are morphological indices helpful in describing the local frequency regime of sub-daily rainfall extremes?

2. Methods

The present Section describes the methodologies adopted to set up all the regional models considered in the study. It is divided into two parts: the first briefly summarizes the storm index model with the L-moments approach, that is considered as baseline; the second illustrates the theoretic principles of the ANN approach that originates the four AI-based models. More detail on the models' set-up is given in Section 4.

2.1. Storm index method and L-moments approach

The storm index method is one of the most commonly adopted models for RFA (Dalrymple, 1960; Brath et al., 2003). Given a certain target station where the rainfall depth needs to be estimated, and a group of gauged stations as homogenous region, the storm index method is described by Eq. (1).

$$h(d, T) = m_d \cdot h'(d, T) \quad (1)$$

The rainfall depth associated with a given duration d and return period T , $h(d, T)$, is the product of a scale factor m_d , that is called the storm index, and the dimensionless rainfall depth $h'(d, T)$, that is called growth factor. The scale factor is site dependent and is usually estimated by averaging the available measurements at the target station or exploiting regional information through interpolation techniques. The growth factor is derived from a regional relation that is assumed to be valid for the entire homogenous group of sites, which requires to be defined.

Following the original version by Dalrymple (1960), several different applications of the storm index method have been proposed. The MAP-Lm model set up in the present study strictly follows the methods in Di

Baldassarre et al. (2006), which in turn strongly relies on the findings of Schaefer (1990) and Alila (1999). These authors studied the extreme dimensionless rainfall depth, which is obtained by dividing the dimensional data by the mean depth associated with the same duration at the same station. They found that statistical moments and L-moments of extreme dimensionless rainfall are in relation to MAP. Thus, homogeneous groups of stations can be identified according to their values of MAP, which can be used as a substitute of the geographical location and as proxy of extreme precipitation. Based on the findings by Hosking and Wallis (1997) and Hosking and Wallis (1993), L-moments should be preferred for RFA to traditional moments as they are more robust to outliers, can characterize a wider range of distributions and are less subject to estimation bias.

In particular, the MAP-Lm model aims to estimate the regional growth factor of the dimensionless rainfall by means of a Gumbel distribution (i.e., generalized extreme value distribution with zero-value shape parameter; see cumulative density function $F(x)$, Eq. (3)) for each one of the considered durations (i.e., 1, 3, 6, 12, 24 h). It is assumed that the whole study area can be described by the same regional laws between local MAP value and statistical moments of rainfall extremes; accordingly, the procedure can be summarized in three major steps: (1) group gauged stations by their MAP values through a moving window of 100 stations; (2) plot the average MAP and L-CV (i.e., L-coefficient of variation); (3) fit the regional function L-CV(MAP) as solution of a least squares problem. This procedure is applied for each duration to all stations with more than five years of measurement.

In detail, the relationship between L-CV and MAP is modelled by adopting a Horton-type curve (Eq. (2)).

$$L - CV = a + (b - a) \bullet \exp(-c \bullet MAP) \quad (2)$$

Once the a , b and c regional parameters are found, L-CV can be obtained in the target location as a function of MAP with Eq. (2). Then, using the relations described in Hosking and Wallis (1997), i.e., Eqs. (4) and (5) for the Gumbel distribution as in this case, the scale and location parameters (respectively, α and ξ) of the desired local frequency distribution are estimated.

$$F(x) = \exp\left(-\exp\left(-\frac{x-\xi}{\alpha}\right)\right) \quad (3)$$

$$\alpha = \lambda_2/\ln(2) = (L - CV) \bullet \lambda_1/\ln(2) \quad (4)$$

$$\xi = \lambda_1 - \gamma \bullet \alpha \quad (5)$$

Where λ_1 , λ_2 and γ are in this order the L-moments of the first and second order (i.e., mean and the L-scale of the dimensionless AMS observed) and the Euler's constant (i.e., 1.504).

Despite relying on a single proxy for extreme rainfall, this model has been shown to be rather accurate (Di Baldassarre et al., 2006) over the region for which it was proposed and tested. Moreover, the general framework is rather flexible since local MAP values can be easily retrieved for various regions of the world. For these reasons, it is selected as the baseline for evaluating the performance of the AI-based approaches.

2.2. ANN approach

Artificial intelligence has been profitably used by several authors for enhancing RFA for floods (e.g., Msilini et al., 2020; Ouali et al., 2016b; Ouarda and Shu, 2009; Shu and Ouarda, 2007), and the accurate choice of models and manipulation of data shows encouraging results. However, the investigation of artificial intelligence techniques for rainfall RFA is not yet well documented.

In the present study, artificial intelligence is used to enhance efficiency in data exploitation and combination. We consider ensembles of unsupervised ANNs; four different models are set up: while fed with

different input data, they are based on the same macro-structure. All the ANN models follow four general guidelines: (1) exploiting simultaneously the AMS with all the available durations; (2) using short timeseries as well as long ones; (3) minimizing negative logarithmic likelihood as objective function (see below); (4) predicting Gumbel distributions as target. Thus, differently from the MAP-Lm model, the timeseries from all the stations are pooled together for training and validating the ANN models. The aim of these ANNs, as for the MAP-Lm, is to estimate the growth factor of the storm index framework. This is done by finding the best parameters for Gumbel probability distributions of the dimensionless extreme rainfall for any duration at any location, that correspond to the minimum negative logarithmic likelihood.

ANNs are among the most common machine learning models (Hastie et al., 2009). They consist of successive layers, each one containing a given number of neurons. Referring to each single i -th neuron, the output y_i is a function f_i of the linear combination of the n input factors x_{ij} , as follows:

$$y_i = f_i(\sum_{j=1}^n w_{ij} \bullet x_{ij}) \quad (6)$$

Where w_j are the weights, or coefficients of the input factors, and f_i is assumed in the present study as a sigmoid function, as it is commonly done (e.g., Han and Moraga, 1995). The input factors for each neuron are the outputs from the previous layer, while for the first layer, the input are the descriptors themselves.

Before the training, the available dataset is divided into a "training/testing set" (the larger one), and a "validation set" (the smaller one, used only for the validation). During the training phase, the training/testing set is randomly divided into two subsets: a larger one that will be referred to as "training set", and a smaller one that is referred to as "testing set". Common proportions of data from the original training/testing dataset are 80% for the training and 20% for the testing set (e.g., see Xu and Goodacre, 2018). The best hyperparameter set (i.e., the weights w_{ij} for each i -th neuron) is searched while observing the training set) and minimizing the negative logarithmic likelihood function ($\log LH$, see eq. (7)) computed on the remaining data included in the testing set.

$$\log LH = \log \left(\prod_{k=1}^m p(x_k) \right) = \sum_{k=1}^m \log(p(x_k)) \quad (7)$$

where m is the testing set, and $p(x_k)$ the predicted probability distribution.

Since ANNs are complex and accurate models, they are likely to learn how to perfectly reproduce the training set while being inaccurate with other datasets. This is referred to as overfitting. To improve the generalization ability and stability of a single ANN, an ANN ensemble can be used. It consists of a number of ANNs whose individual results are combined to generate a unique output. Many approaches have been proposed for generating single ANNs for the ensemble models (Breiman, 1996a) and integrating the multiple outputs (Breiman, 1996b). In the present study, ANNs are generated through bagging, and averaging is used for merging the results. Thus, for each single ANN, the same initial training/testing set is randomly split, as discussed above, so that the optimal hyperparameter set is searched by training the model on the training set while optimizing the objective function for the testing set. This method is a simple and effective way to obtain ANN ensemble models (see e.g., Shu and Burn, 2004; Shu and Ouarda, 2007).

3. Study region and morphoclimatic descriptors

3.1. Study region

The methods explained above are applied on a dataset of 2338 gauged locations. These are located in a wide geographical area in northern and central Italy (Fig. 1), where a variety of climatic and

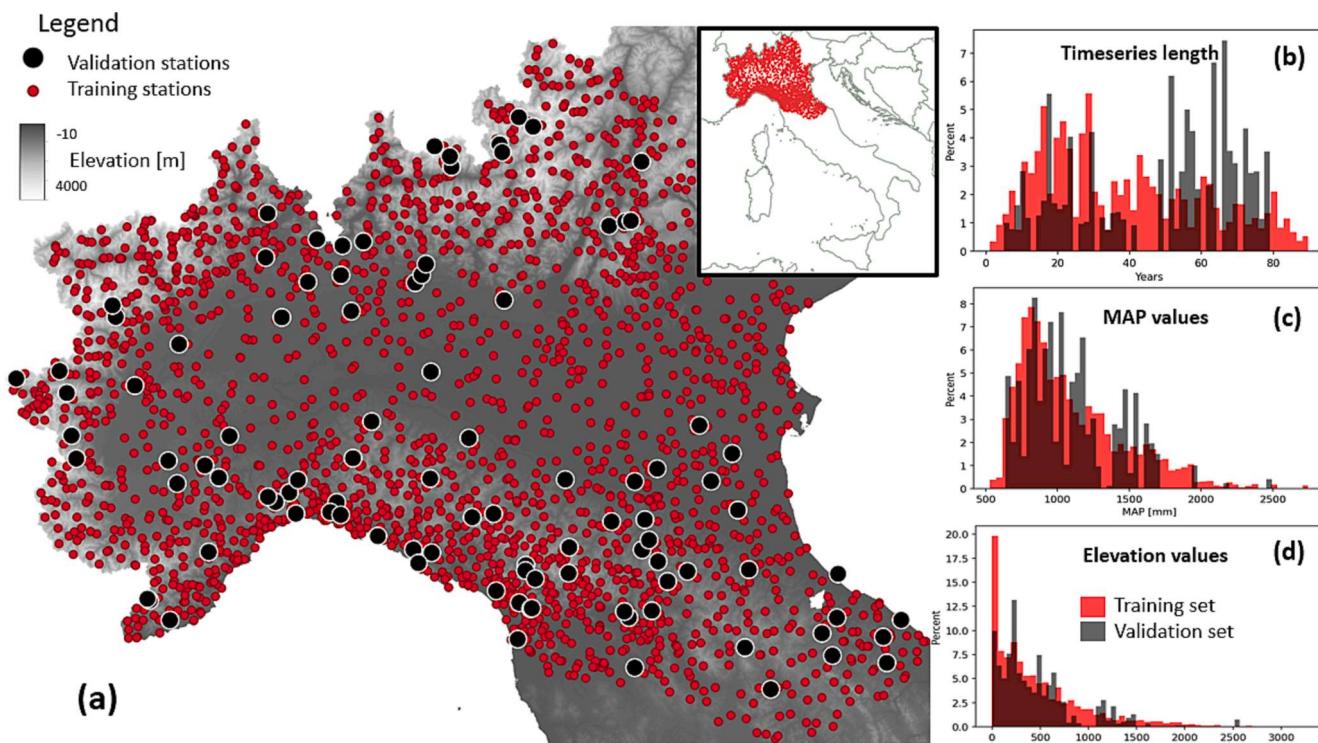


Fig. 1. Study area, training/testing (red dots) and validation (black dots) raingauges (panel a); sample frequency distribution (%) of several characteristics for the training/testing (red bars) and validation (grey bars) raingauges: timeseries length (panel b), mean annual precipitation (or MAP, panel c), and elevation (panel d). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

morphological systems can be found. The north is dominated by the Alps, the highest Italian mountain chain, with a mean elevation of 2500 m.a.s.l., and highest peaks until 4800 m.a.s.l. The largest Italian plain, the Po plain, stretches in the southern border of the Alps, following the course of the Po river from the northwest to the northeast, where low coasts are located. The southern border of the Po plain is marked by the Northern Apennines, whose maximum peak is 2165 m.a.s.l. Within the study area, one of the factors that seem to have the largest effect on the precipitation regime is altitude (Allamano et al., 2009; Marra et al., 2021; Mazzoglio et al., 2022).

Three different datasets (Table 1) are used in the present study to derive the input information and set up the models described in Section 2. First, the Annual Maxima Series (AMS), the variable whose probability distribution needs to be estimated, are retrieved from the dataset I²-RED (Mazzoglio et al., 2020). It includes annual maximum rainfall depths for 1, 3, 6, 12 and 24 consecutive hours from 2338 weather stations across the study area, recorded between 1916 and 2019. While 2238 stations have been selected and used as training/testing set for the five models (i.e. 80% for training and 20% for testing, see Section 2.2), the remaining 100 are used as a validation set (see Fig. 1). The selection of the validation set is based on three main guidelines: (1) to identify a significant number of timeseries, so to have an informative validation set for evaluating the RFA models' performance; (2) to have gauges that are representative of the entire dataset in terms of location, local climate and morphological conditions (see Fig. 1c and d); (3) to have long timeseries for using as reference validation values the at-site predictions of rainfall quantiles (Fig. 1b). Training/testing and validation sets include timeseries with at least 50 years of measurement (i.e., 248 for the training/testing set and 29 for the validation set, see Fig. 1b).

The second dataset used in our study is the multi-error removed improved terrain model (MERIT, see Yamazaki et al., 2017) DEM (see Fig. 1), that was used to derive the morphological descriptors (lines 1 to 14 of Table 1).

A third dataset was used in the analysis for retrieving the climatic

information, that is the ISPRA BIGBANG dataset (i.e., version 4.0, see Braca et al., 2019), that contains, among other variables, a 1 km raster representation of the annual totals of cumulative liquid and snow precipitation over the 1951–2019 time interval.

3.2. Morphoclimatic descriptors

The descriptors adopted in the present study were selected based on Mazzoglio et al. (2022), which explored the influence of climatic and morphological descriptors on the statistics of rainfall extremes. They can be divided into three groups: morphological, climatic and geographical. The first group includes descriptors of the elevation, slope and aspect within a distance of 1 km from the station of interest (i.e., from 1 to 6 of Table 1), and of the orography and distance between the station and the sea coastlines (i.e., from 7 to 14 of Table 1). The climatic descriptors include mean annual rainfall and snow, and their multi-year standard deviation (i.e., from 15 to 18 of Table 1). The geographic descriptors consist of longitude and latitude (i.e., 19 and 20 of Table 1).

The study descriptors can be obtained through GIS processing procedures of freely available datasets: a digital elevation model (DEM) is needed for retrieving the morphological descriptors, a precipitation dataset for the climatic group, and the coordinate of the gauged stations themselves for the geographical descriptors.

Some descriptors show significant inter-correlation, as illustrated by the correlation matrix depicted in Fig. 2. Each element X_{ij} of this square and symmetric matrix is the Pearson's correlation coefficient (PCC) of descriptors X_i and X_j . The PCC varies between -1 and 1, and the higher its absolute value is, the higher is the correlation between the two variables to whom it is referred.

$$PCC(X_i, X_j) = \frac{cov(X_i, X_j)}{\sqrt{var(X_i) \cdot var(X_j)}} \quad (8)$$

Several groups of highly inter-correlated variables are evident (e.g., 1–4, see Fig. 2), and the mean altitude (descriptor 1) is strongly

Table 1

Input descriptors and source dataset used for retrieving them (references for MERIT DEM, BIGBANG and I2-RED datasets are Yamazaki et al., 2017, Braca et al., 2018, Mazzoglio et al., 2020, in this order).

Descriptor	Description	Information origin
1	mean altitude within a distance of 1 km from the gauged location	MERIT DEM
2	standard deviation of the altitude within a distance of 1 km from the gauged location	MERIT DEM
3	mean slope (i.e., ration between vertical and horizontal distance) within a distance of 1 km from the gauged location	MERIT DEM
4	standard deviation of the slope within a distance of 1 km from the gauged location	MERIT DEM
5	mean aspect (i.e., direction of maximum slope) within a distance of 1 km from the gauged location	MERIT DEM
6	the standard deviation of the aspect within a distance of 1 km from the gauged location	MERIT DEM
7	minimum distance from the Adriatic coast	MERIT DEM
8	mean elevation within the distance between the gauged location and the Adriatic coast	MERIT DEM
9	standard deviation of elevation within the distance line between the gauged location and the Adriatic coast	MERIT DEM
10	Maximum elevation within the distance line between the gauged location and the Adriatic coast	MERIT DEM
11	minimum distance from the Tyrrhenian coast	MERIT DEM
12	mean elevation within the distance between the gauged location and the Tyrrhenian coast	MERIT DEM
13	standard deviation of elevation within the distance line between the gauged location and the Tyrrhenian coast	MERIT DEM
14	Maximum elevation within the distance line between the gauged location and the Tyrrhenian coast	MERIT DEM
15	Mean annual precipitation	BIGBANG dataset
16	Mean annual snow precipitation	BIGBANG dataset
17	Standard deviation of annual precipitation within the 1919–2019 record period	BIGBANG dataset
18	Standard deviation of annual snow precipitation within the 1919–2019 record period	BIGBANG dataset
19	Longitude	I ² -RED
20	Latitude	I ² -RED

correlated with most of the other descriptors. This characteristic of the dataset is common, and several authors, as Di Prinio et al. (2011), showed that pre-processing input datasets by means of Principal Component Analysis (PCA, Jolliffe, 2002) or Canonical Correlation Analysis (CCA, Hotelling, 1935), and removing redundant information may improve the training efficiency of data-driven methods and may result in better predictions. To test the convenience of preprocessing techniques for the present study case, PCA and CCA are adopted for two out of the four ANN models trained (see Section 4).

Finally, climate indexes used in our study (i.e., descriptors 15, 16, 17 and 18) are long-term averages referring to a given climate time-window (i.e. 1951–2019). Hence, we do not consider possible non-stationarities in our study, which is an interesting subject for future developments (see e.g. Persiano et al. (2020) for indications of signals on non-stationarity in sub-daily rainfall extremes for the study region).

3.3. Gumbel target frequency distribution

Fig. 3 illustrates the L-moments ratio diagram (see e.g. Hosking and Wallis, 1997) of the study area for the rainfall annual maximum series at 1 h and 24 h durations. Regardless of the duration, a high variability of sample L-moments is evident, as it was expected; the location of the weighted average is a point on the GEV line, very near to the point indicating the theoretical L-moments of the Gumbel distribution.

Thus, the choice of a regional distribution type is restricted to be

either a Gumbel or a GEV. Several studies (e.g., Koutsoyiannis et al., 2000; Koutsoyiannis et al., 2004; Papalexiou and Koutsoyiannis, 2013) showed how the latter should be preferred for a better estimation of the upper tail in some geographical and climatic contexts. However, these studies also made it clear that the estimation of the shape parameter of a GEV distribution is affected by high uncertainty, especially when short time series are used. Due to this reason, several recent studies proposing RFA approaches still resort to the more robust Gumbel distribution (e.g., Svensson and Jones, 2010; Maity, 2018; Ouarda et al., 2019; Caldas-Alvarez, 2022). Thus, our study thoroughly assesses the viability of the Gumbel target distribution for the proposed ANN RFA models. Nevertheless, for the sake of generalization, we also briefly present a possible adaptation of the proposed approach to the GEV distribution.

4. Regional ANN models

4.1. ANN models with Gumbel target frequency distribution

We set up and analyze four different regional ANN models, all consisting of ensembles of ANNs. As the MAP-Lm model (see Sect. 2.1), they all aim to produce a regional estimate of the local growth factor for the dimensionless extreme rainfall depth associated with a given duration (or time aggregation interval). Therefore, they are trained on dimensionless AMS of rainfall depths, that are obtained by dividing each original annual sequence by its sample mean, specific at each site for each duration (see details later on).

The first model is referred to as MAP-ANN and is fed exclusively on the MAP descriptor. Its application and validation allow it to investigate the effect of exploiting the same input information as in the MAP-Lm method (i.e., MAP) with a different model (i.e., ANN) that exploits all available timeseries of the training set and is simultaneously trained with all available durations (i.e., 1 h, 3 h, 6 h, 12 h and 24 h). The second model, EXT-ANN, is fed on the extended dataset composed of all the descriptors considered (see Table 1). It is an example of both a multivariate approach to RFA of rainfall extremes and a machine learning-based way to exploit multiple input information. The third and fourth models (EXT-PCA-ANN and EXT-CCA-ANN, respectively) make use of preprocessed versions of the same input descriptors of EXT-ANN through PCA and CCA, respectively.

PCA is a statistical technique for reducing the dimensionality of a dataset (Jolliffe, 2002, Di Prinio et al., 2011). This is accomplished by linearly transforming the data into a new coordinate system where most of the variation in the data can be described with fewer dimensions than the initial dataset. This consists of a change of basis of the original data matrix, and the new dimensions are the principal components (or PCs). Given a set of r variables ($X = (X_1, X_2, \dots, X_r)$), the covariance matrix is computed, where each element a_{ij} is the covariance between the i -th and j -th variables. The PCs are eigenvectors of the covariance matrix of the original data. The higher the number of selected PCs, the higher amount of variance of the initial matrix is caught. In this study, the variance chosen for reducing the dimensionality of the descriptors' space is 80%, that corresponds to five PCs: these are the input covariates for the EXT-PCA-ANN model. We select this threshold to allow the model to represent most of the variability of the system, yet eliminating the noise affecting the signal.

CCA (Hotelling, 1935, Di Prinio et al., 2011) is a multivariate analysis technique used to identify the possible correlations between two groups of variables. It consists of a linear transformation of two groups of random variables into pairs of canonical variables, which are established in such a way that the correlations between each pair are maximized. With specific reference to our case, the set of right-hand random variables X consists of the r input descriptors ($X = (X_1, X_2, \dots, X_r)$, where r equals twenty). As the left-hand set of variables, the s L-coefficients of variation of the AMS for each station for the analyzed durations are selected ($Y = (L-CV_{1h}, L-CV_{3h}, L-CV_{6h}, L-CV_{12h}, L-CV_{24h})$, where s is five). The objective of CCA is to construct linear combinations

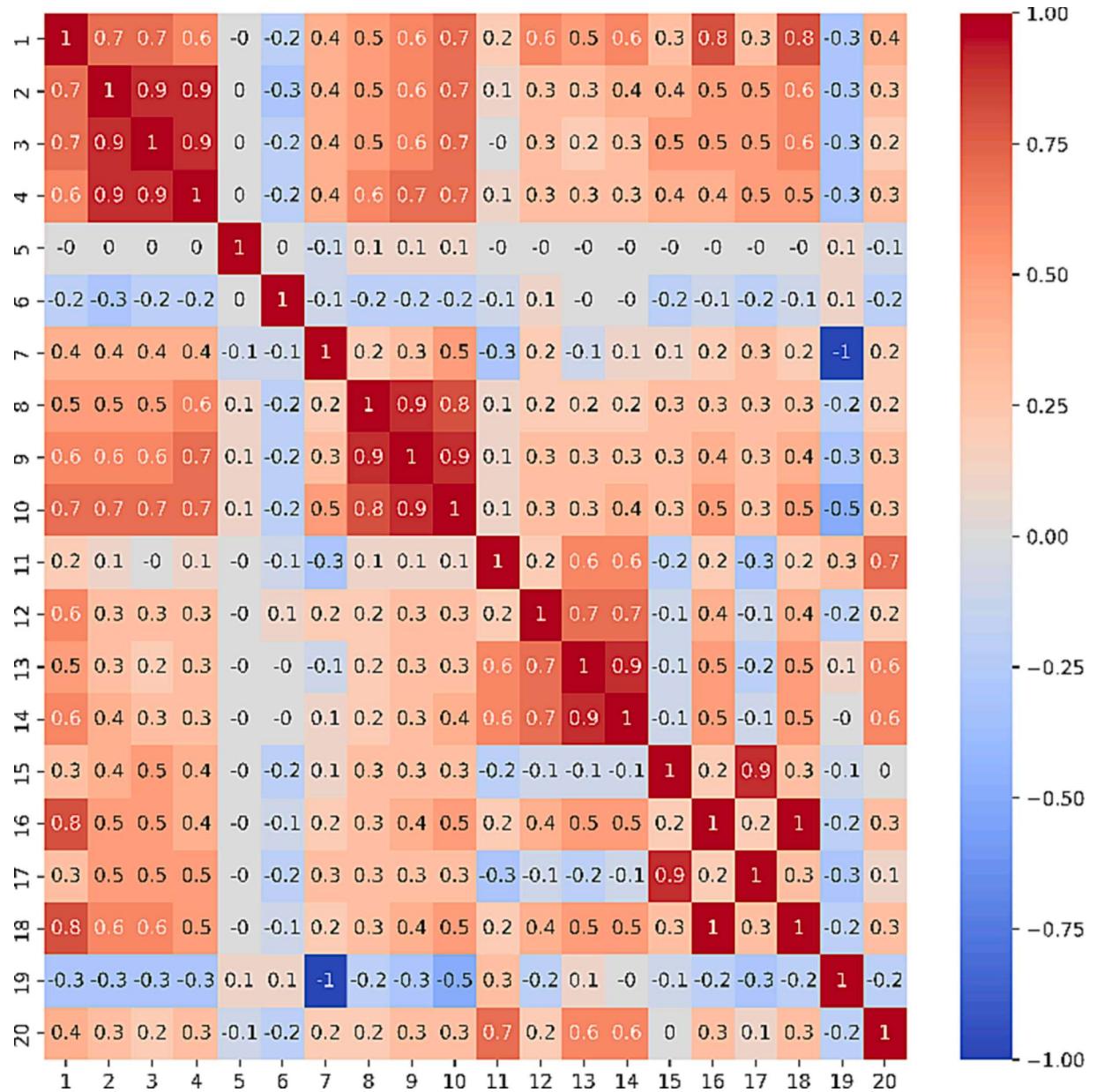


Fig. 2. Correlation matrix (i.e., matrix whose elements are PCC) of input descriptors, reported in the same order as in Table 1.

V_i and W_i (called canonical variables) of the variables X and Y , as follows:

$$V_i = A_{i1} \bullet X_1 + A_{i2} \bullet X_2 + \dots + A_{i20} \bullet X_{20} \quad (9)$$

$$W_i = B_{i1} \bullet L - CV_{1h} + B_{i2} \bullet L - CV_{3h} + \dots + B_{i5} \bullet L - CV_{24h} \quad (10)$$

where $i = 1, \dots, p$, with $p = \min(r, s)$. The first weights vectors A_1 and B_1 maximize the correlation coefficients between resulting canonical variables, under constraints of unit variance. Once the first pair of canonical variables is identified, other pairs can be obtained under the constraint that the correlation between V_i and W_j is 0 (where $i \neq j$). The five canonical variables derived from the canonical transformation of the twenty input descriptors are used as input covariates of the EXT-CCA-ANN model.

The EXT-PCA-ANN and EXT-CCA-ANN models provide the opportunity to assess the effect of PCA and CCA pre-processing techniques. It will be discussed whether the preprocessed input descriptors are able to effectively reproduce the variability of the problem, while the noise in

the real signal is absent.

The workflow for setting-up and validating the ensemble ANN models is summarized in Fig. 4. For a stable training of the ANNs, the input to the four regional ANN models, which include the dimensionless AMS and the considered morphoclimatic descriptors, is standardized (see e.g., Milligan and Cooper, 1988; Jain et al., 2005). Each standardized input ($X_{i,st}$) is obtained from the original input (X_i) by subtracting the regional mean (μ_i , mean of X_i over the whole training/testing set) and by dividing this difference by the regional standard deviation (σ_i , standard deviation of X_i over the whole training/testing set).

$$X_{i,st} = \frac{X_i - \mu_i}{\sigma_i} \quad (11)$$

The same relations used for standardizing the training/testing set (with the same μ_i and σ_i , see T1 in Fig. 3), are used also for the validation set (process V1 in Fig. 3). Accordingly, the ensemble ANN models predict Gumbel parameters for standardized and dimensionless distributions at the 100 validation sites, and these parameters need to be back-

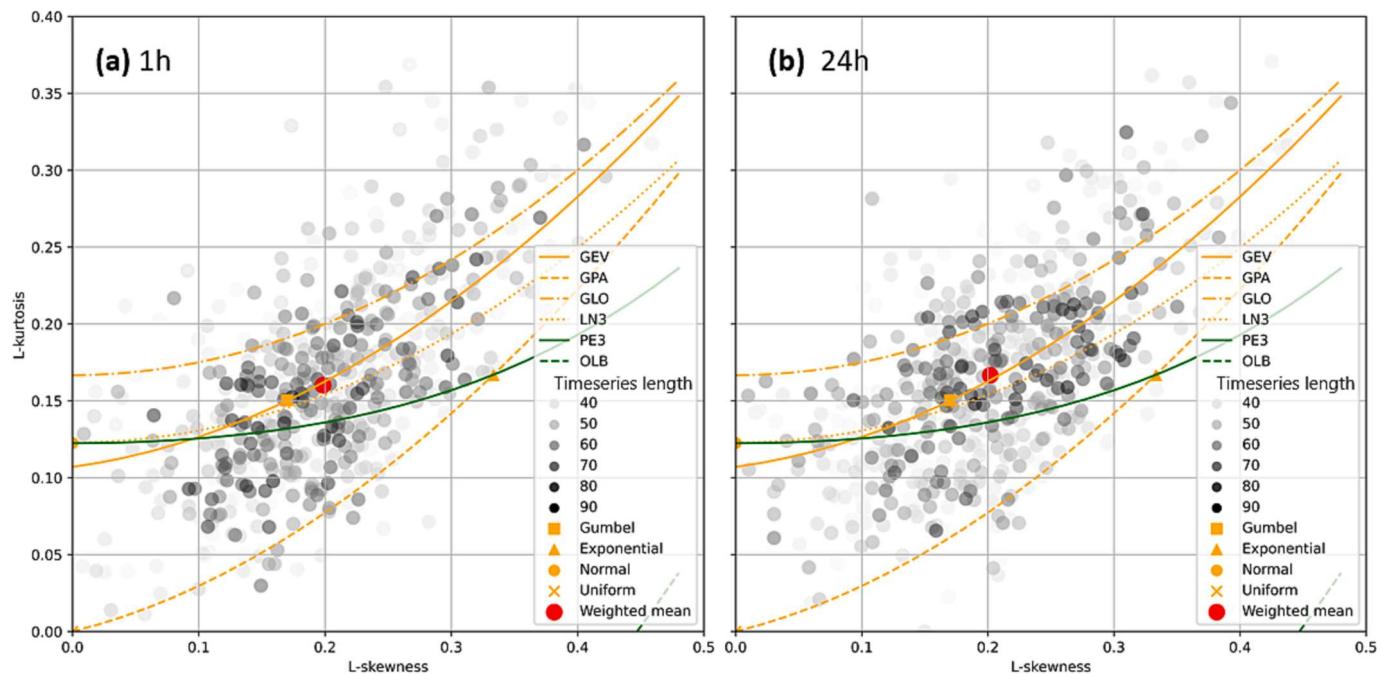


Fig. 3. L-moments ratio diagram of the 2338 gauged stations (i.e., training/testing and validation set) for annual maximum series with 1 h (a) and 24 h (b) duration.

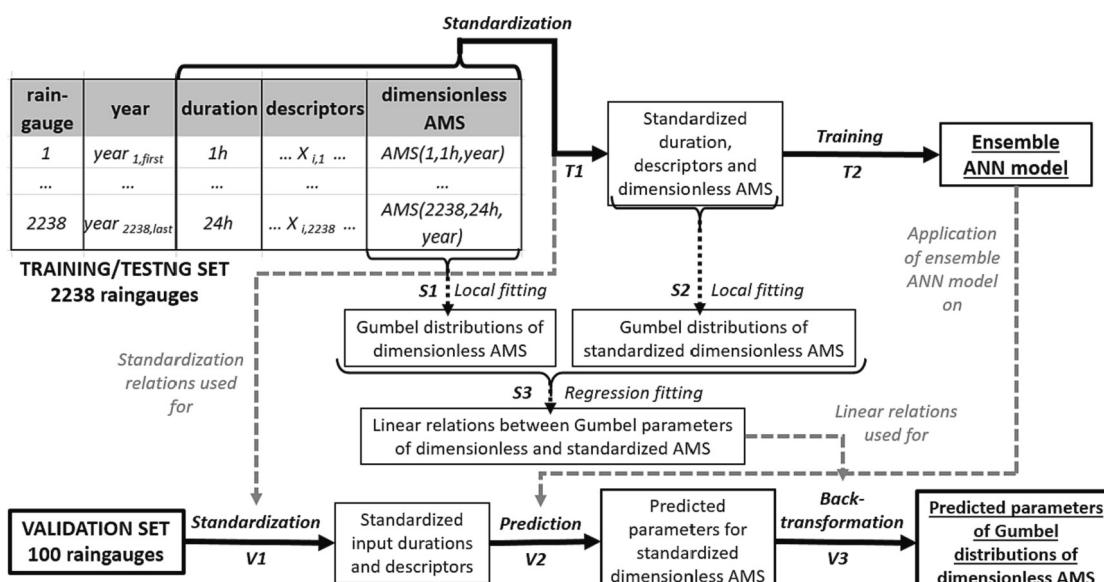


Fig. 4. Workflow for setting-up and validating the ANN models. Main processes for training (T1 and T2) and validation (V1, V2 and V3) are marked with solid black arrows; side processes (i.e., S1, S2 and S3) marked with dotted black arrows. Models and relations defined in the training phase and used for validation are marked with dashed arrows.

transformed to the dimensionless space (see process V3, Fig. 3). This is done through two empirical linear relations (i.e., one for the location parameter and one for the scale parameter) between the parameters of the locally fitted Gumbel distributions that model the frequency of the dimensionless AMS and the standardized AMS (see S1, S2 and S3 of Fig. 3).

After some preliminary experiments with different structures of the models, ensembles of 15 ANNs, each one with four layers, were found to be a good balance between prediction accuracy and computational resources required for training. We tested different proportions for splitting the training/testing set into the training and testing set (e.g., 70%-30% and 80%-20%), but we did not observe significant variations in the

results. Thus, we opted for the 80%-20% configuration, since it used a larger amount of data for the training set.

4.2. ANN models with GEV target frequency distribution: A preliminary assessment

The approach we propose is general, and can be adapted to any distribution. In particular, the usage of more flexible, 3-parameter probability distributions may allow a better representation of the highest rainfall percentiles. Nevertheless, the estimation of three parameters might be highly uncertain, even when state-of-the-art fitting methods are adopted for at-site frequency analysis (see Section 3.3).

Hence, in case the proposed ANN approach makes use of a 3-parameter distribution (or even a 4-parameter one), the usage of very short timeseries should be carefully considered both for training the models and for validating them. The findings of these test experiments may be different from case to case. Thus, in this Section we show a preliminary adaptation of the ANN models with the GEV distribution, that has to be intended as a demonstration of the flexibility of the proposed approach.

The GEV distribution is characterized by three parameters. The third parameter (i.e., the shape) controls the upper tail of the distribution, and is directly linked to the skewness of the distribution, e.g. expressed in terms of L-coefficient of skewness, or L-CS (see Hosking and Wallis, 1997). The remaining two parameters, location and scale, depend on the third one, but are also linked to the first and second L-moments, similarly to the Gumbel case. The GEV distribution and the mathematical relationships between its parameters and the L-moments can be found in Hosking and Wallis (1997) and are not reported here for the sake of brevity.

We adapted our EXT-ANN model to the GEV distribution and validated it for the same 100 validation sites by following four steps:

1. The sample L-CS is computed for the timeseries within the training/test set having at least 30 years of data.
2. The sample L-CS values from step 1 are regionalized across the study area by a geostatistical interpolation technique (Hengl, 2007).
3. For any given raingauge belonging to the 100-sites validation set:
 - a. the shape parameter of the GEV is estimated based on the regionalized L-CS value
 - b. the local L-CV value is obtained from the Gumbel scale parameter predicted by EXT-ANN with Eq. (4).
 - c. the L-CV value from step 3.b and shape parameter from step 3.a are used for deriving the remaining parameters of the GEV distribution
4. The resulting ANN model is compared with a GEV distribution whose shape parameter results from previous step 3.a, while location and scale parameters are fitted using an at-site maximum likelihood procedure

Step 2 adopts the ordinary kriging (OK) method as preliminary analyses using more complex approaches (i.e., kriging with external drift, universal kriging, see Hengl, 2007) did not improve our results. Details on the OK method can be found in several studies (e.g., Shehu et al., 2023; Hengl, 2007).

This application of the EXT-ANN model will be hereinafter referred to as EXT-ANN-GEV.

5. Performance metrics used in validation

The evaluation of the five models (i.e., baseline, MAP-ANN, EXT-ANN, EXT-PCA-ANN, EXT-CCA-ANN) is conducted by considering three aspects of the models' output Gumbel distributions: the scale parameters, and the 80th and 99th percentiles. The true values are the ones related to the Gumbel probability distributions fitted on the validation dataset with the maximum likelihood method.

Three metrics are computed to evaluate the models' performance from a global point of view: relative BIAS (BIASr), root mean squared error (RMSE), and Pearson's correlation coefficient (PCC, see Section 3). The first two are commonly used in literature (e.g., see Msilini et al., 2020; Ghamariadyan and Imteaz, 2021; Shu and Ouarda, 2007), and quantify the systematic error of the models (BIASr) and the gap between the predicted and expected values of the considered variables (RMSE). Differently, the PCC does not take into account the actual values of the variables, since it simply measures the degree of linearity of the relationship between the empirical reference values and the corresponding model's predictions.

$$BIASr = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - y_{i,pred}}{y_i} \right); \quad n = \sum_{j=1}^{n_t} n_j \quad (18)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_{i,pred})^2} \quad (19)$$

Where n is the total number of validation observations (i.e., for a given duration, the sum of the total years of annual maxima records n_j over the n_t validation stations), and each annual maximum is considered as a single i -th observation. Thus, any station is counted as many times as its timeseries length (see also Fig. 4), and the final metrics mainly depend on the longest timeseries. While y_i is the true value of the output variable (i.e., the one related to the fitted Gumbel distributions) for the i -th observation, $y_{i,pred}$ is the value of the output variable obtained with the regional model.

One more metric is computed to evaluate models' accuracy for each single station and each single duration. This is herein referred to as percent relative error (PRE) and defined as in eq. 14.

$$PRE [\%] = \frac{y_{pred} - y}{y} \cdot 100 \quad (20)$$

Positive values of PRE represent overestimation with respect to the true values (y), while negative ones account for underestimation.

6. Validation of the regional models

After training, the five models described in Sections 2 and 4 (i.e., MAP-Lm, MAP-ANN, EXT-ANN, EXT-PCA-ANN and EXT-CCA-ANN) are used to predict Gumbel distributions for the dimensionless annual maximum rainfall at the locations of the validation stations. We considered the performance metrics associated with the estimation of the Gumbel scale parameters and two dimensionless rainfall quantiles, namely the quantiles associated with the 0.8 and 0.99 non-exceedance

Table 2

Performance metrics for estimated scale parameter for Gumbel distributions of dimensionless annual maxima associated with different durations at 100 validation raingauges. For each duration, the best case for each metric is marked in bold, while the worst is in italic.

Metrics	MAP-Lm	MAP-ANN	EXT-ANN	EXT-PCA-ANN	EXT-CCA-ANN
1 h					
BIASr	-0.075	-0.038	-0.051	-0.025	-0.046
RMSE	0.050	0.049	0.050	0.048	0.048
PCC	0.120	0.105	0.242	0.252	0.275
3 h					
BIASr	-0.083	-0.046	-0.053	-0.044	-0.047
RMSE	0.050	0.047	0.044	0.044	0.044
PCC	-0.031	0.015	0.422	0.387	0.394
6 h					
BIASr	-0.081	-0.048	-0.047	-0.060	-0.038
RMSE	0.050	0.048	0.043	0.045	0.044
PCC	-0.037	0.176	0.453	0.387	0.384
12 h					
BIASr	-0.068	-0.046	-0.029	-0.051	-0.030
RMSE	0.049	0.048	0.042	0.045	0.045
PCC	-0.063	0.204	0.463	0.337	0.350
24 h					
BIASr	-0.079	-0.061	-0.045	-0.050	-0.045
RMSE	0.048	0.047	0.039	0.042	0.042
PCC	0.030	0.139	0.548	0.421	0.420

probabilities.

Concerning the metrics for the scale parameters (see Table 2), it is possible to observe that RMSE and PCC present a similar behavior across durations and models, which differs from the outcomes in terms of BIASr. Indeed, RMSE and PCC tend to show optimal values for the same regional model, which is often different from the regional model characterized by the smallest BIASr. The value of BIASr is generally between a few % and 9% for the scale parameters. Overall, the estimation of a regional model for the 1 h duration seems to be more complex than for higher durations, as it is pointed out by higher values of the PCC for 12 and 24 h. The MAP-Lm model is the least accurate according to the RMSE and PCC metrics, while the MAP-ANN model is the second-to-least accurate, but it is still slightly better than the MAP-Lm. Even though the EXT-ANN, EXT-PCA-ANN and EXT-CCA-ANN models have similar performances, the preprocessing of data leads to a better performance of EXT-PCA-ANN and EXT-CCA-ANN compared to the EXT-ANN model for the 1 h duration, while the EXT-ANN is the best one for durations of 6, 12 and 24 h.

The metrics computed for the 80th and 99th percentiles are reported in Table 3.

It is generally possible to observe a good agreement between the metrics listed in Table 2 (scale parameter) and in Table 3 (rainfall percentiles). The EXT-ANN, EXT-PCA-ANN and EXT-CCA-ANN models show a similar performance according to all metrics. As for the scale parameter, BIASr values are discordant with the other two metrics, but always very small in all cases (a few % at most). For longer durations (i.e., 6, 12 and 24 h) EXT-ANN is the best performing model according to RMSE and PCC; differently, for short durations (i.e., 1 or 3 h), the best performing model depends on the metric being considered, but the models with preprocessing usually outperform the EXT-ANN.

Since the overall best performing model is in general achieved by the EXT-ANN model, we carried out a detailed analysis of its behavior. Fig. 5 reports the geographical distribution of the PRE of the 99th percentile predicted for 1 and 24 h durations (panels a and b, respectively). First,

no clear geographical pattern of the prediction error is visible: the goodness of the prediction for both 1 and 24 h does not seem to be linked to elevation, nor geographical location, and shows similar geographical variability for both durations. Absolute values of PRE ($|PRE|$) are higher than 50% in one case only for the set of 100 validation locations. Most of stations have low $|PRE|$ values (i.e., $|PRE| < 5\%$ for 44 and 43 validation locations for 1 h and 24 h duration, respectively). The number of validation locations showing $20\% < |PRE| < 50\%$ for a duration of 1 h (i.e., 9) is larger than for a 24 h duration (i.e., 6).

The same analysis of PRE for the 99th percentile obtained from the MAP-Lm model is presented in Fig. 6. As for the EXT-ANN model, no clear geographical pattern of the PRE is observed and most of the stations have PRE between -20% and 20% (i.e., 17+37+35 for 1 h, 19+33+37 for 24 h). In particular, the stations with PRE between -5% and 5% are 37 for 1 h and 33 for 24 h (lower numbers when compared to Fig. 5), while the stations with $PRE > 20\%$ or $< 20\%$ are 11 for both time-intervals (i.e., 3+7+1 for 1 h and 6+4+1 for 24 h, higher values when compared to Fig. 5).

Finally, the results of model EXT-ANN are used to obtain the location and scale parameters of a GEV distribution (see Section 4.2). Table 3 also reports the global metrics obtained for the 80th and 99th percentiles of dimensionless rainfall depth (columns “EXT-ANN-GEV”). These are relative to empirical predictions of the same percentiles adopting a GEV distribution and the hybrid local/regional estimator described in Section 4.

While BIASr and RMSE are very similar to the previous case of application with the Gumbel ANN (columns “EXT-ANN” of Table 3), PCC values are significantly lower, with the exception of 1 h, which has the highest PCC.

7. Interpolation across space and time-aggregation interval

One of the most innovative and useful aspects of our AI-based approach is its capability to provide predictions of the dimensionless

Table 3

Performance metrics for predicted 80th and 99th percentiles of dimensionless rainfall depth associated with different durations at 100 validation raingauges. For each duration, the best case among the models MAP-Lm, EXT-ANN, EXT-PCA-ANN and EXT-CCA-ANN is marked in bold for each metric, while the worst is in italic. The column EXT-ANN-GEV reports the metrics for a demonstration of the adaptability of EXT-ANN to the GEV distribution.

80th percentile							99th percentile						
metrics	MAP-Lm	MAP-ANN	EXT-ANN	EXT-PCA-ANN	EXT-CCA-ANN	EXT-ANN-GEV	metrics	MAP-Lm	MAP-ANN	EXT-ANN	EXT-PCA-ANN	EXT-CCA-ANN	EXT-ANN-GEV
1 h							1 h						
BIA Sr	-0.009	-0.003	-0.004	0.001	-0.004	0.008	BIA Sr	-0.02	-0.007	-0.013	0	-0.011	0.022
RMSE	0.044	0.043	0.043	0.042	0.042	0.049	RMSE	0.198	0.194	0.196	0.191	0.189	0.197
PCC	0.103	0.077	0.24	0.268	0.251	0.237	PCC	0.105	0.101	0.242	0.257	0.272	0.378
3 h							3 h						
BIA Sr	-0.012	-0.009	-0.011	-0.009	-0.01	0.031	BIA Sr	-0.021	-0.019	-0.02	-0.019	-0.02	0.060
RMSE	0.045	0.04	0.038	0.037	0.038	0.088	RMSE	0.201	0.186	0.172	0.172	0.172	0.276
PCC	-0.096	0.034	0.405	0.375	0.361	0.202	PCC	-0.082	0.02	0.42	0.387	0.389	-0.015
6 h							6 h						
BIA Sr	-0.007	-0.009	-0.009	-0.011	-0.006	0.030	BIA Sr	-0.013	-0.019	-0.02	-0.025	-0.014	0.057
RMSE	0.045	0.041	0.038	0.04	0.039	0.096	RMSE	0.206	0.187	0.171	0.178	0.175	0.279
PCC	-0.152	0.175	0.395	0.319	0.307	0.284	PCC	-0.129	0.175	0.442	0.375	0.369	0.082
12 h							12 h						
BIA Sr	-0.006	-0.008	-0.005	-0.008	-0.004	0.028	BIA Sr	-0.011	-0.017	-0.01	-0.02	-0.01	0.053
RMSE	0.045	0.042	0.038	0.041	0.04	0.098	RMSE	0.205	0.188	0.169	0.181	0.178	0.275
PCC	-0.087	0.146	0.394	0.242	0.255	0.235	PCC	-0.092	0.194	0.45	0.318	0.331	0.142
24 h							24 h						
BIA Sr	-0.006	-0.01	-0.007	-0.008	-0.007	0.011	BIA Sr	-0.014	-0.025	-0.018	-0.02	-0.017	0.020
RMSE	0.044	0.041	0.036	0.039	0.039	0.062	RMSE	0.198	0.185	0.156	0.17	0.169	0.205
PCC	-0.07	0.103	0.474	0.319	0.316	0.444	PCC	-0.065	0.134	0.534	0.401	0.4	0.201

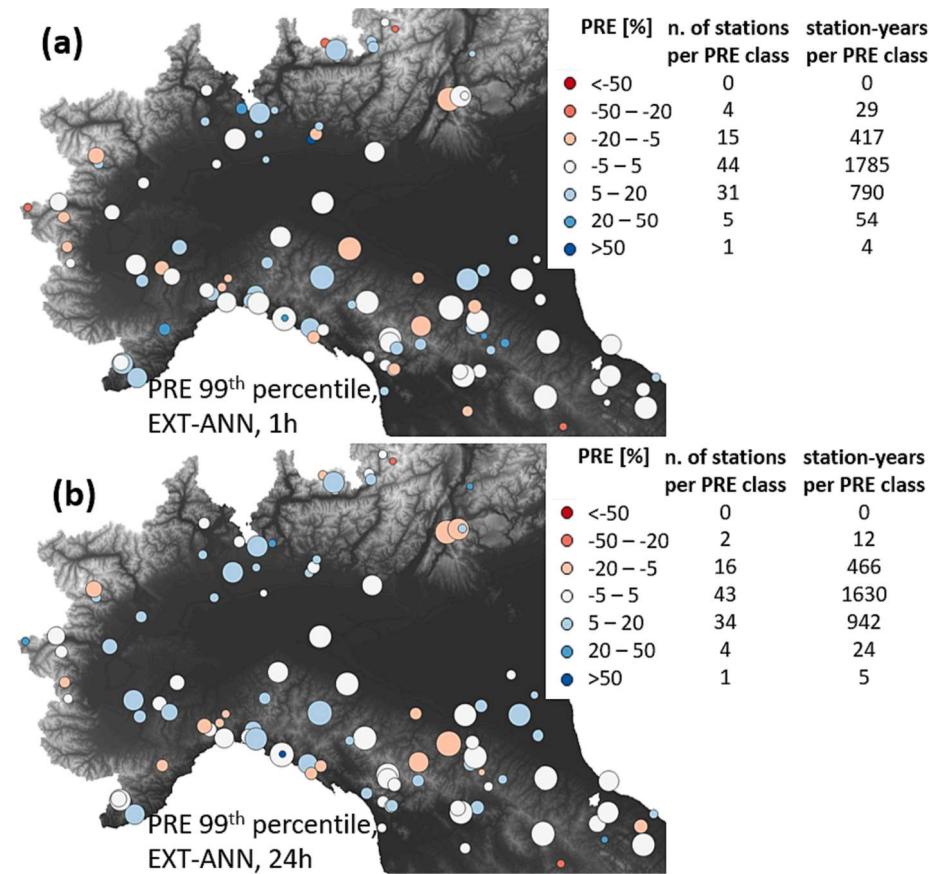


Fig. 5. Percent relative error (PRE) of EXT-ANN dimensionless 99th percentiles at 100 validation raingauges for 1 h (a) and 24 h (b) durations; larger circles represent longer annual maximum series. The number of test raingauges, and the corresponding station-years of data, is reported for each PRE category.

rainfall distribution in any location (spatial interpolation) and for any time-aggregation interval (i.e., duration) between 1 and 24 h (time-aggregation interpolation).

As an example, four stations in different geographic and climatic contexts (Table 4) are selected for time-aggregation interpolation. Fig. 7 shows the Depth Duration Frequency (DDF) curves obtained with the EXT-ANN model in the four raingauges. Dimensionality was reintroduced, and consistency among percentiles ensured, by multiplying the predicted dimensionless percentiles by the mean extreme precipitation for each duration (Eq. (1)). The latter was obtained by applying the scale-invariance hypothesis to the mean extreme precipitation (see Burlando and Rosso, 1996) and using a power scale law between time-aggregation and mean precipitation. Since the focus of the present study is regional modelling of the growth factor of the storm index method, estimation of the index rainfall (i.e., the mean extreme rainfall depth, see Sect. 2) with multiple scaling is not discussed.

Some observations can be highlighted. First, in stations 9086, 5143 and 16126, the EXT-ANN model has greater accuracy at 1 h and 24 h, while it is not fully capable of reproducing the fitted model at 6 h and 12 h; this confirms the metrics in Tables 2 and 3. Station 17020 is a case of underestimation of the EXT-ANN (see also same station in Fig. 4, with PRE in $-20\% - -5\%$ range), where the traditional MAP-Lm model performs better than the ANN-based one. The MAP-Lm approach has similar performances in all four stations: greater errors for longer return periods and longer durations.

Regional ensemble ANN models presented here can produce a spatial interpolation of estimated frequency distributions (i.e., Gumbel distribution in this study) based on the gridded discretization of the study area used for retrieving the local values of the morphoclimatic indices. Panels (a) and (b) of Fig. 8 show the scale parameters predicted by the EXT-ANN model for the dimensionless distributions of 1 h and 24 h

annual maximum rainfall depths over the drainage area of an Apennine catchment in north-central Italy (i.e., Panaro river basin, drainage area $\sim 2300 \text{ km}^2$). It can be observed that the average value is higher for 1 h duration than 24 h (i.e., 0.28 against 0.25). The relation between the predicted scale and the elevation is directly proportional to the MAP and is further explored in panels (c) and (d). Here, a rather evident decrease of the scale parameter for 1 h duration is observed when the elevation is growing, while no clear regression is obtained for 24 h. This seems to be in agreement with recent findings of Marra et al. (2021), who observed a significant decreasing trend of the Weibull scale parameters with elevation for sub-hourly durations, while no significant trend was detected for longer time-intervals.

8. Discussion

The proposed approach aims at improving the predicting ability of the traditional L-moments storm index model with the action of three combined strategies: (1) exploiting complex non-linear regional functions (i.e., through ANN ensembles), (2) increasing the amount of data used for training the regional models, and (3) increasing the number of proxies for extreme precipitation. The first two points are discussed through the comparison between MAP-Lm and MAP-ANN (subsection 8.1), while the third one regards the EXT-ANN, EXT-PCA-ANN and EXT-CCA-ANN models (subsection 8.2).

8.1. Comparing MAP-Lm with MAP-ANN

Metrics in Tables 2 and 3 show that BIAS_r and RMSE are very similar for MAP-Lm and MAP-ANN, while PCC has significant variations and, more in general, rather low values (i.e., lower than 0.60). This is due to the different nature of the three metrics (see Section 4): a small

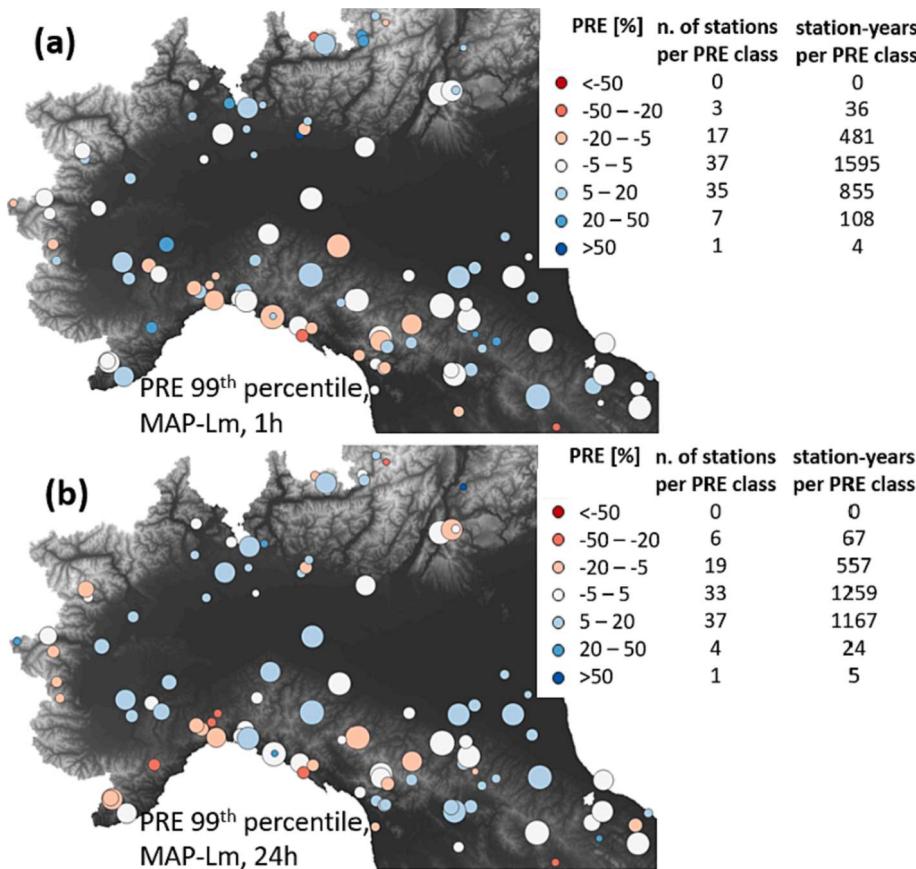


Fig. 6. Percent relative error (PRE) of MAP-Lm dimensionless 99th percentiles at 100 validation raingauges for 1 h (a) and 24 h (b) durations; larger circles represent longer annual maximum series. The number of test raingauges, and the corresponding station-years of data, is reported for each PRE category.

Table 4

Main characteristics of the four stations adopted for the time-aggregation interpolation application through EXT-ANN model.

Code	Location	Record length	Mean elevation [m a.s.l.]	Minimum distance from the Adriatic coast [km]	Minimum distance from the Thyrrenic coast [km]	MAP [mm]	MASnow [mm]
9086	Codogno	64	60.28	202.49	98.18	802.58	7.28
17020	Folgaria	58	1114.7	109.60	225.73	1219.16	193.83
5143	Isola di Palanzano Centrale	70	723.13	167.62	43.18	1416.89	39.28
16126	La Verna	76	1080.65	62.29	124.64	1174.25	163.38

difference in BIAS_r and RMSE confirms that prediction errors from the two approaches are very similar. Anyway, the higher PCC values (with the exception of 1 h time-interval) suggest that changing model type and data management strategy introduces some benefits (i.e., positive correlation between expected and predicted variables). Overall, it is evident that the MAP-ANN model has unsatisfactory accuracy (i.e., maximum PCC 0.204 for scale with 12 h time-interval). However, this could be considered a good result when carefully looking at the regional relation developed within the MAP-Lm framework (see Fig. 9 for 1 h and 24 h cases). In particular, the L-CV(MAP) relations found in the present study (see red lines in Fig. 9) do not show clear and strong relationships between L-CV and MAP, especially for the 24 h case. The comparison with an empirical relation reported in the literature (i.e., Di Baldassarre et al., 2006, see dotted red lines in Fig. 9) suggests that when a large and morphologically and climatically complex region is considered, the classical MAP-Lm approach may not be a viable regionalization strategy. The discrepancies between the relationship identified in this study and that of Di Baldassarre et al. (2006) is more evident for long time aggregation intervals and less pronounced for short durations. In fact, 24 h annual maxima are generally associated with frontal disturbances in the

study region, and show a complex geographical variability of annual maxima statistics (see e.g. Mazzoglio et al., 2022). Differently, 1 h annual maxima are mostly the result of convective storms, which have more spatially homogeneous statistics (see e.g. Schaefer, 1990, Alila, 1999, Di Baldassarre et al., 2006).

8.2. EXT-ANN, EXT-PCA-ANN and EXT-CCA-ANN

For any time-aggregation interval (i.e., duration), the EXT-ANN, EXT-PCA-ANN and EXT-CCA-ANN models outperform the MAP-Lm and MAP-ANN when referring to the global metrics (Tables 2 and 3). The difference is particularly significant for the PCC, which shows that the EXT- models (that make use of an EXTended set of descriptors) are much more effective in capturing the overall trend of the variables (i.e., location, scale, 80th and 99th percentiles) within the study area. However, the maximum value of PCC for the scale parameter is still very low for 1 h (i.e., 0.275), and lower than 0.6 for 24 h. This clearly shows that the EXT- models are more accurate when modelling rainfall phenomena with longer durations, as it is also evident from the higher number of high absolute PRE values (i.e., Percent Relative Error) for 1 h (i.e., 10

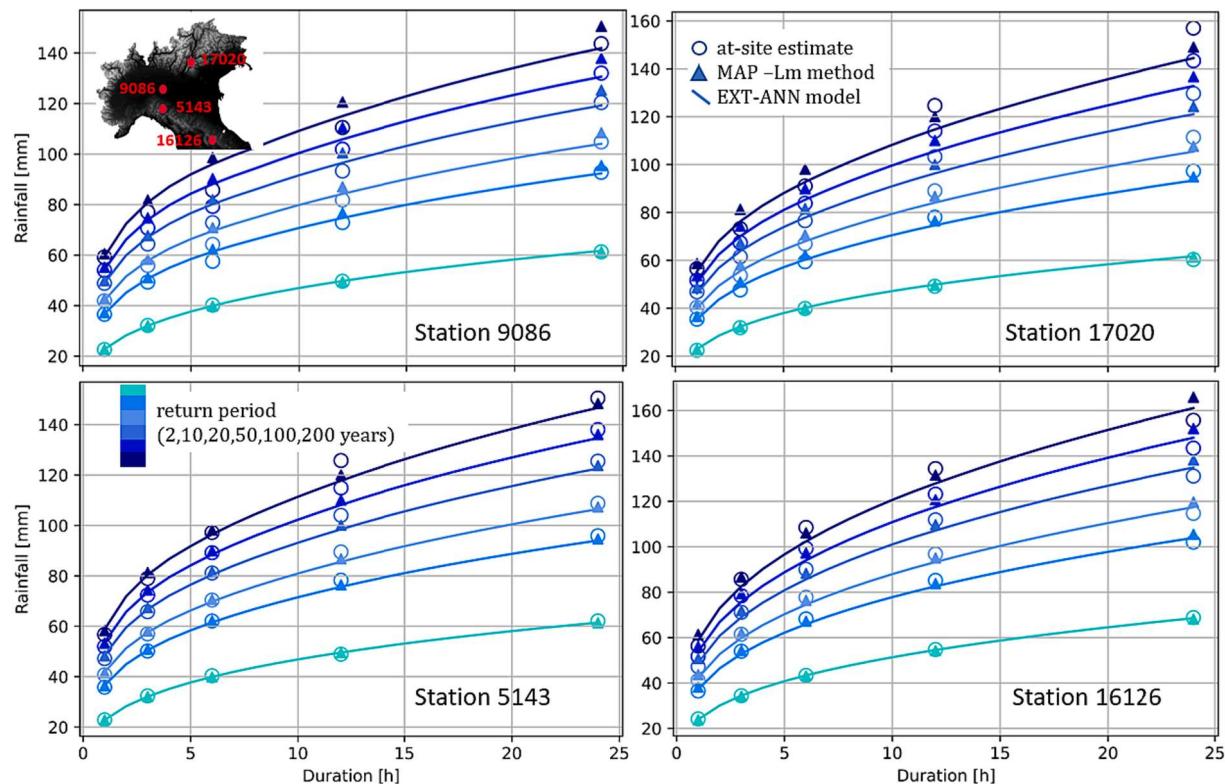


Fig. 7. Depth-duration-freency (DDF) curves obtained with EXT-ANN and MAP-Lm models for stations 9086, 17020, 5143, and 16126 (see also Table 4).

cases in total out of 100) than for 24 h (i.e., 7 cases) for the EXT-ANN (Fig. 5). A possible reason for these results is that convection phenomena are less influenced by the morphology (see, e.g., Schaefer, 1990, and Alila, 1999), leading to lower predictive power of most of the input descriptors. Anyway, it is worth highlighting that PRE values $< -20\%$ or $> 20\%$ are mainly observed in Fig. 5.a in stations where long data series are not available and the expected values (i.e., locally fitted Gumbel distributions) are less reliable, while numerous stations with longer AMS show good results (i.e., PRE values between -5% and 5%).

Also, it is interesting to note that percentiles obtained from the MAP-Lm approach are rather similar to the ones from local data frequency analysis in some cases (see Fig. 7), even if the model is very simple. However, when looking at the PRE values for the dimensionless 99th percentile (see Figure 6), and comparing it with the one predicted with EXT-ANN, the lower accuracy of the former MAP-Lm is evident. First, gauged locations associated with high PRE values for the 99th percentiles are more for MAP-Lm (i.e., 11 for 1 h, and 11 for 24 h, see Fig. 6) than for the EXT-ANN (see above). Accordingly, locations associated with lower PRE values (i.e., 37 for 1 h and 33 for 24 h) are less (i.e., for EXT-ANN, 44 for 1 h and 43 for 24 h).

Regarding the gridded EXT-ANN predictions (Fig. 8), it is difficult to objectively compare our results with previous knowledge, as studies on the link between parameters of Gumbel distributions of dimensionless annual maxima and orography are lacking. Marra et al. (2021) observed the relations between scale parameter of Weibull distributions of ordinary events and elevation in Israel. This seems to be aligned with the trends of the linear and non-linear relations found in the present study between the scale parameter and elevation for 1 h and 24 h (Fig. 8c and d).

Since we use a large number of gauged sites that are described by several correlated indices and annual maximum series of very different length, the real signal of the regional function is expected to be disturbed by a certain noise. Thus, the actual impact of preprocessing morphological and climate descriptors is worth analyzing. Multivariate preprocessing techniques generally allow to reduce the noise, so that the

models train to reproduce the signal. However, it is remarkable here that data preprocessing (i.e., principal component analysis, or PCA, and canonical correlation analysis, or CCA) seems to have a positive impact just for the 1 h duration, while for longer durations the EXT-ANN model performs the best (see Tables 2 and 3). This is a positive result, as it clearly shows how the ensemble ANNs can successfully handle large datasets with several couples of variables that are strongly inter-correlated (see dark colored cells in Fig. 2). In general, it is not possible to affirm whether or not preprocessing the input data should be preferred to using the raw dataset, but some conclusions can be drawn. First, preprocessing can impact the performance of the model, in a positive but also negative way. Second, PCA and CCA have similar performances. Third, in absolute terms, all EXT-regional models have very similar performances and prediction accuracy. Fourth, given the similar performance, it has to be mentioned that reducing the dimensionality of the problem and making the models simpler through preprocessing techniques has significant computational advantages in the training phase.

Concerning the preliminary adaptation of the proposed ANN approach to the GEV (i.e., Generalized Extreme Value) distribution, the results are encouraging, as the BIAS_r and RMSE are very similar to the ones computed for the Gumbel (Table 3). However, a correct estimation of the higher order statistical moments, and therefore of shape parameters of 3- and 4-parameter distributions, remains a critical aspect for the use of more flexible probability distributions. The sampling variability of the predictions of GEV parameters used in this study is probably the main reason for the lower PCC values obtained for the EXT-ANN-GEV relative to the EXT-ANN, which indicates a weaker linearity between ANN predictions of dimensionless GEV rainfall percentiles and their empirical validation counterparts. To further investigate this aspect, we computed the same performance metrics by considering only the validation stations with timeseries longer than 40 years. For the sake of brevity, these results are not reported in Table 3, but an improvement of the performance was observed, leading to PCC values aligned to those associated with the best performing Gumbel ANN models. This confirms

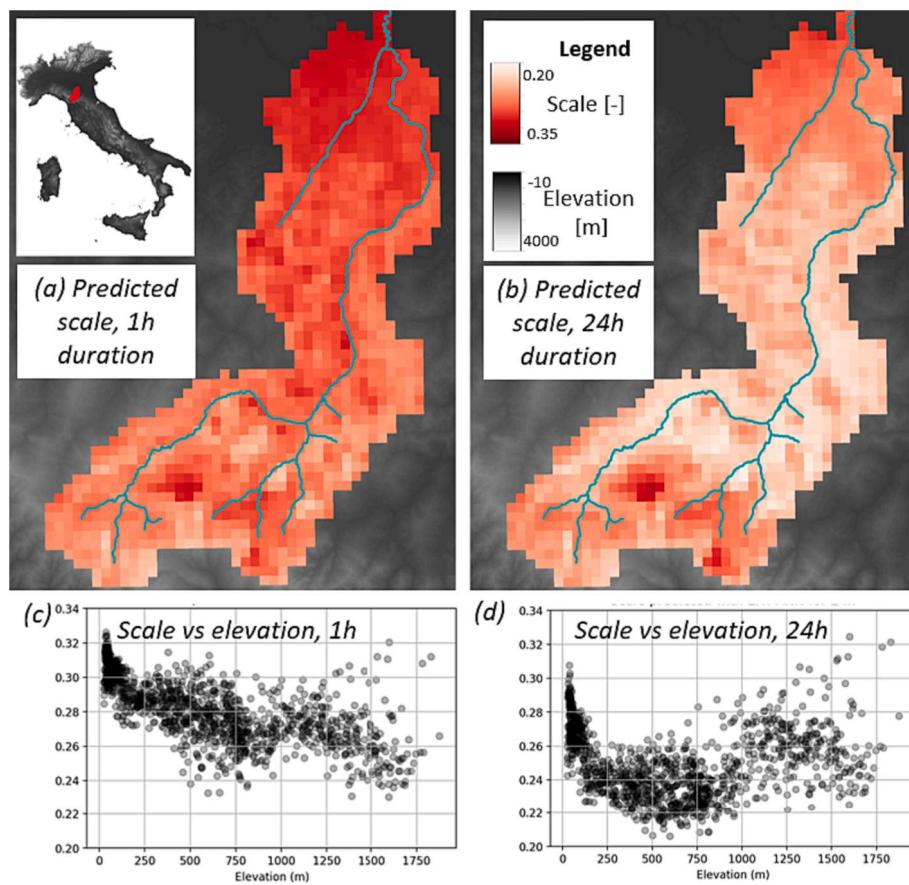


Fig. 8. Raster-based EXT-ANN prediction of Gumbel scale parameters for 1 h (a) and 24 h (b), obtained for an example river catchment in the study area (i.e., Panaro river catchment); the main river network is reported in light blue. Scatterplots of scale parameters against elevation values for 1 h (c), and 24 h (d), from raster-based prediction.

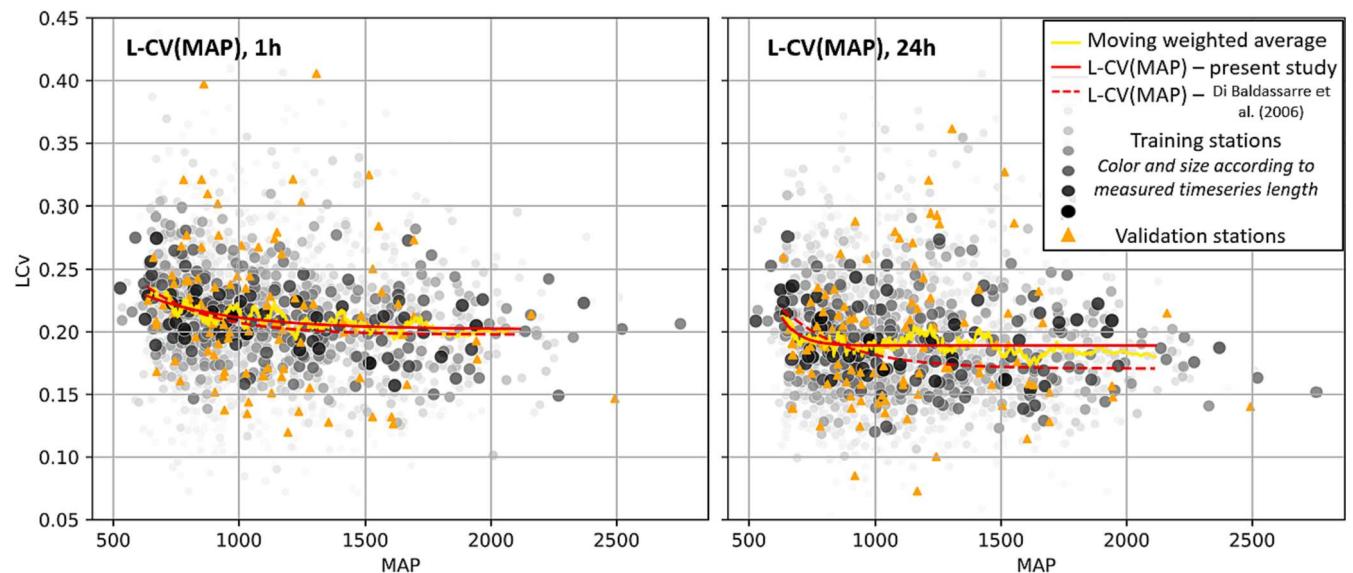


Fig. 9. Empirical L-CV and local MAP values across the study area (dots); moving weighted average (yellow line); Horton-type regional relationship L-CV(MAP) in eq. (2) fitted to the moving weighted average (red solid line) and found by Di Baldassarre et al. (2006) over north-central Italy (red dashed line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that the record length strongly affects the training and validation of the EXT-ANN-GEV approach. Although we showcase the adaptability of the proposed ANN approach to probability distributions with more than two parameters, dedicated studies are needed for assessing the impact and

benefit of short timeseries for the models' training and validation.

In conclusion, the proposed AI-based approach shows satisfactory accuracy relative to classical regionalization methods, and significantly superior performances for time-aggregation intervals equal to or longer

than 12 h. It also has the advantages of being applicable over a very large study area, and allowing to model any time-aggregation interval between 1 and 24 h, which automatizes the construction of duration-depth-frequency curves. However, some drawbacks and margins of improvements are still present. First, the accuracy with 1 h is still low, which could be improved with a more complex model architecture. Second, some of the input variables (i.e., morphological and climatic descriptors) are not easy to retrieve and compute and not always available, which affects model applicability. In particular, descriptors related to the distance from the two coasts (i.e., variables 7–14, Table 1) can require significant GIS-computational resources to be retrieved for a large number of points. This can be a limit specially in a raster-based or gridded application aimed to produce spatially interpolated maps as in Fig. 7. Such a problem could be solved with a reliable ranking of the input descriptors' influence on the final results. However, no direct method exists for input features' ranking in ANNs, and the weights computed for the PCA and CCA are not informative, as in disagreement. Given also the case-specific meaning of this analysis, which is in contrast with the general aims of this study, this point needs to be addressed by future works.

Finally, the present work considered the 2-parameter Gumbel distribution. More flexible statistical models (e.g., 3-parameter GEV, Jenkinson, 1955; 4-parameter Kappa distribution, see e.g. Blum et al., 2017) could eventually be used. Preliminary experiments performed with a GEV distribution highlight strong potential of the approach. However, further testing of the robustness of our approach is needed for models with more than 2 parameters in which for instance the skewness (3-parameter) and the skewness and kurtosis (4-parameter) need to be modelled.

9. Conclusions and further steps

Regional frequency analysis (RFA) is commonly adopted for estimating extreme hydrological variables such as floods or extreme rainfall where local measurements are unavailable or insufficient for at-site frequency analysis. Different approaches have been proposed for the RFA of rainfall extremes, each one characterized by specific advantages and disadvantages (see e.g. Claps et al., 2022). One of the most common drawbacks is that regional models specifically refer to a single duration or a single exceedance probability. Several approaches require the definition of a homogeneous region where the model is trained; this leads to higher accuracy. However, the applicability of the model is then limited to locations that are hydrologically similar to the homogeneous group used in the training. Moreover, most models require filtering the available gauged stations based on the length of the measured timeseries to perform reliable frequency analysis. These aspects lead to discarding a significant amount of data, which could turn out to be detrimental to the accuracy of the regional predictions in some cases.

This study proposes a new approach for estimating the growth factor within the storm index framework for extreme rainfall RFA. This is the dimensionless percentile associated with a given duration and exceedance probability. By multiplying the growth factor by the mean extreme rainfall with the same duration, the dimensional percentile can be obtained (Dalrymple, 1960). Our approach is based on ensemble unsupervised artificial neural networks (ANNs), that are capable of predicting the location and scale parameters of a Gumbel distribution of the dimensionless rainfall for any sub-daily time aggregation interval (duration) in the 1–24 h range.

The study area consists of a large region in north-central Italy, where a wide variety of morpho-climatic contexts are present. From the I²-RED dataset (Mazzoglio et al., 2020), 2338 gauged stations are selected, where measurements of annual maximum rainfall depths are available in the 1931–2019 record period for the 1, 3, 6, 12, 24 h time-intervals (i.e., durations). To train the models 2238 gauged stations are used, while the remaining 100 serve as validation set.

Following one approach proposed in the literature (Di Baldassare

et al., 2006) that showed good results over a significant portion of the selected study area, a baseline regional model is developed (MAP-Lm). This consists of a relationship between the L-coefficient of variation and the mean annual precipitation (MAP). Then, four applications of the ANN-based approach are set up: the first (MAP-ANN) makes use of the MAP as unique input covariate; the second (EXT-ANN) makes use of an extended number of twenty variables, including morphology (e.g., elevation, slope, aspect, distance from the coast), climate (i.e., mean and standard deviation of snow and liquid precipitation) and geographical coordinates of the stations. The fourth and fifth models (EXT-PCA-ANN and EXT-CCA-ANN, respectively) make use of the same extended dataset, but apply two different preprocessing strategies of the input morphoclimatic covariates (or descriptors), namely principal component analysis (PCA) and canonical correlation analysis (CCA).

This method is innovative for several reasons. First, it does not require the identification of a homogeneous group of sites for model training and application. Second, it uses all available annual maximum data, regardless of the length of the annual sequence (which can be very short, and even two observations). Third, training is simultaneously performed for all durations. These characteristics lead to high interpolation ability, meaning that a single model can predict Gumbel distributions for the extreme rainfall in every point in the spatial domain, and for any duration in the 1–24 h range.

The performances of the regional models are analyzed through global metrics (e.g., Pearson correlation coefficient, or PCC), that sum up prediction accuracy over all the validation set, and through the percent relative error (PRE) at each single validation station. Results indicate that the classical approach MAP-Lm appears to have low accuracy when applied to a very large and morpho-climatically heterogeneous region (i.e., PCC ~0.1 for 99th dimensionless percentiles of annual maximum depth for a 1 h duration, and PCC ~ −0.1 for a 24 h duration). The MAP-ANN method, which uses the same input information (i.e., MAP) but a different approach (i.e., simultaneous use of durations, and ANNs ensembles) shows a slightly better performance, but still low accuracy. When twenty descriptors of the local morphoclimatic conditions are used, ANN-based models show a significant improvement over the MAP-Lm and MAP-ANN (i.e., PCC ~0.3 for the 99th dimensionless percentile for 1 h, and 0.5 for 24 h). Even if the maximum PCC are still low, when considering the local PRE over the 100 validation stations, the improvement of the new approach is evident, as the number of stations with low relative error (i.e., PRE between −5% and 5%) increases (i.e., 44 and 43 for EXT-ANN at 1 and 24 h durations, respectively, versus 37 and 33 for the MAP-Lm).

Also, ensemble ANNs show good ability to handle complex and heterogeneous datasets, even without data preprocessing. PCA and CCA seem to have a slight positive effect in modelling short duration extremes, while their impact is limited for longer durations.

In conclusion, based on the outcomes of our study we can affirm that using ensemble ANN models with a few traditional descriptors (i.e., local MAP value as in Schaefer, 1990 and similar and more recent regional studies) does not lead to significant advantages over a traditional method (i.e., statistics of extremes rainfall event as empirical functions of local MAP value). However, when combined with multiple morphological and climatic descriptors, the improvement can be remarkable, particularly for annual maximum rainfall depths associated with longer time-aggregation intervals (between 12 and 24 h in this study). Time and space interpolation ability of the ANNs over the 1–24 h range and across the entire study area enable practitioners to directly obtain depth-duration-frequency curves or raster maps of rainfall extremes associated with a given duration and exceedance probability. Future analyses should build on this preliminary study and address some of the current limitations of the approach. First, methods should be further developed in order to improve the accuracy for extremes originated by convective events. Second, more flexible distributions should be considered (e.g., Generalized Extreme Value). Preliminary experiments in this direction produced encouraging results. Third, some additional research should

aim at identifying the most effective and descriptive morphoclimatic indices, including alternative or complementary information to the descriptors considered in this study.

CRediT authorship contribution statement

Andrea Magnini: Conceptualization, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. **Michele Lombardi:** Conceptualization, Software. **Taha B.M.J. Ouarda:** Conceptualization, Methodology, Writing – original draft. **Attilio Castellarin:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing.

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.

Data availability

All the data are referenced (see in the manuscript). AMS are available on request from [Mazzoglio et al., 2020](#).

Acknowledgements

This work was supported by Autorità di bacino distrettuale del fiume Po, under grant nos. L241CASTELLARIN12820, REP.128 PROT.3431. The research would not have been possible without access to the I²-RED dataset ([Mazzoglio et al., 2020](#)), given by Paola Mazzoglio and Pierluigi Claps. The authors also gratefully acknowledge the use of free and open-source software, in particular Python ([Van Rossum and Drake, 1995](#)), Scikit-learn ([Pedregosa et al., 2011](#)), TensorFlow ([Abadi et al., 2015](#)), QGIS (QGIS Development Team, 2021) and GRASS GIS ([GRASS Development Team, 2019](#)).

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jozefowicz, R., Jia, Y., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Schuster, M., Monga, R., Moore, S., Murray, D., Olah, C., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
- Alila, Y., 1999. A hierarchical approach for the regionalization of precipitation annual maxima in Canada. *J. Geophys. Res. Atmospheres* 104, 31645–31655. <https://doi.org/10.1029/1999JD900764>.
- Allamano, P., Claps, P., Laio, F., Thea, C., 2009. A data-based assessment of the dependence of short-duration precipitation on elevation. *Physics and Chemistry of the Earth* 34 (10–12), 635–641. <https://doi.org/10.1016/j.pce.2009.01.001>.
- Blöschl G., 2011. Scaling and Regionalization in Hydrology. In: *Treatise on Water Science*. Elsevier. 519–535. <https://doi.org/10.1016/B978-0-444-53199-5.00013-5>.
- Blum, A.G., Archfield, S.A., Vogel, R.M., 2017. The probability distribution of daily streamflow in the United States. *Hydrol. Earth Syst. Sci.* 21, 3093–3103. <https://doi.org/10.5194/hess-21-3093-2017>.
- Braca, G., Bussetti, M., Ducci, D., Lastoria, B., Mariani, S., 2019. Evaluation of national and regional groundwater resources under climate change scenarios using a GIS-based water budget procedure. *Rendiconti Lincei Sci. Fis. E Nat.* 30, 109–123. <https://doi.org/10.1007/s12210-018-00757-6>.
- Brath, A., Castellarin, A., Montanari, A., 2003. Assessing the reliability of regional depth-duration-frequency equations for gaged and ungaged sites. *Water Resour. Res.* 39 (12), 1367–1379.
- Breiman, L., 1996a. Stacked regressions. *Mach. Learn.* 24, 49–64. <https://doi.org/10.1007/BF00117832>.
- Breiman, L., 1996b. Bagging predictors. *Mach. Learn.* 24, 123–140. <https://doi.org/10.1007/BF00058655>.
- Burlando P., Rosso R., 1996. Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. *Journal of Hydrology*, volume 187, Issues 1–2, pages 45–64. ISSN 0022-1694. [https://doi.org/10.1016/S0022-1694\(96\)03086-7](https://doi.org/10.1016/S0022-1694(96)03086-7).
- Burn, D.H., 1990. An appraisal of the “region of influence” approach to flood frequency analysis. *Hydrol. Sci. J.* 35, 149–165. <https://doi.org/10.1080/026266900942415>.
- Caldas-Alvarez, A., Augenstein, M., Ayyzel, G., Barfus, K., Cherian, R., Dillenhardt, L., Fauer, F., Feldmann, H., Heistermann, M., Karwat, A., Kaspar, F., Kreibich, H., Lucio-Eceiza, E.E., Meredith, E.P., Mohr, S., Niermann, D., Pfahl, S., Ruff, F., Rust, H.W., Schoppa, L., Schwitalla, T., Steidl, S., Thielen, A.H., Tradowsky, J.S., Wulfmeyer, V., Quaas, J., 2022. Meteorological, impact and climate perspectives of the 29 June 2017 heavy precipitation event in the Berlin metropolitan area. *Nat. Hazards Earth Syst. Sci.* 22, 3701–3724. <https://doi.org/10.5194/nhess-22-3701-2022>.
- Camorani, G., Castellarin, A., Brath, A., 2005. Effects of land-use changes on the hydrologic response of reclamation systems. *Physics and Chemistry of the Earth* 30, 561–574. <https://doi.org/10.1016/j.pce.2005.07.010>.
- Castellarin, A., Burn, D.H., Brath, A., 2001. Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. *J. Hydrol.* 241, 270–285. [https://doi.org/10.1016/S0022-1694\(00\)00383-8](https://doi.org/10.1016/S0022-1694(00)00383-8).
- Castellarin, A., Merz, R., Blöschl, G., 2009. Probabilistic envelope curves for extreme rainfall events. *Journal of Hydrology* 378 (3–4), 263–271.
- Claps, P., Ganora, D., Mazzoglio, P., 2022. Rainfall regionalization techniques, in: *Rainfall*. Elsevier, pp. 327–350. <https://doi.org/10.1016/B978-0-12-822544-8.00013-5>.
- Dalrymple, T., 1960. *Flood frequency analysis, U.S. Geol. Surv. Water Supply Pap.* 1543-A, 11–51.
- Di Baldassarre, G., Castellarin, A., Brath, A., 2006. Relationships between statistics of rainfall extremes and mean annual precipitation: an application for design-storm estimation in northern central Italy. *Hydrol. Earth Syst. Sci.* 10, 589–601. <https://doi.org/10.5194/hess-10-589-2006>.
- Di Baldassarre, G., Laio, F., Montanari, A., 2009. Design flood estimation using model selection criteria. *Phys. Chem. Earth Parts ABC* 34, 606–611. <https://doi.org/10.1016/j.pce.2008.10.066>.
- Di Prinzio, M., Castellarin, A., Toth, E., 2011. Data-driven catchment classification: application to the pub problem. *Hydrology and Earth System Sciences* 15 (6), 1921–1935.
- Ghamariyani, M., Imteaz, M.A., 2021. Prediction of Seasonal Rainfall with One-year Lead Time Using Climate Indices: A Wavelet Neural Network Scheme. *Water Resour. Manag.* 35, 5347–5365. <https://doi.org/10.1007/s11269-021-03007-x>.
- GRASS Development Team, 2019. *Geographic Resources Analysis Support System (GRASS) Software*, Version 7.6, Open Source Geospatial Foundation, <https://grass.osgeo.org>.
- Grieser J., Staeger T., Schonwiese C.-D. Estimates and uncertainties of return periods of extreme daily precipitation in Germany. *Meteorol. Z.* 16, 553–564, <https://doi.org/10.1127/0941-2948/2007/0235>.
- Grimaldi S., Kao S.-C., Castellarin A., Papalexiou S.-M., Viglione A., Laio F., Aksoy H., Gedikli A., 2011. Statistical Hydrology. In: *Treatise on Water Science*. Elsevier. 479–517. <https://doi.org/10.1016/B978-0-444-53199-5.00046-4>.
- Han, J., Moraga, C., 1995. The influence of the sigmoid function parameters on the speed of backpropagation learning. In: Mira, J., Sandoval, F. (Eds.), *From Natural to Artificial Neural Computation. IWANN 1995. Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-59497-3_175.
- Hastie T., Tibshirani R., and Friedman J., 2009. *The Elements of Statistical Learning, Springer Series in Statistics*, Springer New York, New York, NY, .
- Hengl, T., 2007. *A Practical Guide to Geostatistical Mapping of Environmental Variables. Office for Official Publications of the European Communities, Luxembourg*.
- Hosking J.R.M., Wallis J.R., 1997. *Regional Frequency Analysis: An Approach Based on L-Moments*, 1st ed. Cambridge University Press. <https://doi.org/10.1017/CBO9780511529443>.
- Hosking, J.R.M., Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. *Water Resour. Res.* 29, 271–281. <https://doi.org/10.1029/92WR01980>.
- Hotelling, H., 1935. The most predictable criterion. *J. Educ. Psychol.* 26, 139–142. <https://doi.org/10.1037/h0058165>.
- Jain, A., Nandakumar, K., Ross, A., 2005. Score normalization in multimodal biometric systems. *Pattern Recognition* 38 (12). <https://doi.org/10.1016/j.patcog.2005.01.012>.
- Jenkinson, A.F., 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* 81 (348), 158–171. <https://doi.org/10.1002/qj.49708134804>.
- Jolliffe I.T., 2002. *Principal component analysis*, 2nd ed. ed, Springer series in statistics. Springer, New York.
- Kidd, C., Becker, A., Huffman, G.J., Muller, C.L., Joe, P., Skofronick-Jackson, G., Kirschbaum, D.B., 2017. So, How Much of the Earth’s Surface Is Covered by Rain Gauges? *Bull. Am. Meteorol. Soc.* 98, 69–78. <https://doi.org/10.1175/BAMS-D-14-00283.1>.
- Koutsoyiannis, D., 2004. Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation/Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: I. Recherche théorique. *Hydrological Sciences Journal* 49 (4), 590. <https://doi.org/10.1623/hysj.49.4.575.54430>.
- Koutsoyiannis, D., Baloutsos, G., 2000. Analysis of a Long Record of Annual Maximum Rainfall in Athens, Greece, and Design Rainfall Inferences. *Natural Hazards* 22, 29–48. <https://doi.org/10.1023/A:1008001312219>.
- Koutsoyiannis, D., Kozonis, D., Manetas, A., 1998. A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology* 206 (1–2), 118–135. [https://doi.org/10.1016/S0022-1694\(98\)00097-3](https://doi.org/10.1016/S0022-1694(98)00097-3).
- Koutsoyiannis D., 2007. A Critical Review of Probability of Extreme Rainfall: Principles and Models, in: Vassilopoulos, A., Ashley, R., Zevenbergen, C., Pasche, E., Garvin, S. (Eds.), *Advances in Urban Flood Management*. Taylor & Francis, pp. 139–166. <https://doi.org/10.1201/9780203945988.ch7>.

- Libertino, A., Allamano, P., Laio, F., Claps, P., 2018. Regional-scale analysis of extreme precipitation from short and fragmented records. *Adv. Water Resour.* 112, 147–159. <https://doi.org/10.1016/j.advwatres.2017.12.015>.
- Maity R., 2018. Statistical methods in hydrology and hydroclimatology, Springer Nature Singapore Pte Ltd., Singapore, <https://doi.org/10.1007/978-981-10-8779-0>.
- Marra, F., Armon, M., Borga, M., Morin, E., 2021. Orographic effect on extreme precipitation statistics peaks at hourly time scales. e2020GL091498 *Geophys. Res. Lett.* 48. <https://doi.org/10.1029/2020GL091498>.
- Mazzoglio, P., Butera, I., Claps, P., 2020. I2-RED: A Massive Update and Quality Control of the Italian Annual Extreme Rainfall Dataset. *Water* 12, 3308. <https://doi.org/10.3390/w12123308>.
- Mazzoglio, P., Butera, I., Alvioli, M., Claps, P., 2022. The role of morphology in the spatial distribution of short-duration rainfall extremes in Italy. *Hydrol. Earth Syst. Sci.* 26, 1659–1672. <https://doi.org/10.5194/hess-26-1659-2022>.
- Milligan, G.W., Cooper, M.C., 1988. A study of standardization of variables in cluster analysis. *Journal of Classification* 5, 181–204. <https://doi.org/10.1007/BF01897163>.
- Modarres, R., Sarhadi, A., 2011. Statistically-based regionalization of rainfall climates of Iran. *Glob. Planet. Change* 75, 67–75. <https://doi.org/10.1016/j.gloplacha.2010.10.009>.
- Msilini, A., Masselot, P., Ouarda, T.B.M.J., 2020. Regional Frequency Analysis at Ungauged Sites with Multivariate Adaptive Regression Splines. *J. Hydrometeorol.* 21, 2777–2792. <https://doi.org/10.1175/JHM-D-19-0213.1>.
- Msilini, A., Ouarda, T.B.M.J., Masselot, P., 2022. Evaluation of additional physiographical variables characterising drainage network systems in regional frequency analysis, a Quebec watersheds case-study. *Stoch. Environ. Res. Risk Assess.* 36, 331–351. <https://doi.org/10.1007/s00477-021-02109-7>.
- Ngongondo, C.S., Xu, C.-Y., Tallaksen, L.M., Aleman, B., Chirwa, T., 2011. Regional frequency analysis of rainfall extremes in Southern Malawi using the index rainfall and L-moments approaches. *Stoch. Environ. Res. Risk Assess.* 25, 939–955. <https://doi.org/10.1007/s00477-011-0480-x>.
- Ouali, D., Chebana, F., Ouarda, T.B.M.J., 2016a. Quantile Regression in Regional Frequency Analysis: A Better Exploitation of the Available Information. *J. Hydrometeorol.* 17, 1869–1883. <https://doi.org/10.1175/JHM-D-15-0187.1>.
- Ouali, D., Chebana, F., Ouarda, T.B.M.J., 2016b. Non-linear canonical correlation analysis in regional frequency analysis. *Stoch. Environ. Res. Risk Assess.* 30, 449–462. <https://doi.org/10.1007/s00477-015-1092-7>.
- Ouarda, T.B.M.J., Girard, C., Cavadias, G., Bobée, B., 2001. Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology.* 254 (1–4), 157–173.
- Ouarda, T.B.M.J., Shu, C., 2009. Regional low-flow frequency analysis using single and ensemble artificial neural networks: REGIONAL LOW-FLOW ANALYSIS. *Water Resour. Res.* 45 <https://doi.org/10.1029/2008WR007196>.
- Ouarda, T.B.M.J., Yousef, L.A., Charron, C., 2019. Non-stationary intensity-duration-frequency curves integrating information concerning teleconnections and climate change. *Int. J. Climatol.* 2019 (39), 2306–2323. <https://doi.org/10.1002/joc.5953>.
- Papalexiou, S.M., AghaKouchak, A., Foufoula-Georgiou, E., 2018. A diagnostic framework for understanding climatology of tails of hourly precipitation extremes in the United States. *Water Resources Research* 54, 6725–6738. <https://doi.org/10.1029/2018WR022732>.
- Papalexiou, S.M., Koutsoyiannis, D., 2013. Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resour. Res.* 49 <https://doi.org/10.1029/2012WR012557>.
- Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Müller A., Nothman J., Louppe G., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., and Duchesnay É., 2011. Scikit-learn: Machine Learning in Python, arXiv [preprint], J. Mach. Learn. Res. 12. arxiv: 1201.0490.
- Persiano, S., Ferri, E., Antolini, G., Domeneghetti, A., Pavan, V., Castellarin, 2020. A Changes in seasonality and magnitude of sub-daily rainfall extremes in Emilia-Romagna (Italy) and potential influence on regional rainfall frequency estimation. *Journal of Hydrology. Regional Studies* 32, 100751.
- Piper, D., Kunz, M., Ehmele, F., Mohr, S., Mühr, B., Kron, A., Daniell, J., 2016. Exceptional sequence of severe thunderstorms and related flash floods in May and June 2016 in Germany. Part I: Meteorological background. *Nat. Hazards Earth Syst. Sci.* 16, 2835–2850. <https://doi.org/10.5194/nhess-16-2835-2016>.
- Requena, A.I., Ouarda, T.B.M.J., Chebana, F., 2017. Flood frequency analysis at ungauged sites based on regionally estimated streamflows. *Journal of Hydrometeorology* 18 (9), 2521–2539. <https://doi.org/10.1175/jhm-d-16-0143.1>.
- Requena, A.I., Chebana, F., Ouarda, T.B.M.J., 2018. A functional framework for flow-duration-curve and daily streamflow estimation at ungauged sites. *Advances in Water Resources* 113, 328–340. <https://doi.org/10.1016/j.advwatres.2018.01.019>.
- Schaefer, M.G., 1990. Regional analyses of precipitation annual maxima in Washington State. *Water Resour. Res.* 26, 119–131. <https://doi.org/10.1029/WR026i001p0019>.
- Shehu, B., Willems, W., Stockel, H., Thiele, L.-B., Haberlandt, U., 2023. Regionalisation of rainfall depth–duration–frequency curves with different data types in Germany. *Hydrol. Earth Syst. Sci.* 27, 1109–1132. <https://doi.org/10.5194/hess-27-1109-2023>.
- Shu, C., Burn, D.H., 2004. Artificial neural network ensembles and their application in pooled flood frequency analysis: ARTIFICIAL NEURAL NETWORK ENSEMBLES. *Water Resour. Res.* 40 <https://doi.org/10.1029/2003WR002816>.
- Shu, C., Ouarda, T.B.M.J., 2007. Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space. *Water Resour. Res.* 43 <https://doi.org/10.1029/2006WR005142>.
- Soltani, S., Helfi, R., Almasi, P., Modarres, R., 2017. Regionalization of Rainfall Intensity–Duration–Frequency using a Simple Scaling Model. *Water Resour. Manag.* 31, 4253–4273. <https://doi.org/10.1007/s11269-017-1744-0>.
- Svensson, C., Jones, D.A., 2010. Review of rainfall frequency estimation methods: Review of rainfall frequency estimation methods. *J. Flood Risk Manag.* 3, 296–313. <https://doi.org/10.1111/j.1753-318X.2010.01079.x>.
- Van den Besselaar, E.J.M., Klein Tank, A.M.G., Buishand, T.A., 2013. Trends in European precipitation extremes over 1951–2010. *Int. J. Climatol.* 33, 2682–2689. <https://doi.org/10.1002/joc.3619>.
- Van Rossum G., and Drake Jr, F.L., 1995. Python reference manual, Centrum voor Wiskunde en Informatica Amsterdam. <https://ir.cwi.nl/pub/5008> (last access: 13 June 2023).
- Velázquez, J.A., Antil, F., Ramos, M.H., Perrin, C., 2011. Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures. *Adv. Geosci.* 29, 33–42. <https://doi.org/10.5194/adgeo-29-33-2011>.
- Xu, Y., Goodacre, R., 2018. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *J. Anal. Test.* 2, 249–262.
- Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J.C., Sampson, C.C., Kanae, S., Bates, P.D., 2017. A high-accuracy map of global terrain elevations: Accurate Global Terrain Elevation map. *Geophys. Res. Lett.* 44, 5844–5853. <https://doi.org/10.1002/2017GL072874>.