REVIEW

Analysis, characterization, prediction, and attribution of extreme atmospheric events with machine learning and deep learning techniques: a review

Sancho Salcedo-Sanz[1] · Jorge Pérez-Aracil[1] · Guido Ascenso[2] · Javier Del Ser[3] · David Casillas-Pérez[4] ·
Christopher Kadow[5] · Dušan Fister[1] · David Barriopedro[6] · Ricardo García-Herrera[6,7] · Matteo Giuliani[2] ·
Andrea Castelletti[2]

Abstract

Atmospheric extreme events cause severe damage to human societies and ecosystems. The frequency and intensity of extremes and other associated events are continuously increasing due to climate change and global warming. The accurate prediction, characterization, and attribution of atmospheric extreme events is, therefore, a key research field in which many groups are currently working by applying different methodologies and computational tools. Machine learning and deep learning methods have arisen in the last years as powerful techniques to tackle many of the problems related to atmospheric extreme events. This paper reviews machine learning and deep learning approaches applied to the analysis, characterization, prediction, and attribution of the most important atmospheric extremes. A summary of the most used machine learning and deep learning techniques in this area, and a comprehensive critical review of literature related to ML in EEs, are provided. The critical literature review has been extended to extreme events related to rainfall and floods, heatwaves and extreme temperatures, droughts, severe weather events and fog, and low-visibility episodes. A case study focused on the analysis of extreme atmospheric temperature prediction with ML and DL techniques is also presented in the paper. Conclusions, perspectives, and outlooks on the field are finally drawn.

1 Introduction

Atmospheric extreme events (EEs, either weather, or climate-related) gravely impact societies (Horton et al. 2016), causing hundreds of thousands of deaths every year (De et al. 2004; Pörtner et al. 2022), and producing important collateral effects, such as migrations (Marchiori et al. 2012; Carrico and Donato 2019), infrastructure damages (May and Koski 2013), transportation problems (Trinks et al. 2012; Stamos et al. 2015), and damages to agriculture (Ciais et al. 2005; van der Velde et al. 2012; Lal et al. 2012) or ecosystems (Seneviratne et al. 2012; Knapp et al. 2008; Van Oijen et al. 2013; Woodward et al. 2016).

As the number and intensity of EEs have been increasing in the last few decades (likely as a consequence of climate change processes (Mitchell et al. 2006; Herring et al. 2015; Grant 2017)), so has the number of scientific studies on

them. In this context, some classical problems associated with EEs are their analysis (Herring et al. 2015), detection (Zscheischler et al. 2013; Easterling et al. 2016), and causation/attribution to human activities (Stott et al. 2016; Hannart and Naveau 2018; Runge et al. 2019; Madakumbura et al. 2021). Also, different authors have focused their research on studying compound EEs (combinations of multiple EEs that contribute to societal or environmental risk) (Zscheischler and Seneviratne 2017; Zscheischler et al. 2020, 2018; Raymond et al. 2020), the relationship of EEs with different processes such as carbon cycle (Reichstein et al. 2013; Frank et al. 2015; van der Molen et al. 2011) or soil moisture (Hirschi et al. 2011; Whan et al. 2015), and the effects of EEs on economics (Chavez et al. 2015; Ackerman 2017) and their impact on human systems (Zscheischler et al. 2014), to name just a few.

Different mathematical and computational methods have been used to analyze and forecast EEs, including numerical weather methods (NWM) (Lavers and Villarini 2013; Yucel et al. 2015; Vitart and Robertson 2018), statistical, and probability-based methods (Ferro 2007; Naveau et al.

✉ Jorge Pérez-Aracil
  jorge.perezaracil@uah.es

Extended author information available on the last page of the article

2020; Sapsis 2021), non-linear physics and chaos theory (Ghil et al. 2011; Farazmand and Sapsis 2019; Chowdhury et al. 2022), and, in the last decade, an important number of machine learning (ML) and related techniques, a field with an exponential presence in climate and atmospheric sciences (Monteleoni et al. 2013; Cohen et al. 2019), climate change studies (Rolnick et al. 2019), and Earth system science in general (Reichstein et al. 2019; Karpatne et al. 2018; Camps-Valls et al. 2019; Salcedo-Sanz et al. 2020; Bonavita et al. 2021; Irrgang et al. 2021). In the last years, deep learning (DL) algorithms, a particularly promising branch of ML, have also been applied to climate science problems (Kurth et al. 2018; Ardabili et al. 2019), where they have shown great potential to deal with different EE-related problems (Liu et al. 2016; Ren et al. 2020; Qi and Majda 2020; Fang et al. 2021).

In this paper, we discuss the most important ML methods applied to atmospheric EE-related problems, including DL approaches. It is possible to classify atmospheric EEs in terms of their physical characteristics and impact on human society and ecosystems. In addition, different ML techniques have been associated with specific problems in EEs, for example, feature selection/extraction problems in ML have been usually associated with the detection of EEs, in such a way that the ML algorithms are able to select the most important feature which triggers an EE. If we include specific drivers to train the ML problems, we can deal with the attribution of atmospheric EEs. The attribution of EEs with ML involves the application of the algorithms in data (measurements, reanalysis or simulations) from different periods and/or forcings. Finally, ML approaches have been also used to deal with prediction problems related to EEs. This is maybe the most common application of ML in EEs, and it is possible to find a prediction of problems related to EEs in different prediction time horizons, from very short-term to seasonal prediction. Having these ideas in mind, we have chosen a number of atmospheric EEs in terms of their impacts on human societies and ecosystems, to carry out the review of ML methods applied to describe them. In this case, we have chosen extreme precipitation and floods, extreme temperatures and heatwaves, droughts, severe weather and low-visibility events. We provide a comprehensive review of the works applying ML and DL algorithms for these EEs problems, and we finally discuss a case study on ML and DL techniques focused on heatwaves prediction, some final perspectives on this research area in the near future.

The rest of the paper is structured as follows: the next section will give a theoretical overview of some of the ML algorithms most commonly used for studying EEs. Section 3 presents a comprehensive analysis of existing literature on ML and DL techniques for atmospheric EEs problems. Section 4 presents a case study on heatwaves prediction with ML and DL techniques, while Sects. 5.2 and 5 provide con-

clusions, final perspectives and a general outlook on future research.

2 Machine learning methods

This section summarizes the most important ML, DL methods, and related techniques commonly used in the analysis and prediction of EEs.

2.1 Feature selection methods and dimensionality reduction in ML and DL

For ML-based methods, using irrelevant or redundant features as inputs during training can be detrimental, not only because these additional features would increase the training time, but also because they may hinder their generalisability (Blum and Langley 1997). In its more general form, the feature selection problem (FSP) in ML problems can be defined as follows: given a set of labelled data samples $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l)\}$, where $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$ (or $y_i \in \{\pm 1\}$ for classification tasks), obtain subset of m features ($m < n$), that produces the lowest prediction (or classification) error in the estimation of the variable y_i.

There are many different approaches to dealing with FSP problems (Zebari et al. 2020). In general, FS algorithms can be classified into three families:

- The *wrapper approach* (John et al. 1994). Wrapper methods use the ML classifier/regressor in order to obtain the best set of features which minimizes an error measure. Figure 1a shows an outline of the wrapper approach. The interested reader can consult classical works on wrapper FSP approaches (Kohavi and John 1997; Yang and Honavar 1998).
- The *filter approach* to the FSP is based on a completely different idea. In this case, the selection of the best features is based on an external measure calculated from the data, and the classifier/regression algorithm is not taken into account. Figure 1b shows an example of a filter approach for an FSP problem. Note that filter methods are usually faster than wrapper methods, but in general, wrappers obtain better results, since they take into account the real performance of the classification/regression algorithm during the search. The interested reader can extend the analysis of filter methods in Torkkola and Campbell (2000); Torkkola (2002).
- Finally, mixed or *hybrid approach*. They are methods which combine wrappers and filter approaches into a single hybrid methodology. They have obtained good results in different specific applications (Ferreira and Figueiredo 2014; Huda et al. 2014; Solorio-Fernández et al. 2016).

Fig. 1 **a** Outline of a wrapper method; **b** outline of a filter method

(a)   (b)
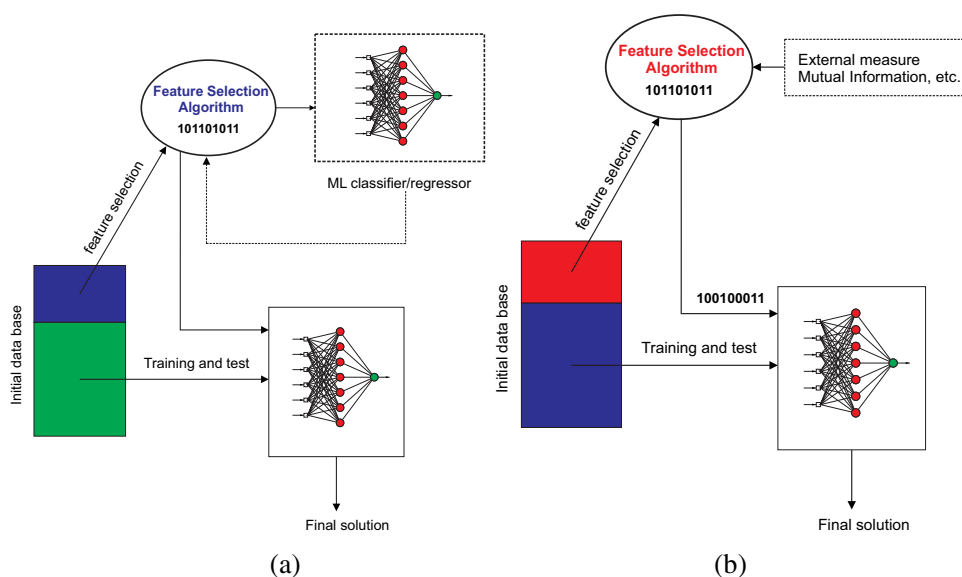
Note that both wrapper and filter methods admit a binary representation for the FSP, where a 1 in the i_{th} position of the binary vector stands for the feature i is considered within the subset of features, and a 0 means it is not. Using this notation there are 2^n different subsets of features to be evaluated (where n is the total number of features), and the problem consists of selecting the best one in terms of a given error measure, either internal (wrapper methods) or external (filter methods) to the classifier/regressor considered. Alternative encodings with integer numbers are also possible. Given the large search space generated by the encoding of the FSP, meta-heuristic approaches are commonly applied to obtain the best set of features, mainly in the wrapper approach (Salcedo-Sanz et al. 2018).

2.1.1 Other dimensionality reduction methods in ML and DL

In addition to classical feature selection methods shown above, there are different traditional dimensionality reduction methods (Ghodsi 2006; Van Der Maaten et al. 2009; Huang et al. 2019; Ghojogh et al. 2023) thought to improve the performance of ML and DL techniques. We review here some of the methods which have been used the most to improve ML and DL techniques in EEs detection, prediction and attribution problems. For instance, the well-known principal component analysis (PCA) (Abdi and Williams 2010), aims to find a linear subspace of low dimension that maintain most of the variability in the data. Also Linear Discriminant Analysis (LDA) (Balakrishnama and Ganapathiraju 1998), is based on the idea of finding a linear combination of features that characterizes or separates two or more classes of objects or events. Another example of a traditional dimensionality reduction technique is locally linear embedding (LLE) (Roweis and Saul 2000), a nonlinear approach

to reduce dimensionality by computing low-dimensional, neighbourhood-preserving embedding of high-dimensional data.

The autoencoder (AE) neural network can also be used for reducing the dimensionality of the data (Pinaya et al. 2020). They aim to reproduce the input in the output (Goodfellow et al. 2016). It is composed of two different parts: the encoder and the decoder. The intermediate representation is called latent space. It can be understood as a meaningful representation of the data. The data is decoded to reconstruct as similar as possible the input data, Fig. 2. A probabilistic framework was introduced with variational AE (VAE) Kingma and Welling (2013). One of the main differences between AEs and VAEs is the latent space representation (Fig. 3). The AE learns a continuous latent space representation for the input data. Thus, a unique encoding of the input is found for each point in the latent space. In the latent space of the VAE, the points follow a probability function. Thus, for each point of the latent space, a sample from the distribution is found. Another difference is related to the loss function. While the AE minimizes a reconstruction loss between the input and the output, the VAE aims to optimize two different terms. The first one refers to the reconstruction loss, whilst the second one is based on the Kullback–Leibler divergence
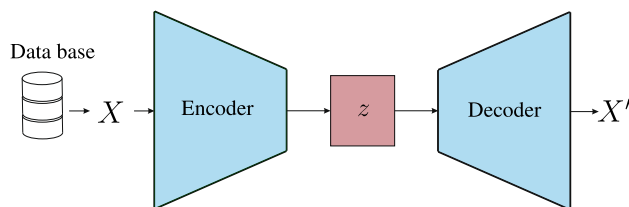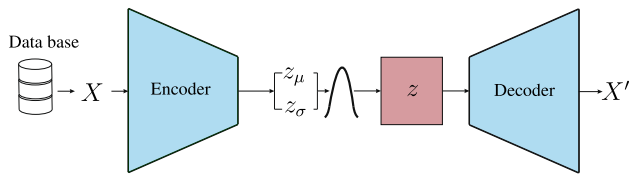

Fig. 2 AE structure

Fig. 3 VAE structure

loss. It aims at the latent space to follow the desired probability distribution. In some applications it is important to note these significant differences between AE and VAE.

2.2 Multi-layer perceptrons

A multi-layer perceptron (MLP) is a particular class of artificial neural network (ANN), which has been successfully applied to solve a large variety of non-linear problems, mainly classification and regression tasks (Haykin and Network 2004; Bishop 1995). The multi-layer perceptron consists of an input layer, a number of hidden layers, and an output layer, all of which consist of a number of special processing units called *neurons*. All the neurons in the network are connected to other neurons by means of weighted links (see Fig. 4). In a *feedforward* MLP, the neurons within a given layer are connected to those of the previous layer. The values of these weights are related to the ability of the MLP to learn the problem, and they are learned from a sufficiently long number of examples. The process of assigning values to these weights from labelled examples is known as the training process of the perceptron. The adequate values of the weights minimize the error between the output given by the MLP and the corresponding expected output in the training set. The number of neurons in the hidden layer is also a hyperparameter to be optimized (Haykin and Network 2004; Bishop 1995).
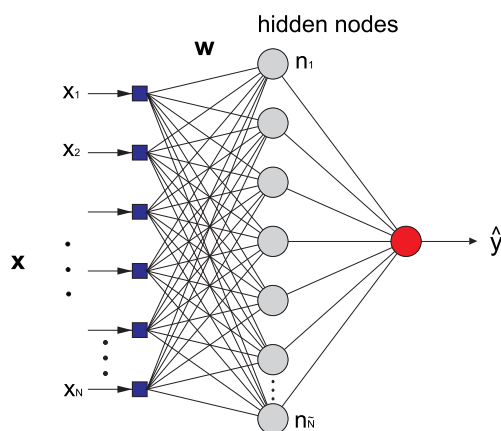

Fig. 4 Structure of a multi-layer perceptron neural network, with one hidden layer

The input data for the MLP consists of a number of samples arranged as input vectors $\{\mathbf{x}^i \in \mathbb{R}^n\}_{i=1}^N$, with each input vector $\mathbf{x}^i = (x_1^i, \cdots, x_n^i)$. Once an MLP has been properly trained, it can be tested on data it did not see during training to evaluate its performance, in terms of how well the learned weights can transform the given input into a desired output $\vartheta \in \mathbb{R}$. The relationship between the output ϑ and a generic input signal $\mathbf{x} = (x_1, \cdots, x_n)$ of a neuron is given by:

$$\vartheta(\mathbf{x}) = \varphi\left(\sum_{j=1}^n w_j x_j - b\right), \tag{1}$$

where ϑ is the output signal, x_j for $j = 1, \ldots, n$ are the input signals, w_j is the weight associated with the j-th input, b is the bias term (Haykin and Network 2004; Bishop 1995), and φ is some function chosen based on the type of layer to which it needs to be applied, for example the logistic function (among other possibilities):

$$\varphi(x) = \frac{1}{1 + e^{-x}}. \tag{2}$$

The well-known stochastic gradient descent (SGD) algorithm is often applied to train MLPs (Rumelhart et al. 1986). There are also alternative training algorithms for MLP which have shown excellent performance in different problems, such as the Levenberg-Marquardt algorithm (Hagan and Menhaj 1994), or the ADAM and RMSProp optimizers for training deep versions of the networks (Zhang 2018; Zou et al. 2019).

2.2.1 Extreme learning machines

An extreme learning machine (ELM) (Huang et al. 2006) is a type of training method for multi-layer perceptrons, characterized by being computationally faster than traditional gradient backpropagation (Hecht-Nielsen 1992). In the ELM algorithm, the weights between the inputs and the hidden nodes are set at random, usually by using a uniform probability distribution. Then, the output matrix of the hidden layer is established and the Moore-Penrose pseudo-inverse of this matrix is computed. The optimal values of the weights belonging to the output layer are directly obtained by multiplying the computed pseudo-inverse matrix with the target (see Huang et al. (2011) for details). The ELM obtains competitive results with respect to other classical training methods, while its training computation efficiency overcomes other classifiers or regression approaches such as SVM algorithms or MLPs (Huang et al. 2011).

Mathematically, the ELM algorithm considers a training set $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ to fit the weights (β_k) associated with each hidden node $\tilde{N}$ to optimally compose an output with mini-

mum mean squared error. The training process is according to the following steps:

1. The input weights $\mathbf{w}_k$ and the bias b_k, where $k = 1, \ldots, \tilde{N}$ are randomly chosen following a uniform distribution with support $[-1, 1]$.

2. In the second step, the hidden-layer output matrix H is computed as follows:

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_{\tilde{N}} \mathbf{x}_1 + b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ g(\mathbf{w}_1 \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_{\tilde{N}} \mathbf{x}_N + b_{\tilde{N}}) \end{bmatrix}_{\tilde{N}} \quad (3)$$

where $g(\cdot)$ is the activation function.

3. The training problem is reduced to a $\boldsymbol{\beta}$ parameter optimization problem, which can be defined as:

$$\min_{\boldsymbol{\beta}} \|\mathbf{H}\boldsymbol{\beta} - \mathbf{Y}\|, \quad (4)$$

4. The last step consists in obtaining the output layer weights $\boldsymbol{\beta}$ by means of the following expression:

$$\boldsymbol{\beta} = \mathbf{H}^{\dagger} \mathbf{Y}^T, \quad (5)$$

where $\mathbf{Y}^T$ stands for the transpose of the training output vector $\mathbf{Y} = [y_1, \ldots, y_n]$ and $\mathbf{H}^{\dagger}$ refers to the Moore-Penrose pseudo-inverse of the hidden-layer matrix $\mathbf{H}$ (Huang et al. 2006).

5. Then, the predicted or classified output is obtained as: $\hat{Y}(\mathbf{x}) = \mathbf{H}\boldsymbol{\beta}$.

The hidden nodes number $\tilde{N}$ can be tuned for improving the ELM performance.

2.3 Support vector machines

A *support vector machine* (SVM) (Schölkopf et al. 2002, 2000) is a statistical learning algorithm for classification problems defined as follows: given a labelled training data set $\{\mathbf{x}_i, y_i\}_{i=1}^n$, where $\mathbf{x}_i \in \mathbb{R}^N$ and $y_i \in \{-1, +1\}$, and given a nonlinear mapping $\boldsymbol{\phi}(\cdot)$, the SVM method solves the following problem:

$$\min_{\mathbf{w}, \xi_i, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \right\} \quad (6)$$

constrained to:

$$y_i \left(\mathbf{w}^\top \boldsymbol{\phi}(\mathbf{x}_i) \right) + b \geq 1 - \xi_i, \quad 1 \leq i \leq n \\ \xi_i \geq 0, \quad\quad\quad\quad\quad 1 \leq i \leq n \quad (7)$$

where $\mathbf{w}$ and b define a linear classifier in feature space, and ξ_i are positive slack variables enabling to deal with permitted errors (Fig. 5). Appropriate choice of nonlinear mapping $\boldsymbol{\phi}$ guarantees that the transformed samples are more likely to be linearly separable in the (higher dimensional) feature space. The regularization hyperparameter C controls the generalization capability of the classifier, and it must be selected by the user. The core problem (6) is solved using its dual problem counterpart (Schölkopf et al. 2002), and the decision function for any test vector $\mathbf{x}_*$ is finally given by

$$f(\mathbf{x}_*) = \text{sgn} \left(\sum_{i=1}^n y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}_*) + b \right) \quad (8)$$

where α_i are Lagrange multipliers corresponding to constraints in (7), being the support vectors (SVs) those training samples $\mathbf{x}_i$ with non-zero Lagrange multipliers $\alpha_i \neq 0$; $K(\mathbf{x}_i, \mathbf{x}_*)$ is an element of a kernel matrix $\mathbf{K}$ (Schölkopf et al. 2002); and the bias term b is calculated by using the *unbounded* Lagrange multipliers as $b = 1/k \sum_{i=1}^k (y_i - \langle \boldsymbol{\phi}(\mathbf{x}_i), \mathbf{w} \rangle)$, where k is the number of *unbounded* Lagrange multipliers ($0 \leq \alpha_i < C$) and $\mathbf{w} = \sum_{i=1}^n y_i \alpha_i \boldsymbol{\phi}(\mathbf{x}_i)$ (Schölkopf et al. 2002).

2.3.1 Support vector regression

Support vector regression (SVR) (Smola and Schölkopf 2004) is a well-established algorithm for regression and function approximation problems. SVR takes into account an error approximation to the data, as well as the capability to improve the prediction of the model when a new dataset is evaluated. Although there are several versions of the SVR algorithm, we show the classical model (ϵ-SVR) described in
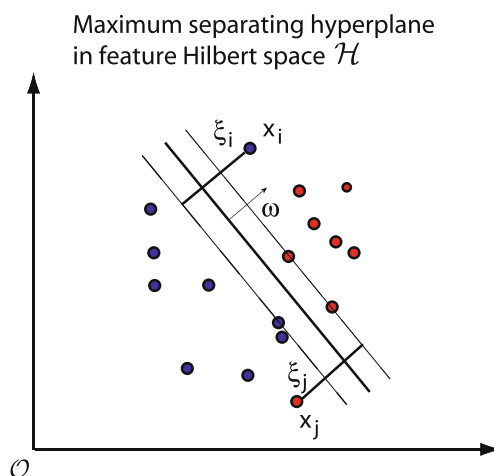

Fig. 5 Illustration of the SVM process: linear decision hyperplanes in a nonlinearly transformed, feature space, where *slack* variables ξ_i are included to deal with errors

detail in Smola and Schölkopf (2004), which has been used for a large number of problems and applications in science and engineering (Salcedo-Sanz et al. 2014).

The ϵ-SVR method for regression starts from a given set of training vectors $\{(\mathbf{x}_i, \vartheta_i)\}_{i=1}^{N}$, where $\mathbf{x}_i \in \mathbb{R}^N$ and $\vartheta_i \in \mathbb{R}$, and model the input–output relation as the following general model:

$$\hat{\vartheta}(\mathbf{x}) = g(\mathbf{x}) + b = \mathbf{w}^T \phi(\mathbf{x}) + b, \tag{9}$$

where $\mathbf{x}_i$ represents the input vector of predictive variables, ϑ_i stands for the value of the objective variable ϑ corresponding to the input vector $\mathbf{x_i}$ and $\hat{\vartheta}(\mathbf{x})$ represents the model which estimates $\vartheta(\mathbf{x})$. The parameters $(\mathbf{w}, b)$ are determined in order to match the training pair set, where the bias parameter b appears separated here. The function $\phi(\mathbf{x})$ projects the input space onto the feature space. During the training, the algorithms seek those parameters of the model which minimize the following risk function:

$$R[\hat{\vartheta}] = \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^{N} L\left(\vartheta_i, \hat{\vartheta}(\mathbf{x}_i)\right), \tag{10}$$

where the norm of $\mathbf{w}$ controls the smoothness of the model and $L\left(\vartheta_i, \hat{\vartheta}(\mathbf{x}_i)\right)$ stands for the selected loss function. We use the L^1-norm modified for the SVR and characterized by the ϵ-insensitive loss function (Smola and Schölkopf 2004):

$$dL\left(\vartheta_i, g(\mathbf{x}_i)\right) = \begin{cases} 0 & \text{if } |\vartheta_i - g(\mathbf{x}_i)| \le \epsilon \\ |\vartheta_i - g(\mathbf{x}_i)| - \epsilon & \text{otherwise.} \end{cases} \tag{11}$$
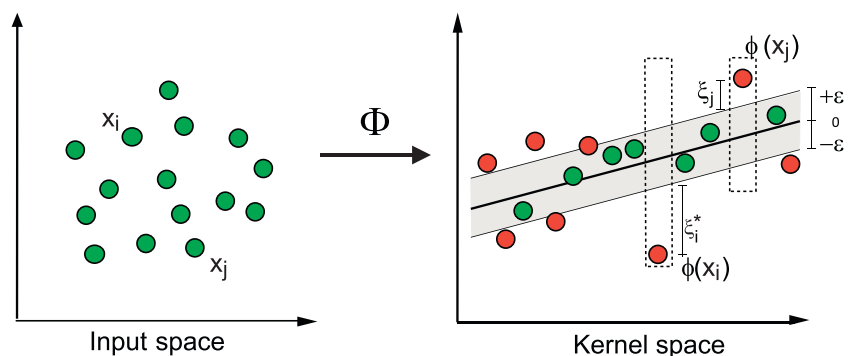
Figure 6 shows an example of the process of a SVR for a two-dimensional regression problem, with an ϵ-insensitive loss function.

To train this model, it is necessary to solve the following optimization problem (Smola and Schölkopf 2004):

$$\min_{\mathbf{w},b,\boldsymbol{\xi}} \quad \frac{1}{2}\|\mathbf{w}\| + C \sum_{i=1}^{N} \xi_i + \xi_i^*,$$

$$\begin{aligned} \text{s.t.} \quad & \vartheta^i - \mathbf{w}^T \phi(\mathbf{x}_i) - b \le \epsilon + \xi_i, & 1 \le i \le N, \\ & -\vartheta^i + \mathbf{w}^T \phi(\mathbf{x}_i) + b \le \epsilon + \xi_i^*, & 1 \le i \le N, \\ & \xi_i, \xi_i^* \ge 0, & 1 \le i \le N. \end{aligned} \tag{12}$$

The dual form of this optimization problem is obtained through the minimization of a Lagrange function, which is constructed from the objective function and the problem constraints:

$$\max_{\boldsymbol{\alpha},\boldsymbol{\alpha}^*} \quad -\frac{1}{2}\sum_{i,j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\mathbf{x}_i, \mathbf{x}_j)$$

$$-\epsilon \sum_{i=1}^{N} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} \vartheta^i(\alpha_i - \alpha_i^*)$$

$$\tag{13}$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0, & 1 \le i \le N, \\ & \alpha_i, \alpha_i^* \ge 0, & 1 \le i \le N, \\ & -\alpha_i, -\alpha_i^* \ge -C, & 1 \le i \le N. \end{aligned}$$

In the dual formulation of the problem, the function $K(\mathbf{x}_i, \mathbf{x}_j)$ represents the inner product $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ in the

Fig. 6 Example of a support-vector-regression process for a two-dimensional-regression problem, with an ϵ-insensitive loss function

feature space. Any function $K(\mathbf{x}_i, \mathbf{x}_j)$ may become a kernel function as long as it satisfies the constraints of the inner products. It is very common to use the Gaussian radial basis function:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \left\| \mathbf{x}_i - \mathbf{x}_j \right\|^2). \tag{14}$$

The final form of the function $g(\mathbf{x})$ depends on the Lagrange multipliers α_i, α_i^* as:

$$g(\mathbf{x}) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}). \tag{15}$$

Incorporating the bias, the estimation of the objective function is finally made by the following expression:

$$\hat{\vartheta}(\mathbf{x}) = g(\mathbf{x}) + b = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b. \tag{16}$$

2.4 Ensemble methods

Ensemble methods overcome the (potential) limitations in the predictive performance of a single learning model by relying on the randomized combination of several of them (Zhou 2012). This paradigm assumes that combinations of several, simple ML models can greatly outperform the performance of a single such model (González et al. 2020), and rival the robustness or generalization capacity of complex ML, such as artificial neural networks, which involve a huge number of parameters.

2.4.1 Bagging

The basic idea behind bagging (*bootstrap aggregating*) is to train a set of simple models and combine their individual

predictions as shown in Fig. 7. Bagging reduces the variance of the ML performance techniques and helps avoid overfitting, which is usually more severe in complex ML methods. Bagging establishes that all the base ML models which compose the ensemble have the same architecture, which results in the same topology, number of input–output variables and number of parameters to train. As an example, a set of decision trees trained with the bagging technique assumes that all trees have the same branches, with the same number of parameters and the same input–output variables (see Fig. 7). The individual models of the ensemble differ in the values that are learned for the model parameters, which are trained with different training sets.

The mathematical description of the bagging technique is as follows: Let $\mathcal{D} = \{lbrace(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ be a given training set of n input–output pairs. The procedure of bagging, shown in Fig. 7, generates N new training sets, of size n', composed of samples from the set $\mathcal{D}$, which can be repeated in each $\mathcal{D}_i$. This sampling used for the creation of the sets is known as a *bootstrap sample*. Then, the parameters of N equal models $\{\mathcal{M}_i\}_{i=1}^{N}$ are learned by training each model $\mathcal{M}_i$ on the respective subset $\mathcal{D}_i$. Finally, the ensemble model combines the individual outputs of each model by averaging their outputs (in the case of regression problems) or by majority voting (if dealing with classification problems) (Mohandes et al. 2018).

Bagging models can be deemed as the simplest way to create ensembles. Note that each base model $\mathcal{M}_i$ is trained independently with no influence between each other. This property allows to train each base model in parallel, which drastically reduces the training time of the ensemble.

Random forests (RF) (Breiman 2001) are among the most commonly applied bagging techniques for classification and regression problems. They specifically use decision or regression trees as learners and differ from pure bagging techniques in that the topology of the trees is not universally

Fig. 7 Diagram of the bagging technique used for classification or regression problems in ML
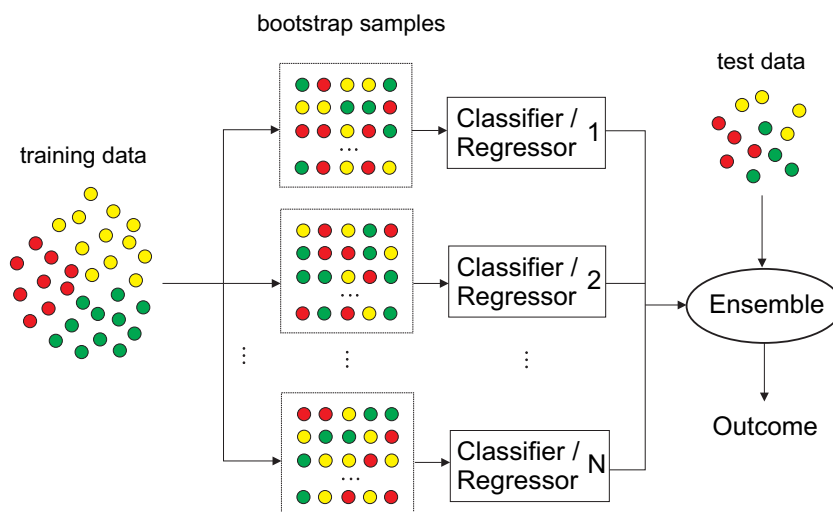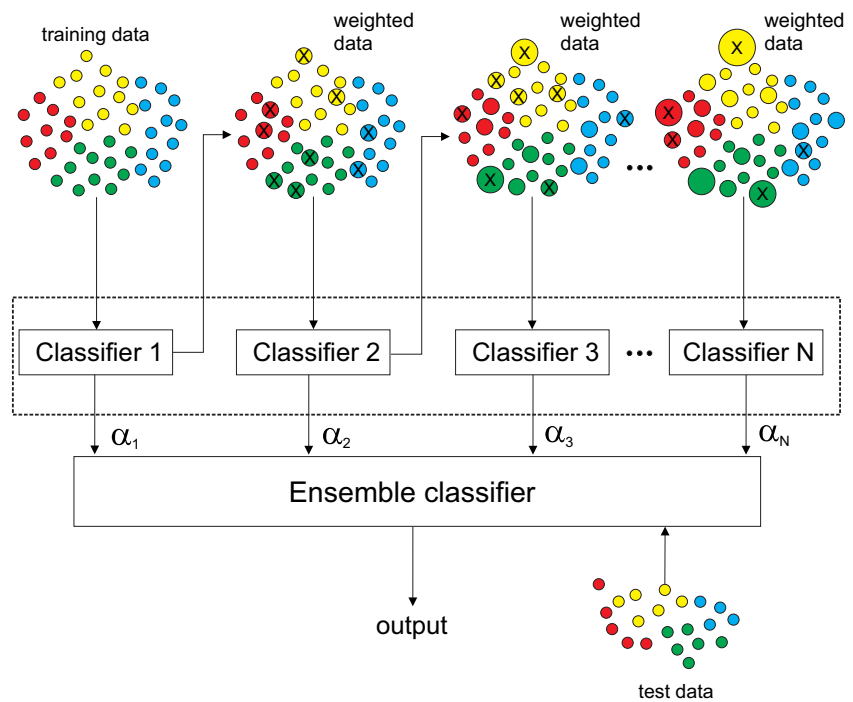
fixed. Trees of the ensemble (the forest) may have different lengths, and topology, or use different input variables, which greatly increases the variability of the learners, but differs from the bagging paradigm from a theoretical viewpoint. The main advantage of RFs over traditional bagging is that by adopting slightly different models in their ensemble, the limitations of each are averaged out, resulting in improved generalization capacity (Breiman 2001).

The RF training procedure consists of the following steps. Let $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ be the training dataset. The main hyper-parameters to be adjusted are: N, which is the number of estimators (namely, the number of tree learners composing the forest); and $maxDepth$, which is the maximum number of features to be explored as a node splitting criterion, which is often set to the square root of the number of features. Once these parameters are set, the method works as follows:

1. Initialize each one of the N decision or regression trees for the classification or regression problem respectively.
2. For each tree $\mathbf{T}_t$, select n_t samples with replacement, by using the bootstrapping technique.
3. Only a subset of maximum $maxDepth$ features shall be considered for the construction of each tree.
4. Each tree $\mathbf{T}_t$ will give a solution.
5. The ensemble output of the random forest method will be computed by majority voting in the case of classification:

$$\hat{Y}(\mathbf{x}) = \arg \max_l \sum_{t=1}^{N} [\mathbf{T}_t(\mathbf{x}) = l]. \tag{17}$$

or averaging for regression problems:

$$\hat{Y}(\mathbf{x}) = \frac{1}{N} \sum_{t=1}^{N} \alpha_t \mathbf{T}_t(\mathbf{x}). \tag{18}$$

2.4.2 Boosting

Boosting approaches are an alternative family of ensemble algorithms which perform well in both classification and regression problems (Ferreira and Figueiredo 2012). Similarly to bagging, boosting follows the learning paradigm of using simple (or "weak") ML models (classifiers/regressors), named learners, to form a powerful final model that combines their outputs. Also similarly to bagging, boosting establishes the same topology for all the learners involved in the ensemble (same architecture, number of input–output variables, and number of parameters to train). The most evident difference from bagging lies in the procedure for training weak learners. In bagging, the weak learners are trained in parallel using different subsets of data $\mathcal{D}_i$ randomly sampled from the whole training dataset $\mathcal{D}$. In boosting, the learners are trained sequentially (see Fig. 8). In this way, subsequent learners are dependent on previously trained ones, contrary to the learners in bagging methods. Furthermore, in boosting all the learners use the whole set of training data for computing their parameters, i.e, there is no bootstrap sample step.

Another important difference is that in bagging all input–output pairs are equally weighted to train each learner; each learner equally contributes to determine the final output of the ensemble model. In boosting, training input–output pairs are

Fig. 8 Diagram of the AdaBoost algorithm exemplified for multi-class classification problems. Different size circles stand for samples with more associated weight (w) due to mis-classification in the previous step (marked with X)

weighting according to the accuracy for being predicted by the previous learner (except for the first learner in the queue, which uses equally weighted samples). Consequently, learners are more specialized as soon as they are placed into the final locations along the queue. Furthermore, the contribution of each learner to the output of the ensemble is usually weighted according to its accuracy, which does not happen in bagging. This is the general scheme for all boosting methods, but there do exist different boosting strategies depending on the kind of weighting policy applied to each training sample, and/or the output of each learner.

A widely used boosting technique is Adaptive Boosting (AdaBoost). AdaBoost trains each weak learner in such a way that each learner focuses on the data that was misclassified by its predecessor so that learners further down the queue iteratively learn to adapt their parameters and achieve better results (Ferreira and Figueiredo 2012; González et al. 2020). Multiple variants of the AdaBoost algorithm exist, starting from the original one (Freund and Schapire 1997) designed to tackle binary classification problems, regression, or multiclass classification options. Figure 8 shows an outline of the AdaBoost algorithm for multi-class classification. The pseudocode for AdaBoost can be described as follows:

1. Let $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ be the training dataset. The first step is to initialise each base learner $\{\mathbf{T}_t \mid 1 \leq t \leq N\}$, and assign the set of sample weights $\{w_i \mid 1 \leq i \leq n\}$ corresponding to the input–output pairs $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ according to the uniform distribution: $w_i = \frac{1}{n}$.
2. For each base learner $\mathbf{T}_t$, the training dataset is used with the distribution of weights w_i for training.
3. After this training process, for each base learner $\mathbf{T}_t$, the estimation error ϵ_t is computed as:

$$\epsilon_t = \sum_{\mathbf{T}_t(\mathbf{x}_i) \neq y_i} \frac{w_i}{\sum_{\mathbf{x}_i} w_i}, \quad 1 \leq i \leq n \tag{19}$$

4. From this error is derived the weight of the current base learner for the ensemble output α_t:

$$\alpha_t = \log \frac{1 - \epsilon_t}{\epsilon_t} \tag{20}$$

5. Finally, the distribution of the weights w_i corresponding to each $\mathbf{x}_i$, which will be used in the next learner, is proportionally adjusted to the probability that a sample is correctly estimated, and inversely proportional to the error of the learner ϵ_t.
6. The final output, provided by the algorithm globally, will be:

$$\hat{Y}(\mathbf{x}) = \arg \max_l \sum_{t=1}^N [\alpha_t(\mathbf{T}_t(\mathbf{x}) = l)]. \tag{21}$$

This final function refers to the boosting method for classification problems, which simply integrates the weighted output of individual learners by voting. In regression problems, the output consists of computing a weighted average of the outputs:

$$\hat{Y}(\mathbf{x}) = \frac{1}{N} \sum_{t=1}^N \alpha_t \mathbf{T}_t(\mathbf{x}). \tag{22}$$

The main difference of this algorithm with the multi-class variant AdaBoost.M1 (Freund and Schapire 1997) is that only the weight values of the correctly classified samples are lowered ($w_i = w_i \frac{\epsilon_t}{1-\epsilon_t}$).

2.5 Deep learning algorithms

When used for predictive modelling, machine learning revolves around modelling the statistical correlation between variables with respect to the target variable to be predicted. In problems dealing with spatial and/or temporal data (such as image classification or time series forecasting), such a correlation emerges from the relationship among data points over such domains. As a result, machine learning models can be either used in their seminal form to tackle spatiotemporal modelling tasks (by, e.g., extracting tabular features from data) or, instead, specialised into archetypes capable of supporting the modelling requirements stemming from such tasks (invariance to spatial transformations of the input or the characterization of long-term correlations over sequential data). Furthermore, continued advances in massively parallel computing and the explosion of non-relational databases containing information of assorted nature (e.g., image, video, audio, text) have spurred research efforts towards the derivation of neural network models of ever-growing modelling complexity, capable of efficiently discovering relevant predictors from highly dimensional data and endowing mechanisms to meet the requirements mentioned previously. Advances over the past 2 decades have blossomed into what is now known as *deep learning* (LeCun et al. 2015), which crystallizes in two main neural architectures: convolutional neural networks (CNNs (O'Shea and Nash 2015)) and recurrent neural networks (RNNs (Sherstinsky 2020)). Figure 9 illustrates two typical applications of these deep learning architectures in the context of EEs.

When the correlation is held in the spatial domain, any model should be made invariant with respect to transformations of the input data that should not affect the prediction. This is the case of translational invariance in image classification, by which visual features relevant to the target to be predicted should retain their predictive importance no matter where they are located in the image. The way the human visual cortex operates to satisfy this requisite was the

Fig. 9 Examples of typical use cases related to extreme atmospheric events that can be tackled by deep learning models: **a** CNNs; **b** RNNs

(a) CNN

(b) RNN
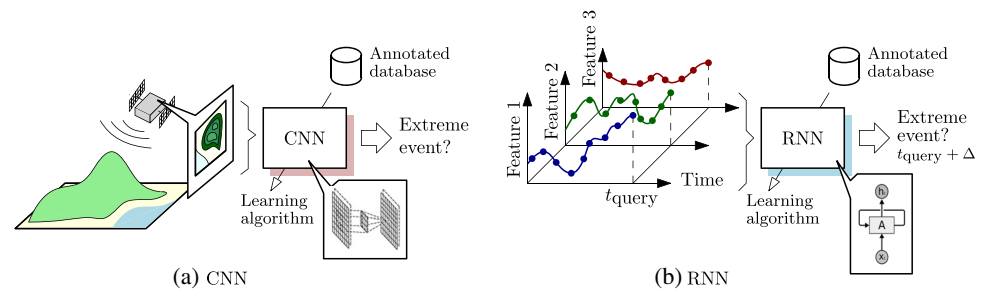
inspiration behind the design of CNNs, which, in their seminal form, comprise a series of hierarchically arranged neural processing layers. Layers closer to the input contain several convolutional neurons (also referred to as convolutional filters or kernels), which extract features from the input data by performing a convolution between the data themselves and the weights at their core. A CNNs for complex modelling tasks may stack several convolutional layers, one after another, so that each layer processes through its filters the output produced by the preceding layer. Some further processing layers can be placed in between convolutional ones, such as pooling layers, which serve to create information bottlenecks that help distil more high-level information while drastically reducing the number of parameters. After the convolutional part of the network, additional layers may be added depending on the application. For instance, in image classification a fully connected multi-layer perceptron is often attached to the end of a CNN to map this output to the target variables to be predicted. Analogously to MLPs, trainable parameters (weights and biases) of the CNN network can be learned by backpropagating error gradients through the network, which also holds for the weights of the convolutional kernels. Since gradients can be computed also for these special neural processing units, their weight values can be adjusted by means of different stochastic gradient descent solvers.

Beyond their benefits in terms of spatial invariance, learnable convolutional layers in CNNs provide several other advantages. First, the fact that gradients can be propagated allows for a massively parallel iterative update of their weights and biases, paving the way for implementations deployable on graphical processing units (GPU) and tensor processing units (TPU). Another advantage of CNNs is the hierarchy of visual features learned by the network, which becomes progressively more specialized for the task at hand as more convolutional layers are stacked on top of each other. This offers a more structured interpretability of the knowledge captured by the layers, which can be disentangled by using deconvolutional filters or local explainability techniques (Zhang and Zhu 2018). But perhaps most interestingly, coarse visual features modeled in the first convolutional layers (edges, primitive shapes, etc.) learned on one task can be useful for others. Such tasks could lever-

age this general-purpose learned knowledge by importing pretrained weights and biases of such layers into their CNN architectures, so that the requirements in terms of learnable parameters or annotated data can be reduced. This simple yet effective knowledge exchange mechanism is referred to as *transfer learning* (Zhuang et al. 2020; Weiss et al. 2016) and has helped the adoption of CNNs in environments with scarcely annotated data or limited computational resources.

Sophisticated CNN architectures nowadays constitute the state-of-the-art for image and video classification modelling tasks, incorporating new ideas that boost even further their performance and/or efficiency. This is the case of capsule networks (Hinton et al. 2011), attention mechanisms (Vaswani et al. 2017), or patch-based learning in visual transformers (Han et al. 2020). When it comes to efficiency, the inner working of spiking neural networks (Grüning and Bohte 2014) has been investigated to alleviate the consumption of computing resources of these models. It is worth noting that the number of trainable parameters in CNNs may amount up to several tens of millions in very deep models, leading to problematically long training times, large storage requirements, and energy consumption footprints (Anthony et al. 2020). Finally, an important area of research is on the development of interpretability techniques for CNNs, which aim to dissect the knowledge captured by the layers of an already trained CNN (Arrieta et al. 2020). The result of this dissection, which can take many forms (e.g., attribution maps, counterfactual explanations, or simplified rule sets) is offered as an interpretable interface for the user to understand how and why the CNN provides its output. We will later elaborate on the plethora of possibilities of explanation techniques for CNNs used in EEs modelling and characterization tasks.

Different from CNNs, RNNs are built for modelling relationships in sequential data, including text and time series. Modelling such correlations requires that the network be capable of modelling, exploiting, and maintaining information (memory) at their neural processing steps, such that long-term relationships over the sequence can be exploited effectively when solving modelling tasks. In RNNs, this is accomplished by formulating a recurrent form of a neural processing unit, in which part of the output of the neuron is fed back to its input to realize a sort of neural mem-

ory. This new recurrent formulation of a neuron endows it with the possibility to learn and store information about the past that is relevant to the problem under consideration. For instance, this property of RNNs is key in time series forecasting, where the temporal lags to be predicted can be affected by data occurring far back in time. When RNNs are used for this task, the memory conferred to the neurons permits to model correlations over the sequence at different time scales. As the convolutional filters in a CNN, the parameters controlling how much of the output of a neuron is fed back to its input or stored in the hidden state vector can be learned via gradient backpropagation. The history of RNNs dates back to the work by Jordan (1997) and Elman (1990). Thereafter, the well-known long short-term memory networks (LSTM (Hochreiter and Schmidhuber 1997)) and the more recently proposed gated recurrent units (GRU (Cho et al. 2014)) became the standard in recurrent neural computation. LSTMs rely on several trainable parameters (gates) to control which parts of the sequence flow into the neuron by releasing or retaining information inside the hidden state vectors of neurons. GRU networks can be regarded as a variant of LSTMs that features small architectural modifications that permit to reduction the number of trainable parameters. In both cases, recurrent neural processing units can be arranged in a hierarchical structure comprising several stacked layers, in such a way that correlations are captured at different scales and levels of granularity. Several RNN approaches have been proposed in the literature over the years to overcome the drawbacks of the training process of these models. Attention mechanisms for instance (also applied in other types of deep networks such as CNN), make networks focus on certain parts of the input when predicting its output, discarding information that is not relevant for that specific input. Similarly, bidirectional RNNs aim at considering future steps of the sequence in the output of the neuron (Schuster and Paliwal 1997)). Recurrent networks that do not hinge on gradient backpropagation have also been developed in recent years, with reservoir computing and particularly echo state networks (Lukoševičius and Jaeger 2009; Gallicchio and Micheli 2017) being at the frontline. Finally, recent studies have emphasized that specialized CNNs for sequence modelling such as Temporal Convolutional Networks (TCN (Lea et al. 2017)) demonstrate longer and more effectively trained memory capabilities over diverse tasks and datasets, showcasing the potential of convolutional architectures also to address problems over sequential data.

3 Review of existing literature

This section critically analyzes and discusses the existing literature related to ML in atmospheric EEs. The method-ology applied has been the following: we perform a large number of search queries in well-known scientific publication databases, including Google Scholar, Scopus, and Web of Science. We systematically introduce a specific set of query strings in order to discover published works related to ML in atmospheric EE. We have used the term ML together with *extreme atmospheric events*, plus *extreme rainfall*, *flood prediction*, *heatwaves prediction*, *extreme temperature prediction*, *droughts prediction*, *convective systems*, *tropical cyclones prediction*, *hail and hailstorms*, *extreme wind gusts*, or *low-visibility prediction*, among many other terms linked to atmospheric EE. Once all results were retrieved from the aforementioned databases, we removed duplicates and performed an exhaustive analysis and discussion on a paper-by-paper basis, towards ascertaining their alignment with the topic under study. This systematic review process gave rise to the review and analysis that we present in the subsequent sections.

Figure 10 summarizes the hierarchical categorization of the state-of-the-art methods for atmospheric EEs problems. We classify the works according to the atmospheric event they predict, and then, using the type of ML methods they involve. Some works are included in several boxes since they apply several ML methods in EEs prediction problems.

3.1 Extreme rainfall and floods

Destructive extreme precipitation events and flooding episodes are a real threat to human settlements in different parts of the world (Madsen et al. 2014; Berghuijs et al. 2017). Extensive research on the monitoring, prediction and analysis of these events has been carried out in the literature. We analyze here those works dealing with ML techniques. Note that a first review on ML for flood prediction can be found in Mosavi et al. (2018), where the state of the art in this topic can be found, up to 2018. In Moon et al. (2019), a ML-based early warning system for short-term heavy rainfall is proposed for Korea. The system is formulated as a binary classification problem, where a logistic regression has been implemented over predictive variables from meteorological data obtained from automatic weather stations, which have been previously preprocessed by applying a principal component analysis algorithm. A comparison against early warning systems formed by alternative classifiers is carried out. An important amount of meteorological variables measured at different locations feed the classifiers in real-time, in order to improve the performance of the classification output. In Diez-Sierra and del Jesus (2020), a number of ML methods (SVM, k-nearest neighbours, RF, k-means clustering and neural networks) are applied to a problem of long-term rainfall prediction, using the atmospheric synoptic patterns as predictive variables. Neural networks are
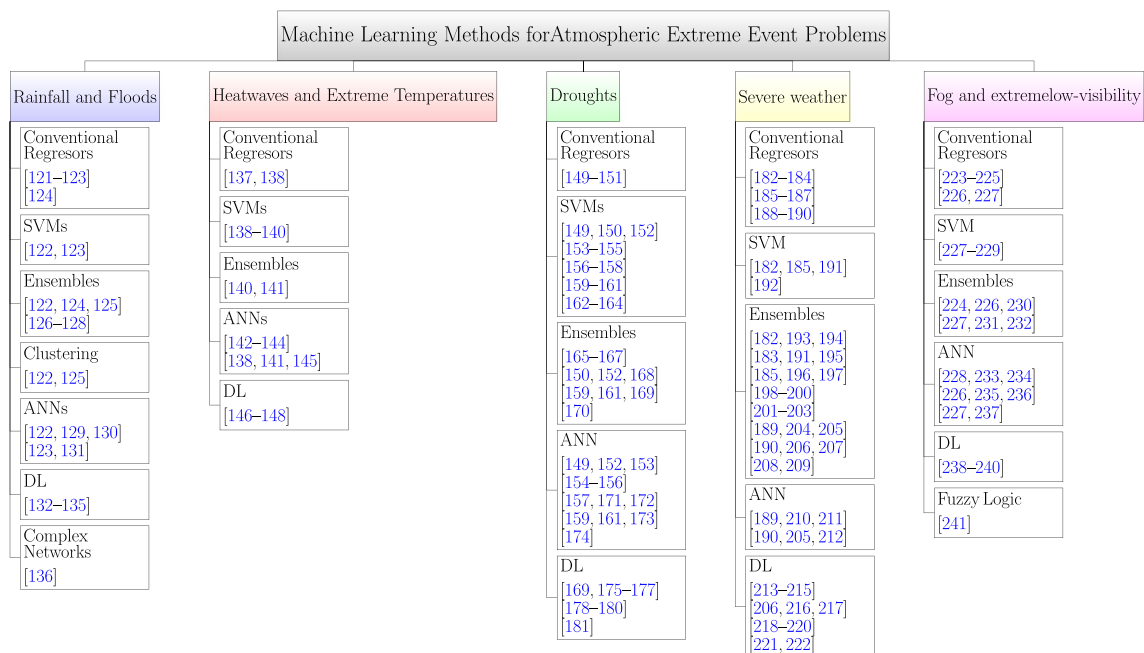
Fig. 10 Summary of recent works dealing with ML, DL, and related techniques in atmospheric extreme event problems

reported as the most accurate method, but surprisingly, the work reports the generalized linear method with gamma-distributed errors as the best method to predict the extreme of the series, improving the performance of the ML approaches. Note that supervised and non-supervised methods (k-means) are tested together, and depending on the method, a classification or regression problem is considered, which is an unusual procedure in the application of ML techniques. Results considered as ground truth rain gauges measurements from Tenerife (Canary Islands, Spain), are discussed. In Schlef et al. (2019), a self-organized map is used to obtain clusters of synoptic situations leading to extreme floods across USA. Then the flood characteristics of each synoptic situation are analyzed, identifying four primary categories of circulation patterns with different flood potential hazard. This methodology also allows identifying regions where extreme floods occur outside the normal flood season, and other regions where multiple extreme flood events occur within a single year, mainly due to tropical cyclones.

In Nayak and Ghosh (2013), a support vector machine is applied to short-term prediction of extreme precipitation in Mumbai, India. The prediction time horizon has been set in this case between 6 and 48 h. The predictive variables consist of mesoscale and synoptic scale weather patterns. The work identifies specific weather patterns for extreme precipitation events, finding out that they are different for nighttime precipitation or daytime extreme precipitation events. The SVM is then used to obtain extreme rainfall classification and prediction.

In Vandal et al. (2019), a problem of extreme precipitation statistical downscaling of GCM is tackled with ML algorithms. Five-ML methods are compared in this task: ordinary least squares, elastic-net, and support vector machine, sparse structure learning (MSSL) and autoencoder neural networks. Experiments with data from Northeastern United States suggest that the direct application of ML techniques does not improve the results of simpler statistical-based methods in the downscaling of extreme precipitation events.

In Grazzini et al. (2020), the classification of precipitation extreme events in northern-central Italy is carried out by means of K-means clustering and RF algorithm. The study reports the importance of integrated water vapour transport variable in the correct detection of extreme precipitation events in this region. This work has been complemented with a second study for the same zone, where the authors investigate the connection between precipitation extremes and Rossby wave packets (Grazzini et al. 2021). In Jahangir et al. (2019), an ANN algorithm is applied for the prediction of discharge values and spatial modelling of floods in Kan River Basin, Iran. Similarly, in Yeditha et al. (2020), different ML models (mainly neural networks) with a previous data treatment by wavelets are applied to forecast extreme precipitation from satellite measurements. The proposed approach has been tested in the prediction of floods in Vamsadhara river basin, India.

In Hosseini et al. (2020), a problem of flash flood forecasting with ML algorithms is tackled. The paper analyzes an ensemble of boosted generalized linear models random

forest, and Bayesian generalized linear models algorithms. A pre-processing step for reducing the number of input variables with a Simulated Annealing algorithm is carried out. These approaches are tested in the prediction of flash floods in the North of Iran. In Hu and Ayyub (2019), a Gradient Boosting Tree algorithm is applied to perform projections of precipitation intensity over short durations events, using outputs from GCMs. The algorithm performance has been tested in observational data (25 years of data) across USA. In Bui et al. (2019), an approach for flash flood susceptibility modelling is proposed. The algorithm combines tree-based ensemble with a pre-processing step of feature selection using a fuzzy-rule method and a Genetic Algorithm. These approaches have been combined with different tree-based ensembles such as LogitBoost, Bagging and AdaBoost algorithms. The performance of the systems was tested in data from Lao Cai Province (Northeast Vietnam). In Choi et al. (2018), different ML classification techniques such as decision trees, bagging, RF or boosting have been applied to the prediction of heavy rain damages at Seoul (South Korea). The work uses data on the occurrence of heavy rain damages in the city from 1994 to 2015, obtaining accurate results specially with the boosting technique. In Yang et al. (2023), a RF approach was applied in a problem of monthly extreme precipitation prediction from meteorological variables in Southern China. Data from 99 measuring stations near the Yangtze River are considered in this problem. The intrinsic RF feature importance is used to describe the physical mechanisms of extreme precipitation. In Pirone et al. (2023), a short-term precipitation prediction based on ML algorithms (ANNs) is proposed. The model employs cumulative rainfall fields from different stations data in Italy as inputs for the neural network and the idea is to predict rainfall interval and the corresponding probability of occurrence. In Lin et al. (2023), an ensemble method based on ML approaches RF, eXtreme Gradient Boosting (XGB) and ANNs is proposed to spot the key contributing variables to monthly extreme precipitation intensity and frequency in six different regions over the United States. In Vitanza et al. (2023), the Affinity Propagation algorithm, a clustering algorithm based on ML, was applied to a problem of extreme rainfall areas in Sicily, Italy. This approach does not require the number of clusters to be determined or estimated before running the algorithm, and it works based on the concept of "message passing" between data points. In this case, it was applied over a high-frequency, large dataset collected in the zone of study from 2009 to 2021, confirming the presence of recent anomalous rainfall events in eastern Sicily.

DL-based approaches have been recently applied to flood prediction, and it is expected that they are predominant in the years to come. In Shi (2020), convolutional neural networks (CNN) are used to carry out a smart dynamical downscaling of extreme convective precipitation from Global Climate Models (GCM). This work shows that when trained with data for three subtropical/tropical regions, CNNs are able to retain between 92 and 98% of extreme precipitation events. In Moishin et al. (2021), a CNN with LSTM Network has been introduced to forecast the future occurrence of flood events. The performance of this deep learning approach has been tested in 9 different rainfall datasets of floods that occurred in Fiji. In Xie et al. (2021), a problem of short-term intensive rainfall prediction was tackled with deep learning approaches. ECMWF forecast data and ground observation station data were taken into account, and K-means, generative adversarial nets and deep belief networks were applied to obtain the prediction as a classification model. Experiments in data from the Fujian Province (southeastern China) in the period 2015–2018, showed a good performance of the proposed prediction approaches, improving the results of LSTM and Stacked Sparse AE networks. In Manna and Anitha (2023), the integration of Rough Set on Fuzzy Approximation Space (RSFAS) with a deep learning (DL) technique is proposed in a problem of precipitation level in India. The idea is that RSFAS handles the uncertainty of the prediction, and the DL technique (an LSTM network) solves the associated classification and prediction problem. In Badrinath et al. (2023), a CNN is proposed to capture complex spatial precipitation patterns of precipitation, trying to identify and reduce biases affecting predictions of the dynamical model. The method is specifically based on a modified U-Net CNN, to postprocess daily accumulated precipitation over the United States West Coast. In Folino et al. (2023), an ensemble of deep neural networks is proposed for a problem of precipitation prediction in Italy, using heterogeneous data sources such as rain gauge measurements, radar and geostationary satellites. In Choudhary and Ghosh (2023), different types of DL networks such as RNN and LSTM have been applied to model monthly rainfall intensity and other climatic variables, such as temperature, in Jodhpur, India. The study shows that the LSTM obtains the best prediction results in this particular problem. In Chen et al. (2023), a DL model called weighted U-Net (WU-Net) is proposed for the problem of extreme precipitation prediction in China. This approach incorporates sample weights from different precipitation events to improve the forecasts of other intensive precipitation events over China. In Barnes et al. (2023), an approach combining ECMWF SEAS5 seasonal forecasts with CNNs is proposed to improve the forecasting of total monthly regional rainfall across Great Britain. An explainable analysis of the synoptic situations leading to specific CNN results is carried out.

Finally, in close connection with ML approaches, Complex Networks (CN) have also been used to analyze problems of extreme precipitation. In Boers et al. (2019), the teleconnections of extreme events over the world are studied,

using the CN paradigm over high-resolution satellite data. The CN methodology confirms Rossby waves as the physical mechanism behind global teleconnection patterns in extreme precipitation events.

3.1.1 Analysis

As a final note on the application of ML models to EEs related to rainfall and floods, we have found ML approaches in very different applications, including short-term and long-term detection and prediction problems, tackled with different ML frameworks (classification and regression) and considering very different prediction (or detection) time horizons. It is also remarkable the different ways in which many of these approaches introduce the physics of the problem within their approaches. In some cases, mainly in short-term prediction problems, the revised works consider real-time meteorological variables to feed ML algorithms, such as in Moon et al. (2019). In other cases, the ML extract information from synoptic patterns, mainly in problems of long-term rainfall and flood prediction (Diez-Sierra and del Jesus 2020; Schlef et al. 2019). In other cases, the output of GCM are treated with ML approaches in order to obtain improvements on the prediction of heavy precipitation events (Shi 2020; Vandal et al. 2019; Hu and Ayyub 2019). Other ML approaches rely on specific variables from reanalysis data but include in the studies variables with physical sense, such as sensitivity to flow conditions and other representatives of thermodynamic conditions for extreme precipitation events modelling, such as (Grazzini et al. 2020). A final group of works have been revised which only rely on measurements or set of data, without any specific consideration of the physics of the problem, especially when DL has been applied (Moishin et al. 2021; Xie et al. 2021), but also with shallow ML approaches (Choi et al. 2018). In these last cases, the works analyzed seem to focus on the ability of ML approaches to extract information and obtain accurate predictions, evaluated from different metrics, and compared against other ML approaches, with very few references to the physical processes causing the EE. It is possible to see how, in the last years, the amount of DL-based approaches has increased a lot, and it is expected that in the near future, DL techniques will dominate the research on extreme precipitation prediction (Chase et al. 2023). Finally, the work in Boers et al. (2019) analyzes extreme precipitation events from CN paradigm, generating networks which take into account the physics of the problem and the relationship among different variables involved in the problem, including the analysis of teleconnections. This introduces a novel paradigm in the study and analysis of extreme precipitation, which may be hybridized with ML techniques in the near future.

3.2 Heatwaves and extreme temperatures

Extreme temperatures (Barriopedro et al. 2011; Pfleiderer and Coumou 2018), heatwaves (Chapman et al. 2019; Barriopedro et al. 2023) and, in the last decades, mega-heatwaves (Bador et al. 2017; Sánchez-Benítez et al. 2018) are among the extreme atmospheric events potentially most dangerous for people, especially the elderly (Díaz et al. 2002a, b) and with deep societal impact. The detection, prediction and attribution of heatwaves and extreme temperatures is, therefore, a hot topic in atmospheric EEs research (Wang et al. 2017), including the study of natural causes such as circulation patterns (Shi et al. 2018) or anthropogenic contribution (Zwiers et al. 2011). ML methods have been applied to study these and other aspects of extreme temperatures and heatwaves (Cifuentes et al. 2020).

3.2.1 Heatwaves

In Pasini et al. (2017), neural computation is used in a problem of attribution of heatwaves. The study considers the last 160 years, where the attribution to anthropogenic forcings is obtained for the last 50 years, whereas in the period 1910-1975 the main driver is solar irradiation. The study also clarifies the role of aerosols and the Atlantic Multidecadal Oscillation in decadal temperature variability.

In Park and Kim (2018), multivariate adaptive regression splines are used to set appropriate heatwave thresholds, in order to improve early warning systems for these events. The work uses daily data of emergency patients diagnosed with heatstroke and also information on 19 meteorological variables obtained for the years 2011 to 2016. The results obtained show that the combination of heat illness data and average daytime temperature (from noon to 6 PM) can be used as an alternative threshold for heatwaves characterization. Finally, in Chattopadhyay et al. (2020), a hybrid approach combining the Analog prediction method (search of analogue synoptic situations in the past) with deep neural networks (capsule neural networks, CapsNets) is proposed to predict heatwaves and cold spells. The proposed CapsNets outperformed other deep approaches such as CNN and alternative prediction algorithms such as logistic regression techniques. Finally, in a recent work (Weirich-Benet et al. 2023) the performance of linear regressors and RF algorithms in a problem of subseasonal heatwaves prediction is discussed. Different inputs (drivers) are previously chosen by using a correlation-based analysis.

3.2.2 Extreme temperatures

One of the first approaches in the application of ML techniques for extreme temperature prediction was Abdel-Aal

and Elhadidy (1995), where different artificial network models are applied to a problem of daily maximum temperature prediction in Dhahran, Saudi Arabia. In this case, daily data for 18 weather parameters are considered as input variables, to predict the maximum temperature on a given day, with different prediction time horizons up to 3 days in advance. In Paniagua-Tineo et al. (2011), a SVR algorithm is used to forecast daily maximum air temperature with a 24h prediction time horizon. The prediction system relies on a number of input variables such as air temperature, precipitation, relative humidity and air pressure. It also considers the synoptic situation of the day in order to improve its results. The performance of the SVR algorithm has been successfully evaluated with data from a number of European measurement stations. In De and Debnath (2009), the prediction of the maximum (and minimum) air temperature in the summer monsoon season is carried out by using a multi-layer MLP perceptron neural network. The mean temperature of previous months in the period of analysis is considered as input for the system. Data from the Indian Institute of Tropical Meteorology belonging to the years 1901–2003 are considered.In Chithra et al. (2015), neural networks are applied to a problem of monthly mean maximum and minimum temperature in Chaliyar river basin, India. The objective is to evaluate the impact of climate change in the accuracy of the predictions obtained by neural networks. In Ahmed et al. (2020), different ML approaches such as MLP, SVM and relevance vector machine (RVM) or K-nearest neighbour (KNN), are proposed to develop multi-model ensembles from global climate models. The objective is to obtain annual predictions of monsoon and winter precipitation, maximum temperature and minimum temperature over Pakistan. The results obtained have shown that KNN and RVM-based multi-method ensembles show better skills than those developed with MLP and SVM.In Peng et al. (2020), a MLP and a natural gradient boosting algorithm (NGBoost), are applied to improve the prediction skills of the 2-m maximum air temperature, with a prediction time horizon with lead times from 1 to 35 days. The ML prediction approaches have shown better results than the ensemble model output statistics (EMOS) method (which was selected as the benchmark for comparison) in 90% of the cases analyzed. In Oettli et al. (2022), a number of ML algorithms such as neural networks, SVMs, RF, Gradient Boosting or regression trees have been applied to the prediction of surface air temperature two months in advance, with input data two months in advance from SINTEX-F2, a dynamical prediction system. The dynamical prediction system includes the physics of the problem, while the ML algorithms improve the results by a statistical downscaling. The performance of these approaches has been tested in Tokio (Japan), obtaining excellent prediction results.In Gómez-Orellana et al. (2023), a problem of long-term air temperature prediction with eXplainable

Artificial Intelligence (XAI) algorithms is tackled. Specifically, artificial neural networks trained with evolutionary algorithms are tested on this problem. This XAI model architecture has been applied to the long-term air temperature prediction at different sub-regions of the South of the Iberian Peninsula, with good performance results.

Very recently, DL approaches have been applied to long-term extreme temperature prediction problems, such as in Nandi et al. (2022), where an approach called Attention-based Long-term Temperature Forecasting Network is proposed. This approach uses an Encoder-Decoder system similar to that shown in Sect. 2.1.1. The Encoder encodes the relative dependencies of the auto-regressive time series into an *attention* tensor (dimensionality reduction) which is used by the Decoder to produce the prediction. The Encoder is augmented to incorporate a convolution block to recognize the seasonal patterns associated with extreme temperatures. The model was evaluated in real data from five different cities around the world. In Fister et al. (2023), different DL algorithms have been tested in a problem of extreme air temperature forecasting. Different DL prediction approaches have been tested, including a Convolutional Neural Network (CNN) with video-to-image translation, several ML approaches including Lasso regression, Decision Trees and Random Forest, and finally a CNN with pre-processing step using Recurrence Plots, which convert time series into images. Good prediction skills have been obtained for two cases of extreme temperature in Paris and Córdoba, Spain.

3.2.3 Analysis

The works revised in this subsection reveal that there are not many works dealing with heatwave prediction using ML approaches. Only a few specific works on the application of ML techniques to heatwave estimation have been found in the recent literature. In Park and Kim (2018), the work uses data from meteorological variables and emergency patients in order to obtain a characterization of heatwaves. A second approach discussed heatwaves prediction with ML (Chattopadhyay et al. 2020). Here, ML algorithms (DL networks in this case) are merged with the Analog method which introduces the physics of the problem in order to predict heatwaves. A recent paper Weirich-Benet et al. (2023) discusses how linear regression and RF can be successfully sued in a problem of heatwaves prediction.

There are many more works on ML algorithms for extreme temperature prediction problems. Artificial neural networks and statistical ML approaches are the main algorithms applied in the literature to tackle these problems. It is interesting to see how in these works, the inclusion of physics is not as relevant as in the works dealing with ML algorithms for rainfall and flood prediction. The reason for this is that air temperature is in general a variable easier to be predicted

than rainfall, in which the inclusion of the atmospheric state and dynamics is key to obtain good results. Synoptic situations (considered in Paniagua-Tineo et al. (2011)) seem to improve the results of ML algorithms in the prediction of extreme temperatures. In the rest of the articles revised, the prediction is based on existing registers of previous temperatures. The application of ML approaches produces good results in this case in weekly or monthly temperature predictions, where the variation of the extreme temperatures is small.

3.3 Droughts

Droughts are extreme events, stochastic in nature, with a deep impact on society, specifically on water supplies, agriculture, and hydroelectric power production, and associated with forest fires and even forced migrations (Spinoni et al. 2019; García-Herrera et al. 2019). Drought early warning systems provide important information about predicted drought hazards. In many cases, these systems rely on ML and DL algorithms.

In Sutanto et al. (2019), a RF algorithm is used to forecast drought impacts, by relating forecasted hydro-meteorological drought indices to previously reported drought impacts. The proposed model based on ML is able to forecast drought impacts with prediction time horizons of some months ahead. In Khan et al. (2020), different ML classification techniques are applied to develop drought prediction models over Pakistan. They include SVM, MLP and KNN algorithms. Meteorological variables from reanalysis are considered as inputs, whereas the objective variable considers three categories of droughts: moderate, severe, and extreme in different cropping seasons. These classes were estimated using the Standardized Precipitation Evaporation Index (SPEI; Vicente-Serrano et al. (2010)), in order to train and test the proposed ML classifiers. In Rhee and Im (2017), a problem of high-resolution spatial drought forecasting is tackled in Korea from remote sensing and climate indices inputs. The performance of different regression tree algorithms, RF and Extremely randomized trees have been compared. In Park et al. (2016), different ML algorithms such as RF boosted regression trees, and Cubist is applied to model meteorological and agricultural droughts from 16 inputs drought factors obtained from satellite measurements. The SPI and crop data are used as objective variables to model the droughts. RF has been reported as the best performing algorithm in data from arid zones of the United States. In Rahmati et al. (2020), drought hazard is tackled with different ML models: classification and regression trees (CART), boosted regression trees (BRT), RF, multivariate adaptive regression splines (MARS), flexible discriminant analysis (FDA) and SVM. Some Hydro-environmental datasets are used to calculate

the relative departure of soil moisture (RDSM), and this index is used as an objective variable, whereas the inputs are eight environmental factors as potential predictors of drought. Experiments in the southeast part of Queensland, Australia, are carried out to evaluate the performance of the different ML methods proposed. In Feng et al. (2019), three ML algorithms (RF, SVM and MLPs) are used to evaluate whether remotely-sensed drought factors (satellite measurements) are good estimators for drought events prediction in south-eastern Australia. RF is again the ML regression technique which best results obtains in this problem, outperforming SVM and MLPs in this task. In Belayneh and Adamowski (2013), short-term drought prediction in the Awash River Basin (Ethiopia) is considered, by means of SPI prediction. Three ML methods are evaluated for this problem, MLP, SVM and MLP with a previous step of wavelets signal decomposition. The coupled wavelet-MLP algorithm showed the best result in SPI prediction with a prediction time horizon of 1 month and 3 months. New results and further analysis on the same problem were reported in Belayneh et al. (2016). In Belayneh et al. (2014), a long-term drought prediction problem in the Awash River is considered by means of MLPs and SVMs, enhanced with wavelets transforms. The SPI at 12 and 24 months (SPI 12 and SPI 24) are predicted by means of the ML methods. Comparison with ARIMA methods for time series prediction shows a better performance of the ML techniques. The same data from Awash River Basin are used in Belayneh et al. (2016) to test advanced versions of ML algorithms in the same problem of drought prediction. Coupled versions of ML algorithms with wavelet transforms are considered, such as wavelet transforms with Bootstrap and Boosting ensembles together with MLP and SVR models. These coupled models show a better performance than the MLP and SVR algorithms on their own. In Roodposhti et al. (2017), a problem of drought sensitivity mapping based on SPI index and enhanced vegetation index (EVI) is tackled, by using one-class SVMs. Data from both synoptic stations and satellite data are combined in this study in the Iranian province of Kermanshah. In Piri et al. (2023), different ML approaches based on ANNs and SVRs with evolutionary-based feature selection mechanisms are proposed to predict different meteorological drought indices for different measurement stations in Iran. In Mokhtari and Akhoondzadeh (2021), ML algorithms such as ANN, SVR, DT or RF are applied to a problem of drought prediction for monthly periods, using inputs derived from the active and passive sensors of different satellite sensors. In Deo and Şahin (2015), the performance of the ELM algorithm is evaluated in a problem of Effective Drought Index prediction in eastern Australia. Predictive variables composed of meteorological variables and climate indices are considered. The ELM approach outperformed the results of different neural network models. In Aghelpour et al. (2020), different ML

approaches are evaluated in the problem of forecasting the precipitation joint deficit index (JDI) and the multivariate standardized precipitation index (MSPI), both of them related to severe droughts. Different ML methods are considered, such as group method of data handling (GMDH), generalized regression neural network (GRNN), least squared support vector machine (LSSVM), adaptive neuro-fuzzy inference system (ANFIS) and ANFIS optimized with meta-heuristics algorithms. Experiments in data from 10 measuring stations in Iran are considered. The GMDH method is reported as the most accurate algorithm. In Zhang et al. (2019), artificial neural networks and XGB algorithms with feature selection by means of a cross-correlation function and a distributed lag nonlinear model (DLNM) are considered in a problem of drought prediction. Data from 32 stations from 1961 to 2016 in Shaanxi Province, China, are used. The results show that the XGB approach outperforms neural networks and the DLNM works better than the cross-correlation function in the selection of the best features for this prediction problem. In Dikshit et al. (2020), MLP and SVR algorithms are tested in a problem of drought prediction in New South Wales, Australia. SPEI index at 1, 3, 6, and 12 months are used as objective values. The results obtained suggest that the MLP outperforms SVM. The results also discard that sea temperature and climate indices had a real impact on the droughts in New South Wales. In Richman and Leslie (2018), a feature selection problem is considered for attribution of the Cape City drought 2015–2017 with ML algorithms. Wrapper algorithms for FSP are considered, in which the SVM has been used as a classification algorithm, and different evolutionary algorithms look for the best set of features (drought drivers) for predicting the cool season precipitation in the years of the drought. In Pande et al. (2023), different SVM versions were tested in a problem of drought prediction in the upper Godavari River basin, India. The SPI index was used as the objective variable to predict future droughts in the zone. In Li et al. (2021), the role of antecedent SST fluctuation pattern (ASFP) as a drought driver is analyzed by using ML techniques such as SVR, RF and ELM. The SPEI is used as an objective to be predicted at different river basins such as Colorado, Danube, Orange, and Pearl Rivers. The obtained results showed that the ASFP-ELM model can effectively predict the space-time evolution of drought events outperforming the rest of the ML algorithms considered. In Prodhan et al. (2022), RF and gradient boosting machine algorithms are applied to characterize future drought metrics and their impact on crops. The magnitude, intensity, and duration of future droughts are characterized by means of the SPEI drought index using CMIP6 (Coupled Model Inter-comparison Phase-6) climate models data. Experimental results on Southern Asia, including countries such as Afghanistan, Pakistan, and India are analyzed.

Very recently, DL algorithms have been also applied to different problems in drought prediction. In Gyaneshwar et al. (2023), a review of the most important DL algorithms with application in drought prediction is presented. The work also includes a number of ML approaches for drought prediction. In Mokhtar et al. (2021), four ML and DL methods (RF, XGB, CNNs and LSTMs) were considered in a problem of SPEI estimation in the Qinghai-Tibet Plateau. Meteorological variables and climate indices are considered predictive variables. In Abbes et al. (2023), a DL-based approach for drought forecasting based on combining Long Short-Term Memory (LSTM) and Multi-Resolution Analysis Wavelet Transform is proposed. Experiments in data from the Sarab region (Iran) based on the standardized precipitation Evaporation index (SPEI) prediction showed a good performance of this DL-based approach. In Kaur and Sood (2020), different ML and DL approaches such as ANN, ANN optimized with Genetic Algorithm and Deep Neural Networks, all hybridized with a SVR algorithm, are tested in a problem of drought prediction. Their performance is compared showing that the deep neural network was the best-performing approach in drought prediction. In Vo et al. (2023), a hybrid model involving DL (LSTM networks) is a climate model for drought prediction. The proposed hybrid DL-based systems were tested in real data from South Korea. In Danandeh Mehr et al. (2022), a hybrid intelligent DL-based model for drought prediction, formed by the combination of CNN and LSTM networks was proposed. This approach was tested in a drought prediction problem with multi-temporal drought indices (SPEI-3 and SPEI-6) as objectives, in the Ankara region, Turkey.

In close connection with drought forecasting, evaporation prediction has been tackled in some cases. For instance, Yaseen et al. (2020) evaluates ML approaches for evaporation prediction in arid regions of Iraq. Four different ML models are considered including classification trees, a cascade correlation neural network, a gene expression programming (GEP), and a SVM algorithm. Another recent work dealing with alternative prediction problems related to drought forecasting is, Tufaner and Özbeyaz (2020) where the Palmer Drought Severity Index (PDSI) is predicted by using different ML algorithms. SVM, MLP and decision trees have been applied to this problem, and their results compared to a Linear Regression algorithm used as a baseline technique. Results in a problem of PDSI prediction in Anatolia (Turkey), have shown that the MLP obtains the best results. Finally, Adikari et al. (2021) evaluates the performance of three different ML algorithms (convolutional neural networks (CNN), long-short term memory network (LSTM), and wavelet decomposition functions combined with the adaptive neuro-fuzzy inference system (WANFIS)) in two different problems of flood and drought forecasting. The results obtained reveal that CNNs is the best-compared

approach for flood forecast and WANFIS outperforms the other two algorithms in drought forecasting.

3.3.1 Analysis

The review of articles about ML techniques for drought and related problems has shown a large number of ML algorithms applied to drought prediction and analysis. Ensemble methods such as RF seem to be strong approaches for prediction problems related to drought, though other algorithms such as neural networks or statistical learning approaches (SVMs) have also shown to be strong possibilities. DL-based algorithms have also been successfully applied to different drought prediction cases, mainly in the last few years. The inclusion of the physics is, in the majority of cases, treated by means of considering climate indices among the predictive (input) variables of the problems, though some approaches such as Dikshit et al. (2020) have discarded that climate indices improve as predictive variables improve the performance of ML algorithms in specific problems of drought prediction. In Vo et al. (2023), a hybrid approach which directly involves a DL algorithm and a climate model is proposed for drought prediction. In general, processes related to atmospheric dynamics seem to dominate this phenomenon, so the inclusion of climate indices as inputs for ML algorithms seems a reasonable election in order to capture the physics of the problem. Regarding the objective variables for defining the problem, the majority of problems analyzed used precipitation indices such as SPI or SPEI, as drought indicators.

3.4 Severe weather

EEs related to severe weather have also been studied and analyzed with ML methods in the last few years. We have divided this subsection into different parts, ML methods in convective systems studies, tropical cyclones, hailstorms and extreme wind and gusts.

3.4.1 Convective systems

There are different works focused on the study of convective clouds and systems formation and related events with ML approaches (Xiu et al. 2016; McGovern et al. 2023).

In Tebbi and Haddad (2016), a problem of convective cloud classification by means of the combination of ANN and SVM, using high-resolution satellite images in northern Algeria is tackled. The proposed system works in two steps. First, the system detects rainy areas in cloud systems, and second, it delineates convective cells from stratiform ones. In Sahoo and Bhaskaran (2019), a problem of storm surge and coastal floods prediction with artificial neural networks is tackled. The work is focused on Odisha state (India), try-ing to simulate the effects of the tide caused by the super cyclone of 1999. Comparison with the ADCIRC prediction model Luettich et al. (1992) shows that the ML-based model is able to obtain significant results in the prediction of storm surge and associated flood of Odisha event. In Guijo-Rubio et al. (2020), a problem of classification of convective situations over Madrid-Barajas airport is tackled, with neuro-evolutionary techniques (neural networks trained with evolutionary computation techniques). The problem is considered a multi-class classification problem, highly imbalanced (there are much less convective situations than clear days). However, the neuro-evolutionary approaches are able to obtain an accurate performance in the identification of days with convective cloud formation in Madrid airport. A similar problem is tackled in Guijo-Rubio et al. (2020) by considering ordinal regression techniques instead of classification. Another study is presented in Jergensen et al. (2020), where a problem of thunderstorms classification is tackled with different ML approaches, such as logistic regression algorithms, RF, gradient-boosted forests and SVMs. The problem has been formulated as a multi-class classification problem, in which the gradient-boosted forest algorithm obtained the best classification results. In Hill et al. (2020), the RF algorithm is evaluated in problems related to convective systems. The study includes different EEs from convective systems such as the presence of tornadoes, large hail (over 1 inch) or induced wind gusts over 58 mph. A large number of predictive variables are considered in this study, including different atmospheric fields such as 10-m winds, surface temperature and specific humidity, precipitable water, accumulated precipitation, and wind shear from the surface at different pressure levels or mean sea level pressure, among others. The RF algorithm was able to obtain relationships between predictive atmospheric fields and observations according to the community's physical understanding of severe weather forecasting. Dealing with a similar idea, McGovern et al. (2017) evaluates the performance of RF and Gradient Boosted Regression Trees in a problem of prediction skill for multiple types of high-impact events related to convective systems, such as severe wind, hail or heavy rain, with discussion on the impact of this severe weather in renewable energy or aviation turbulence. In Flora et al. (2021), three ML approaches RF, gradient-boosted trees, and logistic regression algorithms have been proposed to predict whether ensemble storm tracks will produce a tornado, severe hail, and/or severe wind report. The paper describes postprocessing using the ML algorithms of the ensemble output from the National Oceanic and Atmospheric Administration Warn-on-Forecast (WoF) project. The results obtained have shown that the ML-based postprocessing of WoF data improves short-term, storm-scale severe weather probabilistic guidance.

In Stubenrauch et al. (2023), ML techniques are used to improve the construction of an accurate 3D description of upper tropospheric cloud systems, in order to study the relation between convection and cirrus anvils. For this, different ANN models are trained on collocated radar-lidar data to obtain estimations of cloud top height, cloud vertical extent and cloud layering. ML methods are also used to estimate rain intensity classification in upper tropospheric cloud systems. In Shamekh et al. (2023), using a ML approach based on neural networks, it is shown that it is possible to discover the role of the organization of clouds on precipitation, and then include this information to improve precipitation prediction in climate models.

Finally, DL-based approaches have also been tested in prediction problems related to severe convective systems, such as in Zhou et al. (2019), where a CNN is introduced for severe convective weather prediction, including heavy rain, hail, convective gusts, and thunderstorms. The predictive variables are obtained from a numerical weather model (Global Forecasting System), and the performance of the CNN is compared to that of traditional methods and human expert evaluation of the data. The results showed that the CNN obtained results which improved the performance of previous algorithms and human expert results, but with some flaws such as too many false alarms in predicting hail and convective gusts. In Sobash et al. (2023), different DL-based approaches (DNN, CNN and CNN-Gaussian mixtures were used to probabilistically classify CAM storms into one of three different modes: supercells, quasi-linear convective systems, and disorganized convection. The storm mode classification is very useful to provide information about the hazard types of different convective systems.

3.4.2 Tropical cyclones

Other EEs associated with severe weather are tropical cyclones (TC). In addition to their extremely associated gusts, they always come with other severe weather events such as heavy rain, hail, or thunderstorms, in many occasions deriving in catastrophic events such as floods (Chen et al. 2020), storm surges (Xie et al. 2023), ground slides, etc. There is a very recent comprehensive review on ML approaches in TC forecast (Chen et al. 2020). That article covers previous works on ML for TC up to 2020. There have been some works dealing with topics related to ML for TC after that review paper. For example, there is some recent work dealing with ML for TC prediction and characterization, such as Baki et al. (2021) where a multivariate adaptive regression splines (MARS), has been applied to obtain the optimal values of the WRF mesoscale model parameterizations for TC prediction in the Bay of Bengal. In Tan et al.

(2021), a gradient boosting decision tree model has been proposed for TC track forecast at Western North Pacific. A comparison with climatology and persistence is carried out to evaluate the performance of the proposed ML technique in this problem. In Sun et al. (2021), ensemble methods optimized by ML approaches such as Lasso optimization or Ridge regression are proposed to improve preseason prediction of Atlantic hurricane activity. In Pillay and Fitchett (2021), an analysis of the initialization variables affecting TC formation is carried out. RF algorithms are proposed to analyze the importance of each climate variable considered. The RF models are also used to predict intensification magnitudes of the TC based on the state of the input variables.

In Kar and Banerjee (2021), different ML algorithms have been applied to a problem of cloud intensity classification in TC over the Bay of Bengal and the Arabian Sea. Five ML classifiers have been proposed for this problem: Naïve Bayes, SVM, logistic model tree, random tree, and RF. The RF algorithm showed the best performance over the rest of the tested classifiers for this problem. In Kim et al. (2021), a decision-tree algorithm has been proposed for a problem of TC maximum lifetime intensity. The algorithm predicts the probability that a TC reaches a maximum intensity larger than 70 knots. Accurate results are obtained with classification rates over 90% in the considered test set. There have been some works dealing with the estimation of the precipitation produced by TC using ML techniques. In Zhu and Aguilera (2021), the RF method is applied to a problem of prediction of the precipitation associated with TC in Eastern Mexico. In Ngo et al. (2021), a hybrid Quantum PSO algorithm and a Credal Decision Tree (CDT) ensemble have been proposed for spatial prediction of the flash floods in TC. Experiments are carried out in the northwestern mountainous area of Vietnam. Satellite data from Sentinel-1 C-band SAR images are considered in this case to model the objective function. Finally, there are some recent works dealing with ML applications for evaluating the impacts of TC. In Nethery et al. (2021), ML algorithms, mainly Bayesian methods, are used to estimate health problems caused by TC. In Wendler-Bosco and Nicholson (2021), the economic impact of TC is analyzed by means of ML approaches, and in Zhang et al. (2021), the impact of typhoon Lekima on different Chinese forests is evaluated by means of RF over Landsat 8 OLI images. In Meng et al. (2023), different Gradient Boosting approaches have been proposed for probabilistic forecasting of TC intensity from different predictive variables such as sea surface temperature data, satellite bright temperature data, and data from other models and satellite-derived variables. Finally, in Ascenso et al. (2023), a ML framework based on evolutionary computation techniques (genetic algorithms Del Ser et al. (2019)) is applied to the optimization of TC genesis indexes. This approach is shown to obtain an index

which captures the spatial and interannual variability of tropical cyclone genesis.

As in the case of other EEs applications, DL-based algorithms have been profusely used for TC prediction, mainly in the last few years. In Asthana et al. (2021), a CNN was used to predict Atlantic hurricane activity from reanalysis data. Accurate prediction results are reported, in comparison with alternative state-of-the-art models. In Farmanifard et al. (2023), a problem of TC trajectory prediction is tackled with a DL algorithm, formed by a hybrid MLP-LSTM approach. This approach was evaluated using the North Atlantic Ocean TC dataset, and input data such as wind speed, wind direction, and air pressure in the zone of study. Another work dealing with TC trajectory prediction was presented in Wang et al. (2023), where DL approaches (RNN, LSTM, and GRU) were applied to predict TCs trajectories in the northwestern Pacific in the Reanalysis period. In Zhuo and Tan (2023), DL algorithms were applied to a problem of TC size estimation from data infrared imagery in the Western North Pacific. The DL algorithms developed were then applied to a homogeneous satellite database to reconstruct a new historical dataset of TC sizes in the zone. In Chen et al. (2023), a study on rapid intensification of TC with DL-based algorithms (LSTM networks) is carried out. The results show that the LSTM network is able to improve the enhanced intensity and rapid intensification prediction performance in Western Pacific TC by using information from satellite images.

3.4.3 Hailstorms

Hail is an atmospheric EEs which causes important economic problems in many countries, mostly in agriculture and crop losses. Though it is not a frequent EEs (returning periods of severe hailstorms have been set around 20 years, depending on the zone, according to different studies (Fraile et al. 2003)) there are some works on prediction and characterization of this EE, including the use of ML techniques in the last years. Note, however, that prediction of hailfalls is a difficult task, due to the local spatial characteristic of this EEs and its short duration, which makes that prediction approaches should be developed separately for specific geographic areas.

One of the first works dealing with a prediction problem of hailfalls is López et al. (2007), in which the problem is tackled as a binary classification task (hail/no-hail). A logistic regression was then applied, obtaining a probability of Detection of 0.87 with a False Alarm Ratio of 0.18. After this initial work on hailstorm prediction, some more sophisticated ML methods were introduced. In Gagne et al. (2015), a hybrid approach mixing NWM with ML algorithms is proposed for a problem of hailfall forecasting. The NWM identifies potential hail storms and different ML algorithms mainly RF and gradient-boosting trees are used to predict hail occurrence.

Observed hailstorms are used to obtain the ground truth values for this problem.

RF approaches have been recently applied to problems of hailfall prediction. In Gagne et al. (2017), a storm-based probabilistic hail forecasting is proposed, including an RF algorithm in the system. The prediction starts with an identification and tracking algorithm based on radar grid data and a convection-allowing model. Different parameters for characterizing the storm are then obtained and passed to the RF algorithm which has been previously trained with data from observed hailstorms. The RF algorithm uses this information to predict the probability of a storm producing hail, and also provide the hail size estimation. In Czernecki et al. (2019), a RF algorithm has been proposed for a problem of large hail prediction. Different predictive variables such as radar reflectivity, EUCLID lightning detection data, and convective indices from the ERA5 reanalysis are considered. The objective variables are obtained from observational data of large hail reports from Poland in the period 2008–2017. Also dealing with hail prediction using a RF algorithm, Yao et al. (2020) used hail observation data from 41 meteorological stations in the Shandong Peninsula, China, in the period 1998-2013 to train the algorithm. Different thermal factors and variables such as lifted index, Showalter stability index, and total index are used as predictive variables of hailfalls in this work. Another example of the use of RF in hail prediction is Burke et al. (2020), in which different observational datasets were used to train and test the RF approach, such as the maximum estimated size of hail (MESH), and the multi-radar multi-sensor (MRMS) product.

Finally, Some recent works have applied DL approaches to problems of hail prediction. In Pullman et al. (2019), a DL network has been applied to a problem of hailstorm detection. The GOES satellite imagery and MERRA-2 reanalysis data are used as predictive variables in this case. In Gagne et al. (2019), a CNN is applied to the problem of predicting the probability of severe hail (larger than 2.5 mm) in the next hour. Data for this study have been obtained from NCAR convection-allowing ensemble in May 2016. In Leinonen et al. (2023), a DL-based approach is presented for a problem of thunderstorm prediction, using multiple data sources such as data from weather radar, lightning detection, satellite visible/infrared imagery, numerical weather prediction, or digital elevation models. The DL model is able to predict lightning, heavy hail and precipitation probabilistically on a reduced spatial resolution (about 1 km) and with prediction time horizons between 5 min and 1 h. In Kolios (2023), a DNN model for hail detection is proposed. The input data consist of satellite (Meteosat) multispectral infrared (IR) imagery, exclusively. The DNN model was trained using numerous cases of hail events, as they were recorded from the European Severe Weather Database.

3.4.4 Extreme winds and gusts

Extreme wind gusts (EWG) are associated with severe weather. They can have catastrophic effects on crops and buildings and also have an impact on renewable energy facilities such as wind farms. A first review of techniques for WG prediction, including NWM and also ML approaches has been presented in Sheridan (2018). In Sallis et al. (2011), several ML algorithms have been applied to a problem of WG prediction. Logistic regression, MLPs and C4.5 classification trees and CART algorithms are tested in a problem of WG prediction at Kumeu, New Zealand. In Shanmuganathan and Sallis (2014), a similar problem was tackled, also in New Zealand. In this case, the study evaluates the performance of classification trees, MLPs and Self-Organizing Maps (SOM). In-situ measurements and data acquired between 2008 and 2012 at the Kumeu site, have been used for this study. In Lagerquist et al. (2017), a problem of extreme wind prediction in the surroundings of storm cells in the USA is carried out. The problem consists in calculating the probability of extreme winds over 50kt (25.7 m/s) in zones close to storm cells. The problem is formulated as a binary classification problem. The predictive variables considered in this case are based on radar measurements, storm motion and shape, and atmospheric soundings in the near-storm environment. Several ML models have been tested, including, logistic regression, RF, MLPs and Gradient boosting trees ensembles. In Wang et al. (2020), an ensemble model for WG prediction is presented. The proposed ensemble includes RF, a long-short-term memory (LSTM) algorithm and Gaussian processes for regression. A comparison against each model on their own, the persistence and a gradient-boosted decision tree showed the good performance of the ensemble method. Also dealing with ensemble models, in Schulz and Lerch (2021), a comprehensive review and comparison of eight ensemble methods based on ML for WG forecasting is carried out. The proposed algorithm is tested in 6 years of data from a high-resolution ensemble prediction system of the German weather service. In Spassiani and Mason (2021), a SOM is proposed to analyze the meteorological origin of WG in Australia. The SOM is used to establish the origin of the Application of Self-organizing Maps to classify the meteorological origin of WG into convective (from thunderstorms) and non-convective origin (synoptic), with different subclasses in each case.

In Arul et al. (2022), a RF approach is applied to the identification of extreme wind field characteristics and associated wind-induced load effects on structures, via the detection of thunderstorms. The idea is to use large databases containing high-frequency sampled continuous wind speed data and use the shapelet transform to identify individual attributes distinctive of extreme wind events. Experiments using real data from 14 Mediterranean ports, including sites in Italy, Spain and France are carried out.

In Peláez-Rodríguez et al. (2022), a hierarchical classification-regression ML approach is proposed for a problem of extreme wind prediction. The approach starts with the application of clustering algorithms and different balancing techniques to increase the significance of clusters with poorly represented wind gusts data. Then the classification of each sample into the corresponding cluster is carried out, and then, once we have determined the cluster a sample belongs to, a final regression level provides the prediction of the wind speed value. This approach has shown excellent results when enough data are available to train all the ML algorithms involved in the prediction system.

In Chkeir et al. (2023), DL-based approach based on a LSTM network is applied to a problem of extreme rain and wind speed nowcasting in the area of Malpensa airport, by merging different datasets from sensors in the local area of the airport. The results obtained showed extreme wind speed probability detection higher than 90%, with false alarms lower than 2% in this particular problem.

3.4.5 Analysis

The large majority of EEs related to severe weather are meteorological events, in which thermodynamic processes of the atmosphere play a central role. Depending on the EEs considered as severe weather, the period of return of the EEs is extremely high, such as damaging hailstorms, though other EEs classified as severe weather are much more frequent. Techniques to take into account the physics of these EEs in the ML are based on NWM (the ML algorithms are applied to the output of NWM) such as In Gagne et al. (2015), as the most effective method to consider the thermodynamic processes that characterize these EE, together with in-situ measurements, such as radar reflectivity or convective indices (Gagne et al. 2017; Czernecki et al. 2019). However, note that we have classified as severe weather different meteorological events, with specific peculiarities. For example, convective systems and hail storms are related events, quite local, in which thermodynamics and atmospheric state play an important role, very difficult to include as predictive variables in ML approaches. In extreme winds and gusts, however, the dynamics of the atmosphere may have significant importance to describe the phenomenon, and thus the synoptic situation provides information which may be exploited by ML algorithms (Spassiani and Mason 2021), in addition to other local atmospheric variables describing convective systems. It is also relevant the fact that in the last years, the number of DL-based techniques has increased a lot among the techniques applied to severe weather EEs, showing the

research line which will be followed in future applications and problems related to EEs related to severe weather.

3.5 Fog and extreme low-visibility

Low-visibility EE, usually associated with fog formation (Gultepe et al. 2007) or turbidity in the atmosphere due to pollution, deeply affect transportation facilities such as airports (Cornejo-Bueno et al. 2020; Guerreiro et al. 2020) and roads (Peng et al. 2018; Wu et al. 2018). ML algorithms have been successfully applied in the last years to many fog and low-visibility prediction problems.

In Marzban et al. (2007), a hybrid approach involving MLPs and NWM (mesoscale model) is proposed for a problem of ceiling and visibility prediction in the USA. A total of 20 meteorological variables are considered as inputs for the MLP, obtaining a good visibility prediction in 39 measurement stations of the North-West of USA. In Fabbian et al. (2007), MLPs were tested in a problem of fog events prediction at Canberra International Airport (Australia), from meteorological observations. Data from the Australian Bureau of Meteorology were used to train and test the neural networks, obtaining promising results. In Miao et al. (2012), a fog prediction system formed by fuzzy logic-based predictors was proposed and analyzed at Perth Airport (Australia). The fuzzy logic predictor worked on the outputs of mesoscale numerical model (LAPS125) outputs, with the objective of refining the predictions obtained by the numerical model. This fog prediction model was operational at the airport and its outcomes averaged with the outcomes of two other fog forecasting methods by means of a majority voting approach.

In Colabone et al. (2015), the performance of MLPs with back-propagation training procedure in a fog event prediction problem at Academia da Força Aérea (Brasil) is analyzed. In Boneh et al. (2015), a Bayesian network is applied to a fog prediction problem at Melbourne Airport. In this case, the problem is tackled as a prediction time horizon of 8 h, and 34 years of data have been used to train the network. This fog prediction system has obtained better results than previous systems, becoming operational for fog prediction at Melbourne Airport. In Bartoková et al. (2015), a decision tree for short-time fog prediction in Dubai is presented. The decision tree is able to improve the results of mesoscale models such as WRF in short-term prediction time horizons of up to 6 h. In Cornejo-Bueno et al. (2017), different ML regression techniques have been tested over a fog prediction problem at Valladolid airport, Spain. In this case, radiation-type fog events are the most common in the zone, so the prediction problem is restricted to winter months. The authors reported successful results in event prediction by using support vector regression algorithms and extreme learning machines approaches. In Zhu et al. (2017), a deep neural network

has been applied to a problem of low-visibility prediction at Urumqi airport, China. Meteorological variables measured at the airport between 2007 and 2016 are used to feed the deep neural network. In Durán-Rosal et al. (2018), evolutionary neural networks are considered for a problem of fog events classification from meteorological input variables. Several types of evolutionary neural networks are considered, by selecting different basic neuron types (sigmoidal, product and radial). A multi-objective training procedure is considered, obtaining good results in the fog event classification problem considered. In Guijo-Rubio et al. (2018), a problem of low-visibility events due to fog is tackled by applying ordinal classification methods. Three classes were considered (fog, mist and no-fog), and different ordinal classifiers were successfully tested in this problem of fog event prediction. In Dietz et al. (2019), decision trees models and tree-based ensemble with boosting are applied to a problem of very short-term prediction of low-visibility procedures states at Vienna airport, Austria. The work shows that for prediction time horizons under 1 h, the current low-visibility state (persistence), cloud ceiling, and horizontal visibility are the most important variables to take into account. For longer prediction time-horizons visibility information at the airport's surroundings and meteorological variables become relevant.

In Bari and Ouagabi (2020), different ML algorithms (tree-based ensembles, feed-forward neural networks and generalized linear methods) have been applied to the output of a NWM (mesoscale model, WRF), for a problem of low-visibility prediction in Northern Morocco. In Li et al. (2020), a decision tree algorithm (C4.5 approach) has been applied to a problem of low-visibility prediction at Nanjing city. The work has shown that in this case, the variables related to humidity and particle concentrations (relative humidity, PM10 and PM2.5) are the most important factors to obtain accurate predictions of visibility at Nanjing. Finally, in Yu et al. (2021), a hybrid approach mixing Extreme Gradient boosted and NWM has been applied to a problem of visibility prediction in Shanghai, China. A large number of predictive variables are considered such as air pollutants concentration, meteorological observations, aerosol optical depth data and satellite images. The proposed hybrid approach provides a more accurate visibility forecast for prediction time horizons of 24 and 48 h than LGBM algorithms and NWM on its own.

In Cornejo-Bueno et al. (2020), the persistence and ML prediction of low-visibility events is studied in Valladolid airport, Spain. The performance of binary classifiers is evaluated in a problem of radiation for prediction in winter. In Cornejo-Bueno et al. (2021), a problem of low-visibility events prediction due to orographic forcing is analyzed with ML regressors at Lugo, Northwestern Spain. The work includes the statistical analysis of the low-visibility events in this zone. In Castillo-Botón et al. (2022), a thorough comparison of several ML algorithms in fog prediction problems is carried out.

Both classification and regression techniques are analyzed, including balancing techniques and augmented data methods to improve the performance of ML in fog event prediction.

In close connection with low-visibility events, in this case, due to storms, in Ebrahimi-Khusfi et al. (2021), the number of dusty days is predicted with ML techniques in Northern Iran. SVR, RF and Stochastic Gradient boosting are the ML algorithms successfully applied to this problem. In Ding et al. (2022), the prediction of hourly low-visibility events is tackled in 47 Chinese airports, by means of different ML approaches such as MLP, RF, regression trees (CART) and KNN approaches, among others. The results obtained show important differences in performance from different airports, and also at different seasons (better performance in the cold season than in the warm season).

Finally, the application of DL-based techniques has been important recently. In Miao et al. (2020), a long-short term memory (LSTM) neural network has been applied to a problem of fog forecasting in the Anhui province, China. A comparison with K-Nearest Neighbours, AdaBoost and CNN algorithms has shown that the LSTM network is able to obtain better results. In Ortega et al. (2023), the performance of several DL models for visibility forecasting using time series climatological data are evaluated. Different DL models are considered, such as deep neural networks, CNNs and LSTMs. Results in data from two weather stations in Florida (USA) show a good performance of the DL algorithms. In Peláez-Rodríguez et al. (2023), several DL ensembles are discussed for a problem of low-visibility events prediction in Northern Spain (orographic fog). Recurrent neural networks, LSTM networks, Gated Recurrent Units and CNNs are the DL approaches considered in this ensemble approach. The performance of the ensemble was better than all the algorithms on their own, and it was also compared with alternative ML approaches, improving them in all cases. In Wang et al. (2022), a deep learning model implementing PCA and a deep belief network (DBN) is proposed for a problem of low-visibility events prediction. This approach was able to improve the results obtained by different ML and DL alternatives. In Zang et al. (2023), the RNN model is applied to a problem of low-visibility events prediction in Southern China. Comparisons with other DL-based algorithms including CNNs have shown a good performance of this DL-based method.

3.5.1 Analysis

ML analysis of fog events has been intense in the last few years. Fog formation may follow different physical mechanisms (Gultepe et al. 2007). For example, radiation fog, a typical fog of inland areas, usually occurs in winter under anticyclonic conditions, when clear skies and stability of the atmosphere allows the nocturnal radiative cooling required

to saturate the air (Román-Cascón et al. 2012). On the other hand, advection fog occurs when moist, warm air passes over a colder surface and is cooled from below, producing an immediate condensation of water. This kind of fog is very common at sea when moist and unstable warm air moves over cooler waters. If the moist warm air moves up to a hill or slope, the air undergoes an adiabatic expansion which, in turn, cools down the air as it rises, allowing the moisture in it to condense and this way producing fog, usually called orographic or hill fog. Note that the dissipation mechanisms and persistence of these fog events are also different depending on the formation process (Cornejo-Bueno et al. 2020; Salcedo-Sanz et al. 2021). The inclusion of physics in ML approaches should take into account these formation and dissipation mechanisms, depending on the type of fog event considered. The best way of taking into account this is to consider as inputs meteorological variables related to fog formation or dissipation, as in the majority of cases has been done. Also, there have been some works which have used NWM as a previous step before the application of ML algorithms, as in Marzban et al. (2007); Bari and Ouagabi (2020); Yu et al. (2021). Finally, note that the application of DL-based approaches has been very notable in the last years, with different works discussing DL techniques and DL-based ensembles for visibility prediction problems (Ortega et al. 2023; Peláez-Rodríguez et al. 2023).

3.6 Final discussion

As reviewed in previous sections, a large amount of ML algorithms have been applied to a wide class of problems in EEs detection, prediction and attribution. EEs problems in different spatiotemporal scales have been tackled with ML algorithms. In some cases, long-term physical processes related to atmospheric dynamics seem to be predominant (heatwaves, extreme temperatures, droughts and floods in some cases), while in other cases, local short-term processes associated with thermodynamics are the predominant factor of the problem (convective systems, flash floods or extreme fog events).

We have broadly detected three types of approaches using ML in the literature reviewed, in all EEs problems considered in this work. First, there are articles in which ML algorithms have been applied *raw*, i.e. without any reference to the physics related to the problem. Usually, these works proposed approaches based on time series of measured values or involved some signal processing techniques, such as series decomposition, wavelets, etc. In general, these approaches have been exclusively compared against other alternative methods fully based on ML or autoregressive approaches such as ARIMA methods, and a poor discussion on the physical reasons for the good or bad performance of the algorithms is carried out. A second type of approach described in the lit-

erature reviewed is those works which try to take into account the physics of the problem through the input variables considered in the ML methods. Depending on the problem considered, certain input variables may consider physical aspects of the problems, such as atmospheric dynamics (synoptic situations, Rossby waves, climate indices and other variables related to atmospheric dynamics) or thermodynamics processes (convective or stability indices, and other variables related to thermodynamics process, usually from reanalysis data, satellites or direct measurements). Finally, the third type of works revised in this section are those ML approaches which present hybridization with physical or numerical models considering the physics of the problem, or those which present a coupling with physical models in order to improve their outputs. Different versions including hybridization/coupling with numerical models such as WRF, Analogue-based algorithms, and other NWM have been revised in this section. In general, these latter hybrid approaches were successfully compared with physical models and also with other ML approaches. In some cases, future projections based on CMIP6 models have been carried out from ML approaches, in attribution-related problems.

It is also remarkable the fact that different problem encodings and frameworks have been used in the EEs problems revised. Classification and regression frameworks have been used, depending on the specific EE, at very different spatiotemporal scales, from local to synoptic and global scales, at short-term and long-term temporal scales. The number of input variables in ML algorithms is an important issue in many of the approaches revised. In many cases, FS mechanisms are needed in order to improve the results of ML algorithms. In general, the articles reviewed reported successful ML applications to EE, but the comparison with alternative approaches can be biased. For example, those approaches in which physics processes are not taken into account in the ML, are not usually compared to alternative approaches including physical models, but only with other ML methods. In those works in which ML methods have

been hybridized with NWM to include the physics of the EE, an improvement over the NWM has been reported. In many cases, this ML hybridization with NWM is focused on downscaling processes, in order to improve the spatial resolution of NWM, by using ML algorithms.

Finally, we have detected a clear increase of DL-based techniques in the last years, in all kinds of EEs detection and prediction and attribution problems. This trend is much more accused in 2020, and currently (2023) the large majority of works on EEs deal with DL-based techniques. It seems that this trend is unstoppable, due to the better results obtained with DL techniques, their flexibility and ease to work with spatiotemporal time series, better-covering problems in atmospheric EEs than traditional ML approaches.

4 Case study: summer temperature prediction with ML and DL approaches. Results and open problems

ML approaches devoted to characterising and predicting heatwaves and extreme temperatures have been previously discussed in this paper (Sect. 3.2). In this case study, different problem formulations are shown and discussed, also some results and issues related to summer temperature prediction, where heatwaves signals can be detected, based on reanalysis data for France. A final subsection shows an outlook, findings summary and open problems from this case study.

4.1 August mean temperature prediction in France based on ML approaches and synoptic predictive variables from reanalysis

In this first problem definition, the prediction of August mean temperature by using ML approaches is addressed. In order to give a first definition of the problem, a specific case of August mean temperature prediction in central France is considered, where there have been extremely hard summer

Fig. 11 Monthly averaged temperature in Paris, between 1950 and 2021. Mega-heatwave of 2003 is highlighted in the time series
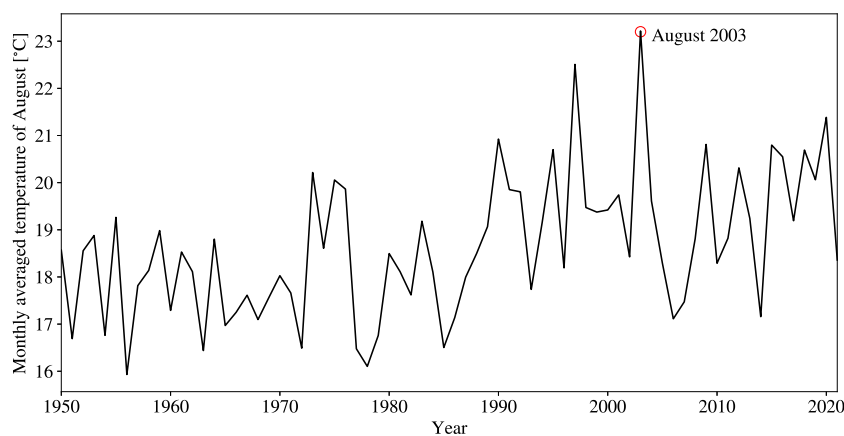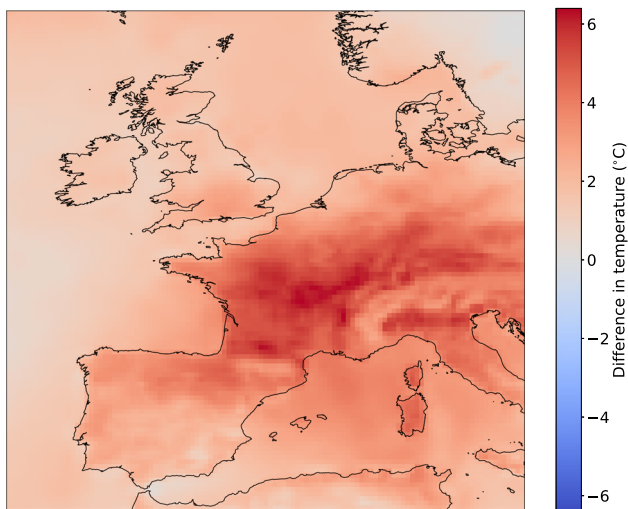
Fig. 12 Anomaly of monthly averaged temperature in August between 2003 and the averaged temperature from 1950 to 2002

heatwaves in the last 20 years (García-Herrera et al. 2010; Ouzeau et al. 2016; Barriopedro et al. 2011). Let $T(t)$ be an objective time series of air temperature (2 m temperature, for instance, or any other similar air temperature variable), obtained at a given point or averaged over a set of known points. In our case, $T(t)$ stands for the mean temperature of a summer month (August) in the location of interest. Air temperature from ERA5 reanalysis data (Hersbach et al. 2020) has been considered in this case, as there are previous works which confirm that reanalysis data can be successfully used in the prediction of extreme temperatures (You et al. 2013). Fig. 11 shows the objective August mean temperature (2 m temperature) in the Paris area (France) from 1950 to 2021. Note that in some cases it is possible to spot heatwave sig-
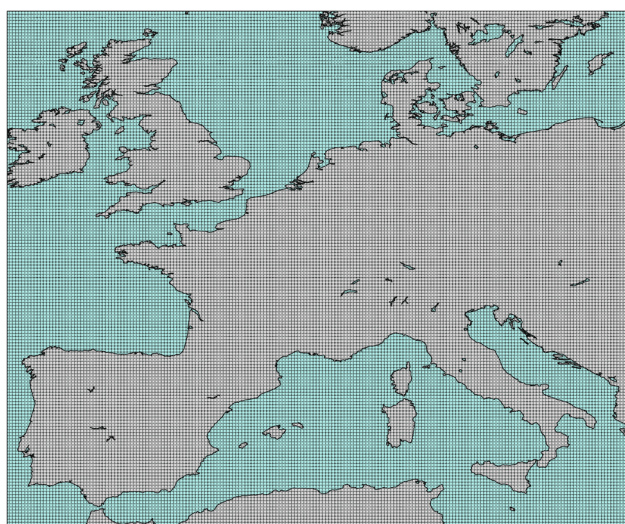
Fig. 13 Synoptic regular grid similar to the one considered in this case study

nal in $T(t)$, such as the mega-heatwave of August 2003 in Europe (Fig. 12).

Let $V(t', \mathbf{x})$ be the set of predictive variables, usually defined in a spatial regular grid $\mathbf{x}$, over time. Note that we have notated t' since it may not match with time t in $T(t)$. In this problem, we consider a synoptic regular grid (Fig. 13), covering France, where we define a number of predictive variables to estimate $T(t)$, also obtained from ERA5 reanalysis (Hersbach et al. 2020). Table 1 shows the predictive and target variables considered in this work.

We consider the problem of predicting the mean temperature of August $T(t)$ (regression problem), by using the value of the predictive variables in the previous months (t' stands for months of July/June, same year) in $V(t', x)$. This approach is similar to that in Oettli et al. (2022), but focused on the summer temperature. Different ML and DL techniques among those described in Sect. 2 are considered to tackle this problem. Specifically, RF, DT, MLP, SVR, LSTM networks, and different dimensionality reduction techniques have been evaluated. We have also included a Linear Regression approach for comparison purposes.

Several research questions arise here: for instance, we want to assess whether there is enough information from variables in $V(t', x)$ to obtain a good quality prediction of $T(t)$ from ML approaches. Regarding extreme values, we aim to know if the model is able to obtain a prediction mechanism which shows a good quality prediction of extreme temperature values, with a prediction time-horizon of one month in advance. Also, the problem of obtaining the best set of features (dimensionality reduction) for the ML algorithms arises here. In order to solve these research questions, we will show different results and we will discuss different open problems found when dealing with this case study.

4.2 Experimental results and research issues

We have structured the results obtained in several subsections, where the results are discussed by considering different input variables from one single reanalysis node (local approach), results from several reanalysis nodes (syn-

Table 1 Predictive and target variables considered in this case study, as obtained by ECMWF (2022)

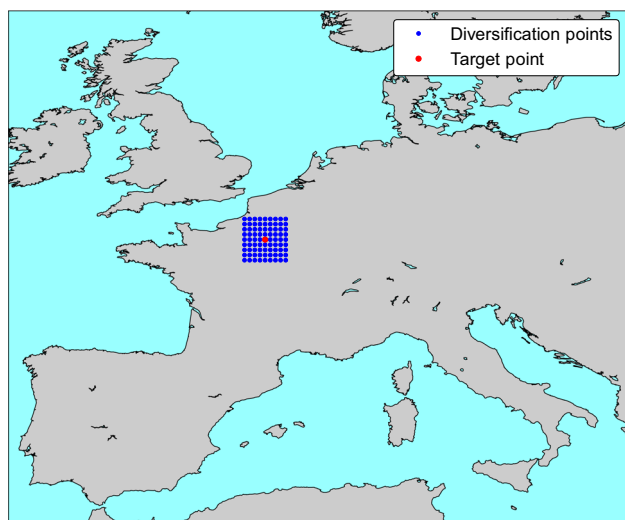| Short name in ERA5	Description
u_{10}	Eastward component of wind at a height of 10 m
v_{10}	Northward component of wind at a height of 10 m
msl	Mean sea level pressure
$swvl_1$	Volumetric soil water at layer 1 (0-7 cm)
t_{2m}	Air temperature at 2 m (target)

Fig. 14 Mean August temperature prediction problem considered. Reanalysis nodes for predictive (blue) and objective variables (red)

optic approach), issues regarding the prediction problems, mainly the number of training samples available, and how to solve them by including new training samples with oversampling approaches. In addition, the feature selection method is shown, whilst a DL approach has been studied.

4.2.1 Input variables from one reanalysis node

A simple regression problem is addressed. In this case, a single node reanalysis field is considered foe extracting the predictor variables. The target node is considered, as above mentioned, in France. From the same point different predictor variables are considered, with the aim of predicting the target (August temperature). Figure 14 shows the considered node, in red. In order to tackle the problem, we first consider a training and test partition of the data. The available data is obtained from 1950 up to 2021. The period 1950–2002 is considered for training, whilst the period 2003–2021 is considered as the test set to evaluate the results. Note that, annual data is considered, thus, only 53 samples are available for training the algorithm, whilst 19 test samples where we

Table 2 MAE and MSE of ML regressors considered in the temperature prediction problem by reanalysis spatial diversity

Method	Single point MAE	MSE
LR	1.34	2.94
RF	1.43	3.25
DT	1.88	4.71
MLP	2.45	9.30
SVR	1.52	3.55

can evaluate the skill of the model. Table 2 (first column) shows the MAE obtained by the different ML algorithms for this simple first case, and Fig. 15 details the predictions obtained by each ML algorithm. As can be seen, the prediction obtained by the ML algorithms is in general not fully accurate. It should be highlighted the different skills shown by the ML models. It can be observed that MLP is the worst approach in this case, with a MAE of 2.45. It is followed by DT, with a MAE of 1.88. In this problem, LR, RF and SVR show better than MLP and DT, with MAE values of 1.34, 1.43 and 1.52, respectively. It can be concluded that the database for training the algorithm is not large enough. It seems that further data is needed to reach a better performance of the models.

4.2.2 Exploiting spatial diversity of reanalysis data to improve ML accuracy

The question that arises at this point is, can we generate additional training samples with the aim of improving the skill of the prediction model? A simple strategy is shown in this section. It allows for increasing the number of training samples by exploiting the spatial diversity of the reanalysis data.

Let us return to the problem tackled above, with a single reanalysis node considered, and 5 predictive variables. If we consider a local approach, note that there are a large number of reanalysis nodes in the neighbourhood of the selected one. In Fig. 14, we have set a number of neighbour reanalysis nodes in blue (81 nodes), around the red point. It is important to note that we have all the predictive (input) and objective (T) variables in all the points considered. Since we are in a local approach, we can assume a similar behaviour of the variables in the selected grid, in such a way that we can use all the variables in the grid as training samples. Surrounding grid points can be considered in the training data set. Thus, an oversampling approach is introduced (Torgo et al. 2015), by exploiting the diversity of reanalysis in a local approach. In this particular case, we finally obtain 4293 training samples (81×53) instead of the initial 53 samples.

Table 3 and Fig. 16 show the new results when a reanalysis of spatial diversity is included to generate oversampling. As it is shown, a better performance of the prediction capability of the different ML models is obtained. In this scenario, the best improvement is for DT (MAE 1.88 → MAE 1.02), which achieves an accurate prediction. The performance of the MLP model is also improved with the oversampling approach by using reanalysis spatial diversity (from MAE 2.45 to MAE 1.54), and the SVR is also improved in this case (from MAE 1.52 to MAE 1.36). The LR and RF do not improve their result when oversampling by reanalysis diversity is considered, but the performance deterioration is not very accused.

Fig. 15 Temperature prediction considering variables from a single reanalysis node
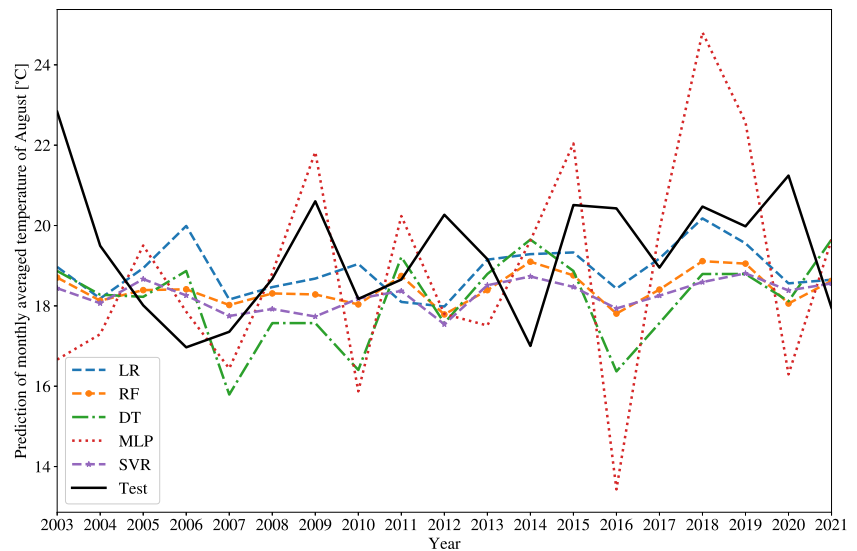
In this way, it is shown that the oversampling approach, by considering reanalysis of spatial diversity, is able to improve the performance of the ML regressors in the temperature prediction problem considered.

4.2.3 Extension to several input reanalysis nodes

Let us consider a second problem, with several reanalysis nodes to carry out the prediction of the heatwaves. We show a case with four reanalysis nodes in Fig. 17.

Note that, in this case, we consider 5 variables per node of reanalysis. The addition of a node implies 5 more predictive variables to the data set. Thus, a total of 20 predictive (input) variables are now considered in the problem, with 53 training samples in this case. It is expected that increasing the number of input variables with just 53 training samples does not lead to better results. Table 4 shows the results obtained with all the ML considered. As can be seen, the prediction of $T(t)$, in general terms, the prediction skill of the models is not better than the case in which a single reanalysis point is considered.

The oversampling can also be introduced when several reanalysis nodes are considered. For that purpose, the spatial

diversity of the ERA5 data is exploited. The diversification points are shown in Fig. 17. Note that, for each reanalysis node, we can generate diversity by randomly selecting a neighbour node in each one. This way we can exploit the fact that the neighbour reanalysis nodes provide similar predictive variables or target values, and we can generate a large number of new training samples. Table 5 shows the results obtained by including oversampling by reanalysis of spatial diversification. As can be seen, the LR improves a lot its result, and the rest of ML algorithms seem to be slightly affected by diversification in this case, obtaining slightly worse results in general. Figure 18 shows the results obtained in the test set, which are, as can be seen, worse than those obtained by considering a single reanalysis node with oversampling.

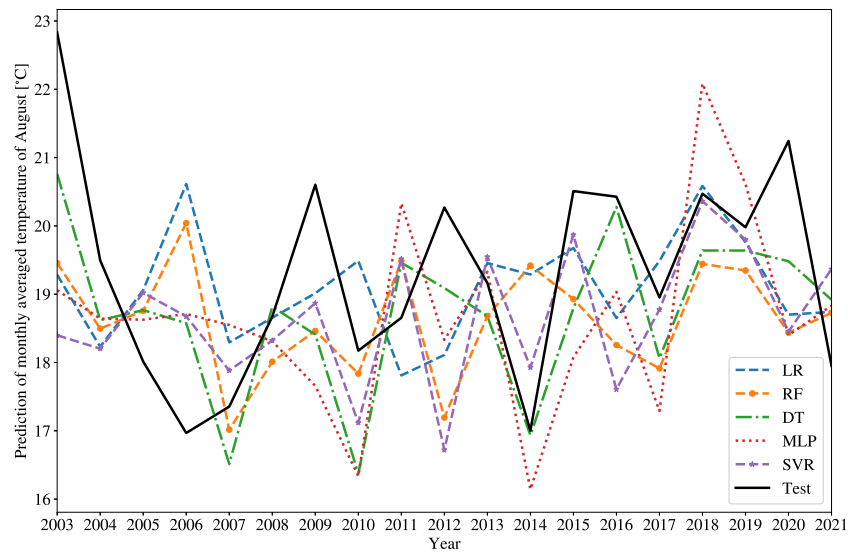### 4.2.4 ML-based oversampling and undersampling approaches

In ML, an *oversampling procedure* consists of increasing the number of observations by generating new data samples, in order to improve the performance of the training algorithms. In a classification problem, it is common the use of oversampling techniques in unbalanced data set problems or in small data sets. There are different oversampling techniques. For a classification task, the most commonly used algorithm is the SMOTE algorithm (Chawla et al. 2002). It creates new samples taking into account the statistics of existing ones, diminishing the risk of creating samples in "wrong" areas. For the regression problems, similar techniques can be encountered, such as the SMOGN algorithm (Branco et al. 2017).

In contrast, the undersampling methods decrease the number of samples. This technique is commonly used in problems

Table 3 MAE and MSE of ML regressors considered in the mean August temperature prediction problem in France, with and without oversampling by reanalysis spatial diversity

Method	Oversampling MAE	MSE
LR	1.35	2.93
RF	1.50	3.26
DT	1.02	1.46
MLP	1.54	3.25
SVR	1.36	3.28

Fig. 16 Temperature prediction considering oversampling by reanalysis spatial diversity (one initial reanalysis node)

with unbalanced data sets, with the aim of reducing the majority class.

In order to test ML-based oversampling approaches, it is addressed a case in which four reanalysis nodes are considered, Fig. 19. The SMOGN algorithm is used to generate ML-based oversampling in the problem. Tables 6 and 7 show the results when the oversampling with SMOGN is and is not considered. As can be seen, the performance of all tested regression models is improved by considering oversampling, but the LR, for which the results are worst when oversampling is considered, Fig. 20.

The first DL-based approach consists of a combination of two different models. The first one is a VAE model. As has been explained above, this type of DL model infer from historical data, by using unlabelled data. In our case, the model is fed with the variables that may drive the event under study (extreme tempgeopotential height at 500 hPa (Z_{500}), the sea surface temperature (sst) and the t_{2m}. Thus, the input data is composed of three channels, each per variable. The variables, periods and regions in which the variables are of considerations. In this scenario, we just focus on the model, but not on the selection of the variables, regions and lag times. Once the VAE model is trained, the encoder part of the model can be used for encoding the input data. The intermediate representation of the data in the VAE may have a lower dimension than the original data. Thus, a reduction in the dimensionality of the data is done. This latent space can be used as the input of the second model. In this case, a MLP is considered, Fig. 21. The prediction of the temperature is made by this model, which uses the latent space as the input data whilst the labelled target data (tempera-

Fig. 17 Temperature prediction problem with input variables from 4 reanalysis nodes

Table 4 MAE and MSE from ML regressors considered in the temperature prediction problem with four input reanalysis nodes

Method	4 Reanalysis nodes	
	MAE	MSE
LR	2.27	6.60
RF	1.44	3.04
DT	1.73	3.90
MLP	2.07	7.45
SVR	1.49	3.30

Table 5 MAE and MSE of ML regressors considered in the mean August temperature prediction problem in France, with and without oversampling by reanalysis spatial diversity

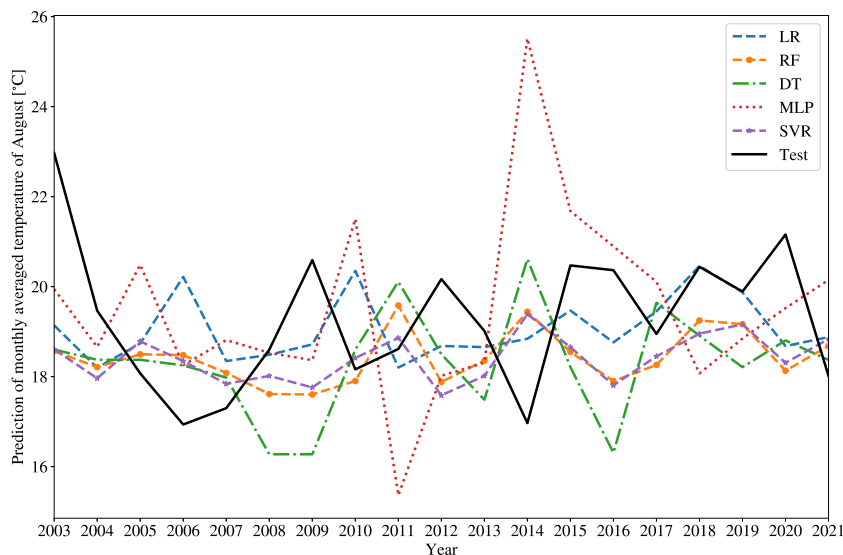| Method	Oversampling(4 nodes)	
	MAE	MSE
LR	1.30	2.82
RF	1.56	3.61
DT	1.90	5.27
MLP	2.07	7.45
SVR	1.53	3.56

Fig. 19 Mean August temperature prediction problem in France, with input variables from 4 reanalysis nodes in the case in which SMOGN method is considered

ture) is for training. It is important to note that two different training processes are developed since the MLP model is not trained until the VAE has been trained. This approach is able to achieve significant results, which are comparable with persistence (operator $x(t) = x(t - T)$) and climatology (operator $x(t) = \frac{1}{N} \sum_{j=1}^{N} x(t - j)$) of the zone. Figure 22 shows the results obtained by the VAE-MLP compared to persistence and climatology. As can be seen, the proposed hybrid VAE-MLP is able to improve both persistence and climatology in this problem, obtaining more accurate results with respect to the ground truth (average weekly temperature). The differences between the VAE-MLP and the persistence are important, and the improvement is more significant for larger prediction-time horizons, as expected. The comparison with the climatology of the zone highlights fewer differences. In general, the VAE-MLP is able to improve the climatology in the cases is the smaller prediction time horizon (1 and 2 weeks in advance); however, in the cases of 3 and 4 weeks in advance prediction, the performance of the VAE-MLP is very similar to that of climatology, though still better than persistence.

There are other alternatives for the feature selection. For example, the wrapper approach (see Sect. 2.1) can be applied to estimate weekly temperature in France, using an evolutionary algorithm for the searching process, together with a fast-training ML approach (ELM), as described in Sect. 2.1. We can include different improvements in this scheme, by considering a previous spatial clustering in the problem. In this way, the evolutionary algorithm must select a variable from each cluster, including a further dimensionality reduction in the process. Figure 23 shows an example of this in the problem of temperature prediction in France. The coloured squares represent different zones which the algorithm must select variables from. This way it is possible to restrict the

Fig. 18 General case with diversification

Table 6 MAE and MSE of ML regressors considered in the temperature prediction problem without and with SMOGN technique applied on the database by reanalysis spatial diversity

Method	4 Reanalysis nodes	
	MAE	MSE
LR	2.13	6.95
RF	1.38	2.86
DT	1.77	4.24
MLP	2.01	5.67
SVR	1.51	3.30

zones to different sizes (synoptic, global), so they describe dynamic processes of different temporal scales.

4.3 Case study outlook, findings, and open problems

In this case study, we have discussed the application of ML algorithms to a problem of mean temperature prediction in August from reanalysis data in France. We have defined the problem in this way in order to extend it to the prediction of a heatwave in France when smaller spatiotemporal scales are considered. In fact, even at a monthly scale, a heatwave signal can be detected in August mean temperature in some cases of meta-heatwaves, as that of 2003 in France (García-Herrera et al. 2010). In addition, we have observed the following issues from the application of the ML algorithms:

- Since the problem definition involves annual samples (temperature in August), the training set has very small number of samples. This point, combined with the fact that we have a large grid with a large number of predictive variables on it, makes the training of the ML an extremely hard task.
- The results obtained in a first problem considering a single reanalysis node, are far from accurate, due to the scarce number of training samples.

Table 7 MAE and MSE of ML regressors considered in the mean August temperature prediction problem in France, without and with SMOGN technique applied on the database by reanalysis spatial diversity

Method	SMOGN applied	
	MAE	MSE
LR	4.67	28.40
RF	1.32	2.71
DT	1.40	3.23
MLP	2.01	5.67
SVR	1.35	2.67

- In order to improve the performance of the ML approaches, we propose to exploit the spatial diversity of the reanalysis data considered. First, we consider a fully local approach, including in the training set a number of neighbour reanalysis nodes to the objective node to generate new training samples. This oversampling approach generates new training samples, which allows a better training of the ML algorithms, improving the results obtained in the prediction of August mean temperature.
- In a second attempt, we consider several reanalysis nodes to make the prediction and oversampling by exploiting local diversity in each node. The prediction obtained in these two cases by the ML algorithms is poorer than in the previous cases, since the number of predictive variables is increased, and much more training samples would be necessary to improve the results.
- We have also shown the performance of ML algorithms in this problem, by considering ML-based oversampling by applying the SMOGN algorithm. SMOGN is especially suited for regression problems. In this problem of August mean temperature prediction in France the SMOGN works fine, producing oversampling which improves the performance of all ML algorithms versus the case without oversampling.
- Thus, we have shown that considering oversampling to expand the training set is a good option in this prediction problem with a scarce number of data. We have proposed a novel oversampling approach by exploiting the spatial diversity of Reanalysis data, and we have also shown that ML-based oversampling also works in the problem.
- We have finally given a note on the possible application of dimensionality reduction techniques, using a hybrid DL-based approach and a wrapper feature selection approach. We have shown that the hybrid DL-based algorithm formed by an AE with a MLP is able to improve the persistence and climatology of the zone when the prediction time horizon is up to 2 weeks in advance, and it works similarly to the climatology for 3 and 4 weeks in advance prediction time horizon. We have also outlined the introduction of a wrapper ML approach for feature selection in the problem, with a further dimensionality reduction using a previous spatial clustering. This approach gives the possibility of choosing different spatial scales for the predictive variables in the problem.

There are several open problems in the prediction of annual mean air temperature from reanalysis data. We summarize them in the following points:

- We have tackled local and synoptic versions of the problem from reanalysis data, with predictive variables back to just one month before. However, it is known that heat-

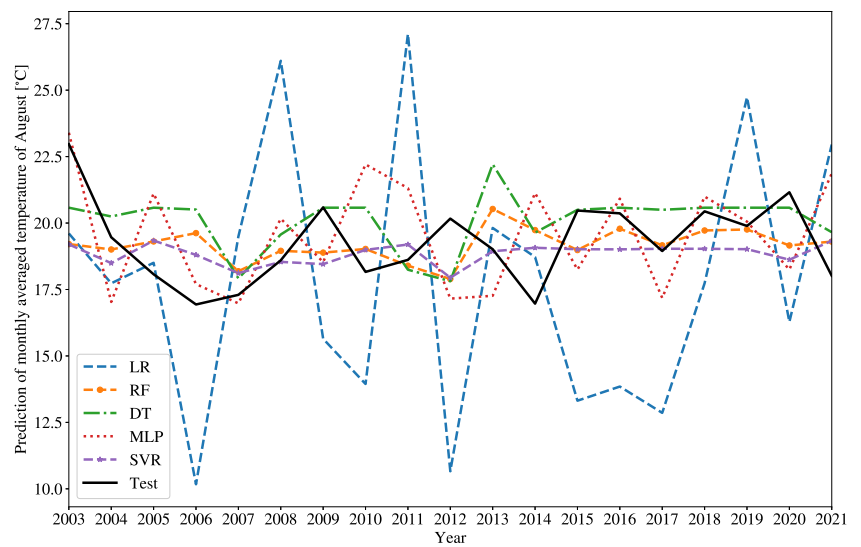**Fig. 20** Temperature prediction by applying SMOGN oversampling method

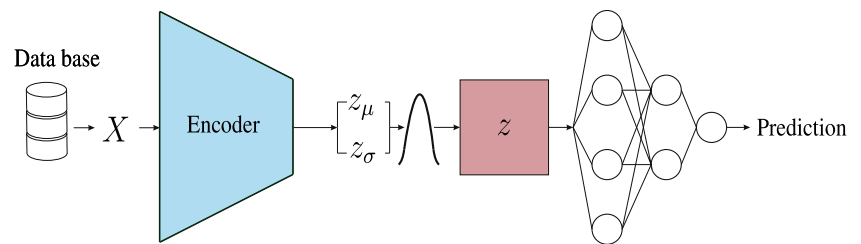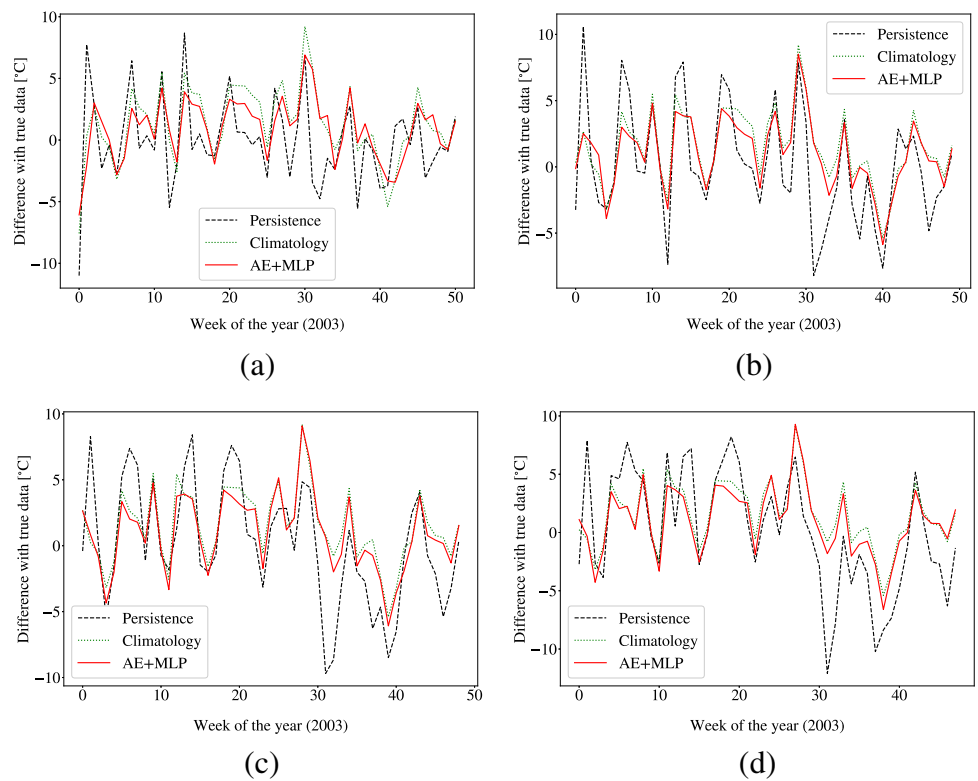**Fig. 21** VAE Structure and MLP

Fig. 22 Error DL model (VAE + MLP) **a** prediction time-horizon 1 week; **b** prediction time-horizon 2 weeks; **c** prediction time-horizon 3 weeks; and **d** prediction time-horizon 4 weeks
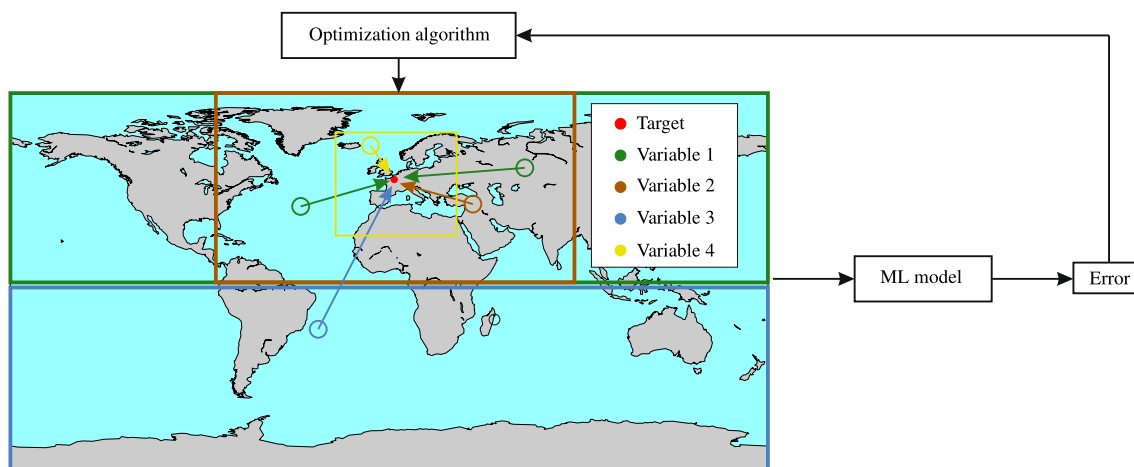

(a)

(b)

(c)

(d)

Fig. 23 Example of wrapper feature selection with previous spatial clustering for the problem of temperature prediction in France

waves detection (signal in mean monthly temperature) may have different drivers, some of them related to climate indices, which points out to a global definition of variables, with time-horizon for these predictive variables back to several months before. In this global definition of the problem, the management of the huge number of features involved will be extremely important to obtain significant results. Also, the generation of enough training samples for the ML algorithms is again a challenging aspect of the global version of the problem.

- Note that there are different possible definitions of this prediction problem, depending on the data considered. We have shown an example with monthly temperature data, but quarterly, weekly or even daily time precision can be chosen and will also contain heatwave signals. It is also possible to directly use heatwaves indices (Awasthi et al. 2022; Nairn and Fawcett 2015) to define the problem, which have been proposed in the past, including other variables in addition to temperature.

- In close relationship with the latter point, note that the problem can be tackled as a regression or classification problem. We have shown here a regression version, where the direct prediction of $T(t)$ is tackled. In a classification problem, $T(t) \rightarrow s[n]$, where $s[n] \in \{0, 1\}$ if we consider a binary classification problem (heatwave signal detected/no heatwave signal). This problem can be extended to a larger number of classes.

- We have shown how the problem cannot be successfully tackled without considering the physics of the phenomenon. In other words, the ML approaches must be coupled with the physics of extreme temperatures, which act at different levels and considering different physical aspects of the problems, such as atmospheric dynamics and thermodynamics processes in order to improve the quality of the prediction.

- There are also open problems related to the ML algorithms. As previously mentioned, it is clear that feature selection (see Sect. 2.1) is key for ML approaches to obtain significant results in the different versions of the problem. Due to the huge number of features involved in the problem, it is probable that wrapper methods on their own do not lead to good results, and a first feature discarding process based on filter approaches is needed. Other possible solutions such as using clustering approaches to reduce the number of features in some specific zones can also provide good results when the number of features is huge, such in the global approach to the problem. We have outlined this possibility in the experimental section of the case study.

- Deep learning (DL) approaches could be used to tackle the problem without taken special care about the huge number of features involved. We have shown a possible DL approach using an AE hybridized with a ML, but other DL schemes are possible. Specifically, in this approach, DL algorithms could be useful to exploit global information and obtain an accurate prediction of heatwaves. Issues related to DL training, such as the number of training samples, and significance of the results obtained are the counterpart of this possible approach with DL algorithms.

5 Conclusions and perspectives for future research

5.1 Conclusions

In this paper, we have carried out a review of ML methods in the analysis, characterization, prediction and attribution

of extreme atmospheric events (EEs). It is currently a hot topic, since EEs are increasing in the current situation of climate change, causing important damages to human and ecosystems. After a brief review of the main ML approaches which have been previously applied to EE-related problems, we have carried out a comprehensive and critical analysis of this topic in the literature, including the main EE, such as extreme rainfall and floods, heatwaves and extreme temperatures, droughts, fog and low-visibility events, and different topics related to severe weather (convective systems, tropical cyclones, hailstorms and extreme winds).

We have shown the application of several ML methods to a case study related to mean summer temperatures prediction in France, from reanalysis (ERA5) data. We have shown the main issues related to this problem using ML, including the scarce number of samples to train the ML approaches, the huge number of input variables and the different possible problem's definitions (regression or classification tasks, prediction time-horizon considered, etc.). We have also shown that the inclusion of the physics is a key point in order to obtain good results for this problem, so it is necessary to couple the ML algorithms with some physical information for the problem in order to improve the results obtained.

Note that these issues associated with the case study considered in this paper can be extrapolated to other similar problems in extreme atmospheric events, which share similar data structure and scarce of events and data. We have also given some solutions to these issues for the case study considered, such as including oversampling techniques from reanalysis diversity, or even using different reanalysis data or global climate models to generate new training samples for the ML algorithms. These proposed solutions can also be applied to other problems related to extreme atmospheric events.

5.2 Perspectives

We also discuss here some final lessons learned, open problems and research possibilities and direction which are currently an option for dealing with EEs using ML algorithms, such as the use of XAI techniques, improving the attribution of EEs with ML techniques, and improving the study of concurrent and compound events, where the lack of data to train the ML algorithms is even more pressing.

- One of the main issue when dealing with ML approaches to EEs prediction problems are the databases. Given the rarity of EEs, there are very few long-enough databases which provide reliable data for EE-Related studies. Even reanalysis data, with more than 70 years of data worldwide with high spatial accuracy may be not enough for some problems (the case study presented before is a

good example of this). In theses cases, oversampling data may be of great help to improve the performance of ML algorithms. Note that only by considering two different reanalysis data (ERA5 and ERA20C, for example (Salcedo-Sanz et al. 2020)), we can duplicate the number of samples in the training set, by considering the output of each reanalysis in the same nodes. This opens the possibility to use climate models (with different parameterizations) to multiply the number of training samples available. Another interesting possibility is the application of different oversampling techniques to increase the number of training samples in a given database. In the case of reanalysis-type data, or data defined in a regular grid, oversampling can be carried out in a natural way by considering neighbor nodes, or with tailor-made techniques depending on the specific problem considered. Yet, the use of model-based data (either reanalysis or climate models' simulations) could potentially limit the ability of ML methods of learning relationships outside the ones already implemented in the model. Moreover, training a ML algorithm on model-based data could overestimate the performance when tested against observational data as model-based data do not perfectly reproduce the real climatic conditions due to modelling errors and assumptions (Hoffmann et al. 2020; Matsuoka 2022). We, therefore, advocate making the most of observational data as they represent a richer ground truth, although sometimes characterized by low data quality and missing values. Here, however, ML can also contribute with advanced methodologies to reconstruct missing climate information (Kadow et al. 2020).

- Another niche of opportunity in the characterization of EEs is the use of explainable AI techniques to gain an informed understanding of the correlations modelled by ML models (Arrieta et al. 2020). Indeed, a large fraction of the ML models used nowadays in this area relies on complex structures and processing units (e.g. deep neural networks) that achieve unrivalled levels of performance at the cost of opaque training and inference processes. This clashes with other models which, by virtue of their transparent internal structure or the way they are trained, elicit interpretable information about what features are relevant for the target at hand (e.g. tree-based bagging ensembles or linear regression). When this interpretability is not provided off-the-shelf, explanations can be generated ad-hoc for already trained models producing, as a result, visualizations, quantitative scores of predictive relevance or alternative what-if hypothesis for the model's input, to mention a few (Montavon et al. 2018). This growing concern with explanatory techniques for ML models has spawned a whole area of research coined as eXplainable Artificial Intelligence (XAI), becoming a topic of central

importance in applied machine learning in almost any discipline.

Very recently, such techniques have started to be explored for extreme events prediction, as early as 2022. This is the case of van Straaten et al. (2022), where XAI was used to verify that a ML model learned to predict high summer temperatures from multiple predictors at different time scales agrees with the theoretical understanding of the underlying physical processes. However, there still prevail several challenges that, in our vision, should congregate the efforts of the community in years to come. Among them, we highlight two differential research directions:

1. The need for stepping beyond correlation-based ML towards data-based causality inference (Peters et al. 2017). Since the goal of decision-making is to avoid – or at least, minimize the consequences of – extreme events, data-based models should guarantee the actionability of the model's input to steer the predicted output in one direction or another. Such interventional tools are being actively investigated nowadays in the context of ML, with models ensuring input–output causality still far from their maturity (see Runge et al. (2019) and references therein) because they often require the introduction of several assumptions (e.g. Gaussian distributions) that might be violated by the processes associated with EEs. At the same time, fewer assumptions are required for identifying the absence of a causal link (Runge 2018), making the findings of non-causality already quite robust in determining when it is unlikely that a cause-effect physical mechanism exists. We expect the use of data-based causality inference to become more and more attractive for supporting the trustability of black-box ML models (Reichstein et al. 2019).

2. The inherent uncertainty of the physical world and the atmosphere propagates to the output of the models devised to characterize extreme phenomena occurring therein. Thereby, a remarkable corpus of literature has striven towards quantifying the confidence of the model in its output considering the modelling (epistemic) uncertainty and the irreducible (aleatoric) uncertainty. While confidence analysis is a well-established area in ML research, the combination of confidence and explainability in a single framework is still to be seen. Indeed, explanations of uncertain models make no practical sense, nor do models that are certain about their predictions but do not explain what they model in the data at hand. The variability and incompleteness of atmospheric data, and the large epistemic uncertainty of deep learning models can, without no doubt, leverage advances such as evidential DL, variational neural networks or model-agnostic techniques such as conformal prediction. Confidence estimations provided by these techniques should be considered when furnishing explanations.

On a summarizing note, we advocate for a focus of the research community steered towards modelling aspects that complement the derivation of more models and performance comparison studies. In other words, we advocate for ML approaches at the end of a pipeline driven by physics, in this review, we have shown very different examples which show that the application of ML techniques without including the physical basis of the problem does not lead to relevant results in the majority of cases.

- Improving attribution of EEs using ML. There are not many works dealing with the attribution of EEs using ML techniques. In this work, we have discussed some works dealing with the attribution of EEs using ML techniques for specific events of heatwaves (Pasini et al. 2017; Zaninelli et al. 2023) and droughts (Richman and Leslie 2018, 2020). There are some recent works dealing with ML in general climate attribution problems (Mamalakis et al. 2022; Trifunov et al. 2021), and also on specific attribution of forced climate change signals over atmospheric fields such as global temperature or precipitation (Barnes et al. 2019, 2020; Hartigan et al. 2020a, b). In Callaghan et al. (2021), a large study on the attribution of climate impacts with ML methods has been recently carried out. However, it is necessary to extend these works to better cope with the attribution of EEs by using ML approaches. The application of novel ML/DL approaches specifically to attribution problems is another line to follow in the years to come. The study of causal inference with ML (Schölkopf 2022) is also a topic fully related to attribution, in which there are some recent works focused on extreme atmospheric events (Nethery et al. 2021; Liu et al. 2021).

- ML for concurrent and compound EEs. The concept of *concurrent event* refers to (atmospheric) EEs of different types occurring within a specific temporal lag, either in different locations or at the same one. This concept can also be used for extremes of the same type occurring in two locations within a specific period (Toreti et al. 2019). On the other hand, *compound events* refer to concomitant (within a given temporal lag) occurrence of events (extremes or not) with severe and harmful consequences of socio-economic relevance. It is possible to see that concurrent events are a subset of compound events. In spite of the work on concurrent and compound events has been intense in the last years (Bresch et al. 2018; Zscheischler et al. 2020; White et al. 2021; Markonis et al. 2021), the

application of ML techniques to prediction or attribution of concurrent or compound events has been minor. There is a very recent work discussing ML techniques applicable to compound events together with statistical and numerical techniques (Zhang et al. 2021), and some white papers and technical reports on the topic (Feng et al. 2021), but in general the application of ML to this topic is an open problem. The most important issue with ML approaches in concurrent and compound EEs is related to the lack of available data to study these types of situations. There have been some intents to generate databases for concurrent and compound events (Feng et al. 2020), but in general further efforts are needed to strengthen this topic, so ML methods can be successfully applied in this area.

Code availability The authors agreed to publish the code in a public GitHub repository.

Declarations

Ethics approval The manuscript is conducted within the ethical manner advised by the Theoretical and Applied Climatology. Permissions or licenses were obtained.

Consent to participate Not applicable.

Consent for publication Authors agree to publish the paper.

Conflict of interest The authors declare no competing interests.

References

Abbes AB, Inoubli R, Rhif M, Farah IR (2023) Combining deep learning methods and multi-resolution analysis for drought forecasting modeling. Earth Sci Inform 1–10

Abdel-Aal R, Elhadidy M (1995) Modeling and forecasting the daily maximum temperature using abductive machine learning. Weather Forecast 10(2):310–325

Abdi H, Williams LJ (2010) Principal component analysis. Wiley Interdiscip Rev Comput Stat 2(4):433–459

Ackerman F (2017) Worst-case economics: extreme events in climate and finance. Anthem Press

Adikari KE, Shrestha S, Ratnayake DT, Budhathoki A, Mohanasundaram S, Dailey MN (2021) Evaluation of artificial intelligence models for flood and drought forecasting in arid and tropical regions. Environ Model Softw 144:105136

Aghelpour P, Mohammadi B, Biazar SM, Kisi O, Sourmirinezhad Z (2020) A theoretical approach for forecasting different types of drought simultaneously, using entropy theory and machine-learning methods. ISPRS Int J Geo Inf 9(12):701

Ahmed K, Sachindra D, Shahid S, Iqbal Z, Nawaz N, Khan N (2020) Multi-model ensemble predictions of precipitation and temperature using machine learning algorithms. Atmos Res 236:104806

Anthony LFW, Kanding B, Selvan R (2020) Carbontracker: tracking and predicting the carbon footprint of training deep learning models. Preprint at http://arxiv.org/abs/2007.03051

Ardabili S, Mosavi A, Dehghani M, Várkonyi-Kóczy AR (2019) Deep learning and machine learning in hydrological processes climate change and Earth systems a systematic review. In: International Conference on Global Research and Education. Springer, pp 52–62

Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, García S, Gil-López S, Molina D, Benjamins R, Chatila R, Herrera F (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115

Arul M, Kareem A, Burlando M, Solari G (2022) Machine learning based automated identification of thunderstorms from anemometric records using shapelet transform. J Wind Eng Ind Aerodyn 220:104856

Ascenso G, Cavicchia L, Scoccimarro E, Castelletti A (2023) Optimisation-based refinement of genesis indices for tropical cyclones. Environ Res Commun

Asthana T, Krim H, Sun X, Roheda S, Xie L (2021) Atlantic hurricane activity prediction: a machine learning approach. Atmosphere 12(4):455

Awasthi A, Vishwakarma K, Pattnayak KC (2022) Retrospection of heatwave and heat index. Theoret Appl Climatol 147(1):589–604

Bador M, Terray L, Boe J, Somot S, Alias A, Gibelin A-L, Dubuisson B (2017) Future summer mega-heatwave and record-breaking temperatures in a warmer France climate. Environ Res Lett 12(7):074025

Badrinath A, Delle Monache L, Hayatbini N, Chapman W, Cannon F, Ralph M (2023) Improving precipitation forecasts with convolutional neural networks. Weather Forecast 38(2):291–306

Baki H, Chinta S, Balaji C, Srinivasan B (2021) Determining the sensitive parameters of WRF model for the prediction of tropical

cyclones in the Bay of Bengal using global sensitivity analysis and machine learning. Geosci Model Dev Discuss 1–46

Balakrishnama S, Ganapathiraju A (1998) Linear discriminant analysis-a brief tutorial. Institute for Signal and Information Processing 18(1998):1–8

Bari D, Ouagabi A (2020) Machine-learning regression applied to diagnose horizontal visibility from mesoscale nwp model forecasts. SN Appl Sci 2(4):1–13

Barnes EA, Hurrell JW, Ebert-Uphoff I, Anderson C, Anderson D (2019) Viewing forced climate patterns through an ai lens. Geophys Res Lett 46(22):13389–13398

Barnes EA, Toms B, Hurrell JW, Ebert-Uphoff I, Anderson C, Anderson D (2020) Indicator patterns of forced change learned by an artificial neural network. J Adv Model Earth Syst 12(9):2020–002195

Barnes AP, McCullen N, Kjeldsen TR (2023) Forecasting seasonal to sub-seasonal rainfall in Great Britain using convolutional-neural networks. Theoret Appl Climatol 151(1–2):421–432

Barriopedro D, Fischer EM, Luterbacher J, Trigo RM, García-Herrera R (2011) The hot summer of 2010: redrawing the temperature record map of Europe. Science 332(6026):220–224

Barriopedro D, García–Herrera R, Ordóñez C, Miralles D, Salcedo–Sanz S (2023) Heat waves: physical understanding and scientific challenges. Rev Geophys 2022–000780

Bartoková I, Bott A, Bartok J, Gera M (2015) Fog prediction for road traffic safety in a coastal desert region: improvement of nowcasting skills by the machine-learning approach. Bound-Layer Meteorol 157(3):501–516

Belayneh A, Adamowski J (2013) Drought forecasting using new machine learning methods. J Water Land Dev 18:3–12

Belayneh A, Adamowski J, Khalil B, Ozga-Zielinski B (2014) Long-term spi drought forecasting in the Awash river basin in Ethiopia using wavelet neural network and wavelet support vector regression models. J Hydrol 508:418–429

Belayneh A, Adamowski J, Khalil B (2016) Short-term SPI drought forecasting in the Awash river basin in Ethiopia using wavelet transforms and machine learning methods. Sustainable Water Resources Management 2(1):87–101

Belayneh A, Adamowski J, Khalil B, Quilty J (2016) Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. Atmos Res 172:37–47

Berghuijs WR, Aalbers EE, Larsen JR, Trancoso R, Woods RA (2017) Recent changes in extreme floods across multiple continents. Environ Res Lett 12(11):114035

Bishop CM et al (1995) Neural networks for pattern recognition. Oxford University Press

Blum AL, Langley P (1997) Selection of relevant features and examples in machine learning. Artif Intell 97(1–2):245–271

Boers N, Goswami B, Rheinwalt A, Bookhagen B, Hoskins B, Kurths J (2019) Complex networks reveal global pattern of extreme-rainfall teleconnections. Nature 566(7744):373–377

Bonavita M, Arcucci R, Carrassi A, Dueben P, Geer AJ, Le Saux B, Longépé N, Mathieu P-P, Raynaud L (2021) Machine learning for Earth system observation and prediction. Bull Am Meteor Soc 102(4):710–716

Boneh T, Weymouth G, Newham P, Potts R, Bally J, Nicholson A, Korb K (2015) Fog forecasting for Melbourne airport using a Bayesian decision network. Weather Forecast 30(5):1218–1233

Branco P, Torgo L, Ribeiro RP (2017) Smogn: a pre-processing approach for imbalanced regression. In: First International Workshop on Learning with Imbalanced Domains: Theory and Applications. PMLR, pp 36–50

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Bresch D, Leonard M, Wahl T, Zhang X (2018) Future climate risk from compound events. Nat Clim Change 8

Bui DT, Tsangaratos P, Ngo P-TT, Pham TD, Pham BT (2019) Flash flood susceptibility modeling using an optimized fuzzy rule based feature selection technique and tree based ensemble methods. Sci Total Environ 668:1038–1054

Burke A, Snook N, Gagne DJ II, McCorkle S, McGovern A (2020) Calibration of machine learning-based probabilistic hail predictions for operational forecasting. Weather Forecast 35(1):149–168

Callaghan M, Schleussner C-F, Nath S, Lejeune Q, Knutson TR, Reichstein M, Hansen G, Theokritoff E, Andrijevic M, Brecha RJ et al (2021) Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. Nat Clim Chang 11(11):966–972

Camps-Valls G, Sejdinovic D, Runge J, Reichstein M (2019) A perspective on gaussian processes for Earth observation. Natl Sci Rev 6(4):616–618

Carrico AR, Donato K (2019) Extreme weather and migration: evidence from Bangladesh. Popul Environ 41(1):1–31

Castillo-Botón C, Casillas-Pérez D, Casanova-Mateo C, Ghimire S, Cerro-Prada E, Gutierrez PA, Deo RC, Salcedo-Sanz S (2022) Machine learning regression and classification methods for fog events prediction. Atmos Res 106157

Chapman S, Watkins NW, Stainforth DA (2019) Warming trends in summer heatwaves. Geophys Res Lett 46(3):1634–1640

Chase RJ, Harrison DR, Lackmann GM, McGovern A (2023) A machine learning tutorial for operational meteorology, part II: Neural networks and deep learning. Weather Forecast

Chattopadhyay A, Nabizadeh E, Hassanzadeh P (2020) Analog forecasting of extreme-causing weather patterns using deep learning. J Adv Model Earth Syst 12(2):2019–001958

Chavez M, Ghil M, Urrutia-Fucugauchi J (2015) Extreme events: observations, modeling, and economics, vol 214. John Wiley & Sons

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357

Chen B-F, Kuo Y-T, Huang T-S (2023) A deep learning ensemble approach for predicting tropical cyclone rapid intensification. Atmos Sci Lett 1151

Chen A, Giese M, Chen D (2020) Flood impact on mainland Southeast Asia between 1985 and 2018- the role of tropical cyclones. J Flood Risk Manag 13(2):12598

Chen R, Zhang W, Wang X (2020) Machine learning in tropical cyclone forecast modeling: a review. Atmosphere 11(7):676

Chen Y, Huang G, Wang Y, Tao W, Tian Q, Yang K, Zheng J, He H (2023) Improving the heavy rainfall forecasting using a weighted deep learning model. Front Environ Sci 11:1116672

Chithra N, Thampi SG, Surapaneni S, Nannapaneni R, Reddy A, Kumar JD (2015) Prediction of the likely impact of climate change on monthly mean maximum and minimum temperature in the Chaliyar River Basin, India, using ANN-based models. Theoret Appl Climatol 121(3):581–590

Chkeir S, Anesiadou A, Mascitelli A, Biondi R (2023) Nowcasting extreme rain and extreme wind speed with machine learning techniques applied to different input datasets. Atmos Res 282:106548

Choi C, Kim J, Kim J, Kim D, Bae Y, Kim HS (2018) Development of heavy rain damage prediction model using machine learning based on big data. Adv Meteorol 2018

Choudhary SS, Ghosh S (2023) Analysis of rainfall and temperature using deep learning model. Theor Appl Climatol 1–16

Cho K, Van Merriënboer B, Bahdanau D, Bengio Y (2014) On the properties of neural machine translation: encoder-decoder approaches. Preprint at http://arxiv.org/abs/1409.1259

Chowdhury SN, Ray A, Dana SK, Ghosh D (2022) Extreme events in dynamical systems and random walkers: a review. Phys Rep 966:1–52

Ciais P, Reichstein M, Viovy N, Granier A, Ogée J, Allard V, Aubinet M, Buchmann N, Bernhofer C, Carrara A et al (2005) Europe-wide

reduction in primary productivity caused by the heat and drought in 2003. Nature 437(7058):529–533

Cifuentes J, Marulanda G, Bello A, Reneses J (2020) Air temperature forecasting using machine learning techniques: a review. Energies 13(16):4215

Cohen J, Coumou D, Hwang J, Mackey L, Orenstein P, Totz S, Tziperman E (2019) S2S reboot: an argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. Wiley Interdiscip Rev Clim Change 10(2):00567

Colabone RO, Ferrari A, da Silva-Vecchia F, Bruno-Tech A (2015) Application of artificial neural networks for fog forecast. J Aerosp Technol Manag 169:1107–1119

Cornejo-Bueno L, Casanova-Mateo C, Sanz-Justo J, Cerro-Prada E, Salcedo-Sanz S (2017) Efficient prediction of low-visibility events at airports using machine-learning regression. Bound-Layer Meteorol 165:349–370

Cornejo-Bueno S, Casillas-Pérez D, Cornejo-Bueno L, Chidean MI, Caamaño AJ, Sanz-Justo J, Casanova-Mateo C, Salcedo-Sanz S (2020) Persistence analysis and prediction of low-visibility events at Valladolid airport. Spain. Symmetry 12(6):1045

Cornejo-Bueno S, Casillas-Pérez D, Cornejo-Bueno L, Chidean MI, Caamaño AJ, Sanz-Justo J, Casanova-Mateo C, Salcedo-Sanz S (2020) Persistence analysis and prediction of low-visibility events at Valladolid airport. Spain Symmetry 12(6):1045

Cornejo-Bueno S, Casillas-Pérez D, Cornejo-Bueno L, Chidean MI, Caamaño AJ, Cerro-Prada E, Casanova-Mateo C, Salcedo-Sanz S (2021) Statistical analysis and machine learning prediction of fog-caused low-visibility events at a-8 motor-road in Spain. Atmosphere 12(6):679

Czernecki B, Taszarek M, Marosz M, Półrolniczak M, Kolendowicz L, Wyszogrodzki A, Szturc J (2019) Application of machine learning to large hail prediction-the importance of radar reflectivity, lightning occurrence and convective parameters derived from ERA5. Atmos Res 227:249–262

Danandeh Mehr A, Rikhtehgar Ghiasi A, Yaseen ZM, Sorman AU, Abualigah L (2022) A novel intelligent deep learning predictive model for meteorological drought forecasting. J Ambient Intell Humaniz Comput 1–15

De S, Debnath A (2009) Artificial neural network based prediction of maximum and minimum temperature in the summer monsoon months over India. Appl Phys Res 1(2):37

De U, Khole M, Dandekar M (2004) Natural hazards associated with meteorological extreme events. Nat Hazards 31(2):487–497

Del Ser J, Osaba E, Molina D, Yang X-S, Salcedo-Sanz S, Camacho D, Das S, Suganthan PN, Coello CAC, Herrera F (2019) Bio-inspired computation: where we stand and what's next. Swarm Evol Comput 48:220–250

Deo RC, Şahin M (2015) Application of the extreme learning machine algorithm for the prediction of monthly effective drought index in Eastern Australia. Atmos Res 153:512–525

Díaz J, Jordán A, García R, López C, Alberdi J, Hernández E, Otero A (2002) Heat waves in Madrid 1986–1997: effects on the health of the elderly. Int Arch Occup Environ Health 75(3):163–170

Díaz J, Garcia R, De Castro FV, Hernández E, López C, Otero A (2002) Effects of extremely hot days on people older than 65 years in Seville (Spain) from 1986 to 1997. Int J Biometeorol 46(3):145–149

Dietz SJ, Kneringer P, Mayr GJ, Zeileis A (2019) Forecasting low-visibility procedure states with tree-based statistical methods. Pure Appl Geophys 176(6):2631–2644

Diez-Sierra J, del Jesus M (2020) Long-term rainfall prediction using atmospheric synoptic patterns in semi-arid climates with statistical and machine learning methods. J Hydrol 586:124789

Dikshit A, Pradhan B, Alamri AM (2020) Temporal hydrological drought index forecasting for New South Wales. Australia using machine learning approaches. Atmosphere 11(6):585

Ding J, Zhang G, Wang S, Xue B, Yang J, Gao J, Wang K, Jiang R, Zhu X (2022) Forecast of hourly airport visibility based on artificial intelligence methods. Atmosphere 13(1):75

Durán-Rosal AM, Fernández JC, Casanova-Mateo C, Sanz-Justo J, Salcedo-Sanz S, Hervás-Martínez C (2018) Efficient fog prediction with multi-objective evolutionary neural networks. Appl Soft Comput 70:347–358

Easterling DR, Kunkel KE, Wehner MF, Sun L (2016) Detection and attribution of climate extremes in the observed record. Weather Clim Extremes 11:17–27

Ebrahimi-Khusfi Z, Nafarzadegan AR, Dargahian F (2021) Predicting the number of dusty days around the desert wetlands in Southeastern Iran using feature selection and machine learning techniques. Ecol Ind 125:107499

ECMWF. https://www.ecmwf.int/ Accessed 2022-03-04

Elman JL (1990) Finding structure in time. Cogn Sci 14(2):179–211

Fabbian D, De-Dear R, Lellyett S (2007) Application of artificial neural network forecasts to predict fog at canberra international airport. Weather Forecast 22(2):372–381

Fang W, Xue Q, Shen L, Sheng VS (2021) Survey on the application of deep learning in extreme weather prediction. Atmosphere 12(6):661

Farazmand M, Sapsis TP (2019) Extreme events: mechanisms and prediction. Appl Mech Rev 71(5)

Farmanifard S, Alesheikh AA, Sharif M (2023) A context-aware hybrid deep learning model for the prediction of tropical cyclone trajectories. Expert Syst Appl 120701

Feng P, Wang B, Li Liu D, Yu Q (2019) Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. Agric Syst 173:303–316

Feng S, Wu X, Hao Z, Hao Y, Zhang X, Hao F (2020) A database for characteristics and variations of global compound dry and hot events. Weather Clim Extremes 30:100299

Feng Y, Maulik R, Wang J, Balaprakash P, Huang W, Rao V, Xue P, Pringle W, Bessac J, Sullivan R (2021) Characterization of extremes and compound impacts: applications of machine learning and interpretable neural networks. Technical report, Artificial Intelligence for Earth System Predictability

Ferreira AJ, Figueiredo MA (2012) Boosting algorithms: a review of methods, theory, and applications. Ensemble Machine Learning 35–85

Ferreira AJ, Figueiredo MA (2014) Incremental filter and wrapper approaches for feature discretization. Neurocomputing 123:60–74

Ferro CA (2007) A probability model for verifying deterministic forecasts of extreme events. Weather Forecast 22(5):1089–1100

Fister D, Pérez-Aracil J, Peláez-Rodríguez C, Del Ser J, Salcedo-Sanz S (2023) Accurate long-term air temperature prediction with machine learning models and data reduction techniques. Appl Soft Comput 136:110118

Flora ML, Potvin CK, Skinner PS, Handler S, McGovern A (2021) Using machine learning to generate storm-scale probabilistic guidance of severe weather hazards in the warn-on-forecast system. Mon Weather Rev 149(5):1535–1557

Folino G, Guarascio M, Chiaravalloti F (2023) Learning ensembles of deep neural networks for extreme rainfall event detection. Neural Comput Appl 1–14

Fraile R, Berthet C, Dessens J, Sánchez JL (2003) Return periods of severe hailfalls computed from hailpad data. Atmos Res 67:189–202

Frank D, Reichstein M, Bahn M, Thonicke K, Frank D, Mahecha MD, Smith P, Van der Velde M, Vicca S, Babst F et al (2015) Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts. Glob Change Biol 21(8):2861–2880

Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55(1):119–139

Gagne II DJ, McGovern A, Brotzge J, Coniglio M, Correia Jr J, Xue M (2015) Day-ahead hail prediction integrating machine learning with storm-scale numerical weather models. In: Twenty-Seventh IAAI Conference. pp 3954–3960

Gagne DJ, McGovern A, Haupt SE, Sobash RA, Williams JK, Xue M (2017) Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. Weather Forecast 32(5):1819–1840

Gagne DJ II, Haupt SE, Nychka DW, Thompson G (2019) Interpretable deep learning for spatial analysis of severe hailstorms. Mon Weather Rev 147(8):2827–2845

Gallicchio C, Micheli A (2017) Deep echo state network (DEEPESN): a brief survey. Preprint at http://arxiv.org/abs/1712.04323

García-Herrera R, Díaz J, Trigo RM, Luterbacher J, Fischer EM (2010) A review of the European summer heat wave of 2003. Crit Rev Environ Sci Technol 40(4):267–306

García-Herrera R, Garrido-Perez JM, Barriopedro D, Ordóñez C, Vicente-Serrano SM, Nieto R, Gimeno L, Sorí R, Yiou P (2019) The European 2016/17 drought. J Clim 32(11):3169–3187

Ghil M, Yiou P, Hallegatte S, Malamud B, Naveau P, Soloviev A, Friederichs P, Keilis-Borok V, Kondrashov D, Kossobokov V et al (2011) Extreme events: dynamics, statistics and prediction. Nonlinear Process Geophys 18(3):295–350

Ghodsi A (2006) Dimensionality reduction a short tutorial. Department of Statistics and Actuarial Science, Univ. of Waterloo, Ontario, Canada 37(38):2006

Ghojogh B, Crowley M, Karray F, Ghodsi A (2023) Elements of dimensionality reduction and manifold learning. Springer Nature

Gómez-Orellana AM, Guijo-Rubio D, Pérez-Aracil J, Gutiérrez PA, Salcedo-Sanz S, Hervás-Martínez C (2023) One month in advance prediction of air temperature from reanalysis data with explainable artificial intelligence techniques. Atmos Res 106608

González S, García S, Del Ser J, Rokach L, Herrera F (2020) A practical tutorial on bagging and boosting based ensembles for machine learning: algorithms, software tools, performance study, practical perspectives and opportunities. Inf Fusion 64:205–237

Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press

Grant PR (2017) Evolution, climate change, and extreme events. Science 357(6350):451–452

Grazzini F, Craig GC, Keil C, Antolini G, Pavan V (2020) Extreme precipitation events over Northern Italy. Part I: a systematic classification with machine-learning techniques. Q J R Meteorol Soc 146(726):69–85

Grazzini F, Fragkoulidis G, Teubler F, Wirth V, Craig GC (2021) Extreme precipitation events over Northern Italy. part II: dynamical precursors. Q J R Meteorol Soc 147(735):1237–1257

Grüning A, Bohte SM (2014) Spiking neural networks: principles and challenges. In: ESANN. Citeseer, pp 1–10

Guerreiro PM, Soares PM, Cardoso RM, Ramos AM (2020) An analysis of fog in the mainland portuguese international airports. Atmosphere 11(11):1239

Guijo-Rubio D, Gutiérrez PA, Casanova-Mateo C, Sanz-Justo J, Salcedo-Sanz S, Hervás-Martínez C (2018) Prediction of low-visibility events due to fog using ordinal classification. Atmos Res 214:64–73

Guijo-Rubio D, Gutiérrez PA, Casanova-Mateo C, Fernández JC, Gómez-Orellana AM, Salvador-González P, Salcedo-Sanz S, Hervás-Martínez C (2020) Prediction of convective clouds formation using evolutionary neural computation techniques. Neural Comput Appl 32(17):13917–13929

Guijo-Rubio D, Casanova-Mateo C, Sanz-Justo J, Gutiérrez P, Cornejo-Bueno S, Hervás C, Salcedo-Sanz S (2020) Ordinal regression algorithms for the analysis of convective situations over Madrid-Barajas airport. Atmos Res 236:104798

Gultepe I, Tardif R, Michaelides SC, Cermak J, Bott A, Bendix J, Müller MD, Pagowski M, Hansen B, Ellrod G, Jacobs W, Toth G, Cober SG (2007) Fog research: a review of past achievements and future perspectives. Pure Appl Geophys 164:1121–1159

Gyaneshwar A, Mishra A, Chadha U, Raj Vincent PD, Rajinikanth V, Pattukandan Ganapathy G, Srinivasan K (2023) A contemporary review on deep learning models for drought prediction. Sustainability 15(7):6160

Hagan MT, Menhaj MB (1994) Training feedforward networks with the marquardt algorithm. IEEE Trans Neural Netw 5(6):989–993

Hannart A, Naveau P (2018) Probabilities of causation of climate changes. J Clim 31(14):5507–5524

Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, Tang Y, Xiao A, Xu C, Xu Y et al (2020) A survey on visual transformer. arXiv e-prints, 2012

Hartigan J, MacNamara S, Leslie LM, Speer M (2020) Attribution and prediction of precipitation and temperature trends within the sydney catchment using machine learning. Climate 8(10):120

Hartigan J, MacNamara S, Leslie LM (2020) Application of machine learning to attribution and prediction of seasonal precipitation and temperature trends in Canberra. Australia. Climate 8(6):76

Haykin S, Network N (2004) A comprehensive foundation. Neural Netw 2(2004):41

Hecht-Nielsen R (1992) Theory of the backpropagation neural network. In: Neural Networks for Perception. Academic Press, pp 65–93

Herring SC, Hoerling MP, Kossin JP, Peterson TC, Stott PA (2015) Explaining extreme events of 2014 from a climate perspective. Bull Am Meteor Soc 96(12):1–172

Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D et al (2020) The ERA5 global reanalysis. Q J R Meteorol Soc 146(730):1999–2049

Hill AJ, Herman GR, Schumacher RS (2020) Forecasting severe weather with random forests. Mon Weather Rev 148(5):2135–2161

Hinton GE, Krizhevsky A, Wang SD (2011) Transforming auto-encoders. In: International Conference on Artificial Neural Networks. Springer, pp 44–51

Hirschi M, Seneviratne SI, Alexandrov V, Boberg F, Boroneant C, Christensen OB, Formayer H, Orlowsky B, Stepanek P (2011) Observational evidence for soil-moisture impact on hot extremes in Southeastern Europe. Nat Geosci 4(1):17–21

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

Hoffmann D, Gallant AJE, Arblaster JM (2020) Uncertainties in drought from index and data selection. J Geophys Res Atmos 125(18):2019–031946

Horton RM, Mankin JS, Lesk C, Coffel E, Raymond C (2016) A review of recent advances in research on extreme heat events. Curr Clim Change Rep 2(4):242–259

Hosseini FS, Choubin B, Mosavi A, Nabipour N, Shamshirband S, Darabi H, Haghighi AT (2020) Flash-flood hazard assessment using ensembles and bayesian-based machine learning models: application of the simulated annealing feature selection method. Sci Total Environ 711:135161

Hu H, Ayyub BM (2019) Machine learning for projecting extreme precipitation intensity for short durations in a changing climate. Geosciences 9(5):209

Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501

Huang G-B, Zhou H, Ding X, Zhang R (2011) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern B (Cybernetics) 42(2):513–529

Huang X, Wu L, Ye Y (2019) A review on dimensionality reduction techniques. Int J Pattern Recognit Artif Intell 33(10):1950017

Huda S, Abdollahian M, Mammadov M, Yearwood J, Ahmed S, Sultan I (2014) A hybrid wrapper-filter approach to detect the source (s) of out-of-control signals in multivariate manufacturing process. Eur J Oper Res 237(3):857–870

Irrgang C, Boers N, Sonnewald M, Barnes EA, Kadow C, Staneva J, Saynisch-Wagner J (2021) Towards neural earth system modelling by integrating artificial intelligence in Earth system science. Nat Mach Intell 3(8):667–674

Jahangir MH, Reineh SMM, Abolghasemi M (2019) Spatial predication of flood zonation mapping in Kan River Basin, Iran, using artificial neural network algorithm. Weather Clim Extrem 25:100215

Jergensen GE, McGovern A, Lagerquist R, Smith T (2020) Classifying convective storms using machine learning. Weather Forecast 35(2):537–559

John GH, Kohavi R, Pfleger K (1994) Irrelevant features and the subset selection problem. In: Machine Learning Proceedings 1994. Elsevier, pp 121–129

Jordan MI (1997) Serial order: a parallel distributed processing approach. In: Advances in Psychology, vol 121. Elsevier, pp 471–495

Kadow C, Hall DM, Ulbrich U (2020) Artificial intelligence reconstructs missing climate information. Nat Geosci 13(6):408–413

Kar C, Banerjee S (2021) Tropical cyclone intensity classification from infrared images of clouds over Bay of Bengal and Arabian sea using machine learning classifiers. Arab J Geosci 14(8):1–17

Karpatne A, Ebert-Uphoff I, Ravela S, Babaie HA, Kumar V (2018) Machine learning for the geosciences: challenges and opportunities. IEEE Trans Knowl Data Eng 31(8):1544–1554

Kaur A, Sood SK (2020) Deep learning based drought assessment and prediction framework. Eco Inform 57:101067

Khan N, Sachindra D, Shahid S, Ahmed K, Shiru MS, Nawaz N (2020) Prediction of droughts over Pakistan using machine learning algorithms. Adv Water Resour 139:103562

Kim S-H, Moon I-J, Won S-H, Kang H-W, Kang SK (2021) Decision-tree-based classification of lifetime maximum intensity of tropical cyclones in the tropical Western North Pacific. Atmosphere 12(7):802

Kingma DP, Welling M (2013) Auto-encoding variational bayes. Preprint at http://arxiv.org/abs/1312.6114

Knapp AK, Beier C, Briske DD, Classen AT, Luo Y, Reichstein M, Smith MD, Smith SD, Bell JE, Fay PA et al (2008) Consequences of more extreme precipitation regimes for terrestrial ecosystems. Bioscience 58(9):811–821

Kohavi R, John GH (1997) Wrappers for feature subset selection. Artif Intell 97(1–2):273–324

Kolios S (2023) Hail detection from Meteosat satellite imagery using a deep learning neural network and a new remote sensing index. Adv Space Res

Kurth T, Treichler S, Romero J, Mudigonda M, Luehr N, Phillips E, Mahesh A, Matheson M, Deslippe J, Fatica M et al (2018) Exascale deep learning for climate analytics. SC18: International Conference for High Performance Computing. Networking, Storage and Analysis. IEEE, pp 649–660

Lagerquist R, McGovern A, Smith T (2017) Machine learning for real-time prediction of damaging straight-line convective wind. Weather Forecast 32(6):2175–2193

Lal R, Delgado JA, Gulliford J, Nielsen D, Rice CW, Van Pelt RS (2012) Adapting agriculture to drought and extreme events. J Soil Water Conserv 67(6):162–166

Lavers DA, Villarini G (2013) Were global numerical weather prediction systems capable of forecasting the extreme Colorado rainfall of 9–16 September 2013? Geophys Res Lett 40(24):6405–6410

Lea C, Flynn MD, Vidal R, Reiter A, Hager GD (2017) Temporal convolutional networks for action segmentation and detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp 156–165

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444

Leinonen J, Hamann U, Sideris IV, Germann U (2023) Thunderstorm nowcasting with deep learning: a multi-hazard data fusion model. Geophys Res Lett 50(8):2022–101626

Li J, Wang Z, Wu X, Xu C-Y, Guo S, Chen X, Zhang Z (2021) Robust meteorological drought prediction using antecedent SST fluctuations and machine learning. Water Resour Res 57(8):2020–029413

Lin X, Fan J, Hou ZJ, Wang J (2023) Machine learning of key variables impacting extreme precipitation in various regions of the contiguous United States. J Adv Model Earth Syst 15(3):2022–003334

Li C, Shi Y, Gao P, Shen Y, Ma C, Shi D (2020) Diagnostic model of low visibility events based on C4.5 algorithm. Open Physics 18(1):33–39

Liu B, He X, Song M, Li J, Qu G, Lang J, Gu R (2021) A method for mining granger causality relationship on atmospheric visibility. ACM Trans Knowl Discov Data (TKDD) 15(5):1–16

Liu Y, Racah E, Correa J, Khosrowshahi A, Lavers D, Kunkel K, Wehner M, Collins W et al (2016) Application of deep convolutional neural networks for detecting extreme weather in climate datasets. Preprint at http://arxiv.org/abs/1605.01156

López L, García-Ortega E, Sánchez JL (2007) A short-term forecast model for hail. Atmos Res 83(2–4):176–184

Luettich RA, Westerink JJ, Scheffner NW et al (1992) ADCIRC: an advanced three-dimensional circulation model for shelves, coasts, and estuaries. Report 1, Theory and methodology of ADCIRC-2DD1 and ADCIRC-3DL. Coastal Engineering Research Center (US)

Lukoševičius M, Jaeger H (2009) Reservoir computing approaches to recurrent neural network training. Comput Sci Rev 3(3):127–149

Madakumbura GD, Thackeray CW, Norris J, Goldenson N, Hall A (2021) Anthropogenic influence on extreme precipitation over global land areas seen in multiple observational datasets. Nat Commun 12(1):1–9

Madsen H, Lawrence D, Lang M, Martinkova M, Kjeldsen T (2014) Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. J Hydrol 519:3634–3650

Mamalakis A, Ebert-Uphoff I, Barnes EA (2022) Neural network attribution methods for problems in geoscience: a novel synthetic benchmark dataset. Environ Data Sci 1:8

Manna T, Anitha A (2023) Precipitation prediction by integrating rough set on fuzzy approximation space with deep learning techniques. Appl Soft Comput 139:110253

Marchiori L, Maystadt J-F, Schumacher I (2012) The impact of weather anomalies on migration in sub-Saharan Africa. J Environ Econ Manag 63(3):355–374

Markonis Y, Kumar R, Hanel M, Rakovec O, Máca P, AghaKouchak A (2021) The rise of compound warm-season droughts in Europe. Sci Adv 7(6):9668

Marzban C, Leyton S, Colman B (2007) Ceiling and visibility forecasts via neural networks. Weather Forecast 22(3):466–479

Matsuoka D (2022) Can machine learning models trained using atmospheric simulation data be applied to observation data? Experimental Results 3:7

May PJ, Koski C (2013) Addressing public risks: extreme events and critical infrastructures. Rev Policy Res 30(2):139–159

McGovern A, Elmore KL, Gagne DJ, Haupt SE, Karstens CD, Lagerquist R, Smith T, Williams JK (2017) Using artificial intel-

ligence to improve real-time decision-making for high-impact weather. Bull Am Meteor Soc 98(10):2073–2090

McGovern A, Chase RJ, Flora M, Gagne DJ, Lagerquist R, Potvin CK, Snook N, Loken E (2023) A review of machine learning for convective weather. Artif Intell Earth Syst 1–61

Meng F, Yao Y, Wang Z, Peng S, Xu D, Song T (2023) Probabilistic forecasting of tropical cyclones intensity using machine learning model. Environ Res Lett 18(4):044042

Miao Y, Potts R, Huang X, Elliott G, Rivett R (2012) A fuzzy logic fog forecasting model for Perth airport. Pure Appl Geophys 169:1107–1119

Miao K-C, Han T-T, Yao Y-Q, Lu H, Chen P, Wang B, Zhang J (2020) Application of LSTM for short term fog forecasting based on meteorological elements. Neurocomputing 408:285–291

Mitchell JF, Lowe J, Wood RA, Vellinga M (2006) Extreme events due to human-induced climate change. Philos Trans R Soc A Math Phys Eng Sci 364(1845):2117–2133

Mohandes M, Deriche M, Aliyu SO (2018) Classifiers combination techniques: a comprehensive review. IEEE Access 6:19626–19639

Moishin M, Deo RC, Prasad R, Raj N, Abdulla S (2021) Designing deep-based learning flood forecast model with ConvLSTM hybrid algorithm. IEEE Access 9:50982–50993

Mokhtar A, Jalali M, He H, Al-Ansari N, Elbeltagi A, Alsafadi K, Abdo HG, Sammen SS, Gyasi-Agyei Y, Rodrigo-Comino J (2021) Estimation of spei meteorological drought using machine learning algorithms. IEEE Access 9:65503–65523

Mokhtari R, Akhoondzadeh M (2021) Data fusion and machine learning algorithms for drought forecasting using satellite data. J Earth Space Phys 46(4):231–246

Montavon G, Samek W, Müller K-R (2018) Methods for interpreting and understanding deep neural networks. Digit Signal Process 73:1–15

Monteleoni C, Schmidt GA, McQuade S (2013) Climate informatics: accelerating discovering in climate science with machine learning. Comput Sci Eng 15(5):32–40

Moon S-H, Kim Y-H, Lee YH, Moon B-R (2019) Application of machine learning to an early warning system for very short-term heavy rainfall. J Hydrol 568:1042–1054

Mosavi A, Ozturk P, Chau K-W (2018) Flood prediction using machine learning models: literature review. Water 10(11)

Nairn JR, Fawcett RJ (2015) The excess heat factor: a metric for heatwave intensity and its use in classifying heatwave severity. Int J Environ Res Public Health 12(1):227–253

Nandi A, De A, Mallick A, Middya AI, Roy S (2022) Attention based long-term air temperature forecasting network: ALTF net. Knowl-Based Syst 252:109442

Naveau P, Hannart A, Ribes A (2020) Statistical methods for extreme event attribution in climate science. Annu Rev Stat Appl 7:89–110

Nayak MA, Ghosh S (2013) Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier. Theoret Appl Climatol 114(3):583–603

Nethery RC, Katz-Christy N, Kioumourtzoglou M-A, Parks RM, Schumacher A, Anderson GB (2021) Integrated causal-predictive machine learning models for tropical cyclone epidemiology. Biostatistics kxab047

Ngo P-TT, Pham TD, Nhu V-H, Le TT, Tran DA, Phan DC, Hoa PV, Amaro-Mellado JL, Bui DT (2021) A novel hybrid quantum-PSO and credal decision tree ensemble for tropical cyclone induced flash flood susceptibility mapping with geospatial data. J Hydrol 596:125682

Oettli P, Nonaka M, Richter I, Koshiba H, Tokiya Y, Hoshino I, Behera SK (2022) Combining dynamical and statistical modeling to improve the prediction of surface air temperatures 2 months in advance: a hybrid approach. Front Clim 4

Ortega LC, Otero LD, Solomon M, Otero CE, Fabregas A (2023) Deep learning models for visibility forecasting using climatological data. Int J Forecast 39(2):992–1004

O'Shea K, Nash R (2015) An introduction to convolutional neural networks. Preprint at http://arxiv.org/abs/1511.08458

Ouzeau G, Soubeyroux J-M, Schneider M, Vautard R, Planton S (2016) Heat waves analysis over France in present and future climate: application of a new method on the Euro-cordex ensemble. Clim Serv 4:1–12

Pande CB, Kushwaha N, Orimoloye IR, Kumar R, Abdo HG, Tolche AD, Elbeltagi A (2023) Comparative assessment of improved svm method under different Kernel functions for predicting multi-scale drought index. Water Resour Manage 37(3):1367–1399

Paniagua-Tineo A, Salcedo-Sanz S, Casanova-Mateo C, Ortiz-García E, Cony M, Hernández-Martín E (2011) Prediction of daily maximum temperature using a support vector regression algorithm. Renew Energy 36(11):3054–3060

Park J, Kim J (2018) Defining heatwave thresholds using an inductive machine learning approach. PLoS ONE 13(11):0206872

Park S, Im J, Jang E, Rhee J (2016) Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. Agric For Meteorol 216:157–169

Pasini A, Racca P, Amendola S, Cartocci G, Cassardo C (2017) Attribution of recent temperature behaviour reassessed by a neural-network method. Sci Rep 7(1):1–10

Peláez-Rodríguez C, Pérez-Aracil J, Fister D, Prieto-Godino L, Deo R, Salcedo-Sanz S (2022) A hierarchical classification/regression algorithm for improving extreme wind speed events prediction. Renew Energy 201:157–178

Peláez-Rodríguez C, Marina CM, Pérez-Aracil J, Casanova-Mateo C, Salcedo-Sanz S (2023) Extreme low-visibility events prediction based on inductive and evolutionary decision rules: an explicability-based approach. Atmosphere 14(3):542

Peláez-Rodríguez C, Pérez-Aracil J, de A L-D, Casanova-Mateo C, Fister D, Jiménez-Fernández S, Salcedo-Sanz S (2023) Deep learning ensembles for accurate fog-related low-visibility events forecasting. Neurocomputing 126435

Peng Y, Abdel-Aty M, Lee J, Zou Y (2018) Analysis of the impact of fog-related reduced visibility on traffic parameters. J Transp Eng A: Syst 144(2):04017077

Peng T, Zhi X, Ji Y, Ji L, Tian Y (2020) Prediction skill of extended range 2-m maximum air temperature probabilistic forecasts using machine learning post-processing methods. Atmosphere 11(8):823

Peters J, Janzing D, Schölkopf B (2017) Elements of causal inference: foundations and learning algorithms. The MIT Press

Pfleiderer P, Coumou D (2018) Quantification of temperature persistence over the Northern Hemisphere land-area. Clim Dyn 51(1):627–637

Pillay MT, Fitchett JM (2021) On the conditions of formation of southern hemisphere tropical cyclones. Weather Clim Extremes 34:100376

Pinaya WHL, Vieira S, Garcia-Dias R, Mechelli A (2020) Autoencoders. In: Machine Learning. Elsevier, pp 193–208

Piri J, Abdolahipour M, Keshtegar B (2023) Advanced machine learning model for prediction of drought indices using hybrid SVR-RSM. Water Resour Manage 37(2):683–712

Pirone D, Cimorelli L, Del Giudice G, Pianese D (2023) Short-term rainfall forecasting using cumulative precipitation fields from station data: a probabilistic machine learning approach. J Hydrol 617:128949

Pörtner H-O, Roberts DC, Poloczanska ES, Mintenbeck K, Tignor M, Alegría A, Craig M, Langsdorf S, Löschke S, Möller V, Okem A (eds) (2022) Summary for policymakers. In: Climate change 2022: impacts, adaptation, and vulnerability. Contribution of working

group II to the sixth assessment report of the intergovernmental panel on climate change. Technical report, Cambridge University Press

Prodhan FA, Zhang J, Sharma TPP, Nanzad L, Zhang D, Seka AM, Ahmed N, Hasan SS, Hoque MZ, Mohana HP (2022) Projection of future drought and its impact on simulated crop yield over south asia using ensemble machine learning approach. Sci Total Environ 807:151029

Pullman M, Gurung I, Maskey M, Ramachandran R, Christopher SA (2019) Applying deep learning to hail detection: a case study. IEEE Trans Geosci Remote Sens 57(12):10218–10225

Qi D, Majda AJ (2020) Using machine learning to predict extreme events in complex systems. Proc Natl Acad Sci 117(1):52–59

Rahmati O, Falah F, Dayal KS, Deo RC, Mohammadi F, Biggs T, Moghaddam DD, Naghibi SA, Bui DT (2020) Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland. Australia. Sci Total Environ 699:134230

Raymond C, Horton RM, Zscheischler J, Martius O, AghaKouchak A, Balch J, Bowen SG, Camargo SJ, Hess J, Kornhuber K et al (2020) Understanding and managing connected extreme events. Nat Clim Chang 10(7):611–621

Reichstein M, Bahn M, Ciais P, Frank D, Mahecha MD, Seneviratne SI, Zscheischler J, Beer C, Buchmann N, Frank DC et al (2013) Climate extremes and the carbon cycle. Nature 500(7462):287–295

Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N et al (2019) Deep learning and process understanding for data-driven Earth system science. Nature 566(7743):195–204

Ren X, Li L, Yu Y, Xiong Z, Yang S, Du W, Ren M (2020) A simplified climate change model and extreme weather model based on a machine learning method. Symmetry 12(1):139

Rhee J, Im J (2017) Meteorological drought forecasting for ungauged areas based on machine learning: using long-range climate forecast and remote sensing data. Agric For Meteorol 237:105–122

Richman MB, Leslie LM (2018) The 2015–2017 cape town drought: attribution and prediction using machine learning. Procedia Comput Sci 140:248–257

Richman MB, Leslie LM (2020) Machine learning for attribution of heat and drought in Southwestern Australia. Procedia Comput Sci 168:3–10

Rolnick D, Donti PL, Kaack LH, Kochanski K, Lacoste A, Sankaran K, Ross AS, Milojevic-Dupont N, Jaques N, Waldman-Brown A et al (2019) Tackling climate change with machine learning. Preprint at http://arxiv.org/abs/1906.05433

Román-Cascón C, Yagüe C, Sastre M, Maqueda G, Salamanca F, Viana S (2012) Observations and WRF simulations of fog events at the Spanish Northern Plateau. Adv Sci Res 8(1):11–18

Roodposhti MS, Safarrad T, Shahabi H (2017) Drought sensitivity mapping using two one-class support vector machine algorithms. Atmos Res 193:73–82

Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500):2323–2326

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323(6088):533–536

Runge J (2018) Causal network reconstruction from time series: from theoretical assumptions to practical estimation. Chaos: An Interdisciplinary Journal of Nonlinear Science 28(7):075310

Runge J, Bathiany S, Bollt E, Camps-Valls G, Coumou D, Deyle E, Glymour C, Kretschmer M, Mahecha MD, Muñoz-Marí J et al (2019) Inferring causation from time series in Earth system sciences. Nat Commun 10(1):1–13

Sahoo B, Bhaskaran PK (2019) Prediction of storm surge and coastal inundation using artificial neural network-a case study for 1999 odisha super cyclone. Weather Clim Extrem 23:100196

Salcedo-Sanz S, Rojo-Álvarez JL, Martínez-Ramón M, Camps-Valls G (2014) Support vector machines in engineering: an overview. Wiley Interdiscip Rev Data Min Knowl Discov 4(3):234–267

Salcedo-Sanz S, Cornejo-Bueno L, Prieto L, Paredes D, García-Herrera R (2018) Feature selection in machine learning prediction systems for renewable energy applications. Renew Sustain Energy Rev 90:728–741

Salcedo-Sanz S, Ghamisi P, Piles M, Werner M, Cuadra L, Moreno-Martínez A, Izquierdo-Verdiguier E, Muñoz-Marí J, Mosavi A, Camps-Valls G (2020) Machine learning information fusion in Earth observation: a comprehensive review of methods, applications and data sources. Inf Fusion 63:256–272

Salcedo-Sanz S, Piles M, Cuadra L, Casanova-Mateo C, Caamaño A, Cerro-Prada E, Camps-Valls G (2021) Long-term persistence, invariant time scales and on-off intermittency of fog events. Atmos Res 252:105456

Sallis PJ, Claster W, Hernández S (2011) A machine-learning algorithm for wind gust prediction. Comput Geosci 37(9):1337–1344

Sánchez-Benítez A, García-Herrera R, Barriopedro D, Sousa PM, Trigo RM (2018) June 2017: the earliest European summer mega-heatwave of reanalysis period. Geophys Res Lett 45(4):1955–1962

Sapsis TP (2021) Statistics of extreme events in fluid flows and waves. Annu Rev Fluid Mech 53:85–111

Schlef KE, Moradkhani H, Lall U (2019) Atmospheric circulation patterns associated with extreme United States floods identified via machine learning. Sci Rep 9(1):1–12

Schölkopf B (2022) Causality for machine learning. In: Probabilistic and Causal Inference: The Works of Judea Pearl. pp 765–804

Schölkopf B, Smola AJ, Williamson RC, Bartlett PL (2000) New support vector algorithms. Neural Comput 12(5):1207–1245

Schölkopf B, Smola AJ, Bach F et al (2002) Learning with Kernels: support vector machines, regularization, optimization, and beyond. MIT press

Schulz B, Lerch S (2021) Machine learning methods for postprocessing ensemble forecasts of wind gusts: a systematic comparison. Preprint at http://arxiv.org/abs/2106.09512

Schuster M, Paliwal KK (1997) Bidirectional recurrent neural networks. IEEE Trans Signal Process 45(11):2673–2681

Seneviratne S, Nicholls N, Easterling D, Goodess C, Kanae S, Kossin J, Luo Y, Marengo J, McInnes K, Rahimi M et al (2012) Changes in climate extremes and their impacts on the natural physical environment. Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC)

Shamekh S, Lamb KD, Huang Y, Gentine P (2023) Implicit learning of convective organization explains precipitation stochasticity. Proc Natl Acad Sci 120(20):2216158120

Shanmuganathan S, Sallis P (2014) Data mining methods to generate severe wind gust models. Atmosphere 5(1):60–80

Sheridan P (2018) Current gust forecasting techniques, developments and challenges. Adv Sci Res 15:159–172

Sherstinsky A (2020) Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D 404:132306

Shi X (2020) Enabling smart dynamical downscaling of extreme precipitation events with machine learning. Geophys Res Lett 47(19):2020–090309

Shi J, Cui L, Ma Y, Du H, Wen K (2018) Trends in temperature extremes and their association with circulation patterns in China during 1961–2015. Atmos Res 212:259–272

Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14(3):199–222

Sobash RA, Gagne DJ, Becker CL, Ahijevych D, Gantos GN, Schwartz CS (2023) Diagnosing storm mode with deep learning in convection-allowing models. Mon Weather Rev

Solorio-Fernández S, Carrasco-Ochoa JA, Martínez-Trinidad JF (2016) A new hybrid filter-wrapper feature selection method for clustering based on ranking. Neurocomputing 214:866–880

Spassiani AC, Mason MS (2021) Application of self-organizing maps to classify the meteorological origin of wind gusts in australia. J Wind Eng Ind Aerodyn 210:104529

Spinoni J, Barbosa P, De Jager A, McCormick N, Naumann G, Vogt JV, Magni D, Masante D, Mazzeschi M (2019) A new global database of meteorological drought events from 1951 to 2016. J Hydrol Reg Stud 22:100593

Stamos I, Mitsakis E, Salanova JM, Aifadopoulou G (2015) Impact assessment of extreme weather events on transport networks: a data-driven approach. Transp Res Part D: Transp Environ 34:168–178

Stott PA, Christidis N, Otto FE, Sun Y, Vanderlinden J-P, van Oldenborgh GJ, Vautard R, von Storch H, Walton P, Yiou P et al (2016) Attribution of extreme weather and climate-related events. Wiley Interdiscip Rev Clim Change 7(1):23–41

Stubenrauch CJ, Mandorli G, Lemaitre E (2023) Convective organization and 3D structure of tropical cloud systems deduced from synergistic a-train observations and machine learning. Atmos Chem Phys 23(10):5867–5884

Sun X, Xie L, Shah SU, Shen X (2021) A machine learning based ensemble forecasting optimization algorithm for preseason prediction of atlantic hurricane activity. Atmosphere 12(4):522

Sutanto SJ, van der Weert M, Wanders N, Blauhut V, Van Lanen HA (2019) Moving from drought hazard to impact forecasts. Nat Commun 10(1):1–7

Tan J, Chen S, Wang J (2021) Western North Pacific tropical cyclone track forecasts by a machine learning model. Stoch Env Res Risk Assess 35(6):1113–1126

Tebbi MA, Haddad B (2016) Artificial intelligence systems for rainy areas detection and convective cells' delineation for the South Shore of Mediterranean Sea during day and nighttime using msg satellite images. Atmos Res 178:380–392

Toreti A, Cronie O, Zampieri M (2019) Concurrent climate extremes in the key wheat producing regions of the world. Sci Rep 9(1):1–8

Torgo L, Branco P, Ribeiro RP, Pfahringer B (2015) Resampling strategies for regression. Expert Syst 32(3):465–476

Torkkola K (2002) On feature extraction by mutual information maximization. In: 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol 1. IEEE, p 821

Torkkola K, Campbell WM (2000) Mutual information in learning feature transformations. In: ICML. Citeseer, pp 1015–1022

Trifunov VT, Shadaydeh M, Barz B, Denzler J (2021) Anomaly attribution of multivariate time series using counterfactual reasoning. In: 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, pp 166–172

Trinks C, Hiete M, Comes T, Schultmann F (2012) Extreme weather events and road and rail transportation in Germany. Int J Emergency Manage 8(3):207–227

Tufaner F, Özbeyaz A (2020) Estimation and easy calculation of the Palmer Drought Severity Index from the meteorological data by using the advanced machine learning algorithms. Environ Monit Assess 192(9):1–14

Van Der Maaten L, Postma E, Van den Herik J et al (2009) Dimensionality reduction: a comparative. J Mach Learn Res 10(66–71):13

van der Molen MK, Dolman AJ, Ciais P, Eglin T, Gobron N, Law BE, Meir P, Peters W, Phillips OL, Reichstein M et al (2011) Drought and ecosystem carbon cycling. Agric For Meteorol 151(7):765–773

van der Velde M, Tubiello FN, Vrieling A, Bouraoui F (2012) Impacts of extreme weather on wheat and maize in France: evaluating regional crop simulations against observed data. Clim Change 113(3):751–765

Van Oijen M, Beer C, Cramer W, Rammig A, Reichstein M, Rolinski S, Soussana J-F (2013) A novel probabilistic risk analysis to determine the vulnerability of ecosystems to extreme climatic events. Environ Res Lett 8(1):015032

van Straaten C, Whan K, Coumou D, van den Hurk B, Schmeits M (2022) Using explainable machine learning forecasts to discover sub-seasonal drivers of high summer temperatures in western and central europe. Mon Weather Rev

Vandal T, Kodra E, Ganguly AR (2019) Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation. Theoret Appl Climatol 137(1):557–570

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. Adv Neural Inf Process Syst 30

Vicente-Serrano SM, Beguería S, López-Moreno JI (2010) A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J Clim 23(7):1696–1718

Vitanza E, Dimitri GM, Mocenni C (2023) A multi-modal machine learning approach to detect extreme rainfall events in Sicily. Sci Rep 13(1):6196

Vitart F, Robertson AW (2018) The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. NPJ Clim Atmos Sci 1(1):1–7

Vo TQ, Kim S-H, Nguyen DH, Bae D-H (2023) LSTM-CM: a hybrid approach for natural drought prediction based on deep learning and climate models. Stoch Environ Res Risk Assess 1–17

Wang Z, Jiang Y, Wan H, Yan J, Zhang X (2017) Detection and attribution of changes in extreme temperatures at regional scale. J Clim 30(17):7035–7047

Wang H, Zhang Y-M, Mao J-X, Wan H-P (2020) A probabilistic approach for short-term prediction of wind gust speed using ensemble learning. J Wind Eng Ind Aerodyn 202:104198

Wang Y, Du J, Yan Z, Song Y, Hua D (2022) Atmospheric visibility prediction by using the DBN deep learning model and principal component analysis. Appl Opt 61(10):2657–2666

Wang L, Wan B, Zhou S, Sun H, Gao Z (2023) Forecasting tropical cyclone tracks in the northwestern Pacific based on a deep-learning model. Geosci Model Dev 16(8):2167–2179

Weirich-Benet E, Pyrina M, Jiménez-Esteve B, Fraenkel E, Cohen J, Domeisen DI (2023) Subseasonal prediction of central european summer heatwaves with linear and random forest machine learning models. Artificial Intelligence for the Earth Systems 2(2):220038

Weiss K, Khoshgoftaar TM, Wang D (2016) A survey of transfer learning. J Big Data 3(1):1–40

Wendler-Bosco V, Nicholson C (2021) Modeling the economic impact of incoming tropical cyclones using machine learning. Nat Hazards 1–32

Whan K, Zscheischler J, Orth R, Shongwe M, Rahimi M, Asare EO, Seneviratne SI (2015) Impact of soil moisture on extreme maximum temperatures in Europe. Weather Clim Extremes 9:57–67

White RH, Kornhuber K, Martius O, Wirth V (2021) From atmospheric waves to heatwaves: a waveguide perspective for understanding and predicting concurrent, persistent and extreme extratropical weather. Bull Am Meteorol Soc 1–35

Woodward G, Bonada N, Brown LE, Death RG, Durance I, Gray C, Hladyz S, Ledger ME, Milner AM, Ormerod SJ et al (2016) The effects of climatic fluctuations and extreme events on running water ecosystems. Philos Trans R Soc B: Biol Sci 371(1694):20150274

Wu Y, Abdel-Aty M, Lee J (2018) Crash risk analysis during fog conditions using real-time traffic data. Accid Anal Prev 114:4–11

Xie H, Wu L, Xie W, Lin Q, Liu M, Lin Y (2021) Improving ECMWF short-term intensive rainfall forecasts using generative adversarial nets and deep belief networks. Atmos Res 249:105281

Xie W, Xu G, Zhang H, Dong C (2023) Developing a deep learning-based storm surge forecasting model. Ocean Model 182:102179

Xiu Y-Y, Han L, Feng H-l (2016) The identification of strong convective weather based on machine learning methods. Electron Des Eng 09

Yang B, Chen L, Singh VP, Yi B, Leng Z, Zheng J, Song Q (2023) A method for monthly extreme precipitation forecasting with physical explanations. Water 15(8):1545

Yang J, Honavar V (1998) Feature subset selection using a genetic algorithm. In: Feature Extraction, Construction and Selection. Springer, pp 117–136

Yao H, Li X, Pang H, Sheng L, Wang W (2020) Application of random forest algorithm in hail forecasting over shandong peninsula. Atmos Res 244:105093

Yaseen ZM, Al-Juboori AM, Beyaztas U, Al-Ansari N, Chau K-W, Qi C, Ali M, Salih SQ, Shahid S (2020) Prediction of evaporation in arid and semi-arid regions: a comparative study using different machine learning models. Eng Appl Comput Fluid Mech 14(1):70–89

Yeditha PK, Kasi V, Rathinasamy M, Agarwal A (2020) Forecasting of extreme flood events using different satellite precipitation products and wavelet-based machine learning methods. Chaos: An Interdisciplinary Journal of Nonlinear Science 30(6):063115

You Q, Fraedrich K, Min J, Kang S, Zhu X, Ren G, Meng X (2013) Can temperature extremes in China be calculated from reanalysis? Glob Planet Change 111:268–279

Yu Z, Qu Y, Wang Y, Ma J, Cao Y (2021) Application of machine-learning-based fusion model in visibility forecast: a case study of Shanghai. China. Remote Sensing 13(11):2096

Yucel I, Onen A, Yilmaz K, Gochis D (2015) Calibration and evaluation of a flood forecasting system: utility of numerical weather prediction model, data assimilation and satellite-based rainfall. J Hydrol 523:49–66

Zang Z, Bao X, Li Y, Qu Y, Niu D, Liu N, Chen X (2023) A modified RNN-based deep learning method for prediction of atmospheric visibility. Remote Sens 15(3):553

Zaninelli PG, Barriopedro-Cepero D, Drouard M, Garrido-Pérez JM, Pérez-Aracil J, Fister D, García-Herrera R, Salcedo-Sanz S, Alvarez-Castro MC (2023) Deep learning techniques applied to an attribution study for heatwaves in the Iberian Peninsula. Technical report, Copernicus Meetings

Zebari R, Abdulazeez A, Zeebaree D, Zebari D, Saeed J (2020) A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. J Appl Sci Technol Trends 1(2):56–70

Zhang Z (2018) Improved ADAM optimizer for deep neural networks. In: 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). IEEE, pp 1–2

Zhang Q-S, Zhu S-C (2018) Visual interpretability for deep learning: a survey. Front Inf Technol Electron Eng 19(1):27–39

Zhang R, Chen Z-Y, Xu L-J, Ou C-Q (2019) Meteorological drought forecasting based on a statistical model with machine learning techniques in Shaanxi Province, China. Sci Total Environ 665:338–346

Zhang X, Chen G, Cai L, Jiao H, Hua J, Luo X, Wei X (2021) Impact assessments of typhoon Lekima on forest damages in Subtropical China using machine learning methods and Landsat 8 OLI imagery. Sustainability 13(9):4893

Zhang W, Murakami H, Khouakhi A, Luo M (2021) Compound climate extremes in the present and future climate: machine learning, statistical methods and dynamical modelling. Front Earth Sci 1122

Zhou Z-H (2012) Ensemble methods: foundations and algorithms. CRC press

Zhou K, Zheng Y, Li B, Dong W, Zhang X (2019) Forecasting different types of convective weather: a deep learning approach. J Meteorol Res 33(5):797–809

Zhu L, Aguilera P (2021) Evaluating variations in tropical cyclone precipitation in Eastern Mexico using machine learning techniques. J Geophys Res Atmos 126(7):2021–034604

Zhu L, Zhu G, Han L, Wang N et al (2017) The application of deep learning in airport visibility forecast. Atmospheric and Climate Sciences 7(03):314

Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q (2020) A comprehensive survey on transfer learning. Proc IEEE 109(1):43–76

Zhuo J-Y, Tan Z-M (2023) A deep-learning reconstruction of tropical cyclone size metrics 1981-2017: examining trends. J Clim 1–42

Zou F, Shen L, Jie Z, Zhang W, Liu W (2019) A sufficient condition for convergences of ADAM and RMSPROP. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp 11127–11135

Zscheischler J, Seneviratne SI (2017) Dependence of drivers affects risks associated with compound events. Sci Adv 3(6):1700263

Zscheischler J, Mahecha MD, Harmeling S, Reichstein M (2013) Detection and attribution of large spatiotemporal extreme events in Earth observation data. Eco Inform 15:66–73

Zscheischler J, Mahecha MD, Von Buttlar J, Harmeling S, Jung M, Rammig A, Randerson JT, Schölkopf B, Seneviratne SI, Tomelleri E et al (2014) A few extreme events dominate global interannual variability in gross primary production. Environ Res Lett 9(3):035001

Zscheischler J, Westra S, Van Den Hurk BJ, Seneviratne SI, Ward PJ, Pitman A, AghaKouchak A, Bresch DN, Leonard M, Wahl T et al (2018) Future climate risk from compound events. Nat Clim Chang 8(6):469–477

Zscheischler J, Martius O, Westra S, Bevacqua E, Raymond C, Horton RM, van den Hurk B, AghaKouchak A, Jézéquel A, Mahecha MD et al (2020) A typology of compound weather and climate events. Nat Rev Earth Environ 1(7):333–347

Zscheischler J, Van Den Hurk B, Ward PJ, Westra S (2020) Multivariate extremes and compound events. In: Climate Extremes and Their Implications for Impact and Risk Assessment. Elsevier, pp 59–76

Zwiers FW, Zhang X, Feng Y (2011) Anthropogenic influence on long return period daily temperature extremes at regional scales. J Clim 24(3):881–892

Authors and Affiliations

Sancho Salcedo-Sanz[1] · Jorge Pérez-Aracil[1] · Guido Ascenso[2] · Javier Del Ser[3] · David Casillas-Pérez[4] · Christopher Kadow[5] · Dušan Fister[1] · David Barriopedro[6] · Ricardo García-Herrera[6,7] · Matteo Giuliani[2] · Andrea Castelletti[2]

Sancho Salcedo-Sanz
sancho.salcedo@uah.es

Guido Ascenso
guido.ascenso@polimi.it

Javier Del Ser
javier.delser@tecnalia.com

David Casillas-Pérez
david.casillas@urjc.es

Christopher Kadow
kadow@dkrz.de

Dušan Fister
dusan.fister@uah.es

David Barriopedro
davidb@fis.ucm.es

Ricardo García-Herrera
rgarciah@ucm.es

Matteo Giuliani
matteo.giuliani@polimi.it

Andrea Castelletti
andrea.castelletti@polimi.it

[1] Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares 28805, Madrid, Spain

[2] Environmental Intelligence Laboratory, Politecnico di Milano, Milano 20113, Milano, Italy

[3] TECNALIA, Basque Research and Technology Alliance (BRTA), Derio 48160, Bizkaia, Spain

[4] Department of Signal Processing and Communications, Universidad Rey Juan Carlos, Fuenlabrada 28942, Madrid, Spain

[5] Deutsches Klimarechenzentrum, GmbH, Hamburg D-20146, Hamburg, Germany

[6] Instituto de Geociencias, Centro Superior Investigaciones Científicas-Universidad Complutense de Madrid, Madrid 28040, Madrid, Spain

[7] Department of Earth Physics and Astrophysics, Universidad Complutense de Madrid, Madrid 28040, Madrid, Spain